TEMA 1. LA ESTADÍSTICA EN EL PROCESO DE INVESTIGACIÓN PEDAGÓGICA EMPÍRICA

1.1.- INTRODUCCIÓN

Son muchas las razones que justifican la presencia de la Estadística en los estudios pedagógicos:
-Comprender los trabajos que se publican en revistas científicas, libros, informes, etc. Ya sean en papel o en digital.
-Entender los procesos implicados en la investigación educativa. Entender las directrices que guían el método científico.
-Facilitar el desarrollo de la investigación socio-educativa. Diseñar y desarrollar estudios empíricos en el campo pedagógico. Nos capacitarán para entender y aplicar estos conocimientos en la propia actividad profesional.

1.2.- CONCEPTO Y FUNCIONES DE LA ESTADÍSTICA

Diversas acepciones. Se puede entender:
→ Como una serie de conjuntos de números: por ejemplo, censos de personas.
→ Como ciencia: estudio de los fenómenos aleatorios, desarrollando métodos, técnicas y modelos que nos ayudan a la resolución de problemas pedagógicos y tomar decisiones.

Por lo tanto, la estadística cumple un papel instrumental de apoyo a la investigación socioeducativa. Nos permite: -Conocer las posibilidades y limitaciones de los trabajos empíricos
-Desarrollar un pensamiento crítico y antidogmático del estudio de la realidad.

Así pues, la Estadística *como conjunto de números* puede aportar una primera aproximación al campo de estudio a partir de los datos que facilita al investigador, con posterioridad se pueden aplicar los principios derivados de la Estadística entendida *como ciencia*, para la resolución de problemas o toma de decisiones. Por ejemplo, para construir escuelas infantiles a corto plazo, tendremos en cuenta el tamaño y el método educativo.

FUNCIONES DE LA ESTADÍSTICA:
A.- Facilitar el manejo de datos amplios y dispersos: nos dan estadísticos de una muestra. Se pretenden reducir a índices o estadísticos (media, mediana, desviación típica, correlación) las características que identifican a un conjunto de datos. Estadística descriptiva.
B.- Inferir desde la muestra a la población. Es decir, desde los estadísticos (proceden de las muestras) se pueden estimar los parámetros (medidas de la población).
C.-Ayudar en la toma de decisiones: ¿existen diferencias significativas en los grupos de tratamiento?

1.3.- TIPOS DE ESTADÍSTICA

Estadística descriptiva

Procesos de análisis que se llevan a cabo con los datos empíricos recogidos en las muestras. Este proceso concluye con la obtención de unos valores numéricos que reciben la denominación de estadísticos, cada uno de los cuales destaca una característica representativa del grupo. Transforma un conjunto de números u observaciones en índices que sirven para describir o caracterizar esos datos dentro de los grupos de sujetos.

Es una parte de la estadística que se ocupa del estudio de los métodos y técnicas necesarios para la descripción gráfica y numérica de los conjuntos de datos numerosos. Así, tendremos una visión global del grupo.

Pero tiene limitaciones: no aporta suficientes argumentos científicos al investigador en la toma de decisiones.

Conclusión: *la Estadística descriptiva utiliza estadísticos procedentes de muestras o de poblaciones con una finalidad eminentemente descriptiva o informativa de las mismas.* Realiza una tarea de síntesis y descripción de las características.

Según la naturaleza de las muestras, se pueden diferenciar:
→ Estadística descriptiva *univariada*: media, mediana, varianza. Intenta descubrir y analizar una distribución de datos que provienen de la medición de una variable en una muestra.
→ Estadística descriptiva *bivariada*: recoge y analiza datos de dos variables. Es el campo de las correlaciones. Ejemplo: coeficiente intelectual y estatura ¿hay relación? Te ayuda a predecir.
→ Estadística descriptiva *multivariada*: cuando intervienen más de dos variables.

Estadística inferencial

Pretende avanzar más en el estudio. Se ocupa de *los métodos estadísticos que nos sirven para realizar inferencias objetivas sobre los datos disponibles y trasladarlos a los grupos más amplios.* Es decir, realiza predicciones sobre la similitud de una muestra con la población de la que fue extraída.

Su finalidad es *obtener una serie de conclusiones sobre algún aspecto o variable presente en una población a partir de las observaciones de comportamientos en una o varias muestras.* Es decir, los valores de la población (parámetros) nos permiten conocer el fenómeno, además de resolver y fijar hipótesis y plantear leyes, valorando los márgenes de error con los que realizamos nuestras afirmaciones (nivel de confianza).

Las decisiones que se toman no se pueden realizar en términos de certeza, sino de probabilidad, por lo que se deben fijar los márgenes de error, que suelen ser de 0.05 a 0.01 (5% y 1% respectivamente). Y llevan asociados los correspondientes niveles de confianza (95% y 90%).

Hay dos grandes campos que forman la estadística inferencial:
→ La estimación de parámetros: se puede llevar a cabo
-Mediante la elección de un solo valor de la muestra que se transforma en parámetro (estimación puntual).
-A través de unos límites entre los cuales se espera se encuentre el verdadero valor del parámetro (estimación por intervalos). En este caso debemos saber que esos límites vienen influenciados por los errores aleatorios y los sistemáticos.
→ El contraste de hipótesis: probar mediante datos empíricos las hipótesis que se plantean en el proceso de investigación, de tal forma que son los datos lo que deben ofrecer una respuesta a los planteamientos iniciales del investigador y no al revés, pues se violaría el sentido de la ciencia.

1.4.- EL PAPEL DE LA ESTADÍSTICA EN EL PROCESO DE INVESTIGACIÓN EMPÍRICA

<u>En el problema de investigación</u>

El punto de arranque de toda propuesta de investigación se ubica en la *identificación y selección del problema.* Aquí debe estar la estadística, no como cálculo de estadísticos sino como garantía para poder establecer relaciones entre las características que se analizan. Ello se lleva a cabo mediante diversos modelos estadísticos, desde los que buscamos si existe alguno que permita la resolución del problema.

Se exige que el problema SEA RESOLUBLE, es decir, que con los datos que se puedan recoger en un futuro sea posible alcanzar las respuestas esperadas. Por ejemplo: ¿Oxford o Cambridge: mejor rendimiento en los estudiantes de 15 años?

<u>En la formulación de hipótesis</u>

Cuando el investigador formula su hipótesis se le exige QUE SEA CONSTRASTABLE y COMPROBABLE. Ello será posible si disponemos de los instrumentos de recogida de datos que nos ofrezcan una información nítida y detallada.

También nos ayuda a determinar si la hipótesis planteada es unilateral o bilateral. En el supuesto de las hipótesis planteadas sobre la existencia o no de *diferencias* entre los grupos, es preciso recurrir a las pruebas estadísticas de contraste que os permitan decidir sobre la significación estadística o no de esas diferencias. Medas, varianzas, porcentajes, relaciones, etc.

-Hipótesis bilaterales: en la mayoría de los casos. (Oxford distinto de Cambridge)
-Hipótesis unilaterales: cuando existen otros trabajos debidamente contrastados o que responden a teorías previas. Se decantan por una de las opciones (Oxford > Cambridge).

Si la hipótesis se decanta de una prueba sobre otra, nos referimos a los valores que alcanzará la correlación, medidos en tiempos diferentes (validez predictiva).

<u>En el control de variables extrañas</u>

Nos permite asegurar la validez de los resultados alcanzados. Garantizar el control en los procesos de investigación empírica supone que se aíslan o minimizan los efectos de las covariaciones y la influencia de las variables extrañas que pudieran llegar a ofrecer explicaciones alternativas a las buscadas por el investigador. Nos permitirá realizar afirmaciones con la suficiente validez y rigor en el campo socioeducativo. Ejemplo: asegurarse de que ambos grupos de tratamiento tienen el mismo nivel de inglés de partida.

<u>En la definición de las variables</u>

En la investigación empírico-experimental, la hipótesis establece una relación de dependencia o causalidad entre las variables. *Variable independiente* (Vi, sobre las que se interviene y modifica) y *Variable dependiente* (Vd, la que recoge los efectos de esa intervención).

Debemos utilizar instrumentos que sean fiables y válidos. En ocasiones, las variables son constructos que no admiten una medida directa, por lo que es preciso definir conductas operativas y medibles que nos indiquen el valor real de las mismas. Esta operación se conoce como *definición operativa de las variables*.

En el contraste de hipótesis o comprobación de objetivos

Para realizar la comprobación empírica de las hipótesis, debemos recurrir a las pertinentes pruebas estadísticas, cuya utilización dependerá de la calidad de los datos recogidos.
-Pruebas paramétricas: Cuando podemos garantizar que los datos recogidos nos ofrecen una medida bastante precisa de la variable.
-Pruebas no paramétricas: cuando los datos no tienen la misma precisión porque se ha tomado como medida la observación. O cuando se recogen datos en categorías o grupos mediante la frecuencia, o sea, el número de sujetos que se adscriben a una categoría o grupo.

En la decisión estadística

La fase anterior concluye con la obtención del denominado valor empírico del estadístico correspondiente, que será diferente según la prueba estadística que se haya seleccionado. Esa interpretación consiste en decidir si la hipótesis de nulidad (Ho) se rechaza, y por consiguiente se acepta la hipótesis alternativa o la hipótesis del investigador (Hi). Esta decisión se hace fijando unos niveles de confianza o márgenes de error.

La decisión compara los dos valores del estadístico correspondiente. Por un lado, tenemos el valor empírico (que se obtienen desde la respuesta de los sujetos que participan en la investigación) y por otro, el valor crítico o teórico (se realiza en las tablas correspondientes).

La regla general asociada al contraste de hipótesis y a la decisión estadística nos dice que *cuando el valor empírico del estadístico es mayor que el valor teórico o crítico se rechaza Ho, ello supone aceptar que las diferencias encontradas son estadísticamente significativas.*

1.5.- LA ESTADÍSTICA Y SU RELACIÓN CON LAS CIENCIAS SOCIALES

Educación, Psicología, Sociología, Economía, Demografía, Administración Pública, Humanidades.

1.6.- POSIBILIDADES Y LIMITACIONES DE LA ESTADÍSTICA

La Estadística nos ayudará en la decisión de rechazar las hipótesis de nulidad (Ho), pero las garantías que han conducido a este punto deben ser revisadas y controladas, es decir, no podemos olvidar que sin un buen control sobre las variables extrañas, podemos estar considerando que los cambios generados son achacables a una variable independiente (Vi), cuando en realidad han sido otros factores o variables no controladas las que han generado los cambios.

TEMA 2. PROBLEMA, HIPÓTESIS/OBJETIVOS, VARIABLES Y DATOS. NIVELES DE MEDIDA

2.1.-INTRODUCCIÓN

Una vez enmarcado el papel de la estadística dentro de la investigación en educación, debemos abordar los aspectos fundamentales para comprender mejor los puntos básicos de este proceso.

2.2.-EL PROBLEMA DE INVESTIGACIÓN

El problema constituye el punto de partida de toda investigación. Se manifiesta cuando nos planteamos una pregunta, porque exista una laguna en los resultados de investigaciones anteriores, porque no exista coincidencia en las conclusiones de varios trabajos, porque existe un fenómeno para el que no tenemos explicación, etc.

La característica fundamental de los problemas es su POSIBILIDAD DE RESOLUCIÓN, es decir, en la investigación empírica no cabe la opción de plantear problemas morales o espirituales. Se deben formular problemas sobre los cuales podamos estudiar comportamientos, analizar hechos y evaluar resultados.

<u>Definición y elección del tema de investigación</u>

Al definir el tema, influyen dos tipos de elementos:
-*Subjetivos*: relacionados con la personalidad, preparación científica en un determinado campo y el conocimiento de otros idiomas.
-*Objetivos*: relacionados con la posibilidad real de acceso a los campos de investigación, es decir, el investigador debe tener la posibilidad de acceso a las fuentes documentales.

Se suele aceptar que "un buen planteamiento es la mitad de la solución".

<u>Estructura y características del problema</u>

Para completar la visión del problema, hay unas características deseables:
-Factibles: que existan los medios apropiados para investigar sobre él.
-Claro: que todos los términos incluidos estén perfectamente definidos y son comprensibles.
-Significativos: indica el nivel de importancia del problema. Supone una valoración del problema en un contexto determinado.
-La formulación de la pregunta debe reflejar la naturaleza del mismo la descripción, la asociación o la intervención.

En todos los casos, *se nos presenta una cuestión a modo de desafío, que mueve nuestro interés y nos invita a buscar una solución.* Así, debemos exigir al problema que sea RESOLUBLE, que se puedan *recoger datos empíricos sobre el mismo que permitan ofrecer una respuesta adecuada a ese interrogante.*

- Investigación basada en la observación.
- Investigación por encuesta
- Investigación correlacional: asociación entre variables.
- Investigación experimental o cuasiexperimental: intervención.
- Investigación descriptiva: descripción.

<u>Criterios para la evaluación de problemas de investigación</u>

Una vez planteado el problema, sería interesante proceder a una primera valoración. Podemos señalar estos criterios:
-*Viabilidad*: relacionado con la posibilidad de resolución.
-*Interés*: que tiene el tema para el investigador, el equipo y la persona que encarga el trabajo. Nivel de implicación personal.
-*Relevancia teórica y práctica*: ¿contribuye al aumento de conocimientos sobre un determinado tema? ¿conduce a una mejor toma de decisiones sobre la sociedad?
-*Coherencia:* con los planteamientos generales de la comunidad científica, es decir, el problema debe enmarcarse dentro de propuestas generales de investigación.
-*Atención al contexto*: exige la incardinación del problema en el contexto en que se va a realizar. Por ello, un tema excesivamente genérico presenta dificultades para ofrecer una respuesta coherente.
-*Otros aspectos:* presentación minuciosa, lenguaje claro y conciso, etc.

2.3.-LA REVISIÓN DE FUENTES Y EL ESTADO DE LA CUESTIÓN

Una vez planteado el problema objeto de investigación, es necesario efectuar una serie de consultas a diferentes fuentes para conocer qué se ha dicho y cómo sobre ese tema. Necesidamos saber si ya se ha solucionado en otras investigaciones y con qué resultados, o necesitamos abordarlo desde la novedad. Según Fox, diferenciamos entre:
-Bibliografía de la investigación: informes sobre trabajos ya realizados, datos empíricos, orientaciones y sugerencias.
-Bibliografía conceptual: textos, artículos, informes, opiniones y sugerencias. Esta información aparece tanto en los escritos como en la red.

Fuentes y bases de documentación

Siguiendo un criterio generalista, diferenciamos dos grupos:
→ Fuentes bibliográficas: textos escritos a los que se acude para recabar la información sobre el problema de investigación:
-Obras generales: diccionarios, enciclopedias, manuales y anuarios, thesaurus.
-Publicaciones bibliográficas periódicas: instituciones y organismos públicos o privados. Boletines bibliográficos, índices y abstracts.
-Revistas de Investigación: especializadas en publicar trabajos teóricos o empíricos sobre temas diversos.
→ Fuentes no bibliográficas:
-Recursos tecnológicos informáticos: el correo electrónico, buscadores como Google, Revistas Electrónicas.
-Centros de Investigación en Educación: IBE.
-Centros de Documentación: CSIC, ISOC, ERIC.
-Asociaciones de Investigación en Educación: AIDIPE, AERA.

Criterios de valoración de la información recogida

-Relevancia: que sea importante.
-Claridad: que sea nítido.
-Actualidad: más próximas en el tiempo.
-Adecuación: que la información tenga relación con el problema y aporte datos para facilitar su resolución.
-Contextualizada: que los contextos en que se desarrolla la investigación de referencia sean similares a los que nos vamos a encontrar en nuestro trabajo.

2.4.-HIPÓTESIS Y OBJETIVOS DE LA INVESTIGACIÓN

Una vez concretado el problema y revisadas las fuentes, el investigador puede aventurar posibles soluciones al mismo, estas son las HIPÓTESIS. Se suele decir “hipótesis directiva”, aunque algunos investigadores hablan de objetivos de investigación. Hemos de tener en cuenta que los objetivos tienen un carácter más descriptivo mientras que las hipótesis buscan la relación causal entre las variables.

Concepto y naturaleza de las hipótesis

Entendemos que la hipótesis es una proposición o conjunto de proposiciones no demostradas, cuyo análisis puede llevar a una conclusión lógica; es un medio o una parte de cualquier investigación y estudio, una explicación razonable sobre el tema a tratar, que debe ser sometida a comprobación empírica.

Bunge propone tres requisitos en la formulación de hipótesis científicas:
-Tiene que ser bien formada y significativa.
-Fundada en los conocimientos previos.
-Ser empíricamente constrastable mediante diversos procedimientos.

A esos criterios básicos, podemos añadir:
-Que exista una conexión con el marco teórico en que se plantean y con otras hipótesis de formulación más sencilla.
-Ser precisas y específicas, expresadas con simplicidad lógica, descriptivas.
-Que traten de explicar el fenómeno, susceptibles de cuantificación.
-Que sean generalizables.

En resumen, lo importante es que estén BIEN FUNDAMENTADAS y que SEAN CONTRASTABLES. La primera exige una profunda revisión de fuentes. La segunda se realiza mediante procedimientos empíricos (si no es contrastable, no es científico).

A estas condiciones, se le puede añadir el establecimiento de relaciones causales entre las variables.

*Conclusión: las hipótesis son los elementos directivos de la investigación científica, pues nos indican qué buscamos, qué variables se deben observar, medir y manipular; así como la realización de la observación, la experimentación y el análisis de datos. (*Ver dibujo 2.1 de página 25).

Debe tener la referencia básica del problema. Esto se basa en el marco teórico y en el conocimiento que posee de forma personal sobre dicho problema.

En algunas investigaciones, nos encontramos con los términos *objetivos / propósito general* (marco general de la investigación) *y objetivos operativo* (se recogen los puntos principales, las tareas que debe realizar el investigador).

Conviene indicar que las hipótesis recogen en sus enunciados la existencia de diferencias o relaciones entre las variables, en cambio los objetivos se dirigen hacia el campo de las descripciones e implicaciones.

Diferentes tipos de hipótesis

Para clasificar las hipótesis, hay varios criterios, no excluyentes, sino complementarios.

A.- Según la posibilidad de generalización

→ *Existenciales:* aquellas que establecen una relación entre dos o más variables para al menos un caso. Ejemplo: "los estudiantes de 2º del Centro Y que utilicen el método X para el aprendizaje de Ciencias, sacarán mejores notas que los que no lo utilicen".

→ *Universales*: válidas para todos los casos. En el campo de la educación resulta prácticamente imposible. Se pueden encontrar resultados similares, pero difícilmente idénticos. Lo más parecido sería: "En un determinado centro, con unas características definidas para los estudiantes, en unas condiciones particulares, es posible que ocurra esto".

→ *Probables*: se establece una gradación en el término universalidad para determinados casos. Debemos ser conscientes de que en nuestras afirmaciones pueden quedar validadas, no en términos de certeza, sino probabilidad. "Los estudiantes de la UNED de Estadística que estudien, tengan tutorías, plataforma, etc. probablemente obtendrán buenas calificaciones. Cada uno de estos apartados debe ser comprobado".

B.- Según su forma de expresión cuantitativa

→ *Sustantiva o científica*: expresa la relación o dependencia entre las variables con definición expresa de las mismas. Esta hipótesis puede venir derivada de la observación y la experiencia (inductiva). Además se puede partir de la realidad para llegar a las teorías. En cambio, cuando se derivan de la teoría (deductivas), el proceso es a la inversa, sirven para comprobar el funcionamiento de las teorías en el campo educativo. "Comprobar que el método de aprendizaje basado en la solución de problemas es más adecuado para favorecer el razonamiento matemático".

→ *Estadística:* definida en términos de relaciones estadísticas deducidas de la anterior. En este caso se establece una diferencia o relación entre los parámetros poblacionales, que debe ser comprobada empíricamente. Pueden enunciarse como Ho, que postula la no existencia de diferencias o relaciones entre los grupos. O puede hacerse como H1, que es la que establece la relación entre las variables, de forma genérica (bidireccional) o decantándose por una de las dos opciones (direccional).

C.- Según el nivel de aproximación a la realidad

→ *Operacional*: en términos observables. "La utilización de subrayado influirá en los resultados académicos alcanzados por los estudiantes en Historia".

→ *Conceptual*: generalizaciones de mayor aplicación en el futuro. "Las técnicas de estudio empleadas por los sujetos influirán sobre los rendimientos alcanzados".

D.- Según el número de variables y sus relaciones

→ *Descriptiva de una sola variable*: existencia de determinadas características, uniformidades o regularidades empíricas en una población o universo. Afirmaciones que deben ser comprobadas. "Los estudiantes de la UNED que van a tutorías probablemente obtendrán buenos resultados".

→ *Descriptiva con dos o más variables y relación de aseveración*: relación simple de asociación o covariación entre ellas; es decir, un cambio en una variable va unido al cambio correlativo en la otra del mismo o de distinto signo. Comprobar los niveles de asociación. "Los que vayan a tutorías, estudien 9 horas semanales y consulten obras, es posible que alcancen nuevos resultados".

→ *Con dos o más variables y relación de dependencia*: son las más interesantes desde el pdv científico. Se dirigen hacia la explicación y la predicción de los fenómenos educativos. Campo propio de las hipótesis causales que pretenden explicar o conocer las razones o motivos de los fenómenos. "Los estudiantes de 1º de Educación Social que realicen el curso sobre técnicas de estudio, mejorarán de forma apreciable sus resultados".

2.5.-IDENTIFICACIÓN Y DEFINICIÓN DE VARIABLES

Concepto y modalidades de variables

Una **variable** es aquella característica que admite diversos valores, es decir, dos o más modalidades.
Una **constante** es un valor numérico que no cambia en un contexto determinado: aquella característica que admite una única forma de manifestarse, una única modalidad de presencia.

→ NIVELES DE MEDIDA

A.- Variables categóricas o atributivas: son las que establecen distintas categorías para cada una de las modalidades. Pueden ser:

-Dicotómicas (sexo, si-no, verdadero-falso): una variable continua se puede dicotomizar en dos categorías.
-Politómicas: más de dos categorías. Niveles de estudios (superiores, secundarios, primarios, sin estudios), clase social (baja, media, alta), etc.

B.- Variables cualitativas: expresan la posesión de una cualidad y con cierta intensidad (rangos). Escala observacional.

C.- Variables cuantitativas: admiten medida numérica, la cuantificación de una determinada cualidad. Pueden ser:
-Discretas: solo admiten números enteros en su definición.
-Continuas: numero infinito de valores potenciales.

→ ENFOQUE METODOLÓGICO

D.- Variables dependientes: Vd. Reciben los efectos de la intervención sobre la variable independiente. Refleja la consecuencia de los cambios que se han producido.

E.- Variables independientes: Vi. Son aquellas sobre las que se interviene o actúa el investigador con el fin de analizar su influencia sobre la dependiente. Suele ser una variable experimental y controlada.

F.- Variables extrañas o intervinientes: Ve. Están presentes y que deben ser controladas para evitar la contaminación de los resultados finales.

→ CONTEXTO TEÓRICO-EXPLICATIVO

G.- Variables estímulo: condiciones externas al sujeto que pueden ser objeto de intervención por parte del investigador, pues afectan directamente al comportamiento del individuo dentro de un determinado contexto. "Recibir premios por tareas realizadas".

H.- Variables respuesta: reflejan el comportamiento manifiesto de los sujetos ante determinadas intervenciones en el campo educativo. Se trata de la respuesta que ofrece el individuo. "Alcanzar unos mayores niveles de integración en el grupo social".

I.- Variables intermediarias: variables que no son objeto de investigación, si bien, están presentes y pueden hacer varias los resultados previsibles. Son una interposición entre el estímulo y la repsuesta.

Definición operativa de las variables

Implica que el concepto analizado debe ser definido en función de las acciones u operaciones que son precisas para poder medirlo y actuar sobre él.

La definición operacional es la traducción de los conceptos teóricos al lenguaje empírico, la sustitución de lo que no se puede observar por aquello accesible a la observación y los instrumentos de medida. Significa la traducción y sustitución de algo inobservable en factores y propiedades observables.

2.6.-LA RECOGIDA DE DATOS Y SU CALIDAD

El proceso de recogida de información y datos parte de la aplicación de aquellas técnicas e instrumentos que faciliten esta tarea. Es un momento importante, pues es necesario disponer de unos buenos instrumentos y de una buena calidad de datos.

Criterios para la selección y elaboración de los instrumentos de recogida de datos

Para tomar la decisión en la **selección de instrumentos** debe contemplar una serie de criterios:
-La identificación total del rasgo o característica que se pretende evaluar (clarificación del mismo).
-El tipo de validez empleado: debe ser coherente con las necesidades del investigador.
-La técnica de fiabilidad: identificada con la precisión de la medida.
-Las características de la muestra con que se validó la prueba, debemos buscar la representatividad de los sujetos seleccionados, comparados con la población global.
-Los niveles de fiabilidad y validez alcanzados por la prueba o instrumento y reflejados en la correspondiente ficha técnica.
-Otros criterios secundarios: tiempo, valoración posterior, facilidad de comprensión etc.

Para la **elaboración de instrumentos**, tenemos los siguientes criterios:
→ Identificación y definición clara y concisa del rasgo o característica que nos conduce a una definición operativa de éste en forma de conducta observable. ¿Qué debemos apreciar? ¿En qué debemos fijarnos?
→ Conocer los objetivos a conseguir con el instrumento: que los objetivos se encuentren categorizados, que indiquen el grado de dominio del sujeto sobre los mismos.
→ Seleccionar los items o elementos más apropiados.
→ Formular los items o elementos con precisión.
→ Ordenar los items de forma apropiada.
→ El tiempo preciso para responder (duración).

→ Las respuestas que debe dar el sujeto: unívocas y fáciles. Sencilla, para evitar errores de interpretación.
Para alcanzar garantías en los puntos contemplados debemos proceder a su aplicación definitiva, someter dicho instrumento a **aplicación piloto,** en la que se trabaja con una muestra reducida de sujetos pero que sea lo más representativa posible. Con los datos suministrados por esta, podemos determinar los niveles de fiabilidad y validez de la prueba.

Los datos y sus niveles de medida

Escalas	**Definición**	**Ejemplos**	**Operaciones**
Nominal	Datos categóricos: Moda. Atribuir números a las categorías.	Colores, sexo, categoría profesional, nivel de estudios.	Igual o no igual. Sólo se trabaja con frecuencias. Moda.
Ordinal	Datos ordenados por rangos con orden creciente o decreciente (rangos) Mediana.	Altos/bajos, Pesados/ligeros, Interesados/desinteresados.	> o < No hay unidad constante de medidas, orden selectivo. Mediana.
Intervalo	Intervalos iguales siendo el cero arbitrario. Media.	Tiempo de ejecución de tareas, test, pruebas objetivas.	Se trabaja directamente con las puntuaciones. Media.
Razón o proporción	Intervalos iguales, el cero se define como ausencia de la característica.	Temperatura, peso, longitud.	Todas las operaciones matemáticas.

→ *Nivel nominal*

Se discute si es un verdadero nivel de medida. Es muy pobre, atribuye números o símbolos a las diferentes categorías en las que se ha dividido un conjunto, las cuales han de ser excluyentes y agotar todas las posibilidades del conjunto inicial. Por ejemplo: nivel de estudios. 1 = primarios, 2= secundarios, 3 = superiores. El numero no tiene ningún valor operativo, pero señala la pertenencia o no a ese grupo o categoría.

Aquí se llevan a cabo las operaciones de igualdad, desigualdad y equivalencia. No se trabaja directamente con números, sino con sus frecuencias (número de veces que se presenta un hecho en el grupo objeto de investigación, y en cada una de las categorías definidas con anterioridad).

Entre los estadísticos que pueden calcularse con datos nominales se encuentran: la moda, la frecuencia, el coeficiente de asociación (C) o de contingencia, la prueba de Ji cuadrado con sus diferentes modalidades.

→ *Nivel ordinal*

Toma enunciados < o >, lo que supone que se puede establecer una escala ordenada creciente o decreciente con los objetos evaluados. Esto nos permite afirma que un objeto ocupa el sexto lugar o es inferior a otro. Hemos de tener en cuenta que la distancia entre dos posiciones sucesivas no es comparable, por no disponer de una unidad constante de medida.

Es el nivel de medida de mayor empleo en la investigación en sociales, y también en educación, pues establecemos criterios de comparación. No nos interesa el número en sí mismo, sino la posición o el rango que ocupa dentro del grupo. Así, las puntuaciones se transforman en rangos para efectuar operaciones matemáticas.

Para ello disponemos de la mediana, la correlación ordinal de Spearman y gran cantidad de pruebas no paramétricas.

→ *Nivel intervalo*

Implica que además de las características de la ordinal, se añale la igualdad de distancia o intervalo entre dos valores consecutivos, lo que conlleva la existencia de una unidad común y constante de medida. La distancia entre 6 y 7 debe ser la misma que entre 4 y 5.

Usarlo en la educación es complejo, pues gran parte de lo que evaluamos es cualitativo. Vamos a trabajar directamente sobre las puntuaciones alcanzadas por los sujetos en la aplicación de las correspondientes pruebas estadísticas.

Tenemos la media, la desviación típica, la correlación de Pearson y toda las pruebas paramétricas.

→ *Nivel de razón o proporción*

Sería el nivel más perfecto, pues a todas las características anteriores debemos añadir el 0 absoluto, la no presencia del atributo o propiedad medida. Esto nos permite ordenar elementos, establecer diferencias entre los mismos. Peso, longitud y temperatura. Entran todas las operaciones matemáticas: +, -, x, :, doble, triple, mitad.

TEMA 4: Organización de datos. Análisis exploratorio de datos:

1. INTRODUCCIÓN:

En esta parte pasamos ya de los grandes principios conceptuales a la práctica, a la parte empírica: la recogida de datos y su primera ordenación.

¿Cómo proceder a la organización de los datos recogidos con nuestros instrumentos de medida? Entramos en el terreno de la **Estadística Descriptiva**.

"La Estadística Descriptiva comienza con un conjunto de datos: el investigador intenta trasladar las características esenciales de los datos a formas más interpretables: distribuciones de frecuencia, gráficos; y calcula índices numéricos como promedios, percentiles y medidas de variabilidad". **Johnson y Christensen** (2008: 465).

Además de organización de datos, hablaremos de análisis exploratorio de datos: el investigador observa y describe lo que sucede en la realidad; sin entrar en ningún contraste de hipótesis.

2. DELA DEFINICIÓN DEL PROBLEMA Y LAS VARIABLES A LA OBSERVACIÓN Y RECOGIDA DE DATOS:

Uno de los errores más frecuentes de los investigadores es recoger datos sin seguir adecuadamente los pasos del **proceso de investigación**. Por eso es fundamental, antes de iniciar cualquier estudio, tratar de dar respuesta a todas las fases del proceso de investigación. Una vez hecho esto, se procede a realizar la **recogida datos - de la información**, es decir, atribuir valores a las variables que son objeto del estudio. Para ello necesitamos algún procedimiento que nos permita asignar dichos valores (números, rangos, categorías) a las variables conceptuales: necesitamos un instrumento de medida adecuado y una regla de medida bien definida.

Ejemplo: Estudiar si un problema de mejora de la convivencia consigue disminuir el número de partes de amonestación en institutos de educación secundaria y si mejora el clima escolar.

Variables:

1. El programa. (V.I.) – Categórica Dicotómica Nominal (0= no hay programa; 1= si hay programa)
2. Partes. (V.D.) – Cuantitativa Discreta (0 - ∞)
3. Clima Escolar. (V.D.) – Cuantitativa Intervalo. Se podría utilizar un cuestionario estandarizado, donde la puntuación de 25 para abajo es el peor clima y de 75 para arriba es el mejor clima.

Recogida de datos: Toda la ESO. Los grupos A reciben el programa, los grupos B no.

	Clase	Programa	Nº de partes	Clima Escolar
Sujeto 1	1A	1	2	70
Sujeto 2	2A	1	0	60
Sujeto 3	1B	0	8	50
Sujeto 4	3B	0	6	40
…	…	…	…	…

En la **investigación empírica** siempre tenemos que trabajar con datos empíricos, es decir, transformar en números la realidad observada, para ello necesitamos instrumentos de medida válidos y fiables. La precisión (fiabilidad) de los instrumentos podrá tener distintos grados (por ej. en variables como la edad, el cociente intelectual, el rendimiento académico, la altura, la presión arterial, etc.) y eso afectará a la validez de los resultados. En este sentido, la **elección de los instrumentos de medida** o el proceso de construcción de los mismos es una parte fundamental de la investigación.

Una vez realizado el diseño de investigación y elegidos los instrumentos de medida, se puede realizar el **trabajo de campo**, aplicar los instrumentos, recoger los datos.

3. PERMISOS Y ÉTICA EN LA INVESTIGACIÓN Y RECOGIDA DE DATOS:

Todavía no hay un consenso completamente claro sobre cómo debe procederse éticamente en la investigación. Lo que sí parece claro es que no se caerá en uno de estos dos extremos:

- Proteger contra la experimentación humana a toda costa

…frente a…

- Permitir a cualquiera ser objeto de experimentación

En la investigación educativa los **principios éticos que deben seguirse** son:

- **Participación voluntaria** de los sujetos.
- **Informe consentido** = los sujetos conocen la naturaleza de la investigación.
- **Evitación de riesgo de daño físico o psíquico**.
- **Confidencialidad y anonimato**.
- Respetar el lugar, causando **las menos alteraciones posibles**.
- **No manipular los datos obtenidos y reflejar fielmente los resultados** aunque contradigan su hipótesis inicial.
- **Citar adecuadamente los autores y las fuentes** de información durante la revisión del estado de la cuestión.

PRÁCTICAS ÉTICAS = DEONTOLOGÍA PROFESIONAL

4. DE LOS DISTINTOS INTRUMENTOS A LOS DATOS: ELECCIÓN DEL PROGRAMA, LA MATRIZ DE DATOS Y EL LIBRO DE CÓDIGOS.

Una vez que hemos aplicado los instrumentos, debemos trasladar los datos a algún programa que nos permita trabajar con ellos.

Las puntuaciones que vamos a introducir se llaman **puntuaciones directas (Xi)**, que es la puntuación que obtiene el sujeto tras aplicarle un instrumento de medida (una pregunta, un test, una escala, etc.). La **codificación** de datos consiste en la asignación de números o caracteres a los valores de variables.

Antes de empezar a tabular o introducir datos, es conveniente escribir el **libro de códigos**. Se trata de un documento en el que se especifican todas las variables del estudio en el orden en el que serán introducidas en la matriz de datos. La utilidad de dicho libro es decisiva para evitar errores de tabulación.

Ejemplo de libro de códigos:

Ítem	Variable	Etiqueta variable	Código	Etiqueta valores
C.1	C1Ident	Identificación	Asignar un valor numérico a cada sujeto (por ejemplo, 01; 02; 03, ...).	—
C.2	C2Sexo	Sexo	0 1	Hombre Mujer
C.3	C3C_aut	Comunidad Autónoma	1 2 3 4 5	Galicia Extremadura Andalucía Madrid Castilla y León
C.4	C4CI	Cociente intelectual	Cualquier valor entre 50-150	—
C.5	C5Rmat	Rendimiento matemático	Cualquier valor entre 0-100	—
C.6	C6Satis	Satisfacción con el curso	1 2 3 4 5	Muy insatisfecho Muy satisfecho

Es importante definir el valor que asignemos a los **datos perdidos** o ***Messing data***. Por ejemplo, cuando estamos tabulando un cuestionario y nos encontramos que un sujeto no ha contestado a una pregunta, tenemos dos opciones:

1. *Se deja en blanco* en la matriz de datos, de modo que no se contabilice a la hora de calcular los estadísticos. Un error muy común es poner ceros, pero esto puede distorsionar el cálculo de los estadísticos ya que puede estar incluido como un valor válido.

2. *Definir un valor* perdido por el usuario, como por ejemplo 99 (siempre que no sea un valor admisible dentro de la escala de medida de la variable). En este caso, debemos programar dicho valor en el programa que utilicemos, de modo que el programa no utilice ese valor para los cálculos.

La **matriz de datos** es simplemente una tabla de doble entrada en la que las filas representan sujetos y las columnas a las distintas variables medidas. Debemos ordenar y categorizar los datos, para poder apreciar mejor las características del grupo en cada una de las variables. A este primer análisis se le denomina **estadística descriptiva**.

5. ORGANIZACIÓN DE LOS DATOS: DEPURACIÓN DE DATOS Y DISTRIBUCIONES DE FRECUENCIA.

Antes de empezar a realizar los análisis estadísticos, debemos hacer lo que entendemos como **depuración de datos** y exige dos fases:

- **El control de calidad de la tabulación**[1] que consiste en comprobar la fidelidad de la tabulación.
- **La depuración** *de datos* que consiste en verificar si hay valores "fuera de rango" según se había definido en el libro de códigos.

  Hay distintas maneras de hacerlo:
  - Realizar un **análisis descriptivo** solicitando los valores mínimo y máximo
  - Con el libro de códigos debemos ir **comprobando** que los valores están dentro del rango definido.
  - Realizar una **distribución de frecuencias**. Aparecen todas las puntuaciones obtenidas en una variable (puntuaciones directas) y el número de veces que se repite cada puntuación (frecuencia absoluta). También son utilizadas la frecuencia relativa (**frecuencia absoluta entre el número total de puntuaciones: f_a/N**) que, multiplicado por 100 indica el porcentaje de aparición de una puntuación respecto del total y, la frecuencia acumulada, que muestra el número de sujetos a los que supera una puntuación determinada.

Distribuciones de frecuencia y aproximación al concepto de percentil.

Una distribución de frecuencia suele incluir:

- **Puntuaciones directas** obtenidas (**X_i**)
- **Frecuencias absolutas** → número de veces que se repite cada X_i (**f_i**)
- **Porcentaje relativo** → o frecuencias relativas **f_i/N** x100. Es el % de aparición.
- **Frecuencia acumulada** → muestra el nº de sujetos a los que supera una puntuación determinada. Se calcula sumando las frecuencias relativas que queden por debajo del límite superior de cada puntuación.

Puntuación directa	Frecuencia absoluta	Frecuencia relativa x 100	Frecuencia acumulada x 100 (porcentaje acumulado)
1 2 3 4 5 6 7	179 93 107 128 108 131 266	17,7 9,2 10,6 12,6 10,7 12,9 26,3	17,7 26,9 37,5 50,1 60,8 73,7 100,0
Total	N= 1012	100,0	

[1] Tabular los datos es trasladar los datos de los instrumentos de medida a la matriz de datos.

Los porcentajes acumulados (sin decimales) se utilizan mucho en la **construcción de baremos** para interpretar las puntuaciones en los test, donde reciben el nombre de ***percentiles***. El percentil indica el porcentaje de sujetos que deja por debajo de si una puntuación determinada.

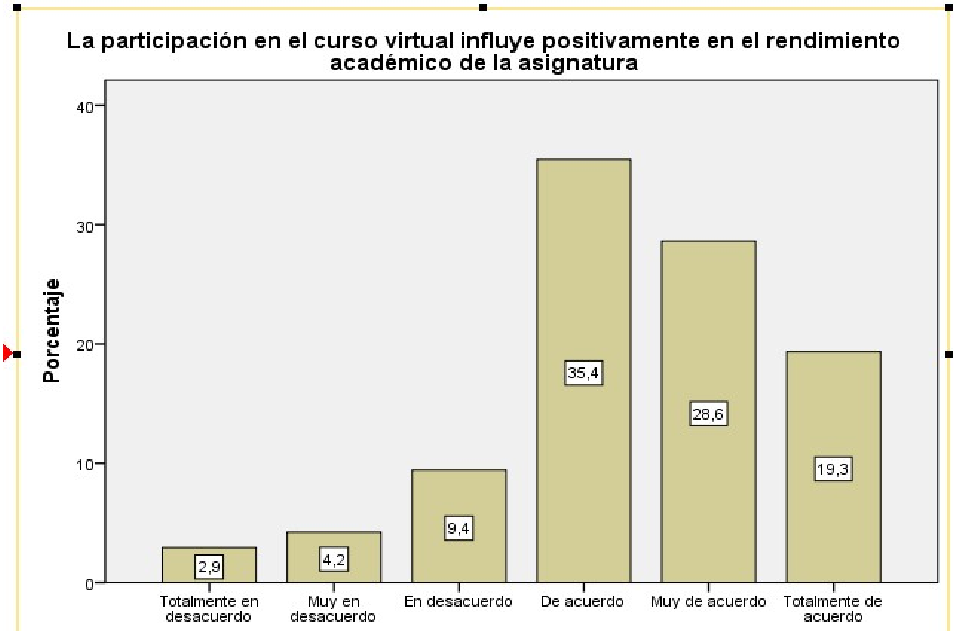
Una distribución de frecuencia suele incluir:

1ª Columna: el número de sujetos o casos que corresponde a cada puntuación directa: **frecuencia absoluta**.

2ª Columna: el **porcentaje relativo** (o frecuencia relativa multiplicada por 100) para código, incluidos los valores perdidos.

3ª Columna: el **porcentaje válido**, que es el porcentaje relativo de cada valor excluidos los valores perdidos.

4ª Columna: el **porcentaje acumulado**, indica el porcentaje de casos que deja por debajo de sí el límite superior de cada puntuación.



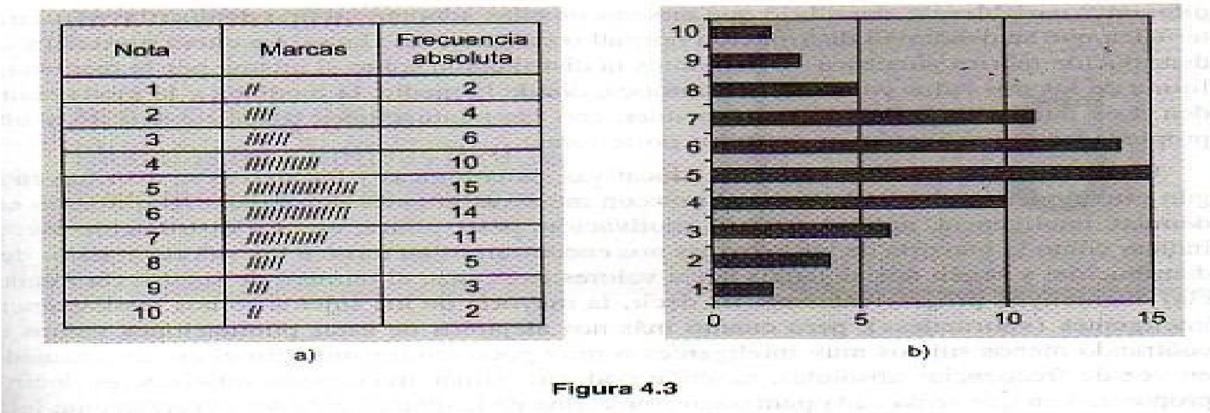
La participación en el curso virtual influye positivamente en el rendimiento académico de la asignatura

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Totalmente en desacuerdo	18	2,9	2,9	2,9
	Muy en desacuerdo	26	4,2	4,2	7,2
	En desacuerdo	58	9,4	9,4	16,6
	De acuerdo	218	35,4	35,4	52,0
	Muy de acuerdo	176	28,6	28,6	80,7
	Totalmente de acuerdo	119	19,3	19,3	100,0
	Total	615	100,0	100,0	

6. APROXIMACIÓN INTUITIVA A LAS REPRESENTACIONES GRÁFICAS Y LA CURVA NORMAL.

Una **representación gráfica** es una forma de ordenar información disponible en la matriz de datos y comprendida con un simple golpe de vista. Para realizar un gráfico necesitamos una distribución de frecuencias.

Hay distintas formas de representar los datos, dependiendo de la naturaleza de las variables: **sectores, barras, polígonos de frecuencia, tallo y hojas, etc**.

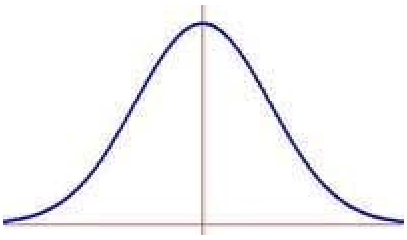


Nota	Marcas	Frecuencia absoluta
1	//	2
2	////	4
3	//////	6
4	//////////	10
5	///////////////	15
6	//////////////	14
7	///////////	11
8	/////	5
9	///	3
10	//	2

a) b)

Figura 4.3

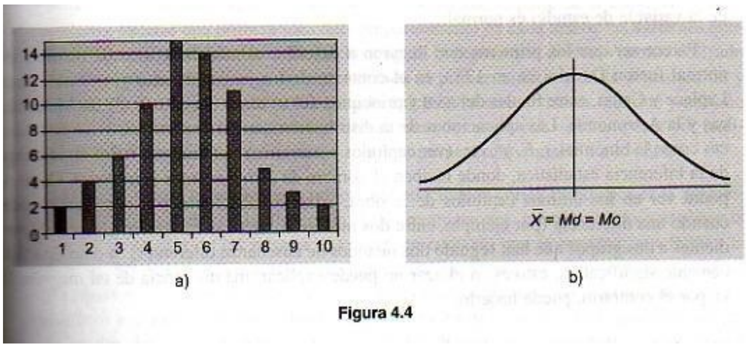
En muchas variables biológicas y psicoeducativas (peso, altura, cociente intelectual, rendimiento), cuando realizamos una distribución de frecuencias, observamos que esta distribución adopta unas características especiales. Entre las distintas formas que pueden adoptar las distribuciones de frecuencias en diferentes variables se descubrió que muchas de ellas adoptan un tipo de distribución característica que se denomina distribución normal o campana de Gauss.



La **curva normal** es una distribución teórica simétrica, donde la media la mediana y la moda coinciden.

$$\boxed{\bar{X} = Md = Mo}$$

La **distribución** viene a indicar cómo la mayoría de las personas nos encontramos en torno a los valores medios.



a) b)

Figura 4.4

TEMA 5: Reducción de datos. Medidas descriptivas básicas y representaciones gráficas.

1. INTRODUCCIÓN. DE LA ORGANIZACIÓN A LA DESCRIPCIÓN DE DATOS:

Normalmente, antes de realizar análisis más complejos, se parte de la **Estadística Descriptiva** para hacerse una primera idea del comportamiento de cada una de las variables contenidas en nuestra matriz de datos.

La **variabilidad**, es el estudio de la dispersión de las puntuaciones, contribuyendo a explicar la magnitud y naturaleza de las mismas. Lo contrario a una variable es una constante, es decir, cuando todos los sujetos u objetos tienen la misma puntuación, cuando no hay diferencias y, por tanto, no hay variabilidad.

Generalmente, primero se estudia la tendencia central del grupo, es decir, hacia qué puntuación tiende el grupo. El índice más conocido es la **media aritmética**. Sin embargo, este índice por sí solo no nos proporciona suficiente información como para hacernos una idea del grupo. ¿La mayoría de los sujetos tiene unas puntuaciones cercanas a la media o hay sujetos con puntuaciones muy altas y muy bajas? En otras palabras, nos estamos preguntando si el grupo es homogéneo en torno a la media o heterogéneo (muchas puntuaciones distan considerablemente de la media, tanto por arriba como por abajo). Dicho de otra forma, ¿hay o no hay dispersión de las puntuaciones en torno a la media aritmética? Para comprobarlo, nos referiremos a las medidas de variabilidad o dispersión, como la **desviación típica**.

Una vez tabulados los datos, comenzamos a trabajar con la matriz de datos y a realizar los análisis pertinente, normalmente se parte de la **Estadística Descriptiva** para hacerse una primera idea del comportamiento de cada una de las variables contenidas en nuestra matriz de datos. La **Estadística Descriptiva** consiste en una serie de procedimientos para organizar, clasificar y resumir conjuntos de datos a través de **índices numéricos** y por medio de **representaciones gráficas**.

- **Indices numéricos:** medidas de tendencia central (**medias, mediana y moda**)
- **Medidas de variabilidad:** (**amplitud, varianza y desviación típica**).
- **Representaciones gráficas:** gráfico de sectores, gráficos de barras, histograma, gráfico de caja...

Para caracterizar un grupo, necesitamos tanto medidas de tendencia central como medidas de variabilidad.

(0, 0, 0, 10, 10, 10) vs. (5, 5, 5, 5, 5, 5)

$$\bar{X} = \frac{\sum X_i}{N} \qquad\qquad s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{N}}$$

0, 0, 0, 10, 10, 10	5, 5, 5, 5, 5, 5
$\bar{X} = \frac{30}{6} = 5$	$\bar{X} = 5$
$s \neq 0$ Valor alto	$s = 0$

2. MEDIDAS DE TENDENCIA CENTRAL: MEDIA, MEDIANA Y MODA. USOS E INTERPRETACIÓN.

La **tendencia central** del grupo indica hacia qué valor tiende ir el grupo. El índice o medida más conocida es la **media aritmética**. Su símbolo es

$$\bar{X} = \frac{suma\ de\ todas\ las\ puntuaciones}{número\ total\ de\ sujetos} = \frac{\sum Xi}{N}$$

Interpretar una media aritmética suele ser muy sencillo. Para interpretarla correctamente, es conveniente conocer la puntuación mínima y máxima de la escala de medida de la variable y situar la media aritmética dentro de ese recorrido.

Las otras medidas de tendencia central son la moda y la mediana.

- **La moda** (**Mo**) es el valor con frecuencia absoluta más alta, **la puntuación que más se repite**. Por tanto, la moda no necesita ningún cálculo. Cuando tenemos dos puntuaciones con la misma frecuencia, diremos que la distribución es bimodal, y entonces diremos que las modas son, por ejemplo, 4 y 5. También podemos encontrarnos distribuciones de frecuencias en las que más de dos puntuaciones tienen la misma. Es este caso hablaremos de distribuciones plurimodales.
- **La mediana** (**Md**) es aquel valor que deja por encima y por debajo de sí al 50% de los sujetos de la muestra, el valor central de la distribución de frecuencias.

Tabla 5.1

X_i	f	f_a
22	2	50
21	4	48
20	1	44
19	5	43
18	1	38
17	2	37
$\boxed{16}$	8	35
15	2	27
14	3	25
13	6	22
12	5	16
11	2	11
10	3	9
9	2	6
8	3	4
5	1	1
	$\Sigma = 50$	

Mo

25

Md 14,5

Tabla 5.3. Actitud.

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	5,00	1	2,0	2,0	2,0
	8,00	3	6,0	6,0	8,0
	9,00	2	4,0	4,0	12,0
	10,00	3	6,0	6,0	18,0
	11,00	2	4,0	4,0	22,0
	12,00	5	10,0	10,0	32,0
	13,00	6	12,0	12,0	44,0
	14,00	3	6,0	6,0	50,0
	15,00	2	4,0	4,0	54,0
	16,00	8	16,0	16,0	70,0
	17,00	2	4,0	4,0	74,0
	18,00	1	2,0	2,0	76,0
	19,00	5	10,0	10,0	86,0
	20,00	1	2,0	2,0	88,0
	21,00	4	8,0	8,0	96,0
	22,00	2	4,0	4,0	100,0
	Total	50	100,0	100,0	

Md 14,5

En puntuaciones pares:

N = 50 Md = 14.5 se realiza la media

$\underbrace{5, 8, 8, 8, 9, \dots 14, 14,}_{25}$ $\underbrace{14, 15, 15, 16\ 22}_{25}$.

25º 26º

50

Siempre que sea posible, se deben calcular Md y Mo porque hay que tener en cuenta que la media aritmética es sensible a las puntuaciones extremas, no así la mediana ni la moda. La mediana puede ser una medida preferible a la media cuando las puntuaciones extremas puedan distorsionar la verdadera tendencia central del grupo.

Podremos utilizar:

- En variables de intervalo o razón: media, mediana y moda.
- En variables de tipo ordinal: mediana y moda.
- En variables de tipo nominal: moda.

Además, como la media se ve arrastrada por puntuaciones extremas, la mediana puede ser una medida preferible a la media cuando existen puntuaciones extremas que distorsionan la verdadera tendencia central del grupo.

(3, 3, 4, 4, 4, 5, 5) vs. (3, 3, 4, 4, 4, 5, 10)

$\bar{X} = 4\,;\, Md = 4 \quad - \quad \bar{X} = 4{,}7\,;\, Md = 4$

$\boxed{1, 1}$, 3, 5, 8, 10, $\boxed{42}$

Mediana: 5

Moda: $\boxed{1}$

Promedio: $\frac{1+1+3+5+8+10+42}{7} = 7$

Valor aberrante: 42

3. MEDIDAS DE VARIACIÓN:

Las medidas de tendencia central por sí solas no nos proporcionan suficiente información como para hacernos una idea de las características del grupo. Nos podemos preguntar si el grupo es homogéneo en torno a la media aritmética (la mayoría de las puntuaciones están cerca de la media) o heterogéneo (muchas puntuaciones distan considerablemente de la media, tanto por arriba como por abajo). Por esta razón, el índice de tendencia central debe ir acompañado por un índice de dispersión o variabilidad que indique en qué medida las puntuaciones de los sujetos se dispersan o varían en torno a la media aritmética.

La desviación media (DM):

Es una medida que implica la **media aritmética de las desviaciones de las puntuaciones directas respecto de la media aritmética**. Es decir, si la DM es alta significa que las puntuaciones se desvían, están alejadas de la media aritmética del grupo (grupo heterogéneo). Si la D.M. es pequeña significa que las puntuaciones del grupo están próximas a la media

$$D.M. = \frac{\text{Valor absoluto de la suma de las diferencias entre cada puntuación directa y la media}}{\text{Número total de sujetos}} \qquad D.M. = \frac{\sum |(X_i - \bar{X})|}{N}$$

aritmética (grupo homogéneo).

La desviación típica y la varianza:

Los índices de variabilidad más conocidos y utilizados son la **desviación típica (s)** y la **varianza (s^2)**.

$$s = \sqrt{\frac{\sum (Xi - \bar{x})^2}{N}}$$

$$s^2 = \frac{\sum (Xi - \bar{x})^2}{N}$$

El **numerador** expresa las distancias de cada puntuación directa a la media aritmética. Dichas distancias se elevan al cuadrado para evitar que el valor siempre fuera 0.

Al elaborar un informe descriptivo, lo usual es incluir la **desviación típica** como medida de dispersión. No siempre es fácil **interpretar una desviación típica,** decir si es grande o pequeña. En muchas ocasiones se utiliza para comparar la dispersión entre grupos distintos:

- Si tenemos un solo grupo, los valores que podemos tomar de referencia son la desviación **típica mínima** (siempre cero) y la **máxima** (puntuación mayor menos puntuación menor entre dos), aunque dichos valores son muy difíciles de obtener, especialmente con variables continuas y muestras grandes.
- Para obtener la **mínima desviación típica**, todos los sujetos deben obtener la misma puntuación.
- Para obtener la **desviación típica máxima**, la mitad de los sujetos deberían obtener la puntuación máxima de la escala y la otra mitad la puntuación mínima (así se conseguirían las distancias máximas respecto de la media y, en este sentido, la máxima heterogeneidad).
- Sin embargo, si bien esto es posible en **variables dicotómicas**, cuanto más grande es la amplitud de la variable, más improbable es que se dé esta situación. En consecuencia, tendremos que ver en estos casos cuánto se aleja de cero la desviación típica.

Hasta ahora nos hemos referido al cálculo de la desviación típica ***sesgada***, que se utiliza frecuentemente al trabajar con muestras. No obstante, es igualmente frecuente trabajar con la desviación típica ***insesgada***, que no es más que la estimación de la desviación típica de la población a la que pertenece la muestra. Por lógica, se entiende que la desviación típica en una población será más elevada que en una muestra (al haber más sujetos, será más probable encontrar mayores diferencias interindividuales), por lo que si disminuimos el denominador, el cociente será mayor. Para calcular la desviación típica insesgada, simplemente se le resta una unidad al denominador:

$$s = \sqrt{\frac{\sum (Xi - \bar{x})^2}{N-1}}$$

$$s^2 = \frac{\sum (Xi - \bar{x})^2}{N-1}$$

La amplitud o recorrido:

Como decíamos, existen otras medidas de dispersión. La más básica es la **amplitud o recorrido de una variable**. Se calcula como la diferencia entre la puntuación mayor y menor más uno:

$A = X_{i\,mayor} - X_{i\,menor} + 1$ $\qquad X_{max} = 10$

$A = 10 - 0 + 1 = 11$ $\qquad X_{min} = 0$

Esta medida se utiliza, sobre todo, para organizar los datos de una distribución de frecuencias en intervalos y realizar gráficos.

La amplitud se utiliza como media de dispersión únicamente cuando no es posible calcular otra o como complemento de la moda (cuando el nivel de medida de la variable es nominal). Al igual que sucedía con la moda como índice de tendencia central, es un índice muy simple, basado en tan sólo dos puntuaciones, lo que puede dar lugar a malas interpretaciones si no se valora con precaución, sobre todo cuando existen puntuaciones extremas.

Desviación semi-intercuartílica (Q):

Es una medida que indica **la dispersión en el 50% central de la distribución,** es adecuada cuando el **nivel de medida de la variable es ordinal** (en este caso será el complemento de la mediana). También es adecuado su uso cuando la existencia de puntuaciones extremas pueda distorsionar en exceso la desviación típica.

$$Q = \frac{Q_3 - Q_1}{2}$$

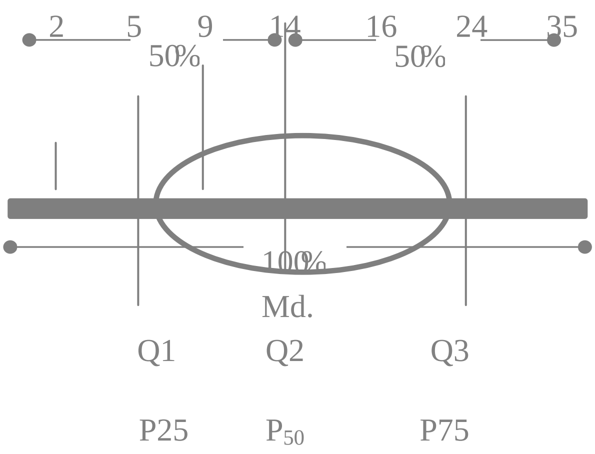
La **Q** prescinde del 25% inferior y del 25% superior de las puntuaciones, calculando la dispersión en el 50% central, entre los percentiles 25 y 75 o, lo que es lo mismo, entre los cuartiles 1 y 3.

	VAR00001	var
1	2	
2	5	
3	9	
4	14	
5	16	
6	24	
7	35	
8		

Estadísticos

VAR00001

N	Válidos	7
	Perdidos	0
Percentiles	25	5,00
Md	50 Q_2	14,00
	75	24,00



El coeficiente de variación

El **coeficiente de variación** (**V**) nos permite comparar la variabilidad entre variables que tienen distinta amplitud.

V se expresa en términos porcentuales, y se calcula así:

$$\boxed{V = \left(\frac{S}{\bar{x}}\right) 100}$$

Dos desviaciones típicas procedentes de instrumentos con distinto recorrido o distinta escala de medida no son directamente comparables.

4. MEDIA Y DESVIACIÓN TÍPICA PARA VARIABLES DICOTÓMICAS:

Son variables que sólo pueden tomar **dos valores**. Si sus dos posibles valores, como suele hacerse, se codifican con ceros y unos (ceros para noes o respuestas incorrectas y unos para síes o respuestas correctas), entonces la media aritmética representa la proporción de unos. Por ejemplo, supongamos que hemos hecho la siguiente pregunta en una encuesta:

¿Te gustan las matemáticas? *Sí* □ NO □

Si tenemos una muestra de 50 sujetos, y 30 han contestado que *sí* (sabiendo que los síes se han codificado como unos y los noes como ceros), ¿cuál es la media?: Aplicando la fórmula, el numerador es la suma de todas las puntaciones directas, es decir, la suma de 30 unos y 20 ceros, y el denominador el número total de sujetos ($N = 50$). Por tanto, la media será 30/50 = 0,6. Esta media indica la proporción de sujetos que ha respondido con un 1 (SÍ) a esta pregunta y se representa por «***p***». En otras palabras, que el 60% de la muestra (0,6 multiplicado por 100) ha respondido que *sí*, que al 60% le gustan las matemáticas y al 40% no le gustan. Por otra parte, la proporción de sujetos que contesta NO, se corresponde con la proporción de ceros, que lógicamente será 1 — *p*, en este caso 1 — 0,6 = 0,4. Si el 60% contestó que sí, queda el 40% que contestó que no. La proporción de noes o errores se representa por ***q***, de modo que:

$$\boxed{p + q = 1}$$

Por otra parte la **varianza** es muy fácil de calcular.   Y por lo tanto, la **desviación típica** es:

$$\boxed{S^2 = p \cdot q}$$

$$\boxed{S = \sqrt{p \cdot q}}$$

La **media aritmética representa la proporción de unos**.

p = nº de "1" / nº total de respuestas **q = nº de "0" / nº total de respuestas**

Media **p + q = 1**

Ítem
0
0
1
1
1

$\bar{X} = \frac{\sum X}{N} = \frac{3}{5} = 0{,}6 = p \rightarrow 60\%$ Contestaron correctamente.

$p + q = 1 \rightarrow 100\%$

$s = \sqrt{p \cdot q} = \sqrt{0{,}24} = 0{,}48$

Cuando:

$p = 1$ $q = 0$ o $p = 0$ $q = 1$ $\longrightarrow$ $s = 0$ $s^2 = p \cdot q = 0$

$p = 0{,}5$ $q = 0{,}5$ $\longrightarrow$ $s^2 = 0{,}25\ (max)$ $s = \sqrt{0{,}25} = 0{,}5\ (max)$

Las medias con variables dicotómicas son muy usadas en pruebas objetivas con preguntas de acierto/error. En este caso, la media nos dirá la proporción de sujetos que ha contestado correctamente a la pregunta, que viene a denominarse **"índice de dificultad del ítem"**.

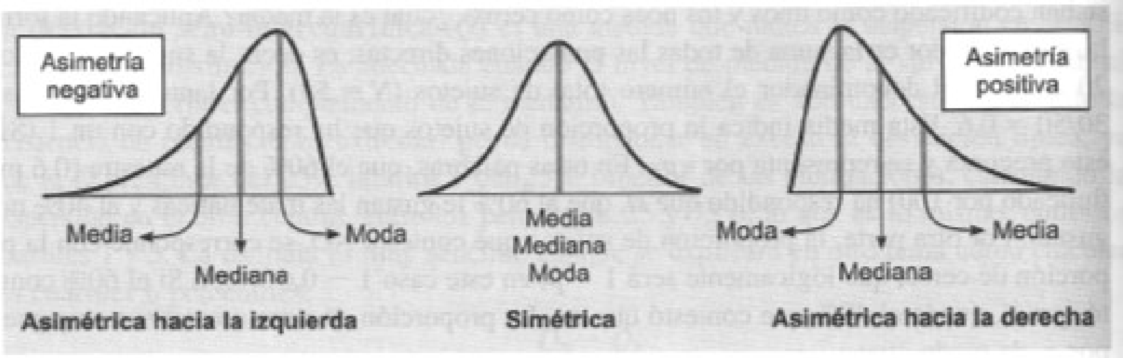
La desviación típica variará entre 0 (ausencia de variabilidad) y 0,5 (máxima variabilidad, es decir, cuando la mitad contesta sí y la otra mitad contesta que no).

5. ASIMETRÍA Y APUNTAMIENTO: RELACIÓN CON LA CURVA NORMAL.

La asimetría y el apuntamiento son dos características relativas a la forma gráfica de la distribución de frecuencias. Muchas variables educativas, psicológicas y biológicas se distribuyen según la llamada **distribución normal = curva normal = campana de Gauss** (rendimiento académico, inteligencia, aptitud, motivación, peso, altura).

La distribución viene a indicar cómo la mayoría de las personas nos encontramos en torno a los valores medios de la distribución y según nos alejamos hacia valores extremos el número de sujetos disminuye progresivamente.

Las representaciones gráficas de las variables, sobre todo cuando trabajamos con muestras grandes, tienden a ser curvas que, por su **grado de asimetría,** pueden asemejarse a una de estas tres.



La **asimetría negativa** indica que los sujetos tienden a agruparse en torno a las puntuaciones altas de la distribución.

La **curva normal** es una distribución teórica, simétrica y asintótica, donde la media, la mediana y la moda coinciden. (Tiene índice de asimetría igual a cero y apuntamiento igual a cero).

Como se ve, la **asimetría positiva** indica que la mayoría de los sujetos tiende a concentrarse en la parte baja de las puntuaciones de la distribución de frecuencias.

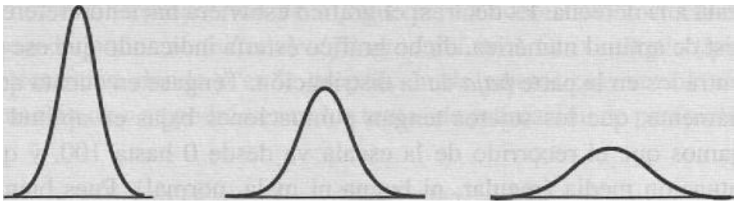
El hablar de asimetría positiva o negativa se debe simplemente al signo del cálculo del

$$As = \frac{\bar{X} - Mo}{s}$$

índice.

El **apuntamiento** o la **curtosis** indica el grado en el que la distribución es más o menos «picuda», es decir, el grado en el que la distribución resulta más abierta o dispersa respecto a la media (las puntuaciones están poco concentradas respecto a la media).

- **Mesocúrtica**: la curva normal. Curtosis = 0
- **Leptocúrtica**: más apuntada. Concentración de las puntuaciones alrededor de la media. Curtosis > 0
- **Platicúrtica**: más achatada. Dispersión de las puntuaciones alrededor de la media. Curtosis < 0



Leptocürtica Mesocúrtica Platicúrtica

Para el cálculo de la curtosis, puede utilizarse la siguiente fórmula:

$$g_2 = \left(\frac{1}{N} \cdot \frac{\sum fi\,(Xi - \bar{x})^4}{\sigma^4}\right) - 3$$

6. REPRESENTACIONES GRÁFICAS:

Gráfico de sectores o ciclograma o gráfico de tarta: variables con nivel de medida nominal. Dicotómicas. Cada porción representa la proporcional de cada uno de los niveles de la variable.

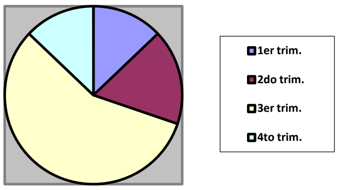
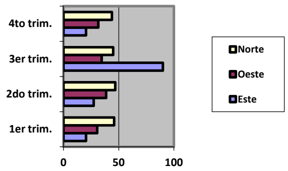
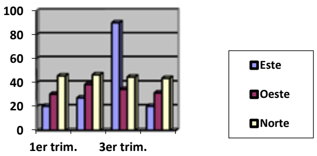


Gráfico de barras: variables con nivel de medida ordinal, aunque se puede utilizar cuando el nivel de medida es nominal y para realizar comparaciones de variables clasificatorias o categóricas.





Histograma: es muy parecido al de barras, pero se utiliza para variables cuantitativas continuas, con nivel de medida de intervalo o de razón. **Agrupa las puntuaciones en intervalos.**

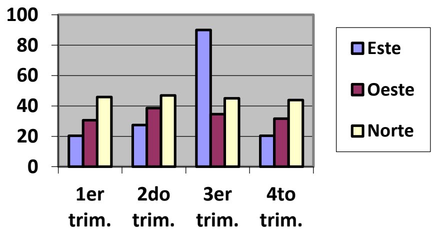
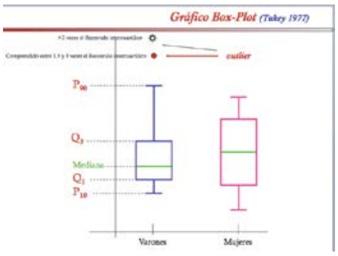


Gráfico de caja, caja y patillas o caja y bigotes: para hacerse una idea rápida de la distribución de las puntuaciones en la zona central (Comprende desde el Q_1 al Q_3, o lo que es lo mismo, desde el percentil 25 al 75) Es útil para comparar visualmente grupos medidos en distintas variables. No son adecuados para un nivel nominal.



$$Md \pm 1{,}5 \cdot (Q_3 - Q_1)$$

Gráfico de tallo y hojas: combina la representación numérica y gráfica:

Tallo	Hoja
3	
3	5 6 8
4	0 2 2 4
4	5 5 7 8 9
5	0 0 0
5	

TEMA 6: Medidas individuales.

1. INTRODUCCIÓN:

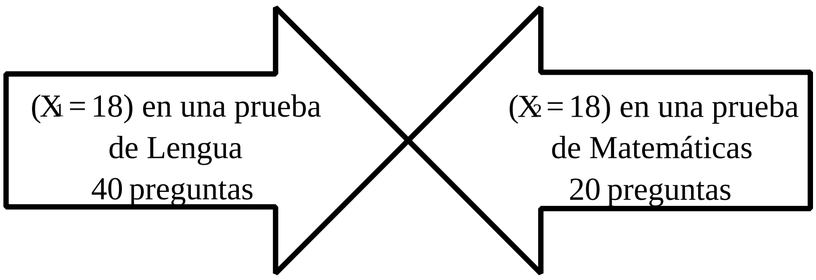
Habitualmente la estadística se ocupa de estudiar conjuntos de datos más que los datos particulares, pero en el campo educativo, las puntuaciones individuales de sujetos concretos son de vital importancia puesto que la persona es el objeto fundamental de la educación.

“El conocimiento de las puntuaciones individuales de cada persona y su correcta interpretación resultan esenciales para la intervención educativa”.

En este tema hablaremos de las **puntuaciones individuales, de su interpretación y de la necesidad de su transformación**, para su mejor comprensión.

2. PUNTUACIONES DIRECTAS: Problemas de interpretación y transformaciones permisibles:

Una **puntuación directa** (X_1) es la puntuación que obtiene un sujeto al realizar una prueba o aplicarle un instrumento de medida.



¿Son comparables estas 2 puntuaciones directamente? NO
Necesitamos transformar X_1 y X_2 para poderlas comparar correctamente.

Puntuaciones proporcionales y porcentuales:
Una forma rápida y sencilla de comparar X_i es convertirla en una **proporción de respuestas correctas** (n.° de respuestas correctas/nº total de preguntas) o un **porcentaje** de respuestas correctas (multiplicando la proporción por 100).

Lengua	Matemáticas
X1 = 18	X2 = 18
nº = 40	nº = 20
18/40 = 0,45	18/20 = 0,9
45%	90%

Sin embargo, no siempre es posible ni pertinente hacer esta transformación, no sirven para cuestionarios (de opinión, estimación… ya que no hay respuestas correctas e incorrectas) ni tienen en cuenta el rendimiento del resto del grupo.

Si tenemos una escala de actitud hacia el estudio con un recorrido de O a 50 puntos (cuyos ítems indican el grado de acuerdo o desacuerdo con distintas afirmaciones), una X_i de 30 puntos manifiesta cierta actitud positiva hacia el estudio, pero no es transformable en un porcentaje de respuestas correctas.

Un sujeto que obtiene un 45% de respuestas correctas en una prueba objetiva se corresponde con un resultado medio-bajo pero dicha afirmación es discutible. Podría ocurrir que el test hubiese sido tan difícil, que el sujeto que obtuvo un 45% de aciertos fue el mejor de su grupo. Entonces, quizás podríamos interpretar que esta puntuación no fue tan mala.

Las puntuaciones diferenciales

Una puntuación diferencial (x) es una puntuación individual relativa a la media aritmética del grupo de referencia. Por tanto, para calcular una puntuación diferencial es necesario haber aplicado un instrumento de medida a un grupo de sujetos. Para calcularla, simplemente se le resta a la puntuación directa del sujeto la media aritmética del grupo al que pertenece.

$$x = X_i - \bar{X}$$

Siguiendo nuestro ejemplo…

Lengua	Matemáticas
$\bar{X} = 13$	$\bar{X} = 19$
$x = 18 - 13 = +5$	$x = 18 - 19 = -1$

Ahora bien, una puntuación diferencial sólo nos permite saber si una puntuación está por encima o por debajo de la media aritmética. Es evidente que no es lo mismo separarse dos puntos de la media cuando tenemos un recorrido de 5 puntos que cuando el recorrido es de 100. Necesitamos, por tanto, una puntuación que permita situar a un sujeto con respecto a su grupo de referencia y hacer comparaciones independientemente de la amplitud del instrumento del que procedan las puntuaciones directas que deseamos comparar.

Las puntuaciones típicas:

Una **puntuación típica (z)** indica el número de desviaciones típicas que se desvía una puntuación directa de la media aritmética. Las dos propiedades más importantes de las puntuaciones típicas son que **la media de dichas puntuaciones es igual a 0 y la desviación típica igual a 1**.

$$z = \frac{X_i - \bar{X}}{s}$$

Ejemplo: Un grupo de sujetos tiene una media en CI de 100 puntos y una desviación típica s = 10

$$X = 110 \quad z = \frac{110 - 100}{10} = 1 \; \textit{Se ha apartado una desviación típica por encima de la media}$$

$X = 120 \quad z = \frac{120 - 100}{10} = 2$

En nuestro ejemplo:

Lengua

$\bar{X} = 23$

$s = 5$

$z = \frac{18 - 23}{5} = -1$

Matemáticas

$\bar{X} = 9$

$s = 3$

$z = \frac{18 - 9}{3} = 3$

Las puntuaciones típicas son muy utilizadas porque **nos permiten comparar cualquier puntuación entre sí**, independientemente del instrumento de medida o de la amplitud de la escala utilizada.

A toda puntuación directa superior a la media le corresponderá una puntuación típica positiva, y si es inferior a la media, negativa. **La media aritmética siempre coincide con una z = 0**.

Debe entenderse que una **puntuación típica (z)**, como hemos dicho, es una puntuación que depende tanto de la media como de la variabilidad del grupo. En consecuencia, las puntuaciones típicas variarán dependiendo de la homogeneidad o heterogeneidad del grupo. Si un grupo es muy homogéneo, su desviación típica será pequeña y, por tanto, a una puntuación que no se aleje mucho de la media le puede corresponder una puntuación típica mucho mayor que la que le correspondería si el grupo fuese heterogéneo.

Por ejemplo: examen de estadística:

Grupo Homogéneo

$\bar{X} = 5,5$

$s = 0,4$

$X_1 = 7$

$z = 3,75$

Grupo Heterogéneo

$s = 2,5$

$X_1 = 7$

$z = 0,6$

Puntuaciones tipificadas o escalas derivadas:

Las puntuaciones tipificadas consisten en una simple transformación de las puntuaciones típicas, creadas con el ánimo de evitar las puntuaciones decimales y las negativas. La transformación se reduce a multiplicar por una constante a la puntuación típica, valor que se convertirá en la nueva desviación típica y sumarle otra constante, valor que se convertirá en la nueva media de las puntuaciones tipificadas:

T=a.z+b Donde: **b = media $\bar{X}$** **a = desviación típica s** **z = puntuación típica**

Entre las puntuaciones tipificadas más usadas, se encuentran las siguientes:

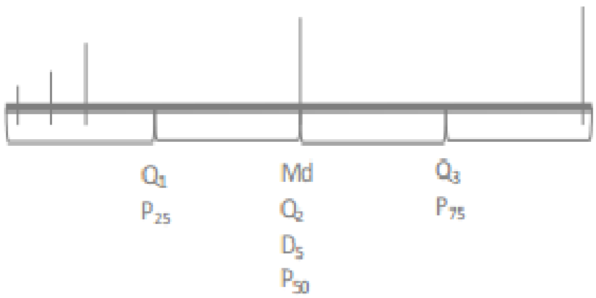
T*= *10z* + 50** ***S **= 2z + 5**

Entonces, un sujeto con una puntuación directa igual a la media tendrá una puntuación ***T*** **50 y una *S* = 5**.

Las puntuaciones cuantiles:

Un cuantil indica el porcentaje de sujetos que deja por debajo de sí una puntuación determinada. Las puntuaciones cuantiles más utilizadas son los ***percentiles***, que dividen una distribución de frecuencias en 100 partes. (Por ejemplo, P_{85} = puntuación directa que deja por debajo de sí al 85% de los sujetos de su grupo).

Otros cuantiles utilizados son los **deciles** (D, diez divisiones) y los **cuartiles** (Q, cuatro divisiones).



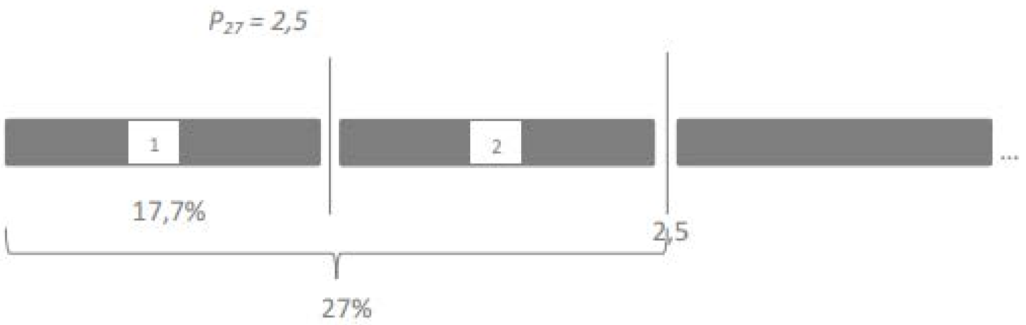
$P_{75} = Q_3$

$P_{50} = D_5 = Q_2 = Md$

Los percentiles son utilizados para construir los **baremos de los tests estandarizados**.

Ejemplo:

Puntuación directa	Frecuencia absoluta	Frecuencia relativa × 100	Frecuencia acumulada	Porcentaje acumulado	Percentil P_c
7	266	26,3	1.012	100,0	100
6	131	12,9	746	73,7	74
5	108	10,7	615	60,8	61
4	128	12,6	507	50,1	50
3	107	10,6	379	37,5	38
2	93	9,2	272	26,9	27
1	179	17,7	179	17,7	18
Total	1.012	100,0			

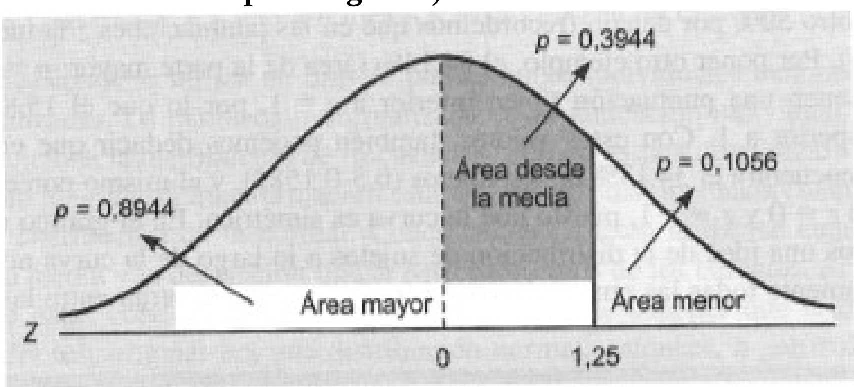


$$C_m = L_{inf} + \frac{\left(\frac{C}{100}n\right) - f_{a(I-1)}}{f_i} a_1, \text{ donde } a \text{ es el valor del intervalo.} \longrightarrow P_{25} = 2,3$$

3. LAS PUNTUACIONES INDIVIDUALES EN LA CURVA NORMAL:

En casi todos los textos de Estadística se pueden encontrar las ***tablas de la curva normal*** con las probabilidades asociadas a cada puntuación típica.

La **interpretación de las tablas** es muy sencilla. Tenemos que pensar que una puntuación típica divide la curva normal en dos partes: **una grande y otra pequeña**
(**excepto z = 0 que la divide en dos partes iguales**)



En este caso, hemos utilizado la **puntuación típica 1,25**. Si trazamos una línea vertical a partir de dicha **z**, el área de la curva normal queda dividida en dos partes. El área mayor, lógicamente, es el área más grande que queda al realizar dicha división, es decir, desde la línea gruesa hasta - ∞. Por tanto, una puntuación típica igual a 1,25 tiene una probabilidad acumulada de aparición de *p* = 0,8944 (ver tablas de la curva normal), esto es, el 89,44% de los sujetos obtienen una puntuación z 1,25. Consecuentemente, en este caso, el área menor indica la probabilidad de obtener una puntuación z 1,25 (*p* = 0,1056).

Las puntuaciones individuales normalizadas:

La puntuación normalizada es la puntuación individual que le corresponde a un sujeto si la distribución de frecuencias original es una distribución normal. Concretamente suelen utilizarse cuando en un test conocemos la media y la desviación típica, pero carecemos de los baremos.

Ejemplo: **¿qué puntuación corresponde a una z?** Sabemos que se trata del percentil P10, con una $\bar{X}$= 100 y S = 15. Deja por debajo al 10 % de los sujetos, es decir, una proporción del 0,01. Buscamos en el formulario el valor de p = 0,01 en la columna C y encontramos que corresponde a una z = -2,33 (es negativa porque estamos trabajando en la parte izquierda de la curva, por debajo de la $\bar{X}$). Ahora aplicamos la fórmula y despejamos:

$$-2.33 = \frac{Xi - 100}{15} \Rightarrow Xi = 65{,}05 \rightarrow P_{10}$$

Ejemplo: ¿a cuántos sujetos superaría un alumno que obtiene una Xi = 65, si la muestra es de 150 sujetos?

$$z = \frac{65 - 100}{15} = -2{,}33$$

Entrando en tablas con 2,33 y la columna C →p = 0,01
entonces →N. p = 150.0,01 = 1,5 →2 sujetos

TEMA 7: Relación entre variables. Las correlaciones y la regresión:

1. INTRODUCCIÓN:

Al hacer investigación en educación nos interesará en muchas ocasiones conocer la posible relación entre 2 o más variables.

En este capítulo estudiaremos:
- Las relaciones entre las variables que intervienen en el proceso educativo: correlación.
- Las posibilidades y limitaciones de la predicción de puntuaciones en una variable, conociendo los valores de otra: regresión.

Estudiaremos los coeficientes de correlación más importantes en el campo educativo:
- El coeficiente de **correlación de Pearson** (**r**).
- El coeficiente de **correlación ordinal de Spearman** ($\mathbf{r_s}$).
- El coeficiente de **contingencia** (**C**).
- El coeficiente de **correlación biserial-puntual** ($\mathbf{r_{bp}}$).

2. EL CONCEPTO DE CORRELACIÓN:

La correlación nos, indica la tendencia de dos o más conjuntos de datos a variar de forma conjunta. Para cuantificar la intensidad de la correlación contamos con el **coeficiente de correlación** que nos mide **el índice de covariación o variación conjunta** de dos, o más, series de datos.

Por ejemplo: podemos estar interesados en conocer qué variables pueden influir sobre el rendimiento académico y cuál en su peso relativo en la predicción, como pueden ser: el tiempo de estudio, las técnicas de trabajo, los recursos materiales disponibles, la motivación, etc.

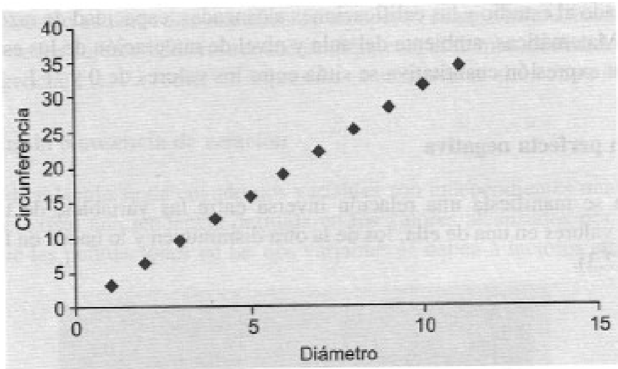
Las situaciones que se pueden analizar pueden ser diversas:
- Estudiar la relación entre **dos o más variables** medidas en un mismo grupo de sujetos.
- Con dos o más grupos de individuos comprobar el **grado de relación** entre dichas muestras de los grupos en una sola variable.
- También se puede dar el caso de **una misma variable medida en dos momentos distintos** en una misma muestra.

*Una relación simple entre dos series de datos, se identificará con la *correlación entre dos variables.*

En el análisis de la correlación entre dos variables, teniendo en cuenta la intensidad y el sentido de la relación, se presentan **diferentes posibilidades** que se expresan mediante un **diagrama de dispersión**:

A) Relación perfecta positiva

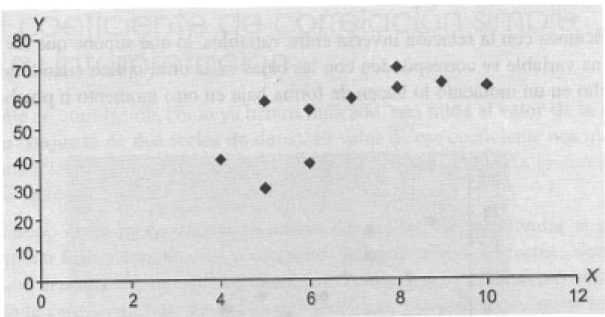
Se manifiesta en aquellos supuestos en que al aumentar los valores de una de las variables los valores de la otra lo hacen siempre en la misma proporción, gráficamente se expresa así:



Su expresión cuantitativa sería +1.

b) Relación imperfecta positiva

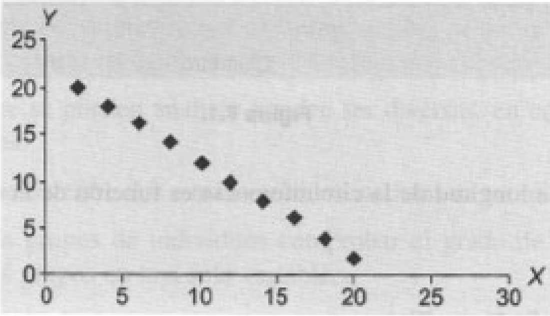
Se la conoce como relación directa de variables, es decir, que a valores elevados en una variable le corresponden valores altos en la otra; de tal forma que si se mide un grupo en dos variables distintas el sujeto que puntúa alto lo hace en las dos variables y a la inversa.



Su expresión cuantitativa se sitúa entre los valores 0 y +1.

c) Relación perfecta negativa

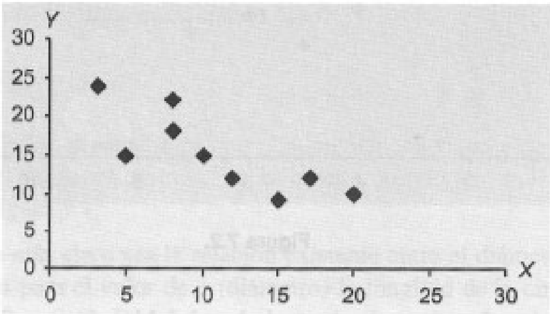
En este caso se manifiesta una relación inversa entre las variables, de tal forma que al aumentar los valores en una de ella, los de la otra disminuyen y lo hacen en la misma proporción.



Su expresión cuantitativa es -1.

d) Relación imperfecta negativa

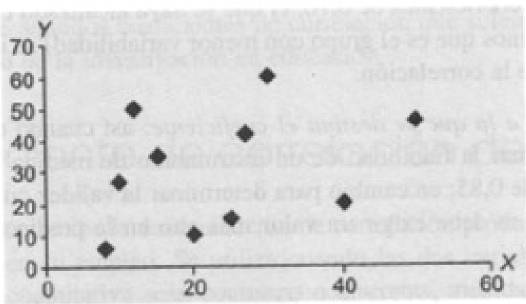
La identificamos con la relación inversa entre variables, lo que supone que las puntuaciones altas en una variable se corresponden con las bajas en la otra; o bien cuando los sujetos que puntúan alto en un momento lo hacen de forma baja en otro momento o prueba distinta.



Su forma de expresión cuantitativa se sitúa entre los valores 0 y -1.

e) Relación nula o ausencia de relación

Esta ausencia de relación se da cuando dos variables son independientes una de la otra (ello implica que no existe una tendencia definida en los valores alcanzados por los sujetos).



La forma de expresión cuantitativa sería 0 (ausencia de relación).

3. EL COEFICIENTE DE CORRELACIÓN SIMPLE Y SU INTERPRETACIÓN:

El coeficiente de correlación, como ya hemos indicado, **nos mide el valor de la covariación o variación conjunta de dos series de datos.** El valor de ese coeficiente nos marca la existencia de una relación **directa de variables** (valores positivos) o **inversa** (valores negativos), es decir, los valores cuantitativos del coeficiente se sitúan **entre + 1 y - 1**.

Aunque existen diversas directrices en su interpretación, en general, la mayor parte de los investigadores tienden a identificar tres aspectos:

1. ***El tipo de variables que se relacionan:*** cuando se da una similitud entre el valor del coeficiente encontrado en el estudio empírico y el encontrado en el mismo grupo en trabajos previos, también cuando se han encontrado unos valores elevados en anteriores estudios y se repite en la actualidad la misma tendencia en los datos.
2. ***La variabilidad del grupo:*** cuanto mayor es la variabilidad del grupo mayor será el valor del coeficiente de correlación. Así cuando tenemos dos coeficientes de 0,70; el que se haya alcanzado en el grupo más homogéneo (el de menor variabilidad) se identifica con una mayor intensidad de la correlación.
3. ***La finalidad a la que se destina el coeficiente:*** así cuando el coeficiente se emplea para determinar la fiabilidad de un instrumento de medida debe tener unos valores por encima de 0,85; en cambio para determinar la validez con 0,60 puede ser aceptable, también se debe exigir un valor más alto en la predicción que en la prueba de hipótesis.

Aunque no existe acuerdo entre los diferentes autores consultados se pueden establecer unos intervalos para categorizar y valorar los coeficientes de correlación. En la mayoría de las ocasiones se suelen aceptar las interpretaciones que *se* muestran en la tabla siguiente:

Valor del coeficiente	**Interpretación**
Entre 0,00 y + o — 0,20	Correlación muy baja, indiferente, despreciable
Entre 0,21 y + o — 0,40	Correlación baja
Entre 0,41 y + o — 0,70	Correlación media, marcada, notable
Entre 0,71 y + o — 0,90	Correlación alta, elevada, fuerte
Entre 0,91 y + o — 1	Correlación muy alta, muy elevada

*Coeficiente de Determinación (d) = Coeficiente de correlación (r) al cuadrado por 100.

$$\mathbf{d = r^2 . 100}$$

(Se interpreta como el porcentaje de la varianza de una variable explicada por la otra).

Ejemplo: (Pág. 134)

$\mathbf{r_{xy} = 0,70}$ $\mathbf{d = (0,7)^2 . 100 = 49\%}$

La interpretación de los coeficientes de correlación se suele completar con su significación estadística. Se trata de poder afirmar que la correlación entre 2 variables es real y no se puede explicar por efecto del azar.

4. TIPOS DE COEFICIENTE DE CORRELACIÓN:

- El coeficiente de **correlación de Pearson** (**r**): las dos variables son cuantitativas y medidas a nivel de intervalo y además se atribuyen normalmente.
- El coeficiente de **correlación ordinal de Spearman** ($\mathbf{r_s}$): las variables son cuantitativas pero sólo puede garantizarse un nivel de medida ordinal.
- El coeficiente de **contingencia** (**C**): para hallar el grado de asociación entre 2 variables nominales (categóricas).
- El coeficiente de **correlación biserial-puntual** ($\mathbf{r_{bp}}$): entre una **variable cuantitativa** (continua o discreta) y una auténticamente **dicotómica**.

❖ *El coeficiente de correlación de Pearson (r):*

Este coeficiente también recibe la denominación de correlación producto-momento y es uno de los más exigentes en su empleo. Se utiliza cuando las dos variables que se relacionan tienen una naturaleza cuantitativa (números).

Para llevar a cabo su cálculo empírico existen diferentes fórmulas, según el tipo de puntuaciones con el que se trabaje: puntuaciones directas, diferenciales y típicas.

Puntuaciones directas:

$$r_{xy} = \frac{N.\sum XY - \sum X \sum Y}{\sqrt{N.\sum X^2 - (\sum X)^2}.\sqrt{N.\sum Y^2 - (\sum Y)^2}}$$

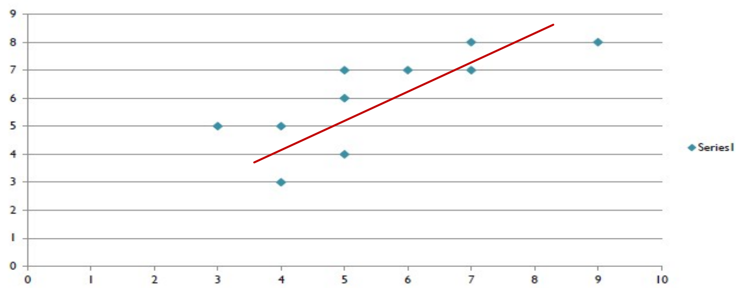
Puntuaciones diferenciales:

$$r_{xy} = \frac{\sum x.y}{\sqrt{\sum x^2}.\sqrt{y^2}}$$

Ejemplo: calificaciones en matemáticas (variable X) y en física (variable Y) de 10 sujetos.

D1 =COEF.DE.CORREL(A1:A10;B1:B10)

	A	B	C	D	E	F
1	7	8		0,77145428		
2	5	4				
3	9	8				
4	7	7				
5	4	3				
6	6	7				
7	3	5				
8	5	6				
9	4	5				
10	5	7				
11						



(Pág. 136).

Suj.	X	Y	X^2	Y^2	X*Y	x	x^2	y	y^2	x*y
1	7	8	47	64	56	1,5	2,25	2	4	3
2	5	4	25	16	20	…	...	…	…	…
…	…	…	…	…	…	…	…	…	…	…
Total	55 $\bar{X}=5{,}5$	60 $\bar{Y}=6$	331	386	351		28,5		26	21

$$r_{xy} = \frac{\sum x \cdot y}{\sqrt{\sum x^2 \cdot \sum y^2}} = \frac{21}{\sqrt{28{,}5 \cdot 26}} = 0{,}77$$

❖ ***El coeficiente de correlación ordinal de Spearman*** ($\mathbf{r_s}$):

En muchas de las variables que utilizamos con frecuencia en el campo educativo no es posible alcanzar unos niveles precisos de medida. Al no disponer de información precisa se suelen emplear los puestos que ocupan las puntuaciones en una serie ordenada: Nos interesa conocer la posición, orden o rango que ocupan en una serie ordenada de valores (1º, 2º, 3º, 4º…)

Ante esos casos en que las dos variables tienen la misma naturaleza, y además con esos datos, solamente podemos garantizar que alcanza un nivel de medida ordinal, importa el lugar que ocupa no la puntuación obtenida, debemos recurrir al coeficiente de correlación de Spearman.

Para proceder a la *transformación de las puntuaciones obtenidas de la aplicación directa del instrumento de recogida de datos en* ***rangos*** se suele comenzar asignando el rango o posición 1 a la puntuación más alta, la siguiente tendrá el rango 2 y así sucesivamente de tal forma que el último rango que se asigne debe coincidir con el número de sujetos de la muestra.

$$r_s = 1 - \frac{6 \cdot \sum D^2}{n(n^2 - 1)}$$

"n" nos indica el número de sujetos o de pares de puntuaciones y ***D*** es la diferencia, de rangos o posiciones que ocupa un mismo sujeto en dos variables distintas.

40	1
38	3

- En aquellos casos en que existan varias puntuaciones directas coincidentes, la asignación del rango se realiza calculando el promedio de las posiciones que ocupan.
- El criterio de asignación de rangos debe ser el mismo para ambas variables.

38	3
38	3
5	5

Ejemplo: Correlación entre “comprensión oral” y “comprensión escrita” para una muestra de 15 sujetos.

D1 =COEF.DE.CORREL(A1:A15;B1:B15)

	A	B	C	D	E	F
1	45	38		0,915261867		
2	39	35				
3	50	45				
4	33	30				
5	47	50				
6	25	28				
7	40	35				
8	30	35				
9	20	15				
10	40	40				
11	25	32				
12	15	15				
13	23	14				
14	35	30				
15	17	12				
16						
17						
18						

Correlaciones

		VAR00003	VAR00004
VAR00003	Correlación de Pearson	1	,915**
	Sig. (bilateral)		,000
	N	15	15
VAR00004	Correlación de Pearson	,915**	1
	Sig. (bilateral)	,000	
	N	15	15

**. La correlación es significativa al nivel 0,01 (bilateral).

Correlaciones

			VAR00003	VAR00004
Rho de Spearman	VAR00003	Coeficiente de correlación	1,000	,923**
		Sig. (bilateral)	.	,000
		N	15	15
	VAR00004	Coeficiente de correlación	,923**	1,000
		Sig. (bilateral)	,000	.
		N	15	15

**. La correlación es significativa al nivel 0,01 (bilateral).

En Excel:

X Y

50	1	45	2
47	2	50	1
45	3	38	4
40	4,5	40	3
40	4,5	35	6
39	6	35	6
35	7	30	9,5
33	8	30	9,5
30	9	35	6
25	10,5	32	8
25	10,5	28	11
23	12	14	14
20	13	15	12,5
17	14	12	15
15	15	15	12,5

-1

+1

-1

+1,5

… …

$$r_s = 1 - \frac{6\cdot\sum D^2}{n(n^2-1)} =$$

$$= 1 - \frac{6\cdot 43}{15\cdot(15^2-1)} = 0,923$$

❖ ***El coeficiente de contingencia (C)***:

En el caso de **variables nominales o atributos**, se suele hablar de **"grado de asociación"** (no de "grado de correlación").

Se utiliza aquellos supuestos en que se recogen datos de **variables clasificadas en categorías**.

El coeficiente de contingencia (C).

$$C = \sqrt{\frac{\chi^2}{n + \chi^2}} \qquad \chi^2 = \sum_{g=1}^{G}\sum_{c=1}^{C}\frac{(f_o - f_e)^2}{f_e} \qquad f_e = \frac{f_f \cdot f_c}{f_t}$$

f_o Frecuencias observada

f_e Frecuencias esperadas

Así, las **tablas de contingencia** reflejan una asignación de sujetos a grupos y categorías en cada una de las variables.

Ejemplo: Grado de asociación entre tipo de estudios y nivel socioeconómico para 320 sujetos.

	CC. SOCIALES Y JURÍDICAS	CIENCIAS	HUMANIDADES	
BAJO	f_o=20 $f_e = \frac{120\ 100}{320} = 37,5$	f_o=40 f_e=37,5	f_o=60 f_e=45	120
MEDIO	40	40	40	120
ALTO	40	20	20	80
	100	100	120	320

$$\chi^2 = \frac{(20 - 37,5)^2}{37,5} + \frac{(40 - 37,5)^2}{37,5} + \cdots$$

$$C_{máx.} = \sqrt{\frac{c - 1}{c}} = \sqrt{\frac{3 - 1}{3}} = 0,81$$

$$\chi^2 = 27,56 \qquad C = \sqrt{\frac{27,56}{320 + 27,56}} = 0,28$$

$$0,28 \leftrightarrow 0,81$$
$$x \leftrightarrow 1$$

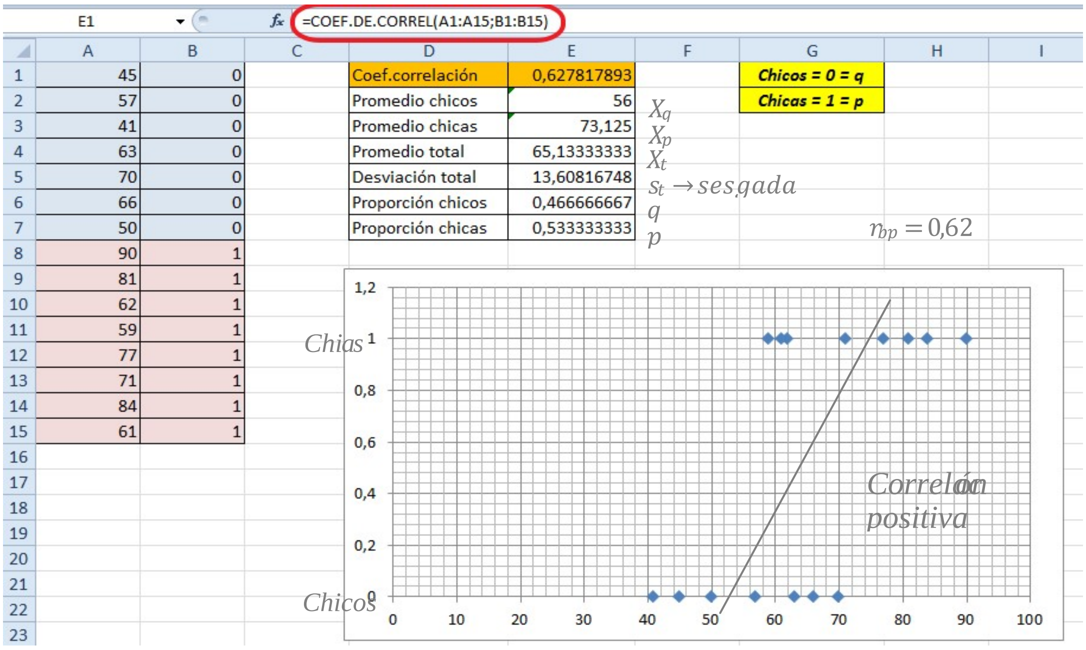
$$x = 0,34$$

❖ **El coeficiente de correlación biserial-puntual (r_{bp}):**

Cuando buscamos el grado de relación que se manifiesta entre una variable cuantitativa, continua o discreta, y otra auténticamente dicotómica, debemos recurrir al coeficiente biserial-puntual.

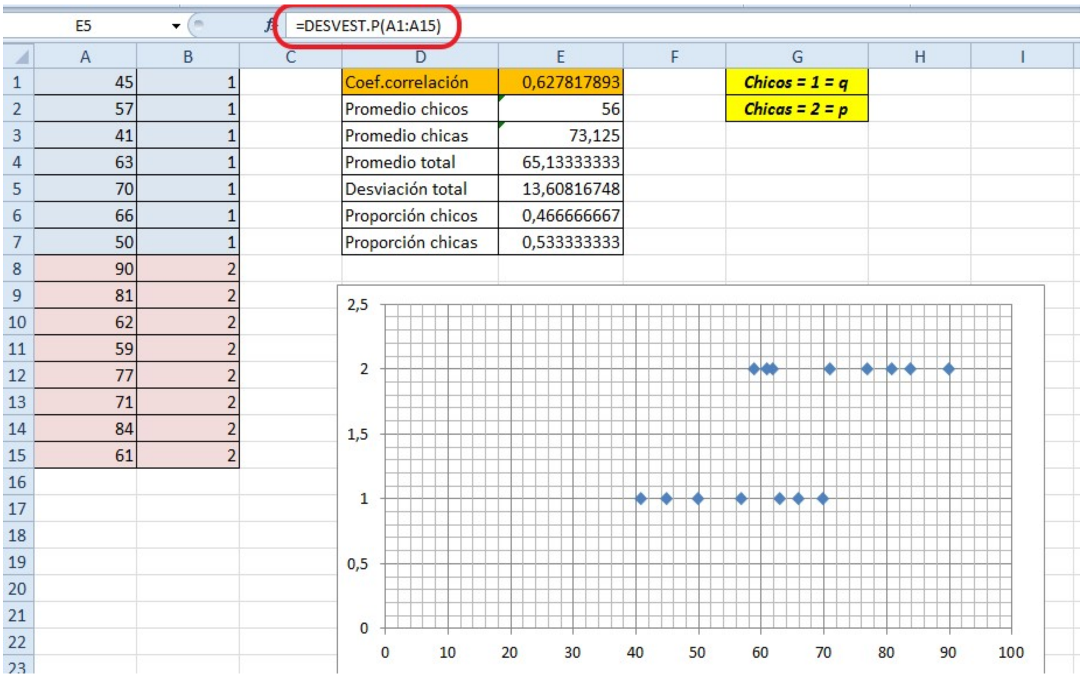
$$r_{bp} = \frac{\bar{X}_p - \bar{X}_t}{s_t} \cdot \sqrt{\frac{p}{q}} \qquad\qquad r_{bp} = \frac{\bar{X}_p - \bar{X}_q}{s_t} \cdot \sqrt{p \cdot q}$$

Ejemplo: nivel de relación entre el rendimiento matemático y el sexo.



=COEF.DE.CORREL(A1:A15;B1:B15)

	A	B	D	E	G
1	45	0	Coef.correlación	0,627817893	Chicos = 0 = q
2	57	0	Promedio chicos	56	Chicas = 1 = p
3	41	0	Promedio chicas	73,125	
4	63	0	Promedio total	65,13333333	
5	70	0	Desviación total	13,60816748	
6	66	0	Proporción chicos	0,466666667	
7	50	0	Proporción chicas	0,533333333	
8	90	1			
9	81	1			
10	62	1			
11	59	1			
12	77	1			
13	71	1			
14	84	1			
15	61	1			



=DESVEST.P(A1:A15)

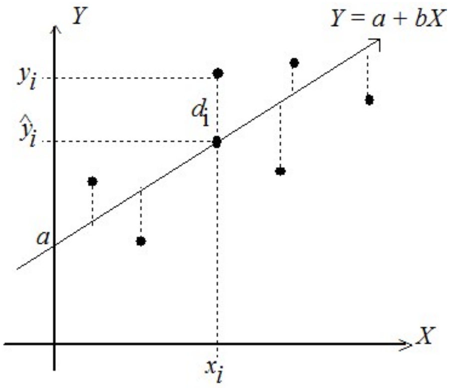
	A	B	D	E	G
1	45	1	Coef.correlación	0,627817893	Chicos = 1 = q
2	57	1	Promedio chicos	56	Chicas = 2 = p
3	41	1	Promedio chicas	73,125	
4	63	1	Promedio total	65,13333333	
5	70	1	Desviación total	13,60816748	
6	66	1	Proporción chicos	0,466666667	
7	50	1	Proporción chicas	0,533333333	
8	90	2			
9	81	2			
10	62	2			
11	59	2			
12	77	2			
13	71	2			
14	84	2			
15	61	2			

❖ **La regresión lineal simple:**

Elevando al cuadrado el coeficiente de correlación obtenemos el coeficiente de determinación, que permite conocer el porcentaje de varianza compartida.

Esto hace posible que se puedan estimar los valores de una variable conociendo los valores en la otra: **regresión lineal simple**.

La recta de regresión (Y = a + bX) nos permite llevar a cabo la predicción o estimación de los valores en una variable (variable criterio) a partir del conocimiento de los valores en la otra variable (variable predictora), con la que mantiene una alta correlación.



TEMA 8: Aplicaciones de la correlación: fiabilidad y validez de las medidas.

1. INTRODUCCIÓN:

En este capítulo veremos las características técnicas de los instrumentos de medida. Es fundamental que los instrumentos sean precisos en sus apreciaciones (fiabilidad) y sirva para lo que realmente se diseñaron (validez).

En este tema se abordan 3 cuestiones:

- Fiabilidad de los instrumentos.
- Validez de los instrumentos (especialmente la validez predictiva).
- Análisis de elementos/ítems de una prueba:
    - Índice de Dificultad.
    - Índice de Homogeneidad.
    - Índice de Validez.

¡Ojo! No se exigirá en el examen el desarrollo paso a paso de los procedimientos de Rulon, KuderRichardson y Cronbach. Los epígrafes 8.2.2, 8.3.6, 8.3.7 y 8.3.8 quedan eliminados como material obligatorio de estudio.

2. EL ESTUDIO DE LA FIABILIDAD:

Entendemos que la **fiabilidad de las medidas se identifica con la *precisión,* de tal forma que decimos que un instrumento es fiable cuando mide algo con precisión, independiente de lo que se esté midiendo.** Aceptar esta definición lleva consigo contemplar la posibilidad de error en el momento de medir. Por ello, una puntuación observada en un sujeto podrá descomponerse en dos partes: la que se corresponde con la puntuación verdadera y el posible error que se comete.

Se acepta que esos *errores* los podemos dividir en ***sistemáticos y aleatorios***. Los primeros vienen asociados a las características internas del instrumento que afectaría a todo lo que se mida con él, mientras que el error de tipo aleatorio, es debido a aquellas variables cuyos efectos nos resultan desconocidos, se rigen por la ley del azar: la magnitud de los errores viene condicionada por el *margen de error o nivel de confianza* con que el investigador realiza sus apreciaciones. Naturalmente, *cuando menor sea el error más fiable es el instrumento*.

También podemos expresar la fiabilidad como la **constancia** *en las puntuaciones de los sujetos o bien la* **concordancia** *entre varias mediciones sucesivas de una misma realidad;* en cada uno de los casos nos vamos a enfrentar a diferentes procedimientos de cálculo de la fiabilidad, aunque en todos ellos se procede a la obtención de dos conjuntos de datos a partir de los cuales se puede determinar el coeficiente de correlación que estime la fiabilidad del instrumento.

❖ Procedimientos para determinar la fiabilidad:

a) La fiabilidad como estabilidad:

También se la identifica como ***procedimiento de la repetición o del test-retest***, dado que se busca la correlación que existe entre las puntuaciones obtenidas por un mismo grupo de sujetos, debidamente identificados, en dos aplicaciones sucesivas de una misma prueba en dos momentos diferentes, entre ambas aplicaciones debe transcurrir un tiempo, que no sea muy corto ni muy largo.

Si transcurre poco tiempo los sujetos pueden recordar las respuestas anteriores. Si transcurre mucho pueden influir los efectos de la maduración, la práctica del aula. Se suelen fijar entre 20-25 días de la primera a la segunda aplicación.

Ejemplo: Prueba de 20 ítems. 2 aplicaciones sucesivas a 10 sujetos.

B14 =COEF.DE.CORREL(B2:B11;C2:C11)

	A	B	C	D	E	F	G
1		X	Y	X^2	Y^2	X*Y	
2		12	13	144	169	156	
3		18	17	324	289	306	
4		15	15	225	225	225	
5		11	10	121	100	110	
6		9	10	81	100	90	
7		17	16	289	256	272	
8		13	15	169	225	195	
9		8	9	64	81	72	
10		18	17	324	289	306	
11		12	12	144	144	144	
12	Suma	133	134	1885	1878	1876	
13							
14	Correlación	0,95901					

$$r_{xy} = \frac{n \cdot \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[n \cdot \sum X^2 - (\sum X)^2] \cdot [n \cdot \sum Y^2 - (\sum Y)^2]}}$$

Puntuaciones directas

$$r_{xy} = \frac{10 \cdot 1876 - 133 \cdot 134}{\sqrt{[10 \cdot 1885 - (133)^2] \cdot [10 \cdot 1878 - (134)^2]}} = 0,959$$

Muy buena fiabilidad

b) La fiabilidad como equivalencia:

También recibe la denominación de ***formas paralelas***; consiste en aplicar dos pruebas diferentes pero que miden el mismo rasgo o característica, de tal forma que los resultados de la aplicación de la primera prueba se correlacionan con los de la segunda, lo que nos permite calcular el coeficiente de correlación o equivalencia entre ambas puntuaciones.

Este procedimiento es difícil y complicado, sobre todo por los problemas que plantea encontrar pruebas equivalentes, que tengan los mismos objetivos y que los contenidos y las condiciones de aplicación sean similares.

También se utiliza el **coeficiente de correlación de Pearson**. Si transcurren más de 20 días entre la aplicación de la prueba y la forma paralela, el coeficiente encontrado puede ser considerado como de equivalencia y de estabilidad.

c) La fiabilidad como consistencia interna:

Los instrumentos de medida están representados por una serie de elementos o ítems bastante amplio; por ello cabe esperar que cada uno de ellos medirá una parte de ese rasgo o característica que mide la prueba en su conjunto; en consecuencia, es lógico suponer que existirá una coherencia o consistencia en las respuestas que ofrecerán un conjunto de sujetos a los diferentes elementos que integran tal instrumento.

Este procedimiento también conocido de ***las mitades nos permite dividir la puntuación total del sujeto en una prueba en dos partes (mitades)***, bien eligiendo como criterio los ítems pares y los impares o bien la primera y la segunda mitad. En realidad obtenemos dos puntuaciones para cada sujeto. En efecto: por un lado está la puntuación que corresponde a los ítems impares y por otro la de los impares; de tal forma que estableciendo una relación entre ambas partes nos dará el coeficiente de fiabilidad como consistencia interna, si bien de un instrumento cuya longitud es la mitad. Somos partidarios de utilizar este criterio de mitades, pues en muchas pruebas, y con el fin de estimular las respuestas de los estudiantes, los ítems más fáciles se colocan al comienzo de la prueba y los más difíciles al final, por ello la división impares/pares recogerá ítems del comienzo y del final y será más equilibrada que la primera y la segunda mitad.

Existen varios procedimientos para determinar la fiabilidad de una prueba por medio de la **consistencia interna**:

- **Procedimiento de Spearman-Brown:**

  Este procedimiento se basa en la correlación entre las mitades, generalmente mediante el *coeficiente de Pearson,* si bien esta correlación referida a las mitades ha de ser corregida mediante una fórmula sencilla para calcular la fiabilidad de la prueba en su conjunto.

  Así pues:

  $$R_{xx} = \frac{2.r_{xx}}{1+\ r_{xx}}$$

  Los subíndices son iguales, pues se trata de una correlación interna, dado que se establece una relación entre dos partes de la misma variable.

  Además r_{xx} se calcula mediante el coeficiente de correlación de Pearson entre las mitades, así llamamos X_1 a las puntuaciones de los ítems impares (1.ª mitad) y X_2 a la suma de los ítems pares (2º mitad).

  Siendo:

$$r_{xx} = \frac{n\ \Sigma\ X_1X_2 - \Sigma\ X_1\ \Sigma\ X_2}{\sqrt{[n\ \Sigma\ X_1^2 - (\Sigma\ X_1)^2]\cdot[n\ \Sigma\ X_2^2 - (\Sigma\ X_2)^2]}}$$

Ejemplo: Prueba de 30 ítems para una muestra de 12 sujetos. (Pág. 156)

B16 f_x =COEF.DE.CORREL(B2:B13;C2:C13)

	A	B	C	D	E	F	G
1		X_1	X_2	X_1^2	X_2^2	$X_1 * X_2$	d
2		12	13	144	169	156	-1
3		8	7	64	49	56	1
4		11	10	121	100	110	1
5		14	15	196	225	210	-1
6		7	6	49	36	42	1
7		9	11	81	121	99	-2
8		13	11	169	121	143	2
9		9	9	81	81	81	0
10		5	6	25	36	30	-1
11		13	12	169	144	156	1
12		6	8	36	64	48	-2
13		11	9	121	81	99	2
14	Suma	118	117	1256	1227	1230	1
15							
16	Correlación	0,8752					
17							

$$R_{xx} = \frac{2 \cdot 0{,}875}{1 + 0{,}875} = 0{,}93$$

$$r_{xy} = \frac{12 \cdot 1230 - 118 \cdot 117}{\sqrt{[12 \cdot 1256 - (118)^2] \cdot [12 \cdot 1227 - (117)^2]}} = 0{,}875$$

- **Proedimiento de Guttan:**
  Este procedimiento se basa en la varianza de las mitades, de tal forma que a menor valor las de varianzas más elevada será la fiabilidad de la prueba. Así pues la fórmula de cálculo es la siguiente:

$$r_{xx} = 2\left(1 - \frac{s_{1a}^2 + s_{2a}^2}{s_t^2}\right)$$

Dado que trabajamos con los mismos datos y ya tenemos calculada la varianza total, solamente debemos obtener los valores de las varianzas de las mitades (impares/pares).

$$s_{1a}^2 = \frac{\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{n}}{n-1} =$$

$$s_{2a}^2 = \frac{\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{n}}{n-1} =$$

3. EL ESTUDIO DE LA VALIDEZ:

Podemos afirmar que **un instrumento es válido cuando mide lo que dice medir y no otra cosa distinta.** Ahora bien, debemos indicar que ningún instrumento va a ser absolutamente válido, sino que tendremos que *matizar el grade y la naturaleza de la validez.*

En la investigación socioeducativa existen diferentes enfoques o tipos de validez de los instrumentos de medida y recogida de datos, en cada caso debemos recurrir a aquella validez que guarda una relación más estrecha con el propio instrumento y la forma en que está configurado, así como la información que nos suministra. Seguidamente pasamos a analizar de forma más detallada las principales.

❖ **Tipos de validez:**

a) La validez de contenido:

La muestra ha de ser ***suficiente y representativa.*** La ***suficiencia*** se relaciona con el número mínimo de elementos que debe incluir el instrumento, para tener garantías de que comprende aquellos aspectos que son esenciales para el estudio. La ***representatividad***, exige un conocimiento profundo sobre el tema, de tal forma que esas tareas sean definidas con precisión y claridad, en ese caso será más sencillo proceder a la selección de los ítems. En el caso de que la relación entre las tareas no sea tan directa, es conveniente elaborar una ***tabla de especificaciones*** y ***juicios de expertos***. A este proceso se le denomina ***validación.***

b) La validez predictiva:

Nos permite conocer la capacidad que tiene la prueba para avanzar las expectativas sobre ***futuros hechos o fenómenos***; se pretende establecer predicciones sobre la variable que se mide o bien sobre los diversos aspectos relacionados con ella. Su cálculo se realiza estableciendo una correlación entre las puntuaciones alcanzadas en la prueba a validar por un determinado número de sujetos y los obtenidos en otra prueba denominada criterio. Es preciso que todos los sujetos que responden *estén previamente identificados*, pues se trata de relacionar las puntuaciones de cada sujeto en las dos pruebas. En la gran mayoría de los casos, recurriremos al coeficiente de correlación de Pearson.

c) La validez concurrente:

Este tipo de validez es una *modalidad de la predictiva*, pues se calcula mediante una correlación entre las puntuaciones de los sujetos en la prueba a validar y el criterio externo. La diferencia estriba en que ambas mediciones se llevan a cabo en el mismo tiempo, además los resultados permiten realizar pronósticos a corto plazo, es decir, de utilización inmediata. Mientras que en la predictiva no coinciden la aplicación de la prueba y la recogida de datos. En cada caso el investigador tendrá que revisar y definir con claridad el objetivo del instrumento de medida para así elegir aquél tipo de validez que es el más apropiado en cada situación.

d) La validez de constructo:

También se denomina *de elaboración o de construcción*, se refiere al objeto mismo de la medición, acude a la base del problema de la validez. Se trata de analizar la conexión que se manifiesta entre la teoría en la que se basa la prueba y los ítems que la componen. Ello implica que se deben establecer una serie de hipótesis iniciales, que, una vez comprobadas empíricamente, nos ofrecerá tanto la validez del instrumento como la teoría subyacente al instrumento. Cuando las hipótesis no sean confirmadas, se puede deber a que la prueba carece de validez o a que es insuficiente, también puede ocurrir que la hipótesis no esté bien planteada o que la propia investigación no permita validar la hipótesis.

e) La validez aparente:

Validez didáctica o validez superficial: son aspectos externos cuya influencia puede ser relevante para alcanzar *participación y respuestas válidas*. Que no haya nada que lo invalide.

4. EL ESTUDIO DE LOS ELEMENTOS O ÍTEMS DE UNA PRUEBA:

Para tener las suficientes garantías científicas de que la prueba que vamos a aplicar a un grupo de sujetos es la más adecuada, es preciso conocer el comportamiento de cada uno de los ítems que la componen. Así se podrán eliminar aquéllos que no pasen el control o añadir otros que sí cumplan esos requisitos. Estos índices no son aplicables a las pruebas de medida de actitudes.

- **El índice de dificultad (ID):**

La dificultad de los elementos depende del grupo de sujetos que lo conteste y se expresa numéricamente por el número de ellos que los resuelven satisfactoriamente, de tal forma que cuando es contestado bien por un número pequeño de sujetos ese ítem será muy difícil y, a la inversa, si prácticamente todos lo responden acertadamente ese elemento será muy fácil.

"El índice de dificultad (ID) de un elemento o ítem nos indica la proporción de los sujetos que lo resuelven de forma correcta en relación con el total de los que lo contestan".

Cuando se pretende evocar una respuesta, la calificación final será igual al número de aciertos de cada sujeto, en este caso los errores no penalizan, por ello el índice de dificultad se calcula de la siguiente forma:

$$ID = \frac{A}{n}$$

A nos indica el número de sujetos que aciertan el ítem.
n el número de sujetos que lo intentan.

En el supuesto de que la prueba esté compuesta por ítems es decir (entre varias opciones) elegir la correcta, los errores penalizan, pues se deben corregir las respuestas ofrecidas al azar.
Puntuación de un sujeto:

$$= A - \frac{E}{na - 1}$$

E se refiere al número de errores
n_a es el número de alternativas de respuesta que se ofrecen.

Para este caso de elementos de varias alternativas de respuesta la fórmula es:

$$ID = \frac{A - \frac{E}{na - 1}}{n}$$

Para proceder a la interpretación de los índices de dificultad de los ítems hay cinco categorías:

Muy fáciles (ID >0,75)	**10%**
Fáciles (0,55 < ID < 0,75)	**20%**
Normales (0,45 < ID < 0,55)	**40%**
Difíciles (0,25 < ID < 0,45)	**20%**

Muy difíciles (ID < 0,25)	**10%**

0 = muy difícil

1 = muy fácil

Lo más pedagógico es que los ítems fáciles se sitúen al principio de la prueba, y los difíciles al final.

❖ **El índice de homogeneidad (IH):**

La homogeneidad de los elementos pone de manifiesto **la coherencia de cada uno de ellos con el total de la prueba**: si esta mide un rasgo o característica es lógico que cada uno de los ítems mida hasta cierto punto lo mismo.

Se calcula **a partir del valor de la corrección entre cada elemento y el conjunto de los demás, así pues, cuanto mayor es el coeficiente entre ambas puntuaciones, mayor será la homogeneidad**.

Ejemplo: (Página 174) $r_{bp} = r_{AB}$ $\quad r_{bp} = 0{,}52;\ s_A = 3{,}25;\ s_B = 0{,}495$

$$I.H. = \frac{0{,}52 \cdot 3{,}25 - 0{,}495}{\sqrt{(3{,}25)^2 + (0{,}495)^2 - 2 \cdot 0{,}52 \cdot 3{,}25 \cdot 0{,}495}} = 0{,}395$$

$$I.H. = \frac{r_{AB} \cdot s_A - s_B}{\sqrt{s_A^2 + s_B^2 - 2 \cdot r_{AB} \cdot s_A \cdot s_B}} \qquad r_{bp} = \frac{\bar{X}_p - \bar{X}_t}{s_t} x \sqrt{\frac{p}{q}} \qquad r_{bp} = \frac{\bar{X}_p - \bar{X}_q}{s_t} x \sqrt{p \cdot q}$$

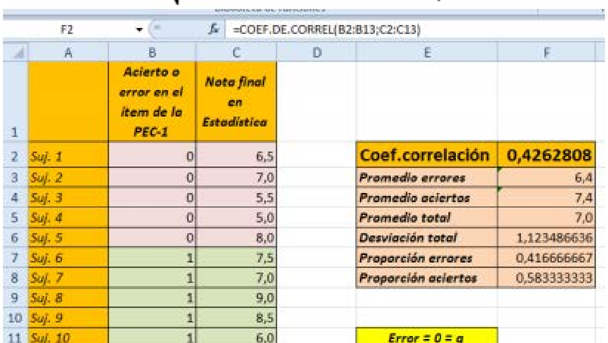
Donde r_{AB} es la correlación entre el ítem y el total, s_A la desviación típica de las puntuaciones en la prueba, s_B la desviación típica en el ítem que se calcula $s_B = \sqrt{p \cdot q}$.

❖ **El índice de validez (IV):**

Los elementos de una prueba se consideran válidos cuando realmente miden lo que dicen medir y no otra cosa distinta. Ello exige la existencia de una clara **relación entre las puntuaciones obtenidas por los sujetos de la muestra en cada uno de los ítems y las alcanzadas en el criterio de validez: cuanto mayor sea esa relación mayor será el índice de validez (IV)**.

Ejemplo (página 175): media aciertos: 6,32; media total: 5,79; $s_t = 1{,}86;\ p = 0{,}61;\ q = 0{,}39$

$$r_{bp} = \frac{|\bar{X}_p - \bar{X}_t|}{s_t} \sqrt{\frac{p}{q}} = \frac{|6{,}32 - 5{,}79|}{1{,}86} \cdot \sqrt{\frac{0{,}61}{0{,}39}} = 0{,}36$$



F2 =COEF.DE.CORREL(B2:B13;C2:C13)

	A	B	C	D	E	F
1		Acierto o error en el ítem de la PEC-1	Nota final en Estadística			
2	Suj. 1	0	6,5		Coef.correlación	0,4262808
3	Suj. 2	0	7,0		Promedio errores	6,4
4	Suj. 3	0	5,5		Promedio aciertos	7,4
5	Suj. 4	0	5,0		Promedio total	7,0
6	Suj. 5	0	8,0		Desviación total	1,123486636
7	Suj. 6	1	7,5		Proporción errores	0,416666667
8	Suj. 7	1	7,0		Proporción aciertos	0,583333333
9	Suj. 8	1	9,0			
10	Suj. 9	1	8,5			
11	Suj. 10	1	6,0		Error = 0 = q	

$$r_{bp} = \frac{\bar{X}_p - \bar{X}_t}{s_t} x \sqrt{\frac{p}{q}}$$

$$r_{bp} = \frac{\bar{X}_p - \bar{X}_q}{s_t} x \sqrt{p \cdot q}$$

TEMA 9: Aplicaciones de la correlación: fiabilidad y validez de las medidas.

1. MODELOS:

Modelo es una representación de la realidad, con frecuencia de las relaciones que se dan en la realidad representada; la forma en que la representa puede ser muy variada: desde una **representación icónica** (un cuadro, una escultura...), a la **matemática**, en la que, mediante fórmulas, se establece una igualdad, más o menos compleja pasando por la analógica, en la que, mediante esquemas, diagramas y representaciones de diverso tipo, nos es permitido acercarnos a la comprensión de realidades abstractas.

Todo modelo es una representación simplificada de la realidad a la que se refiere. Este hecho debe ser tenido en cuenta, aún en el caso de modelos icónicos, como una estatua o una fotografía, no queda representada la realidad compleja del ser representado sino una parte, la externa, y solamente en su aspecto observable.

No obstante esta limitación, si el modelo es adecuado, y dada su relación con la teoría en que se sustenta, las propiedades del modelo pueden ser aplicadas a la realidad empírica. Y aquí reside su importancia y su gran utilidad: *siempre que se cumpla que una realidad es razonablemente bien representada por un modelo, las cualidades de este, con una base teórico-científica, se pueden aplicar a aquella,* haciendo posible su comprensión más allá de lo externamente visible.

- **Utilidad de los modelos:**

La complejidad de los fenómenos humanos es sumamente superior a la de los físicos; por ello, cualquier predicción a partir de los modelos está limitada por todo lo que representa esa complejidad.

Si bien las predicciones a partir de modelos difícilmente pueden aplicarse a sujetos concretos, sí pueden predecirse las tasas de incidencia de un determinado fenómeno en términos de porcentajes y, siempre, debidamente moderadas a través de la probabilidad. Así, no tiene sentido decir si una persona concreta morirá o no en la carretera en un determinado fin de semana, pero Tráfico puede estimar, con el determinado riesgo de error, por lo general bastante reducido, cuántas personas morirán en un fin de semana.

- **Modelos matemáticos y modelos estadísticos:**

Por lo general, se considera que un modelo estadístico es un tipo de modelo matemático el que se incorpora como componente fundamental la probabilidad.

Los modelos matemáticos se expresan mediante igualdades que reflejan la relación existente entre los componentes de la realidad. La **función** es un modelo matemático.

Un modelo matemático clave en nuestro campo es **la campana de Gauss**, esta fórmula puede escribirse del siguiente modo, conocido como distribución normal tipificada [z = $(X_i$ — media aritmética)/desviación típica]:

$$P(x) = \frac{1}{\sqrt{2\Pi\sigma}} . e^{\frac{-1}{2} . \left(\frac{\mu - \bar{x}}{\sigma}\right)^2} = \frac{1}{\sqrt{2\Pi\sigma}} . e^{\frac{-1}{2} . z^2}$$

Los modelos estadísticos parten de un supuesto: las relaciones entre ambos términos de la igualdad no son determinísticas o necesarias, sino estocásticas. En consecuencia, toda predicción asume un cierto riesgo de error, un error que se considera aleatorio. Los errores aleatorios tienden a compensarse y su magnitud puede estimarse.

Si un modelo matemático como el siguiente establece una relación funcional necesaria:

$$(Y_1, Y_2, Y_3, \ldots, Y_i) = f\,(Y_1, Y_2, Y_3, \ldots, Y_i)$$

En el caso de un modelo estadístico nos encontraríamos con:

$$(Y_1, Y_2, Y_3, \ldots, Y_i) = f\,(Y_1, Y_2, Y_3, \ldots, Y_i) + e$$

Aquí, e es el error que afecta a la predicción. Como podemos comprender, el interés de los estadísticos es reducir a su mínima expresión tal valor, siendo la meta alcanzar $e = 0$. La anterior expresión se conoce como Modelo Lineal Clásico.

2. PROBABILIDAD:

Hay una diferencia esencial entre los fenómenos determinísticos o necesarios y los aleatorios y estocásticos, regidos por las leyes de la probabilidad.

- **La Estadística:**

Cuando los resultados de una prueba estadística vayan más allá de lo esperado por puro se aceptará que el fenómeno en cuestión no se explica por azar sino por la acción del investigador, sometida a contraste en condiciones controladas y analizada mediante tal prueba. Se comprenderá así la importancia del conocimiento de los fenómenos aleatorios.

- **La probabilidad**:

Como podemos apreciar, *aleatoriedad y probabilidad* son dos conceptos íntimamente ligados. A los fenómenos determinísticos no se les aplica la probabilidad: son fenómenos que ocurren porque tienen que ocurrir necesariamente. Pero los fenómenos aleatorios, que ocurrirán o no, antes de que ocurran pueden ser más o menos probables. Todos entendemos que es más fácil que nos toque la lotería si jugamos 1.000 números que si solo lo hacemos con uno. Pero también sabemos que una persona que haga una sola columna o juegue un solo número, le puede tocar el “gordo”.

He aquí la diferencia entre dos conceptos de probabilidad:

- Probabilidad ***a priori*** → antes de que ocurra un fenómeno podemos estimar las probabilidades de que ocurra o de acertar una predicción. Se establece sobre la base del número de casos favorables dividido por el número de casos posibles, y constituye la base de la probabilidad matemática, de la probabilidad en términos teóricos.
- Probabilidad ***a posteriori*** → cuando los fenómenos ya han ocurrido, podemos establecer la probabilidad de ocurrencia de tal fenómeno. Esta probabilidad se calcula empíricamente y se traduce en la frecuencia relativa con la que ocurre tal fenómeno cuando se repite un elevado número de veces en las mismas condiciones. Cuando estudios reiterados vienen a arrojar resultados compatibles, es posible establecer la realización de estimaciones sobre la probabilidad de que, en ocasiones sucesivas, ocurra un determinado acontecimiento.

3. CALCULO DE LA PROBABILIDAD:

ESPACIO MUESTRAL → es el conjunto de todos los resultados posibles de un fenómeno. Cuando dos conjuntos, A y B, no tienen elementos en común, decimos que $A \cap B = \varnothing$; y se lee "A intersección con B, es igual al conjunto vacío". Entonces:

$$\boxed{P(A \cup B) = P(A) + P(B)}$$

Pero esto no es lo normal, pues se da con bastante frecuencia que dos ó más conjuntos tengan elementos en común y entonces decimos que $A \cap B \neq \varnothing$, luego

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

El fenómeno de la ***EXHAUSTIVIDAD*** o ***AGOTAMIENTO*** se produce cuando los diferentes subconjuntos que puedan crearse, son subconjuntos del espacio muestral y todos ellos juntos lo agotan, lo completan plenamente.

Hablamos de ***MUTUA EXCLUSIÓN*** cuando dos acontecimientos distintos no tienen ningún elemento en común, es decir, su intersección es el $\varnothing$. Como en tal caso $A \cap B = \varnothing$, las probabilidades de cada subconjunto pueden sumarse.
LA INDEPENDENCIA supone que la probabilidad de que ocurra el fenómeno conjunto, es igual al producto de las probabilidades de cada uno por separado. Entonces:

$$\boxed{P(A \cap B) = P(A) \,.\, P(B)}$$

Esta es una de las condiciones o supuestos que se han de verificar para la aplicación de las pruebas denominadas paramétricas en el contraste de hipótesis. Concepto de ***PROBABILIDAD CONDICIONAL.***

En nuestro ámbito de trabajo es bastante frecuente encontrarnos con fenómenos relacionados, es decir, que no son independientes. Como en este caso, $A \cap B \neq \varnothing$, la probabilidad condicional nos sitúa ante un caso en el que deseamos conocer la probabilidad de un determinado acontecimiento ó suceso cuando la probabilidad del otro es conocida.

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

- **El caso de las variables aleatorias continuas y discretas:**

Una variable cuantitativa es discreta cuando no puede adquirir todos los valores posibles, es decir, cuando el conjunto de valores posibles es numerable.
El número de los alumnos de una clase sólo admite valores enteros: 1, 2, 3... 40; no tiene sentido hablar de 19,3 alumnos, aunque ese valor puede resultar de dividir el número de alumnos de diversas clases por el número de clases, siendo su media aritmética. En el primer caso, la variable aleatoria es continua; en el segundo, estamos ante una variable aleatoria discreta.
Si realizamos gráficos de histogramas, disminuyendo progresivamente la base de cada barra, como se aprecia en la fig. 9.1, pág. 187, llegaríamos a una curva que uniría los infinitos puntos posibles de esa distribución.

Esas sucesivas representaciones gráficas nos permiten desvelar dos valores de la probabilidad: la **función de densidad de probabilidad,** y la **función de distribución.** Tienen estas características:

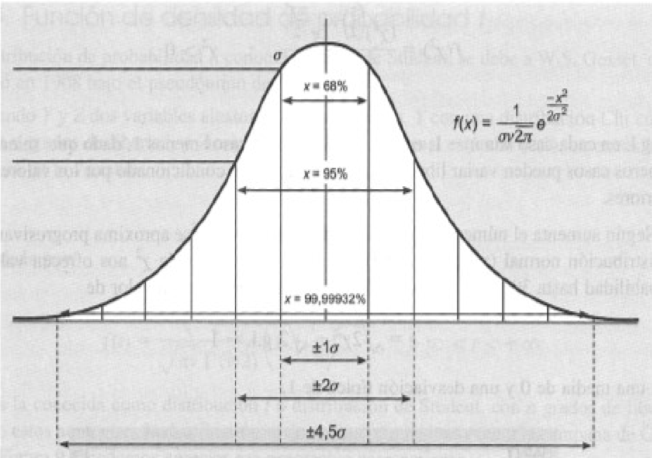
- El área ocupada por ambas representaciones tiene como valor la unidad.
- En todos los casos, las representaciones tienen siempre valores positivos.
- En las representaciones de histogramas de barras, y sea cual sea su base, el eje de abscisas representa una variable aleatoria discreta.
- En las representaciones de histogramas sin barras, cuando la base son puntos y la representación es una curva, estamos ante una variable aleatoria continua.

En el caso de las variables aleatorias discretas, el área que queda entre dos valores A y b, nos indica la PROPORCIÓN de casos del total que se encuentran entre esos valores. En el caso de variables cuantitativas continuas, indica la PROBABILIDAD de que tal variable tome esos valores.

4. FUNCIONES DE DENSIDAD DE PROBABILIDAD:

❖ **Función de densidad de probabilidad normal:**

La función de densidad de probabilidad más frecuentemente utilizada entre nosotros es la denominada normal, conocida como Campana de Gauss.



Algunas de sus características más importantes:

- Porcentaje de casos entre dos valores de σ (desviación típica poblacional):
  - ± σ → 68%
  - ± 2 σ → 95%
  - ± 4, 5 σ → 99,99932%

  Estos valores en porcentajes se convierten en *probabilidades* dividiendo por 100; es decir, el conjunto de los casos situados entre ±σ tiene una probabilidad de ocurrencia de 0,68.
- Mediante la tabla de áreas de la curva normal, podemos atribuir probabilidades a un caso concreto, en sus diversas columnas.

- **Función de densidad de probabilidad χ^2 (Ji cuadrada):**

Las variables aleatorias a las que se les aplica se distribuyen según χ^2 con **n-1** grados de libertad (g.l.), esto es, el número de casos menos 1, puesto que mientras los primeros casos pueden variar libremente, el último viene condicionado por todos los anteriores.

Según aumenta el número de los grados de libertad, la distribución χ^2 se aproxima progresivamente a la distribución normal (ver fig. 9.3 pág. 190). De hecho, las tablas de χ^2 nos ofrecen valores de probabilidad hasta 30 g.l. A partir de ahí, la distribución sigue con un valor de:

$$\boxed{z = \sqrt{2\chi^2} - \sqrt{2(gl) - 1}} \qquad \text{con } \bar{x} = 0 \text{ y } S = 1$$

La importancia de esta distribución radica en sus aplicaciones:

- Como prueba de bondad de ajuste
- Como prueba de independencia
- Como prueba del grado de asociación entre dos conjuntos de variables de atributo, para calcular el coeficiente de contingencia C:

$$\boxed{C = \sqrt{\frac{\chi^2}{N + \chi^2}}}$$

- **Función de densidad de probabilidad t:**

La distribución de probabilidad t se conoce como t de Student. Si tenemos Y y Z, dos variables aleatorias independientes, Y con una distribución χ^2 con n gl., y Z con una distribución normal (0,1), definimos distribución

$$\boxed{t = \frac{z}{\sqrt{\frac{Y}{v}}}}$$

Cuando aumentan los gl., la distribución se aproxima progresivamente a la campana de Gauss.

Esta distribución **se utiliza frecuentemente en pruebas de contraste de hipótesis** para decidir si la diferencia de medias es o no, estadísticamente significativa, a un determinado nivel de confianza.

Cuando las muestras son correlacionadas, la distribución t no sigue el estadístico de contraste aplicado en las muestras independientes. Dos muestras son correlacionadas cuando se forman parejas de sujetos, uno de cada muestra, que gozan de cierta característica común o similar.

- **Función de densidad de probabilidad F:**

Denominada F de Fisher, esta función de probabilidad, junto con la anterior t, son de las más utilizadas en el ámbito del contraste de hipótesis.

$$\boxed{F = t^2}$$

F puede aplicarse además a contrastes con tres o más pares de medias en diseños de tres o más grupos. **Se aplica fundamentalmente en el análisis de varianza (ANAVA).** F nos indica si se dan o no diferencias estadísticamente significativas entre varios grupos de medias. En caso afirmativo, es preciso averiguar entre qué dos pares de medias se concreta tal diferencia, por lo que se hace necesario la continuación del trabajo con las denominadas

pruebas a posteriori. La distribución F se define como la razón entre dos distribuciones χ^2 independientes, dividida cada una de ellas entre sus respectivos gl.

$$F = \frac{\chi_1^2 / \upsilon_1}{\chi_2^2 / \upsilon_2}$$

Si las dos varianzas poblacionales son iguales, la fórmula se reduce a:

$$F = \frac{s_1^2}{s_2^2}$$

Las tablas nos ofrecen los valores de probabilidad que corresponden a diversos gl del numerador (varianza **INTERgrupos en la ANAVA)** y del denominador (varianza **INTRAgrupos en la ANAVA).** La distribución F es no negativa, sesgada hacia la derecha y sus valores oscilan entre 0 e ∞, siendo asintótica al eje de abscisas.

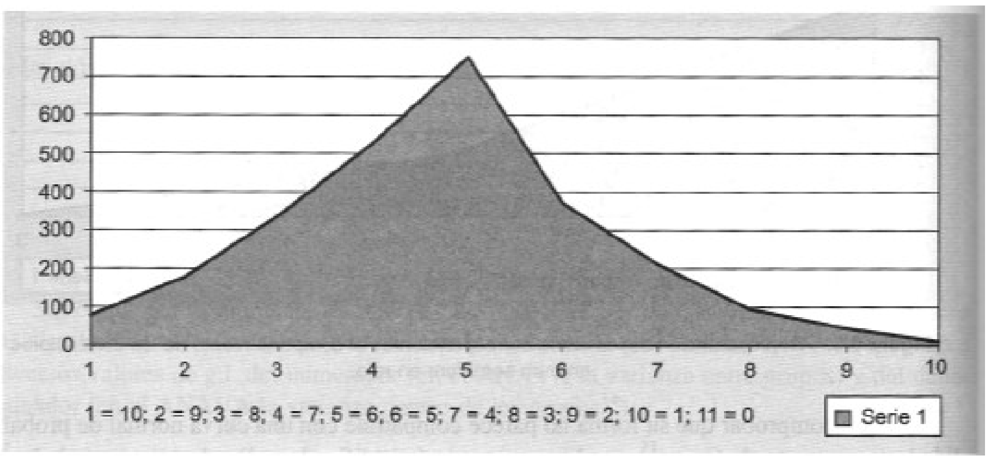
5. LA CURVA NORMAL DE PROBABILIDADES:

En ciencia no es aconsejable hacer apreciaciones subjetivas acerca del grado de acercamiento o parecido entre el modelo y la realidad. Conviene que las apreciaciones sean precisas y para ello tenemos la **prueba de bondad de ajuste.**

Si los datos empíricos se ajustan razonablemente al modelo, es decir, si las discrepancias son compatibles con las esperadas por puro azar, consideraremos estos datos como normales, y les aplicaremos todas las propiedades del modelo. En caso contrario, no podemos afirmar que el modelo sea idóneo.

❖ **Sobre el modelo**:

No hay una única curva normal, sino una por cada par de valores de y de $\bar{X}$S. Todas ellas cumplen las siguientes **características**:



- El valor máximo de la ordenada corresponde a la $\bar{X}$, y por tanto, a una $z = 0$.
- A ambos lados de la $\bar{X}$, se encuentran dos puntos de inflexión que se corresponden con los valores de $z \pm 1$.
- La curva es simétrica respecto a la $\bar{X}$, puesto que = Md = Mo. La ordenada de la $\bar{X}$ divide a la curva en dos partes iguales, cada una con el 50% de los casos.
- La curva es asintótica al eje de abscisas, por lo que nunca se llega a abarcar el 100% de los casos.

❖ La prueba de bondad de ajuste:

Pretendemos poder afirmar que una distribución dada se desarrolla normalmente, o que sus datos se distribuyen normalmente, siguiendo el modelo de la normal (0,1).

Para ello se acude a las pruebas de bondad de ajuste. Nosotros utilizaremos la de χ^2

Esta prueba valora las discrepancias entre los valores de las frecuencias empíricas y las teóricas. Si estas discrepancias no fueran significativas a un determinado nivel de confianza, admitiríamos que los datos empíricos y el modelo son una misma cosa (aceptamos H_0). Admitiríamos que las discrepancias encontradas pueden explicarse por puro azar como consecuencia de errores de muestreo.

$$\boxed{\chi^2 = \sum \left[\frac{(f_0 - f_e)^2}{f_e}\right]}$$

Se distribuye según χ^2 con:
* gl = nº de columnas -1 (si μ y σ son conocidas)
* gl = nº de columnas -3 (si μ y σ son estimadas)

Xi	fo	Lsup	Zi	P(Zi)	Pi	fe	(fo-fe)	(fo-fe)²	$\frac{(fo-fe)^2}{fe}$
	Σ					Σ			Σ

X_i → puntuaciones directas obtenidas por el alumno

f_o → frecuencias observadas

Lsup → límite superior de cada intervalo
Z_i → las puntuaciones típicas de tales límites
(Necesitamos calcular previamente los valores de $\bar{X}$ y de S)
$P(Z_i)$ → probabilidades que corresponden a esas Zi (columna B en las tablas)
P_i → probabilidad de cada intervalo, que se calcula restando de su valor P(Zi), el valor que corresponde al intervalo anterior.
f_e → frecuencias esperadas o teóricas. Para su cálculo:

- hallamos la columna P(Zi)
- hallamos la columna Pi
- hallamos la columna fe →Pi . N

Calculado el valor de χ^2 empíricamente, lo comprobamos con el valor de las tablas, para un nivel de confianza y (n-1) gl.

- **Si $\chi^2 \leq$ valor de tablas** → acepto H_0 → no hay diferencias significativas entre las distribuciones → el ajuste es bueno
- **Si $\chi^2 >$ valor de tablas** → acepto H_1 → si hay diferencias significativas entre las distribuciones → no hay ajuste.

Un ejemplo sería:

Este estadístico se distribuye según la distribución χ^2 para un valor igual al de filas menos 1 cuando μ y σ son conocidas, y con –3 en caso de ser estimadas. Vale la pena recordar aquí que χ^2 es un modelo estadístico.

Esta prueba funciona con las siguientes hipótesis:

$H_0 \rightarrow D_{empírica} = D_{teórica}$ (D = Distribución)

$H_1 \rightarrow D_{empírica} \neq D_{teórica}$

H_0 se rechaza si: $\chi^2_{empírico} > \chi^2_{teórico}$

Se obtiene de la tabla de Ji-cuadrado
Para búscalo se necesita conocer el nivel de confianza (N.C.) y los grados de liberta (g.l.).

11 filas

Límite superior de los intervalos

I	f_o	L_i	z_i	$p(z_i)$	p_i	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$(f_o - f_e)^2 / f_e$
					0,0055	15,48	15,48	239,54	15,48
10	74	10,5	2,54	0.9945	0.0157	44,18	29,82	889,23	20,13
9	175	9,5	2,026	0.9788	0.0443	124,66	50,34	2534,12	20,33
8	219	8,5	1,51	0.9345	0.0956	269,02	50,02	2502,00	9,3
7	340	7,5	0,99	0.8389	0,1563	439,83	99,83	9966,03	22,66
6	528	6,5	0,475	**0.6808**	**0.1968**	**553,79**	**25,79**	**665,12**	**1,2**
5	750	5,5	-0,04	0.4840	0,1963	552,39	197,61	39049,71	70,69
4	370	4,5	-0,56	0.2877	0,1454	409,16	39,16	1533,51	3,75
3	210	3,5	-1,07	0.1423	0,0864	243,13	33,13	1097,6	4,51
2	96	2,5	-1,59	0.0559	0.0385	108,34	12,34	152,27	1,41
1	43	1,5	-2,11	0.0174	0,0132	**37,14**	**5,86**	**34,34**	**0,92**
0	9	0,5	-2,625	0.0043	0,0043	12,10	3,10	9,61	0,79
N = 2814						2814			**171,17**

$p_i \cdot N = f_e$

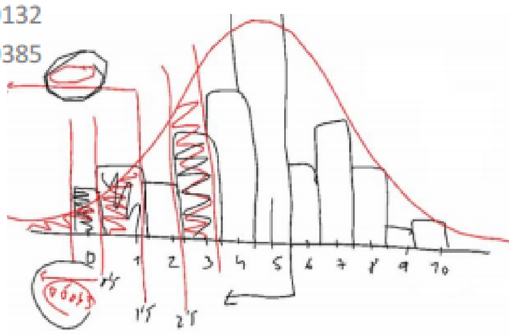
$\bar{X} = 5{,}58$

$s = 1{,}935$

$$z = \frac{X_i - \bar{X}}{s}$$

$$z = \frac{0{,}5 - 5{,}58}{1{,}935} = -2{,}625$$

Se calcula restando a $p(z_i)$ la $p(z_i)$ anterior.
Ej:
0,0174 – 0,0043 = 0,0132
0,0559 – 0,0174 = 0,0385
…
1 – 0,9945 = 0,0055



Se busca en la tabla de la curva normal respecto de z_i:
Si z_i es negativo: área de la parte menor.
Si z_i es positivo: área de la parte mayor.

$$\chi^2 = \Sigma\,[(f_o - f_e)^2/f_e] = 171{,}17$$

Si nuestro valor empírico (171,17) fuera igual o mayor que el de las tablas para los g.l. correspondientes, rechazaríamos H_0 y afirmaríamos que nuestros datos empíricos no son compatibles con el modelo de la curva normal de probabilidades con una probabilidad de tomar una decisión errónea $\leq \alpha$. Pues bien, las tablas de ji cuadrado, para un nivel de confianza del 99% ($\alpha = 0{,}01$), y (11 – 1) g.l. nos dan un valor de 23,209. Para 11 -3 g.l., el valor es de 20,090

Distribución Ji-cuadrado

Nivel de confianza(NC) = 1 - α = 1 - nivel de significación * n = Grados de libertad**

NC					90%	95%	97,5%	99%	99,5%	99,9%	NC
α	0,995	0,975	0,900	0,500	0,100	0,050	0,025	0,010	0,005	0,001	α
n											n
1	0,000	0,000	0,016	0,455	2,706	3,841	5,024	6,635	7,879	10,828	1
2	0,010	0,051	0,211	1,386	4,605	5,991	7,378	9,210	10,597	13,816	2
3	0,072	0,216	0,584	2,366	6,251	7,815	9,348	11,345	12,838	16,266	3
4	0,207	0,484	1,064	3,357	7,779	9,488	11,143	13,277	14,860	18,467	4
5	0,412	0,831	1,610	4,351	9,236	11,070	12,832	15,086	16,750	20,515	5
6	0,676	1,237	2,204	5,348	10,645	12,592	14,449	16,812	18,548	22,458	6
7	0,989	1,690	2,833	6,346	12,017	14,067	16,013	18,475	20,278	24,322	7
8	1,344	2,180	3,490	7,344	13,362	15,507	17,535	20,090	21,955	26,124	8
9	1,735	2,700	4,168	8,343	14,684	16,919	19,023	21,666	23,589	27,877	9
10	2,156	3,247	4,865	9,342	15,987	18,307	20,483	23,209	25,188	29,588	10
11	2,603	3,816	5,578	10,341	17,275	19,675	21,920	24,725	26,757	31,264	11
12	3,074	4,404	6,304	11,340	18,549	21,026	23,337	26,217	28,300	32,910	12

171,17 > 20,090

Se rechaza H_0

Rechazamos la hipótesis de que nuestros datos provienen de una distribución normal. Hay demasiada discrepancia

TEMA 10: Los baremos o normas. Muestreo. Aplicaciones.

1. INTRODUCCIÓN:

Para interpretar una puntuación necesitamos de algún tipo de referencia. En Educación es frecuente acudir a alguna de estas tres:

- **La idiosincrásica o personalizada**: la puntuación alcanzada por una persona en una prueba, se valora atendiendo bien a la puntuación que alcanzó en una prueba anterior, apreciando si aumenta, se mantiene o disminuye; o bien estableciendo previamente un nivel adecuado a sus características.

  El juicio de la comparación suele ser de *satisfactorio / insatisfactorio.*
- **La criterial**: *la referencia es un nivel objetivo*, previamente fijado por las personas competentes como deseable o adecuado para decidir sobre la suficiencia / insuficiencia. Se conoce como un dominio perfectamente definido y quienes lo alcanzan obtienen una valoración de suficiente, frente a la insuficiencia de quienes no llegan.

  En el ámbito del aprendizaje, esta es la referencia más adecuada.
- ***La normativa:*** Para la medida de muchas variables, especialmente en el ámbito de la inteligencia o de la personalidad, ni es fácil ni posible en ocasiones el fijar unos estándares, ni podemos decidir qué o cuánto es lo propio de cada persona, ni resulta factible establecer unas puntuaciones como las que debieran ser alcanzadas.

  Esta es una referencia relativa, ya que, dependiendo del grupo en que uno esté integrado, una misma puntuación puede ser más `positiva o favorable en unos casos que en otros.
  Esta regla para medir e interpretar las puntuaciones, **se conoce como baremo o norma.**

2. NORMAS O BAREMOS:

Según el Diccionario de la Real Academia Española, el término ***baremo*** procede de un matemático francés del siglo XVII e inicios de XVIII, llamado B. F. Barréme.

> "Cuadro gradual establecido convencionalmente para evaluar los méritos personales, la solvencia de empresas, etc., o los daños derivados de accidentes o enfermedades."

Baremar es construir un baremo, esto es una escala de puntuaciones obtenidas con un instrumento de medida que permite su interpretación mediante la atribución a cada una *de* ellas de un determinado valor.

Las puntuaciones brutas alcanzadas por las personas en una prueba resultan de difícil interpretación en sí mismas. Afirmar que un niño ha alcanzado 35 puntos en una prueba de inglés no nos dice nada si no disponemos de algún tipo de referencia. Hay algunos elementos que deben ser conocidos para interpretar una puntuación:

- **El suelo y el techo de las puntuaciones posibles**. Depende del número de ítems y de la propia regla de medida (*Ej. en el caso del tabaco, el suelo está claro: 0 cigarrillos; lo que no podemos saber es cuál es el techo.*)

- **La regla de medida: cómo se atribuye valor a cada uno de los componentes del instrumento de medida, por lo general denominados ítems**. Se debería cuidar de decidir con rigor si cada cuestión tiene la misma dificultad para poder valorar el éxito o fracaso por igual, si todos los errores pueden tener la misma penalización o los hay que deberían restar puntos en mayor medida.
- **El propio contenido del citado instrumento**. No todos los objetos a medir pueden serlo con la misma precisión, fiabilidad y validez. Su naturaleza decide en gran medida hasta qué punto los valores obtenidos con cada uno pueden situarse en uno u otro tipo de las denominadas escalas de medida.

  Es más fácil alcanzar niveles o escalas de medida más perfectos en el ámbito del rendimiento que en el de la inteligencia y, mucho más, que en la medida de determinadas variables de la personalidad, como la competencia emocional, la autoestima, la motivación o la seguridad en sí mismo.
  En todos estos casos, la "distancia" entre la realidad u objeto a medir y los reactivos a utilizar para hacerlo es muy grande; estamos ante lo que se denomina "constructor", es decir, construcciones en torno a algo que nada tiene de visible, pesable, audible... y sí de supuesto: suponemos que existe algo que denominamos inteligencia, lo definimos de una manera y sobre esa definición se construye un instrumento que pretende medirla. La persona que construye el instrumento para medirlos supone que tal variable se define como él lo hace.
- **A quiénes va destinado tal instrumento**. Todos admitimos que existe algo que 'denominamos "inteligencia" pero no parece correcto pensar que es posible elaborar un instrumento para "medirlas" válido en general, para todas las edades, clases sociales, niveles académicos, naciones o momentos históricos.

  Lo que a los 6 años parece adecuado a los 30 puede parecer una burla por infantil.
  Todos estos aspectos tienen que ser tenidos en cuenta no tanto estrictamente para la interpretación de las puntuaciones pero sí para que los números arrojados por la aplicación de los instrumentos de medida sean válidos y podamos proceder adecuadamente a interpretación.

3. CUALIDADES DE LOS BAREMOS O NORMAS:

La construcción de un baremo de calidad depende de la muestra utilizada para servir de referencia. En efecto, si de lo que se trata es de conformar una serie de valores, debidamente ordenados, para valorar los resultados, los valores seleccionados no pueden ser valores cualesquiera sino valores representativos del grupo al que pertenezcan los sujetos cuyas puntuaciones deseamos interpretar.

Ejemplo:

Si deseamos valorar la inteligencia de un alumno español de ingeniería informática, el baremo o norma de comparación, deberá estar integrado por estudiantes de ingeniería informática. Pero no estudiantes cualquiera sino estudiantes representativos del total de estudiantes de ingeniería informática, sin sesgos de ningún tipo, esto es, sin que se trate de los mejores o los peores, sin que determinados subgrupos estén sobre o infra representados.

Para alcanzar la representatividad de una muestra es preciso tomar dos decisiones clave:

- Fijar el *tamaño* de la misma, de forma que sea *suficiente* para que puedan manifestarse las características que definen la población (en el caso de muestras pequeñas esto no siempre es posible).
- Y utilizar un procedimiento de selección *imparcial*, que evite todo tipo de sesgos.

El que ofrece mayores garantías *a priori* es el muestreo aleatorio simple.

Así pues, *tamaño suficiente y selección aleatoria* son las condiciones fundamentales para construir una norma o baremo para la interpretación de las puntuaciones obtenidas en un instrumento de medida.

4. CONTRUCCIÓN DE LOS BAREMOS O NORMAS:

❖ Normas cronológicas o de edad:

Uno de los primeros tipos de normas utilizados fue el de las *cronológicas o de edad.* La denominada Edad Mental (EM) es el ejemplo más conocido. El procedimiento se trata de que realicen la prueba o cumplimenten el instrumento de medida de que se trate unas muestras imparciales y de adecuado tamaño para cada edad o grupo de edad cronológica.

La puntuación media de los sujetos de cada edad se convierte en representativa de la misma. En adelante, las puntuaciones de cualquier sujeto, tenga la edad cronológica (EC) que tenga, se comparan con las del baremo o norma resultante y se le asigna la edad mental correspondiente. Cuando la puntuación obtenida corresponde a una edad mental superior, estamos ante personas que llevan un desarrollo superior, ocurriendo lo contrario en el caso de que su puntuación corresponda a edades inferiores a la cronológica.

❖ Normas cuantiles:

Otro tipo de normas, son las denominadas genéricamente cuantiles, entre las que destacaremos las cuartiles, deciles y centiles o percentiles, esto es: cuantiles de orden cuatro, diez o cien.

Entendemos por cuantil cada una de las partes en que puede dividirse una serie ordenada de puntuaciones. Conocemos ya el más sencillo de todos: en efecto, la *mediana* es el cuantil 1 de orden 2, se trata de aquella puntuación que divide la serie en dos partes con un número de frecuencias igual, en concreto el 50% en cada una.

Los cuantiles más frecuentemente utilizados, además de la mediana, son los de orden 4 o *cuarteles* (suelen representarse como Q_1, Q_2, Q_3), que son aquellas puntuaciones que dejan por debajo de sí el 25, 50 y 75% de los casos, respectivamente; los *deciles* o cuantiles de orden 10 (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9), que dejan cada uno por debajo de sí el 10, 20, 30, ..., 90% de casos; y los *centiles* o *percentiles*, cuantiles de orden 100 -c_x- cada uno de los cuales deja por debajo de sí, el 1, 2, 3, ..., 97, 98, 99%.

Es fácil comprender que el cuantil 1 de orden 2, o mediana, es a la vez el cuantil 2 de orden 4 o cuartil dos (Q_2), el cuantil 5 de orden 10 o decil cinco (d_5) y el centil o percentil 50 de orden 100 *(N_o)*.

- **Construcción de un baremo en cuantiles:**

El procedimiento comienza con la elaboración de una distribución de frecuencias acumuladas, sean estas con puntuaciones directas o agrupadas en intervalos.

Veamos:

Supongamos que disponemos de la siguiente serie de puntuaciones en una prueba de aptitud manual que integran una muestra imparcial y suficiente (*representativa)* de alumnos de una ciudad que no han obtenido el Graduado en ESO.

Aunque no sería imprescindible, sí es recomendable su ordenación, ascendente o descendente:

X_i = 40, 36, 35, 32, 31, 30, 30, 29,29, 28, 28, 28, 28, 26, 26, 26, 26, 26, 26, 25, 24, 23, 23, 23, 23, 23, 23, 22, 22, 22, 22, 22, 22, 21, 21,20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 19, 19, 19, 18, 17, 17, 17, 16, 16, 16, 15,15, 15, 14, 14, 14, 13, 12,11, 11, 9, 7

Como podemos apreciar, el rango total de la serie va de 7 a 40 puntos, esto es, 33, por lo que en total hay 34 posibles puntuaciones diferentes. Si hacemos 9 intervalos, un número razonable, cada intervalo tendrá una amplitud de 4, lo que nos da 36 puntuaciones, lo que representa que habrá dos puntuaciones más de las posibles. Podemos atribuirlas al intervalo superior, al inferior o una a cada uno. En este caso, sin razones de peso, lo hacemos con primero de tales intervalos, como sigue.

Tabla 10.1. Distribución de frecuencias (f_i), frecuencias acumuladas (f_a) y porcentajes acumulados ($\%_a$). Puntuaciones típicas (z_i) y típicas normalizadas (z_{norm}).

I	X_i	f_i	f_a	$\%_a$	z_i	z_{norm}
37-40	38,5	1	70	100	2,75	
33-36	34,4	2	69	98,57	2,12	2,19
29-32	30,5	6	67	95,71	1,48	1,72
25-28	26,5	11	61	83,56	0,84	0,98
21-24	22,5	15	50	71,43	0,21	0,57
17-20	18,5	20	35	50	−0,425	0
13-16	14,5	10	15	21,43	−1,06	−0,79
9-12	10,5	4	5	7,14	−1,70	−1,44
5-8	6,5	1	1	1,43	−2,33	−2,19
		N = 70 Media: 1.482,8 : 70 = 21,18 s = 6,296				

La columna de porcentajes acumulados nos permite interpretar como porcentajes de casos que dejan por debajo cada una de las puntuaciones superiores del intervalo, en realidad auténticos percentiles. No obstante, lo que suelen presentar los baremos de las pruebas no es este tipo de datos sino una serie de puntuaciones a las que corresponden los cuantiles habituales, para cuya construcción utilizamos la Fórmula:

$$Cm = Linf + \frac{\left(\frac{C}{100}.N\right) - f_{a(I-1)}}{f_r}.a$$

El procedimiento más sencillo consiste en calcular, en primer lugar, el intervalo en el que se encuentra el cuantil (mediana, cuartil, decil o percentil) multiplicando el cociente C/100 por n o número de casos de la serie. C/100 se sustituye por D/10, Q/4 o 1/2 si se trata de deciles, cuartiles o mediana.

Este valor nos permite identificar el límite inferior de dicho intervalo (L_{inf}) al que deberemos sumar el resultado de una regla de tres.

- **Normas típicas:**

En las puntuaciones centiles se observa que están muy cercanas unas de otras en el centro del baremo, mientras que las distancias aumentan en los valores extremos.

Una forma de evitar esto consiste en expresar la puntuación de cada sujeto tomando **una unidad constante de medida.** Las variables típicas (*z)* indican la distancia de cada puntuación directa hasta la $\bar{X}$ del grupo, medida en unidades de S. Recordamos la fórmula de las puntuaciones típicas:

$$z = \frac{X - \bar{x}}{s_x}$$

Los sujetos con puntuaciones superiores a la $\bar{X}$, tienen *z* positiva, y la tienen negativa en caso contrario.

- **Puntuaciones típicas normalizadas:**

Una distribución de frecuencias se acerca más a la normal, cuanto mayor sea el número de casos (N) de la serie, siempre que la variable medida se distribuya normalmente en la población.

Cuando nuestros datos empíricos sean compatibles con la curva normal, tras haberlo comprobado con las correspondientes pruebas de bondad de ajuste, es interesante normalizar la distribución, y para ello procedemos a normalizar las puntuaciones típicas (*z).*

Podemos obtener una puntuación típica normalizada mediante las tablas. Para ello:

- calculamos el porcentaje de casos que se encuentren por debajo (o por encima) de cada puntuación
- se busca tal porcentaje en la tabla de áreas de la curva normal y se identifica la *z* normalizada correspondiente.

Ejemplo: (con los datos de la tabla 10.1)

$\bar{x}$ = 21,5

S = 6,29 $\qquad z = \frac{36{,}5-21{,}5}{6{,}29} = 2{,}384$

Xi = 36,5 En la tabla 10,1, una puntuación de 36,5 deja por debajo de sí a 69 de 70 casos → proporción: $\frac{69}{70}$ = 0,9857

buscamos en tablas de la curva normal esta proporción en la columna B, y encontramos z_{norm} = 2,19

Las z_{norm} corresponden a una distribución con $\bar{x}$ = 0 y S = 1. Para evitar los decimales de las puntuaciones típicas, podemos transformarla en otra distribución con $\bar{x}$ = 50 y S = 10, dando lugar a las **puntuaciones T**. O bien en otra con $\bar{x}$ = 50 y S = 20, y tendremos la **escala S**.

- **Estaninas y pentas:**

En EEUU es frecuente una escala de 10 rangos (9 puntos de corte), con la $\bar{X}$ = 5 y S = 2. En nuestro país, se utiliza en ocasiones una escala de 5 rangos (4 puntos de corte), con la $\bar{X}$ = 3 y S = 1. La primera se llama de **estaninos** o **eneatipos,** y la segunda se conoce como **pentas.**
La situación de un sujeto cualquiera en una de estas escalas es muy fácil. Primero calculamos la puntuación *z* del sujeto y posteriormente hacemos la transformación:

estaninos →5 + 2 *z* pentas →3 + *z*

- **El muestreo:**

Hasta ahora, lo que hemos afirmado es que las normas se construyen sobre muestras que, para bien, deben representar con fidelidad a la población de la que han sido extraídas.

Una *muestra* no es sino una parte, un subconjunto, de una población o universo. Una muestra de calidad es aquélla que representa fielmente el conjunto de características de la población.

Una muestra debe cumplir ciertas condiciones o exigencias; las dos fundamentales, resumidas en una, son su adecuado *tamaño* y su selección *imparcial*. Ambas notas dan lugar a muestras *representativas*, esto es, a muestras que reflejan con fidelidad, sin sesgos ni desviaciones, las características de la población.

- **Tamaño de la muestra:**

Para la fijación del tamaño de las muestras deberemos atender, en primer lugar, al tamaño del universo, considerado como infinito a partir de 100.000 casos; junto a ello, se deben tomar en consideración otras tres características que debe fijar el investigador en función de sus intereses:

- El nivel de confianza con el que desea trabajar.
- El error de estimación que considera adecuado asumir.
- La proporción en que la característica a estudiar se encuentra en el total de la población.

Del mismo modo, se supone la distribución normal de la característica muestreada.
En tamaños infinitos no aparece el valor N.

El *nivel de confianza*, decidido por el investigador, dando por hecho que estemos una distribución *normal* de la característica muestreada, se suele fijar en el 95, 99 o 99% admitiendo la distribución normal, y conociendo como conocemos la curva normal de probabilidades, este dato se traduce en las fórmulas en valores de 1,96, 2,58 o 3,3 desviaciones típicas.

En cuanto al *error de estimación* que decide admitir el investigador se fija en té de porcentaje; el investigador es consciente de que entre muestra y población siempre dará alguna diferencia; le cabe la decisión de admitir un error, en términos de porcentaje, más o menos grande. Obviamente, cuanto menor error esté dispuesto a aceptar, mayor no deberá tener la muestra.

Por último, se debe tomar en consideración la *proporción en que una característica encuentra en la población*. En ocasiones podemos conocer este dato por contar con censos ficheros de diferente naturaleza.

Con frecuencia se desconoce este dato, en cuyo caso lo habitual es considerar que tai característica se da al 50% en la población, lo que representará un mayor tamaño para la muestra. En el caso de que la población estuviera organizada en diversos grupos (estratos) esta decisión debería aplicarse a cada uno de ellos. Tal puede ser el caso de seleccionar una muestra de estudiantes universitarios, atendiendo al sexo, pero representando a diferentes subgrupos de carreras; por ejemplo, humanidades, ciencias de la salud, ciencias económicas, ciencias jurídicas, ingenierías.

Para el cálculo del tamaño de la muestra contamos con dos fórmulas diferentes, sepia que el tamaño de la población de origen sea finita o infinita.

Infinita → $\boxed{n = \frac{(z^2.p.q)}{E^2}}$ $\boxed{n = \frac{(z^2.p.q.N)}{[E^2.(N-1)+(z^2.p.q)]}}$

Donde N es el tamaño de la población, n el de la muestra; z es el valor que corresponde al nivel de confianza elegido (número de desviaciones típicas precisas para que la curva normal deje en su interior el 95, 99, 99,9%, etc.) *y* E el correspondiente a la proporción de la característica en la población. Tanto z como E están elevados al cuadrado.

Ejemplo:

Aplicaremos ambas fórmulas al caso de una población con 108.000 sujetos, en el primer caso, y con 35.600 en el segundo, tomando un nivel de confianza del 99% y un error de estimación del 3%. En el primer caso desconocemos la proporción en que se encuentra la característica en la población; en el segundo sabemos que es de 35 a 65 (esta es, aproximadamente, la relación de alumnos que se examinan en la UNED en la primera y en la segunda semana de pruebas presenciales).

En el primer caso tendríamos:

N = ($2{,}58^2$ x 50 x 50)/9 = **1.849** sujetos En el segundo: n = ($2{,}58^2$ x 35 x 65 x 35.600)7(9 x 35.599) + ($2{,}58^2$ x 35 x 65) =

= 539.101.836/320.391 + 15.143,31 = 1.606,69 (redondeando, **1.607** sujetos)

Para obtener el tamaño de las muestras puede acudirse a tablas específicas en las que aparecen diferentes valores de *p y q*, con los correspondientes niveles de confianza y error de estimación.

- **Procedimiento de selección:**

El principal procedimiento de extracción de muestras imparciales es el **muestreo aleatorio simple.** Lo característico de este muestreo es que todos los sujetos tienen, a priori, las mismas posibilidades de ser seleccionados para integrar la muestra.

El **muestreo sistemático** es una modalidad del anterior, que nos permite fijar el primero de los sujetos de la muestra, y a partir de él, seleccionar sistemáticamente el resto sumándole un valor constante, denominado coeficiente de evaluación, que equivale al cociente (N/n).

- **Procedimientos de muestreo.**

Además de los dos anteriores:

Estratificado: los sujetos de cada estrato pueden seleccionarse mediante el sistema aleatorio simple o el sistemático.

Por cuotas: cuando una determinada población está estratificada por nivel de estudios, clase social, consumo de drogas, religión, grupo político al que votan, ---, pueden seleccionarse sujetos representativos de los mismos, a fin de contar con representantes de los diferentes estratos poblacionales.

Incidental o casual: es el más usual por ser el más asequible. Se toman los sujetos disponibles o asequibles.

- **Error muestral:**

La propia teoría de la probabilidad nos va a permitir estimar la magnitud del error muestral para un determinado nivel de confianza. El error muestral puede calcularse mediante dos fórmulas, una se aplica a muestras finitas (<100.000) y la otra para muestras infinitas.

Infinita → $\boxed{E = \sqrt{\frac{(z^2.p.q)}{n}}}$ Finita → $\boxed{E = \sqrt{\frac{(z^2.p.q)}{n}.\frac{(N-n)}{(N-1)}}}$

Ejemplo:

Así, en el caso de las dos muestras anteriores, para poblaciones de 108.000 y 35.600 casos respectivamente, los errores muestrales, para el caso de que p = q = 50 en el primer caso, y de que p = 35 y q = 65 en el segundo, tendríamos:

— $E = \sqrt{(2{,}58^2 \cdot 50 \cdot 50)/1.820} = \mathbf{3{,}02}$

— $E = \sqrt{[(2{,}58^2 \cdot 65 \cdot 35/1.583) \times (35.600 - 1.583)/(35.599)]} = \sqrt{9{,}42 \times 0{,}955} =$
$= \sqrt{8{,}996} = \mathbf{3{,}02}$

— En caso de que las proporciones de p y q fueran, como en el caso anterior, de 50 × 50, el valor resultante sería de **3,17** ($\sqrt{10{,}35 \times 0{,}955} = \sqrt{9{,}88} = \mathbf{3{,}17}$).

Estos valores nos permiten establecer, en torno a los resultados obtenidos con las muestras, un *intervalo de confianza*, esto es, un conjunto de puntuaciones entre las cuales consideramos que se encontrará la verdadera puntuación de la población, si bien no con certeza sino con una determinada probabilidad, establecida por el investigador.

Cuanto menor sea el error muestral, menor será el intervalo de confianza, y ese error muestral será tanto mayor cuanto más elevada sea la seguridad que deseamos tener de que el valor de la característica en la población se encuentra en ese intervalo.

TEMA 11: Estimación de parámetros. Errores de estimación.

1. INTRODUCCIÓN:

En este tema comenzamos con la extrapolación de los resultados obtenidos en nuestras muestras a las grandes poblaciones a las que pertenece.

Es la INFERENCIA ESTADÍSTICA que tiene 2 aplicaciones básicas:

- Estimación de parámetros.
- Contraste de hipótesis.

La Inferencia Estadística se aplica con frecuencia en nuestra vida cotidiana: encuestas de opinión, encuestas electorales, estudios de mercado, etc…

En este tema aprenderemo a partir de los valores obtenidos en nuestras muestras (estadísticos, media = $\bar{X}$), estimar esos mismos valores en la población a la que pertenecen (parámetros, media = μ).

2. APROXIMACIÓN INTUITIVA A LA INFERENCIA ESTADÍSTICA:

La Estadística es la ciencia que se ocupa de la ordenación y análisis de datos procedentes de muestras, y de la realización de inferencias acerca de las poblaciones de las que éstas proceden.

- **POBLACIÓN:** conjunto de todos los elementos que cumplen una o varias características o propiedades. Los valores numéricos que describen a la población se llaman **parámetros (PP).**

- **MUESTRA:** es un subconjunto de los elementos de una población. Los índices numéricos que describen a las muestras se denominan **estadísticos.**

  La técnica para seleccionar a los sujetos que entrarán a formar parte de la muestra se denomina **muestreo.** Siempre que sea posible debe utilizarse el muestreo aleatorio porque nos da mayores garantías de que la muestra sea representativa de la población.

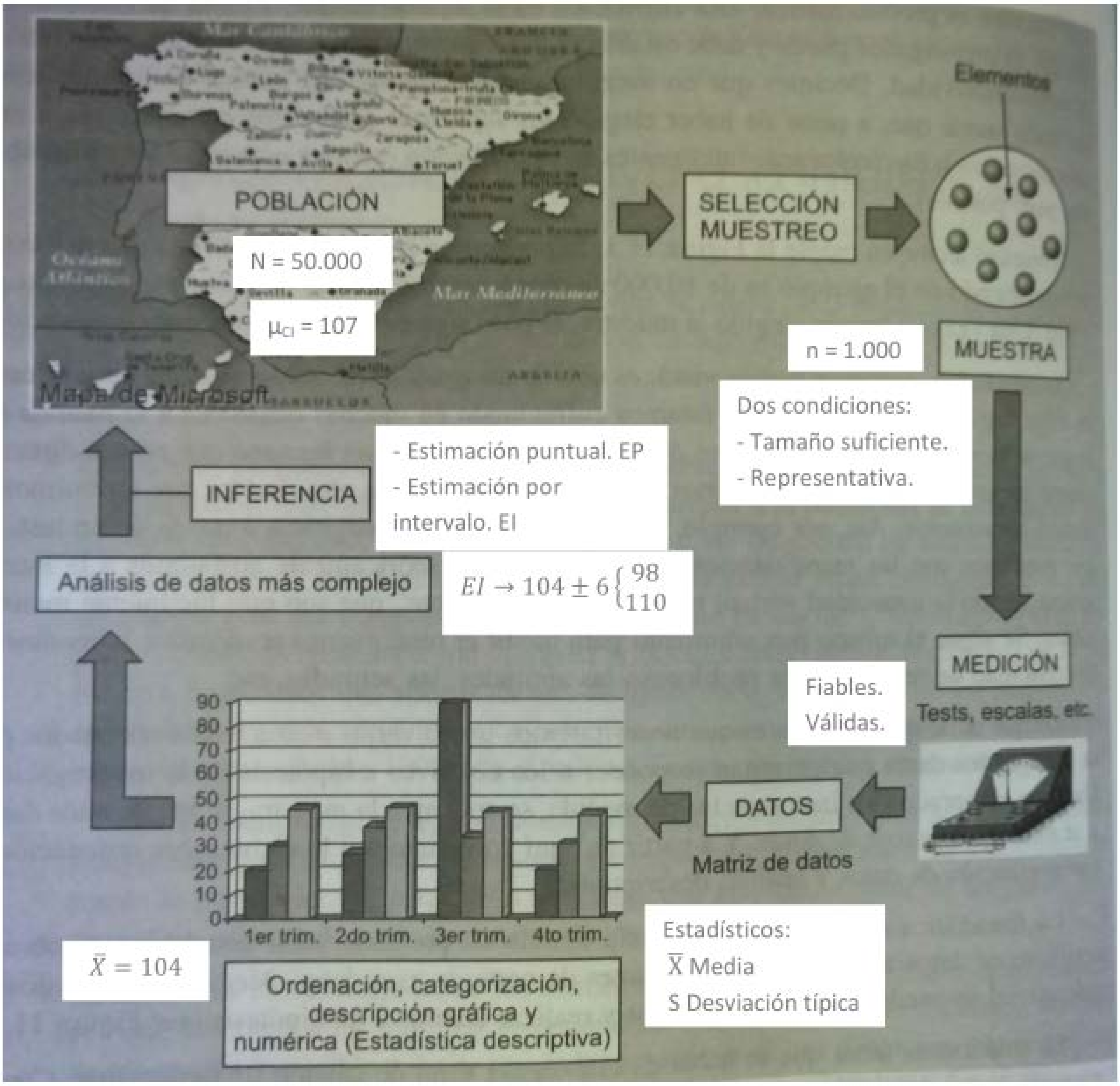
Lo más fácil es seleccionar una muestra suficiente en número (existen procedimientos para calcular el número de sujetos necesarios) y representativa de la población, que no esté sesgada.

La **medición**, es uno de los problemas más graves en Educación y en Psicología. Pensamos que nos estamos enfrentando en muchas ocasiones a la medida de lo que se denominan constructos, es decir, características del ser humano que no son directamente mensurables. Lo que medimos son las manifestaciones observables que atribuimos a dichos constructos. Así, por ejemplo, cuando medimos la inteligencia a través de un test, lo que medimos son las manifestaciones observables que teóricamente atribuimos a la inteligencia, como la capacidad verbal, numérica, abstracta, etc., que son más fácilmente mensurables. Se sigue el mismo procedimiento para medir el rendimiento académico, la memoria, la capacidad de resolución de problemas, las aptitudes, las actitudes, etc.

La **Estadística inferencial** o **inferencia estadística** pretende sacar conclusiones sobre el conjunto de datos a través de observaciones de parte de esos datos. Mediante ella, se pueden estimar parámetros y realizar contraste de hipótesis.

Es importante notar que se habla de estimación y no de cálculo de parámetros.

Dentro del marco de la Estadística inferencial suelen distinguirse dos objetos de estudio: la **estimación de parámetros** y el **contraste de hipótesis**. Un parámetro se estima siempre a partir de un estadístico calculado en una muestra. Hay dos tipos de estimación: **estimación puntual** y **estimación por intervalos**. Básicamente, en la estimación puntual se hace coincidir el estadístico con el parámetro, es decir, se escoge un solo punto (el estadístico) para estimar el valor del parámetro. Por su parte, en la estimación por intervalo se ofrece un



intervalo de puntuaciones en el cual es más probable que se encuentre el valor del parámetro.

3. PROPIEDADES DE LOS ESTIMADORES:

La estimación consiste en la técnica que permite conocer el valor aproximado de un parámetro de una población con una determinada probabilidad a partir de los datos proporcionados por una muestra. Un estimador es un estadístico muestral que permitirá la estimación de un parámetro poblacional. Estas son las características que debe poseer un buen estimador:

- **Carencia de sesgo:** podemos decir que un estimador insesgado es aquél que tiene sesgo igual a cero. La propiedad del insesgamiento nos garantiza que las estimaciones que hagamos con el estimador se encuentran alrededor del parámetro en cuestión, de forma simétrica, es decir; que el promedio de los estadísticos coincide con el verdadero valor del parámetro.
- **Eficiencia:** un estimador es tanto más eficiente cuanto menor es su desviación típica. Cuanto más «estrecha» sea, cuanto menor sea su error típico, más cercanos estarán los estadísticos al valor del parámetro.
- **Consistencia:** un estimador es consistente si a medida que aumenta el tamaño de la muestra, la probabilidad de que el valor del estadístico se acerque al valor del parámetro va siendo mayor. Si el valor de N tiende a infinito, un estimador consistente es, a la vez, insesgado.
- **Suficiencia:** un estimador es suficiente cuando es capaz de obtener de la muestra toda la información que ésta contenga acerca del parámetro.

4. DISTRIBUCIÓN MUESTRAL, ERROR MUESTRAL Y ERROR TÍPICO: ESTIMACIÓN DEL PARÁMETRO MEDIA ARITMÉTICA:

Ya nos hemos referido a un tipo de distribución característica que se denomina distribución normal o campana de Gauss. Pues bien, al igual que la distribución normal, hay otra serie de distribuciones *teóricas* como la binomial, la *t* de Student, la multinomial, la *F*, etc., en las que se conoce la probabilidad de aparición asociada a todos los posibles valores.

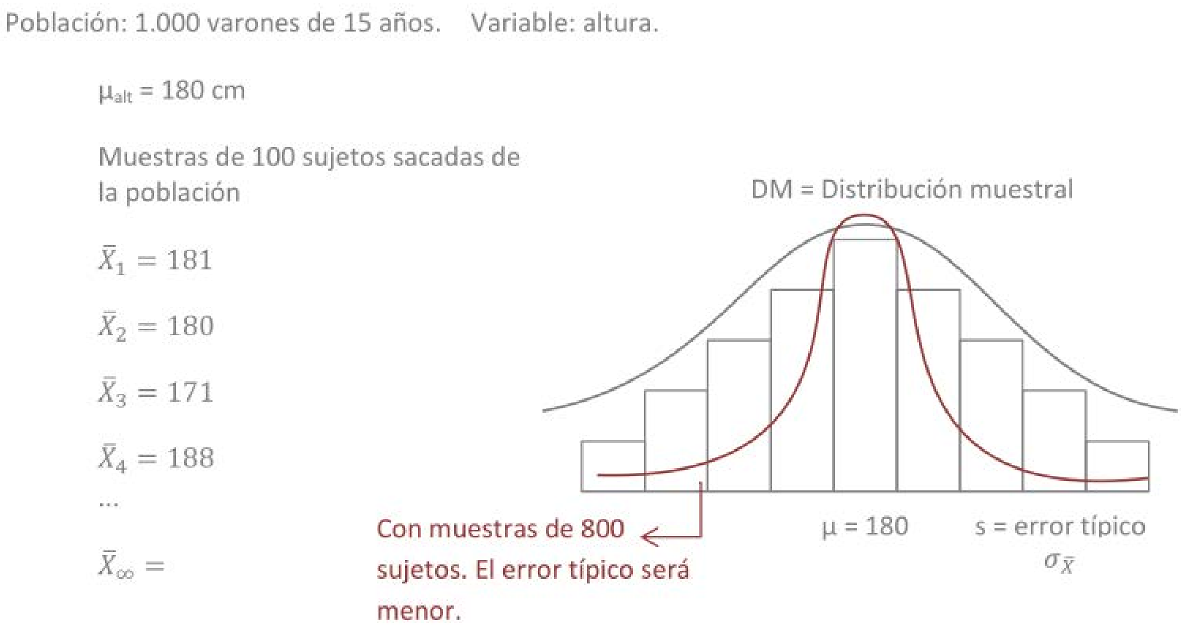
Distribución muestral:

Existe un concepto teórico de distribución *(función de densidad de probabilidad)* que es la **distribución muestral** y que puede definirse como la distribución de *un estadístico en el muestreo*. Está formada por los infinitos valores de un estadístico obtenidos de infinitas muestras aleatorias del mismo tamaño extraídas de la misma población.

Si nos remitimos a la **distribución muestral de la media aritmética (μ),** está demostrado según el teorema central del límite y para muestras grandes (N > 30), que la distribución se asemeja a una distribución normal.

Si el parámetro μ fuese conocido, comprobaríamos como la mayoría de las muestras se encontrarían cerca del valor del parámetro, pero precisamente por ser muestras aleatorias, algunas se alejan un poco y otras, muy pocas, se alejan mucho de ese valor.

Por tanto, confiamos en que la muestra elegida sea una de las que el valor de su esté cercano al valor de μ, pero no lo podemos asegurar. Por esta razón, en inferencia siempre hablamos de **nivel de confianza** (porcentaje de confianza al hacer la estimación; también puede darse en probabilidad y se denomina 1- α) y del **nivel de significación** α (probabilidad de error que estamos dispuestos a asumir en la estimación). Obviamente, se trata de conceptos complementarios que se refieren a lo mismo.



Intervalo confidencial:

Para hallar el **intervalo confidencial,** es decir, los valores entre los cuales es más probable que se encuentre el verdadero valor del parámetro, necesitaremos calcular la $\bar{X}$, a la que sumaremos y restaremos el **error muestral;** es decir, la diferencia más probable entre el estadístico y el parámetro.

$$\boxed{IC = \bar{x} \pm EM}$$

EM es el ERROR MUESTRAL:

Depende del nivel de confianza con el que estemos trabajando.

En el caso de muestras pequeñas y grandes, $EM = t_{(\alpha/2)} \cdot \sigma_{\bar{X}}$

O como alternativa sólo en caso de muestras grandes $EM = Z_{(\alpha/2)} \cdot \sigma_{\bar{X}}$, donde

El error muestral nos da una idea de la precisión de nuestra inferencia estadística. Cuanto más grande sea el error muestral, menor será la precisión en la estimación y menor será la utilidad de la estimación.

Para calcular el valor muestral previamente hemos tenido que definir el **nivel de significación** α con el que vamos a realizar la estimación, **calcular la puntuación** *Z* correspondiente a ese nivel (porque la distribución es normal) y aplicamos la fórmula:

$$\boxed{EM = Z_{\alpha/2} \cdot \sigma_{\bar{x}}}$$

Una distribución muestral tiene variabilidad, por tanto tendrá su propia **desviación típica** σ**,** que **recibe el nombre de error típico**. Es una medida de dispersión con respecto al parámetro, es decir, nos indica la dispersión de las infinitas muestras aleatorias extraídas respecto a la μ.

ERROR TÍPICO DE LA MEDIA ($\sigma_{\bar{X}}$):

Error típico de una distribución muestral de medias (si en el cálculo de s se utilizó la s insesgada):

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{N}}$$

Error típico de una distribución muestral de medias (si en el cálculo de s se utilizó la s sesgada):

$$\sigma_{\bar{X}} = \frac{s}{\sqrt{N-1}}$$

Así mismo en el (EM) hay que tener en cuenta el nivel de significación escogido por el investigador (normalmente α = 0,05, que se corresponde a un nivel de confianza del 95%)

Un ejemplo:

Supongamos que hemos aplicado un test de cociente intelectual a una muestra aleatoria de 1.000 adolescentes, para estimar el cociente intelectual medio de la población de adolescentes.

N=1.000 ; $\bar{X}$=105; S=10.
Estimar el valor del parámetro μ (intervalo confidencial) con un nivel de confianza del 99%.

IC = ± EM

nivel de confianza → 1 - α = 99% →0,99

nivel de significación → α = 0,01

Por tanto, α/2 = 0,005. El área bajo la curva normal de valor 0,005, corresponde a una z = - 2,575 (columna C)

$$\sigma_{\bar{X}} = \frac{S_x}{\sqrt{N-1}} = \frac{10}{\sqrt{999}} = 0{,}316$$

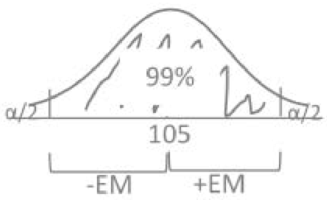
$$IC = 105 \pm Z_{\alpha/2} \cdot \sigma_{\bar{X}} = 105 \pm (-2{,}575)\, 0{,}316 \left\langle \begin{array}{l} 104{,}18 \\ \\ 105{,}81 \end{array} \right.$$

Podemos afirmar con un nivel de confianza del 99%, que el valor de μ se encuentra en el intervalo siguiente: (104,18 ≤ μ ≤ 105,81)

Mismo ejemplo, pero con otra explicación:

N = 1.000 $\bar{X} = 105$ $s_{sesg.} = 10$ $\mu = ?$ Nivel Confianza = 99% = 0,99 α = 0,01

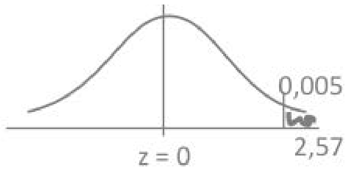
(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
2.25	.4878	.9878	.0122	.0317
2.26	.4881	.9881	.0119	.0310
2.27	.4884	.9884	.0116	.0303
2.28	.4887	.9887	.0113	.0297
2.29	.4890	.9890	.0110	.0290
2.30	.4893	.9893	.0107	.0283
2.31	.4896	.9896	.0104	.0277
2.32	.4898	.9898	.0102	.0270
2.33	.4901	.9901	.0099	.0264
2.34	.4904	.9904	.0096	.0258
2.35	.4906	.9906	.0094	.0252
2.36	.4909	.9909	.0091	.0246
2.37	.4911	.9911	.0089	.0241
2.38	.4913	.9913	.0087	.0235
2.39	.4916	.9916	.0084	.0229
2.40	.4918	.9918	.0082	.0224
2.41	.4920	.9920	.0080	.0219
2.42	.4922	.9922	.0078	.0213
2.43	.4925	.9925	.0075	.0208
2.44	.4927	.9927	.0073	.0203
2.45	.4929	.9929	.0071	.0198
2.46	.4931	.9931	.0069	.0194
2.47	.4932	.9932	.0068	.0189
2.48	.4934	.9934	.0066	.0184
2.49	.4936	.9936	.0064	.0180
2.50	.4938	.9938	.0062	.0175
2.51	.4940	.9940	.0060	.0171
2.52	.4941	.9941	.0059	.0167
2.53	.4943	.9943	.0057	.0163
2.54	.4945	.9945	.0055	.0158
2.55	.4946	.9946	.0054	.0154
2.56	.4948	.9948	.0052	.0151
$\boxed{2.57}$	.4949	.9949	$\boxed{.0051}$	.0147
2.58	.4951	.9951	.0049	.0143



$$\sigma_{\bar{X}} = \frac{10}{\sqrt{999}} = 0{,}32$$

α/2 = 0,005

$$EM = 2{,}57 \cdot 0{,}32 = 0{,}82$$



$$IC = 105 \pm 0{,}82 \left| \begin{matrix} L_{inf} \ 104{,}18 \\ L_{sup} \ 105{,}82 \end{matrix} \right.$$

Estimación del parámetro media aritmética para muestras pequeñas:

Cuando contamos con muestras pequeñas (N < 30), la distribución muestral de la media sigue la distribución ***t* de Student**. La distribución *t* varía en función del número de sujetos (y, en consecuencia, de los grados de libertad), aunque se trata también de una distribución simétrica y asintótica. Cuando *N* tiende a infinito, la distribución *t* tiende a la distribución z.

Solo nos cambia en este caso el estadístico para calcular el error muestral. En lugar de trabajar con $z_{(\alpha/2)}$, lo haremos con $t_{(\alpha/2)}$ (ver tablas).

Ejemplo 2:

Los mismos datos del ejemplo anterior pero ahora con N = 25.

IC = 105 ± EM

α/2 = 0,005. Se distribuye según **t** con N-1 grados de libertad. Buscando en tablas, corresponde a un valor de $t_{\alpha/2} = 2{,}797$

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{24}} = 2.04$$

$$IC = 105 \pm t_{\alpha/2} \cdot \sigma_{\bar{X}} = 105 \pm (2{,}797)\ 2{,}04 \begin{cases} 99{,}29 \\ 110{,}70 \end{cases}$$

Como puede verse, hemos perdido una gran cantidad de precisión al disminuir el tamaño de la muestra, puesto que el intervalo de confianza es ahora mucho mayor.

Otra explicación para el ejemplo 2:

N = 25 $\bar{X} = 105$ $s_{sesg.} = 10$ $\mu = ?$ NC = 99% = 0,99 α = 0,01 α/2 = 0,005

											Cola
α	0,450	0,250	0,200	0,100	0,050	0,025	0,010	$\boxed{0,005}$	0,001	0,0005	una
α	0,900	0,500	0,400	0,200	0,100	0,050	0,020	$\boxed{0,010}$	0,002	0,001	dos
n											n
1	0,158	1,000	1,376	3,078	6,314	12,706	31,821	63,657	318,289	636,619	1
2	0,142	0,816	1,061	1,886	2,920	4,303	6,965	9,925	22,328	31,598	2
3	0,137	0,765	0,978	1,638	2,353	3,182	4,541	5,841	10,214	12,924	3
4	0,134	0,741	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610	4
5	0,132	0,727	0,920	1,476	2,015	2,571	3,365	4,032	5,894	6,869	5
6	0,131	0,718	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959	6
7	0,130	0,711	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7
8	0,130	0,706	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041	8
9	0,129	0,703	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781	9
10	0,129	0,700	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587	10
11	0,129	0,697	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437	11
12	0,128	0,695	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318	12
13	0,128	0,694	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221	13
14	0,128	0,692	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140	14
15	0,128	0,691	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073	15
16	0,128	0,690	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015	16
17	0,128	0,689	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965	17
18	0,127	0,688	0,862	1,330	1,734	2,101	2,552	2,878	3,610	3,922	18
19	0,127	0,688	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883	19
20	0,127	0,687	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850	20
21	0,127	0,686	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819	21
22	0,127	0,686	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792	22
23	0,127	0,685	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767	23
$\boxed{24}$	0,127	0,685	0,857	1,318	1,711	2,064	2,492	$\boxed{2,797}$	3,467	3,745	24

N − 1

$$\sigma_{\bar{X}} = \frac{10}{\sqrt{24}} = 2{,}04$$

$$t_{(\alpha/2)} = 2{,}797$$

$$EM = 2{,}797 \cdot 2{,}04 = 5{,}71$$

$$IC = 105 \pm 5{,}71 \left| \begin{matrix} 99{,}29 \\ 110{,}71 \end{matrix} \right.$$

$$IC = \bar{X} \pm EM \qquad EM = t_{(\alpha/2)} \cdot \sigma_{\bar{X}} \qquad \sigma_{\bar{X}} = \frac{s}{\sqrt{N-1}}$$

Estimación del parámetro proporción(π):

Es un caso particular del anterior, en el que la media oscila entre 0 y 1. En las variables dicotómicas ya vimos que **p** corresponde a la proporción de unos y **q** a la proporción de ceros (q = 1-p). En este caso, la σ_p viene dada por la fórmula:

$$\boxed{\sigma_p = \frac{p.q}{\sqrt{N-1}}}$$

El intervalo de confianza se establece igual que en el caso de la μ, pero ahora partiendo de una proporción. Se emplea el valor de $z_{\alpha/2}$, al nivel de confianza correspondiente.

Ejemplo 3: Estimación del acuerdo con la "dación en pago" de la población de adolescentes de la Comunidad de Madrid.

N = 1.000 $\bar{X} = p = 0{,}65$ $\mu = ?$ NC = 99% = 0,99 α = 0,01 α/2 = 0,005

$$\sigma_p = \frac{\sqrt{0.65 \cdot 0{,}35}}{\sqrt{999}} = 0{,}015 \qquad EM = 2{,}57 \cdot 0{,}015 = 0{,}038 \qquad IC = 0{,}65 \pm 0{,}038 \left| \begin{matrix} 0{,}688 \rightarrow 68{,}8\% \\ 0{,}612 \rightarrow 61{,}2\% \end{matrix} \right.$$

$$IC = \bar{X} \pm EM \qquad EM = t_{(\alpha/2)} \cdot \sigma_{\bar{X}} \qquad \sigma_p = \frac{\sqrt{p \cdot q}}{\sqrt{N-1}}$$

5. ESTIMACIÓN DE LA PUNTUACIÓN VERDADERA EN UNA PRUEBA:

Anteriormente hemos hablado de la *fiabilidad absoluta, que es la que* está más directamente relacionada con lo que conocemos como *error típico de medida* y, en consecuencia, con la teoría de la inferencia estadística. Su utilidad fundamental es la estimación de la *puntuación verdadera* de un sujeto en un instrumento o, dicho de otra forma, entre qué puntuaciones es más probable que se encuentre su verdadera puntuación, ya que toda medida tiene algún margen de error.

De nuevo se trata de hallar el **intervalo de confianza** en el que es probable que se encuentre la verdadera puntuación del sujeto en la prueba. Sabiendo que la distribución muestral es normal, **necesitamos conocer el error típico de medida:**

$$\boxed{\sigma_e = S_t \cdot \sqrt{1 - r_{xx}}}$$

$S_t \rightarrow$ desviación típica total en el instrumento de medida
$r_{xx} \rightarrow$ coeficiente de fiabilidad

$$\boxed{IC = X_i \pm EM}$$

$X_i \rightarrow$ puntuación obtenida en la prueba

$X_1 = 47$ en una prueba de (0 – 100) $\alpha = 0{,}01$ $\alpha/2 = 0{,}005$ Fiabilidad: $r_{xx} = 0{,}85$ $s = 10$

$$\sigma_s = 10 \cdot \sqrt{1 - 0{,}85} = 3{,}87$$

$$IC = 47 \pm z_{(\alpha/2)} \cdot \sigma_s = 47 \pm 2{,}57 \cdot 3{,}87 = 47 \pm 9{,}98 \begin{cases} 56{,}98 \\ 37{,}02 \end{cases}$$

Intervalo de confianza en torno a la puntuación de un sujeto: $IC = X_i \pm EM$

Error típico de medida: $\sigma_s = s_t \cdot \sqrt{1 - r_{xx}}$

6. INTERVALO DE CONFIANZA DE LA PUNTUACIÓN ESTIMADA EN LA REGRESIÓN LINEAL SIMPLE:

Se trata de estimar las puntuaciones en el criterio, conociendo las puntuaciones alcanzadas en la prueba predictora o antecedente, una vez determinado el coeficiente de validez. Esta predicción es más segura y precisa a medida que aumenta el coeficiente de correlación (validez
predictiva o concurrente) entre las variables.

Cuando estimamos las puntuaciones en el criterio (Y) a partir de las puntuaciones en la prueba (X), no tenemos la seguridad total de que la puntuación predicha sea única y siempre la misma. Es decir estamos haciendo una estimación (Y') que conlleva un error:
Error de estimación = Y' – Y

Así, cada predicción lleva asociado un error de estimación.
La desviación típica de los errores de estimación es lo que recibe el nombre de error típico de estimación (σ est).

$$\sigma_{est} = S_y \cdot \sqrt{1 - r_{xy}^2}$$

$$IC = Y' \pm EM$$

*Encontraremos ejemplos en las páginas 170 y 171 del libro.

7. ESTIMACIÓN DEL PARÁMETRO CORRELACIÓN DE PEARSON:

En muchas ocasiones, cuando calculamos una correlación, en realidad estamos más interesados en la relación que existe entre esas variables en la población que en la muestra. Se trata de calcular el intervalo de confianza para la **correlación de Pearson**, de modo que podamos estimar entre qué valores se encuentra dicho coeficiente entre la población. Sabemos que la distribución muestral de la correlación de Pearson se asemeja a la distribución normal con el siguiente **error típico:**

Para muestras grandes (N > 100) → $\sigma_{r_{xx}} = \dfrac{1}{\sqrt{N-1}}$

Para muestras pequeñas → $\sigma_{r_{xx}} = \dfrac{1 - r_{xx}^2}{\sqrt{N-1}}$

Y como ya es habitual $IC = r_{xx} \pm EM$ donde $EM = z_{\alpha/2} \cdot \sigma_{r_{xx}}$

Ejemplo:

Sea N = 20 y r_{xx} = 0,35. Cómo será la correlación en la población con un nivel de confianza del 95%?

$$\sigma_{r_{xx}} = \frac{1-(0{,}35)^2}{\sqrt{20-1}} = \frac{0{,}8775}{4{,}358} = 0{,}201$$

α = 0,05 → $\alpha/2$ = 0,025. En tablas, un área bajo la curva de 0,025 (columna C) corresponde a una z = -1,96

IC = 0,35 ± (-1,96).0,201 < -0,04 ; 0,74

El intervalo tan amplio es debido al pequeño tamaño de la muestra. Cuanto más pequeña sea la muestra, más imprecisa será la estimación.

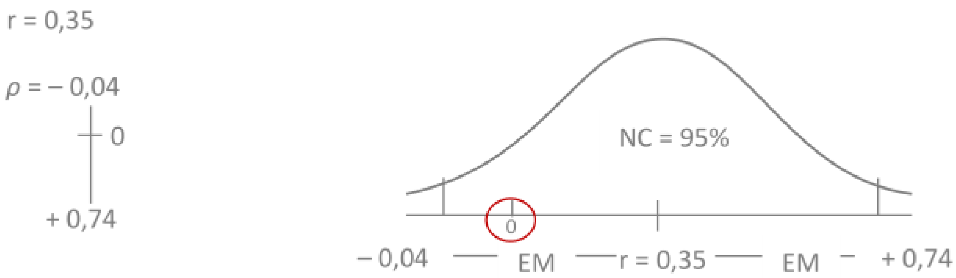
Significación o significatividad estadística:

Hablamos de significación estadística cuando nos referimos al significado de la diferencia entre dos medidas.

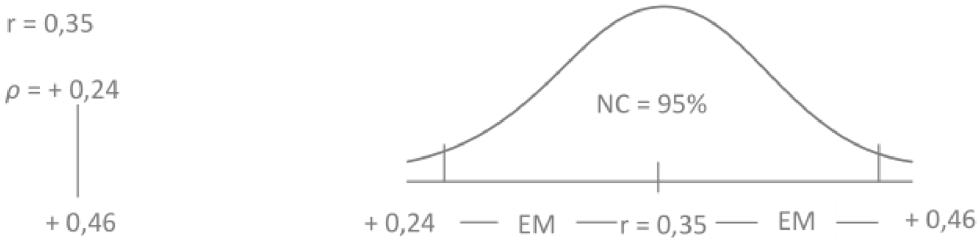
Decimos también que un coeficiente de correlación es estadísticamente significativo cuando es distinto de cero, es decir, cuando el coeficiente de correlación (*estadístico*) es lo suficientemente grande como para decir que la correlación en la población (*parámetro*) de referencia es distinta de cero. Si queremos conocer cuál es su magnitud en la población, entonces calcularemos el intervalo de confianza.

Cuando estimamos el intervalo de confianza en el cual es probable que se encuentre la verdadera correlación en la población, y este intervalo contiene el valor cero (ausencia

…es decir, que estadísticamente hablando, esa correlación es (puede ser) igual a cero y la diferencia con r = 0 que hemos obtenido en la muestra se debe al azar, al error de muestreo.



No es estadísticamente significativo, la correlación obtenida se debe al error de muestreo.



Si es estadísticamente significativo.

absoluta de correlación) diremos que dicha correlación NO es estadísticamente significativa.

Hay que ser cuidadosos en la interpretación, cuando decimos que una correlación no es estadísticamente significativa:

- Una correlación no significativa simplemente es una correlación que no podemos generalizar.
- Una correlación no significativa no es prueba de no relación.

(NO PROBAR QUE HAY RELACIÓN ≠ PROBAR QUE NO HAY RELACIÓN)

8. ESTIMACIÓN DEL PARÁMETRO DIFERENCIA DE MEDIAS ($\mu_1 - \mu_2$):

Si una universidad encuentra que el CI medio de sus estudiantes es de 110, y otra universidad lo tiene de 105, ¿cómo es la diferencia entre estas dos puntuaciones en la población de referencia? ¿La diferencia entre estas dos puntuaciones es estadísticamente igual a cero? Si dicha diferencia es compatible con una diferencia igual a cero, concluimos que ambas muestras tienen el mismo CI, al nivel de confianza fijado, y que la diferencia encontrada es aleatoria.

Si establecemos **el intervalo de confianza** a partir del estadístico diferencia de medias, obtendremos los límites confidenciales entre los cuales es más probable que se encuentre la diferencia de medias en la población.

Si este intervalo **incluye la puntuación cero,** entonces dicha diferencia es compatible con una diferencia de medias igual a cero, y en consecuencia, podremos interpretar que dicha diferencia es estadísticamente igual a cero.

$$\boxed{IC = (\bar{x}_1 - \bar{x}_2) \pm EM} \qquad \boxed{EM = z_{\alpha/2} \cdot \sigma_{(\bar{x}_1 - \bar{x}_2)}}$$

$$\boxed{\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{S_1^2}{N_1 - 1} + \frac{S_2^2}{N_2 - 1}}}$$ para muestras grandes e independientes (N > 100)

$$\boxed{\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\left(\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 + N_2 - 2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$ para muestras pequeñas e independientes (N ≤ 100)

Universidad 1	Universidad 2
$\bar{x}$ = 110	$\bar{x}$ =105
S_1 = 10	S_2 = 12
N_1 = 90	N_2 = 120

Calculamos primero el error típico de la diferencia de medias.

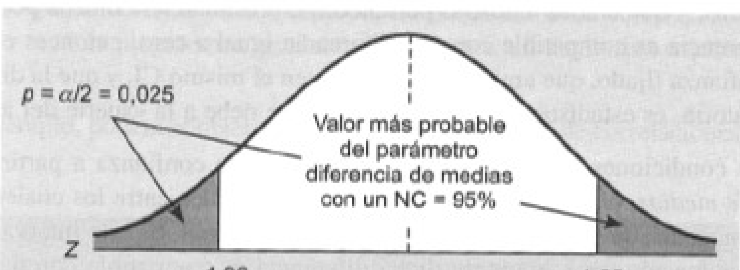
$$\sigma_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{90.100 + 120.144}{90 + 120 - 2}\left(\frac{1}{90} + \frac{1}{120}\right)} = \sqrt{\frac{183\,960}{74880}} = \sqrt{2{,}4567} = 1{,}567$$

La $z_{\alpha/2}$ al 95% de nivel de confianza, ya sabemos que corresponde a una z = - 1,96

EM = (- 1,96) 1,567 = - 3,071 → por tanto:

$$IC = (110 - 105) \pm (-3{,}071) \begin{cases} 1{,}929 \\ 8{,}071 \end{cases}$$

Al ser el intervalo incompatible con el valor cero, concluimos que la diferencia marcada es estadísticamente significativa, es decir, que las diferencias de medias en CI entre ambos grupos no son aleatorias.



Otra explicación diferente del ejemplo:

Ejemplo: página 237 – 238 (Muestras grandes)

$$IC = (\bar{X}_1 - \bar{X}_2) \pm EM$$

$$EM = Z_{(\alpha/2)} \cdot \sigma_{(\bar{X}_1 - \bar{X}_2)}$$

Cociente Intelectual

Harvard	Alcalá
N = 90	N = 120
$\bar{X}_1 = 110$	$\bar{X}_2 = 105$
s = 10	s = 12

NC = 95%

α = 0,05 α/2 = 0,025

$z_{\alpha/2} = 1{,}96$

Muestras grandes

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{N_1 - 1} + \frac{\sigma_2^2}{N_2 - 1}}$$

Muestras grandes o pequeñas

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{90 \cdot 10^2 + 120 \cdot 12^2}{90 + 120 - 2} \cdot \left(\frac{1}{90} + \frac{1}{120}\right)} = 1{,}57$$

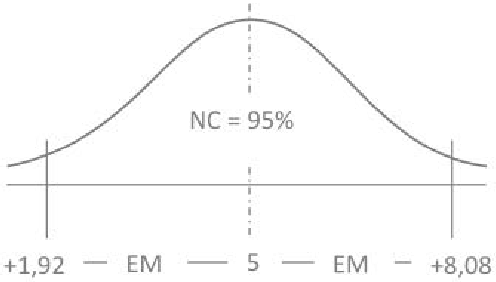
$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{N_1 \cdot \sigma_1^2 + N_2 \cdot \sigma_2^2}{N_1 + N_2 - 2} \cdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

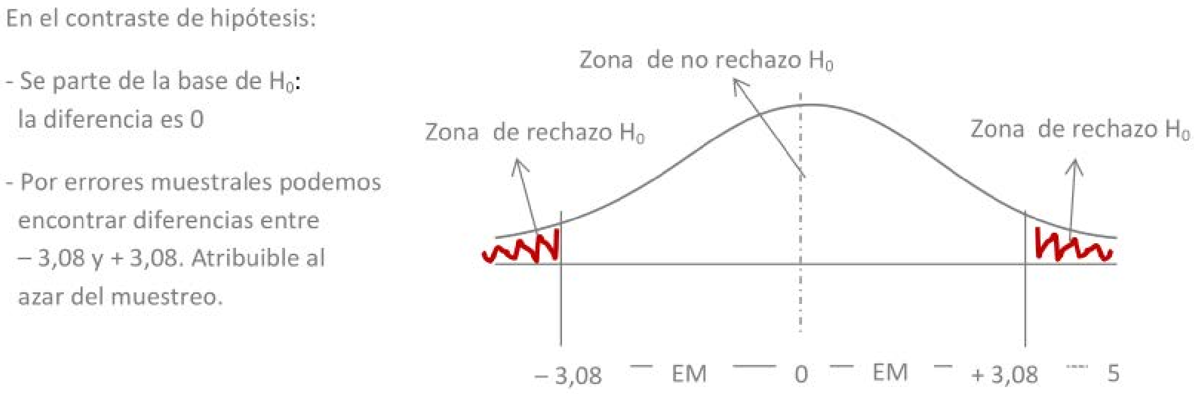
$$EM = 1{,}96 \cdot 1{,}57 = 3{,}08$$

Se debe usar la σ (s) sesgada

$$IC = 5 \pm 3{,}08 \begin{vmatrix} +1{,}92 \\ +8{,}08 \end{vmatrix}$$

- No incluye el valor “0”
- La diferencia hallada sí es estadísticamente significativa.
- Podemos rechazar H_0





Estimación del parámetro diferencia de proporciones ($\pi_1 - \pi_2$):

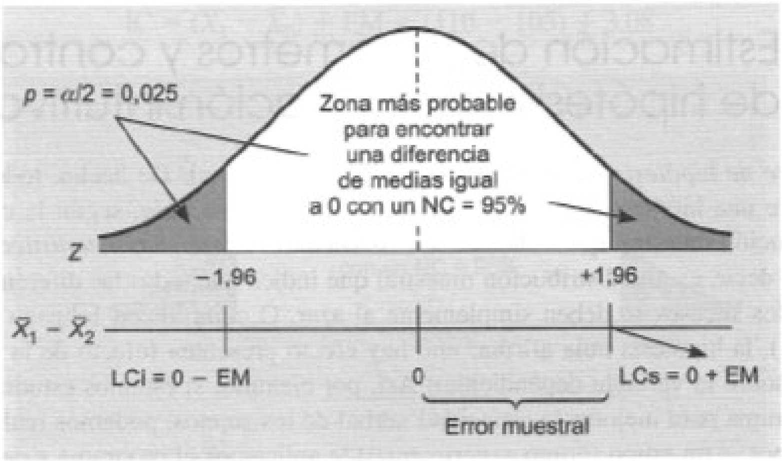
Lo único que varía en este caso es el error típico. Al igual que en el contraste de medias, lo más interesante suele ser la diferencia entre dos proporciones.

El error típico de la diferencia de proporciones es: $\boxed{\sigma_{(p_1-p_2)} = \sqrt{pq.\left(\frac{1}{N_1}+\frac{1}{N_2}\right)}}$

Y la estimación del intervalo de confianza y su interpretación será la misma que en el caso anterior, pero con medias en términos de proporción, que frecuentemente se multiplican por 100 para convertirlos en porcentajes.

9. ESTIMACIÓN DE PARÁMETROS Y CONSTRASTE DE HIPÓTESIS: INTERPRETACIÓN INTUITIVA:

Todo se reduce al contraste de una hipótesis estadística, denominada **hipótesis nula (Ho),** según la cual NO existen diferencias estadísticamente significativas. Se afirma en ella que no hay efecto presente de la variable independiente sobre la variable dependiente.



Se establece una zona central alrededor de la media en la curva normal, en la que es más probable que las diferencias encontradas entre las medias de las muestras se deban efectivamente a los efectos del azar. Es la **zona de aceptación** (ó de no rechazo) **de Ho.** Por tanto, al no poder rechazar la Ho, decimos que no es falsa.

Queda de forma inmediata **otra zona** (bilateral o unilateralmente) en la que las diferencias entre las medias resulta muy improbable (al nivel de confianza establecido) que sean aleatorias, por lo que dichas diferencias se atribuyen a los efectos de la VI.

La H_0 se expresa así:

> **Hipótesis nula → H_0 : μ1 - μ2 = 0** *las diferencias entre las de los grupos o muestras son estadísticamente iguales a cero; o bien, que las diferencias empíricas que existen entre las de las muestras se deben al azar; o bien, que los valores paramétricos son iguales; o bien, que ambas muestras pertenecen a la misma población.*

Las fórmulas para hallar **el intervalo de confianza** son las mismas que el epígrafe anterior, para muestras grandes y pequeñas.

Si resulta que nuestra diferencia empírica de medias se encuentra dentro del intervalo de confianza de la distribución muestral conforme a H_0, diremos que nuestra diferencia de medias es compatible con una diferencia de medias igual a cero, y por tanto, que se trata de una diferencia estadísticamente no significativa o igual a cero.

Lo que haremos después será **calcular un estadístico** (t, *z,* etc.) que nos dirá cuántas desviaciones típicas (errores típicos) se aleja nuestra diferencia de medias de una diferencia de medias igual a cero.

Tema 12 INTRODUCCIÓN AL CONTRASTE DE HIPÓTESIS.

Introducción.

En el tema anterior vimos el proceso para **establecer un intervalo de confianza para estimar el parámetro desconocido de una población a partir de la información conocida de una muestra**…

… ahora veremos como los valores obtenidos en una muestra pueden ser usados para probar una hipótesis en la población, a través del uso de la **prueba de significación de la hipótesis nula**.

Hipótesis nula.

La **Hipótesis Nula** **(H_0)** representa la **no relación** entre las variables que estamos estudiando…

…H_0 es la hipótesis de no diferencias (no hay diferencias en los niveles de la VD en función de la VI). O dicho en otras palabras, que la VI no produce ningún efecto en la VD.

La **Hipótesis Nula** **(H_0)** tiene 2 propósitos básicos:

- Sirve como punto de partida cuando no tenemos conocimiento o no hay razones para creer que existen diferencias entre los grupos que estamos comparando.
- Ayuda a definir un intervalo en el que cualquier diferencia observada entre grupos puede atribuirse a la casualidad o azar; y otro intervalo de valores en que dicha diferencia quizás se deba a otro factor distinto del azar, como el efecto de la VI en la VD.

¿Rechazar la hipótesis nula?

El investigador debe tomar la decisión de si rechazar o no rechazar H_0.

(Y, por tanto, si aceptar o no aceptar la Hipótesis Alternativa H_1).

Esa decisión dependerá del valor de **α** (nivel de significación) que hayamos establecido **a priori**.

RECHAZO H_0 SI…

- Probabilidad ($z_{empírica}$, $t_{empírica}$…) ≤ α.
- $z_{empírica}$, $t_{empírica}$… ≥ $z_{crítica}$, $t_{crítica}$.
- El intervalo de confianza **no** incluye la diferencia empírica.

Estimación de Parámetro diferencias de medias (grupos grandes e independientes).

Ejemplo I:

Grupo I: HARVARD	Grupo II: ALCALÁ
$X_1 = 112$	$X_2 = 104$
$s_1 = 15$	$s_2 = 13$
$N_1 = 150$	$N_2 = 200$

$$IC = (\bar{X}_1 - \bar{X}_2) \pm EM$$

$$EM = Z_{(\alpha/2)} \cdot \sigma_{(\bar{X}_1 - \bar{X}_2)}$$

$N.C. \rightarrow 99\% \quad \alpha = 0{,}01 \quad \alpha/2 = 0{,}005 \quad Z\alpha_{/2} = 2{,}58$

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{\sigma_1^2}{N_1 - 1} + \frac{\sigma_2^2}{N_2 - 1}}$$

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor
2.55	.4946	.9946	.0054
2.56	.4948	.9948	.0052
2.57	.4949	.9949	.0051
2.58	.4951	.9951	.0049
2.59	.4952	.9952	.0048

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\frac{15^2}{150 - 1} + \frac{13^2}{200 - 1}} = 1{,}54$$

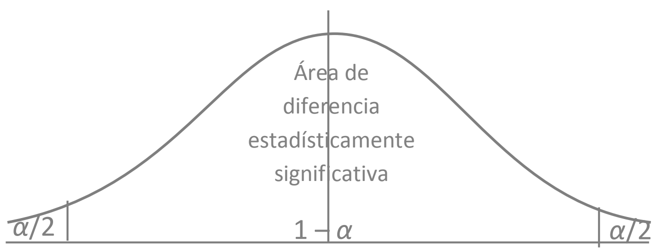
$$EM = 2{,}58 \cdot 1{,}54 = 3{,}97$$

Se podría coger cualquiera de las dos puntuaciones típicas.

$$IC = (112 - 104) \pm 3{,}97 = 8 \pm 3{,}97 \begin{vmatrix} L_i = +4{,}03 \\ L_s = +11{,}97 \end{vmatrix}$$

Como el intervalo no incluye la diferencia nula rechazamos la hipótesis nula.

Dibujo desde el punto de vista del intervalo de confianza:



z — 2,58 0 + 2,58

$X_1 - X_2$ 0 —— + 4,03 —— +8 —— + 11,97

Contraste de Hipótesis diferencia de medias (grupos grandes e independientes).

Ejemplo I:

$H_0 \rightarrow \mu_1 = \mu_2 \qquad \mu_1 - \mu_2 = 0$

$H_1 \rightarrow \mu_1 \neq \mu_2$

$$Z = \frac{|\bar{X}_1 - \bar{X}_2| - 0}{\sigma_{\bar{X}_1 - \bar{X}_2}}$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{15^2}{150-1} + \frac{13^2}{200-1}\right)} = 1{,}54$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{\sigma_1^2}{N_1 - 1} + \frac{\sigma_2^2}{N_2 - 1}\right)}$$

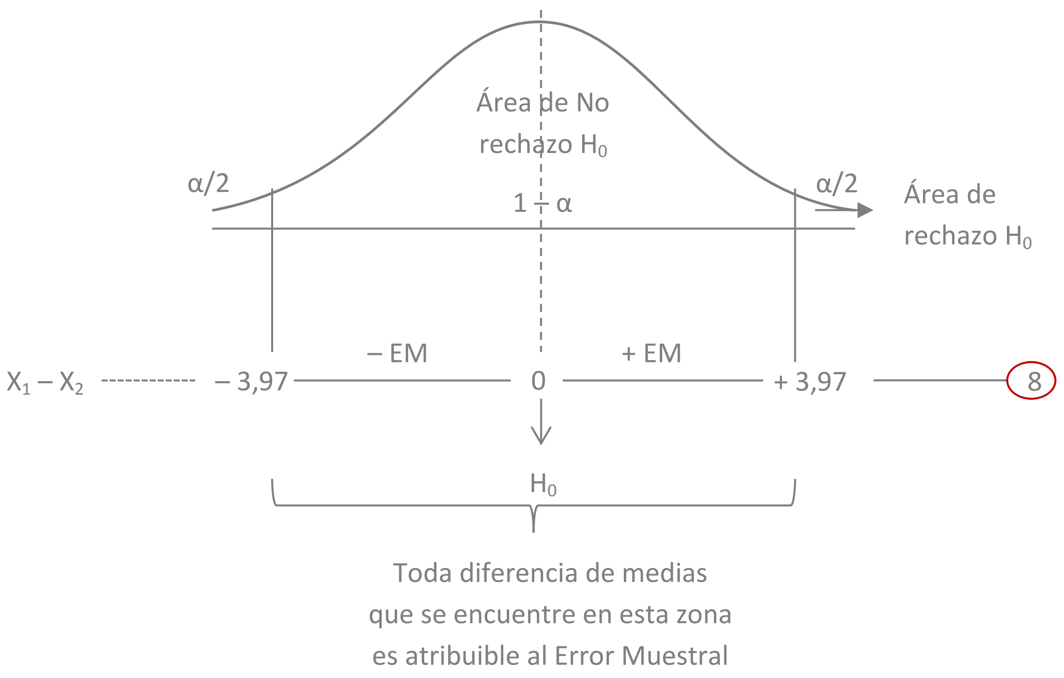
$$Z_{empírica} = \frac{|112 - 104| - 0}{1{,}54} = \frac{8}{1{,}54} = 5{,}19$$

H_0 se rechaza si:

- $Z_{empírica} > Z_{crítica} \rightarrow 5{,}19 > 2{,}58$
- $p(Z_{empírica}) \leq \alpha/2 \rightarrow 0{,}0001 < 0{,}005$

Se sacan de la tabla de las áreas de la curva de distribución normal.

- El IC no contiene la diferencia cero.



Contraste de Hipótesis en los diseños de dos grupos.

1. Formulación del problema
2. Revisión bibliográfica del estado de la cuestión
3. Definición de variables
4. Formulación de hipótesis
5. Diseño de investigación
6. Contraste estadístico de hipótesis:
   6.1. Formulación de las hipótesis estadísticas
   - Hipótesis nula: H_0
   - Hipótesis alterna: H_1

   6.2. Elección de la prueba estadística adecuada
   6.3. Establecimiento del nivel de significación (α)
   6.4. Definición de la distribución muestral respecto de H_0
   6.5. Establecimiento de la región de rechazo de H_0
   6.6. Decisión estadística (rechazo o no rechazo de H_0)
7. Decisión práctica (magnitud del efecto y conclusiones generales)

5. Diseño de la investigación.

Diseño (cuasi) experimental de 2 grupos independientes.

Una única medida post-test.

Variable Independiente (VI): Programa.

Variable Dependiente (VD): Inteligencia Emocional (IE).

6.1. Formulación de las hipótesis estadísticas

- Hipótesis nula: H_0
- Hipótesis alterna: H_1

$$H_0 \quad \rightarrow \quad \mu_E - \mu_C = 0 \qquad \mu_E = \mu_C$$

$$H_1 \quad \rightarrow \quad \mu_E > \mu_C \qquad \mu_E - \mu_C > 0$$

6.2. Elección de la prueba estadística adecuada

Asumimos:

- Independencia de las observaciones. (Muestreo aleatorio).
- La VD está a un nivel de intervalo y se distribuye normalmente.
- Homocedasticidad: igualdad de las varianzas de las poblaciones.

Como esto se cumple: Elegimos la prueba t de diferencia de medias entre dos grupos independientes.

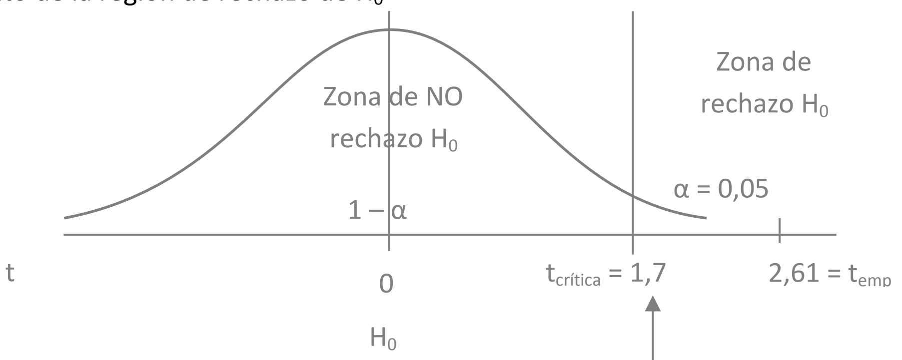
6.3. Establecimiento del nivel de significación (α)

$$\alpha = ? \rightarrow \alpha = 0{,}05$$

6.4. Definición de la distribución muestral respecto H_0

La distribución muestral conforme a la H_0 estaría formada por los infinitos valores de t obtenidos en infinitas muestras aleatorias del tamaño dado extraídas de la misma población.

6.5. Establecimiento de la región de rechazo de H_0



					0,007					Colas
α	0,450	0,250	0,200	0,100	0,050	0,025	0,010	0,005	0,001	0,0005	una
α	0,900	0,500	0,400	0,200	0,100	0,050	0,020	0,010	0,002	0,001	dos
n											n
24	0,127	0,685	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745	24
25	0,127	0,684	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725	25
26	0,127	0,684	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707	26
27	0,127	0,684	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690	27
$\boxed{28}$	0,127	0,683	0,855	1,313	$\boxed{1,701}$	2,048	2,467	2,763	3,408	3,674	28
29	0,127	0,683	0,854	1,311	1,699	2,045	2,462	2,756	3,396	3,659	29
30	0,127	0,683	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646	30

$n \rightarrow g.l.(grados\ de\ libertad) \rightarrow N_1 + N_2 - 2$

6.6. Decisión estadística (rechazo o no rechazo H_0)

Grupo EXPERIMENTAL	Grupo CONTROL
$X_1 = 67,3$	$X_2 = 55,0$
$s_1 = 11,8$	$s_2 = 13,2$
$N_1 = 15$	$N_2 = 15$

$NC = 95\% \quad \alpha = 1 - 0,95 = 0,05$

$t_{(g.l.=28;\alpha=0,05)} = 1,7$

$$t_{emp} = \frac{|\bar{X}_1 - \bar{X}_2| - 0}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{|67,3 - 55|}{4,7} = 2,61 \quad \rightarrow \quad p(t_{emp}) = 0,007 \rightarrow \text{Calculado con excel}$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\left(\frac{N_1 \cdot s_1^2 + N_2 \cdot s_2^2}{N_1 + N_2 - 2}\right) \cdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)} = \sqrt{\left(\frac{15 \cdot (11,8)^2 + 15 \cdot (13,2)^2}{15 + 15 - 2}\right) \cdot \left(\frac{1}{15} + \frac{1}{15}\right)} = 4,7$$

$$H_0 \text{ se rechaza} \quad \begin{cases} t_{emp} > t_{crit} \rightarrow 2{,}6 > 1{,}7 \\ p(t_{emp}) \leq 0{,}05 \rightarrow 0{,}007 < 0{,}005 \end{cases}$$

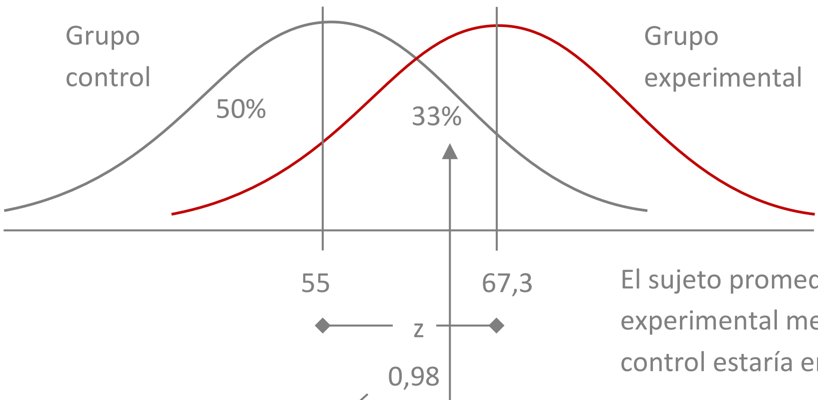
Aceptamos $H_1 \rightarrow \mu_E > \mu_C$

	A	B	C
1	Prueba t para dos muestras suponiendo varianzas iguales		
2			
3		*Grupo Experimental*	*Grupo Control*
4	Media	67,33333333	55
5	Varianza	149,5238095	185,7142857
6	Observaciones	15	15
7	Varianza agrupada	167,6190476	
8	Diferencia hipotética de las medias	0	
9	Grados de libertad	28	
10	Estadístico t	2,608851846	
11	P(T<=t) una cola	0,007207323	
12	Valor crítico de t (una cola)	1,701130934	
13	P(T<=t) dos colas	0,014414646	
14	Valor crítico de t (dos colas)	2,048407142	

7. Decisión práctica (magnitud del efecto y conclusiones generales)

$$d = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{combinada}} = \frac{67{,}3 - 55}{12{,}5} = \frac{12{,}3}{12{,}5} = 0{,}98$$

$$\sigma_c = \sqrt{\frac{(N_1 \cdot s_1^2) + (N_2 \cdot s_2^2)}{N_1 + N_2}} = \sqrt{\frac{[15 \cdot (11{,}8)^2] + [(15 \cdot (13{,}2)^2)]}{15 + 15}} = 12{,}5$$



El sujeto promedio del grupo experimental metido en el grupo control estaría en un percentil 83

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor	(5) y Ordenada en $\frac{x}{\sigma}$
0.90	.3159	.8159	.1841	.2661
0.91	.3186	.8186	.1814	.2637
0.92	.3212	.8212	.1788	.2613
0.93	.3238	.8238	.1762	.2589
0.94	.3264	.8264	.1736	.2565
0.95	.3289	.8289	.1711	.2541
0.96	.3315	.8315	.1685	.2516
0.97	.3340	.8340	.1660	.2492
$\boxed{0.98}$	$\boxed{.3365}$	.8365	.1635	.2468
0.99	.3389	.8389	.1611	.2444

d ≤ .20 efecto *pequeño*
d ≈ .50 efecto *moderado*
d ≥ .80 efecto *grande*

Se ha obtenido un efecto grande.

20) Una diferencia de medias no estadísticamente significativa, junto con un tamaño del efecto elevado es más fácil que se dé cuando:

a) Las muestras son grandes.
b) Las muestras son pequeñas.
c) No se puede dar esta situación.

20) Afirmar que una diferencia es estadísticamente significativa equivale a decir:

a) Que tales diferencias se explican por efecto del azar.
b) Que tales diferencias son relevantes para la investigación.
c) La probabilidad de que se deban al azar es igual o menor que el nivel de significación fijado por el investigador.

11) En un contraste de medias, rechazamos la hipótesis nula cuando:

a) La probabilidad asociada al estadístico que estandariza las diferencias entre los grupos es menor o igual que alfa.
b) La probabilidad asociada al estadístico que estandariza las diferencias entre los grupos es mayor o igual que alfa.
c) Las diferencias entre las medias empíricas sea grande.

<u>Nivel de significación estadística y errores Tipo I y Tipo II.</u>

		Lo que he hecho sucede en la población	
		H_0 es verdadera	H_0 es falsa
Decisión del investigador	Rechazar H_0	Error tipo I (α): nivel de significación Falso positivo	Decisión correcta (1 – β): potencia estadística
	No rechazar H_0	Decisión correcta (1 – α)	Error tipo II (β) Falso negativo

Tabla 1. Resultados en la decisión estadística.

La **<u>potencia estadística</u> (1 – β)** se define como la probabilidad de rechazar la hipótesis nula cuando ésta es realmente falsa (decisión correcta). Es deseable alcanzar una potencia de la prueba de 0,80 (es decir, $\beta \leq 0{,}20$).

Debemos tratar de mantener ambos tipos de errores tan bajos como sea posible:

- El error tipo I (α) lo fija a priori el investigador, siendo los valores convencionales utilizados de forma más frecuente 0,01 y 0,05.
- Dado un valor de α fijado, podemos calcular el tamaño que debe tener nuestra muestra para alcanzar un determinado valor de β.

24) Se denomina "potencia de una prueba" a:

a) La probabilidad de rechazar la hipótesis nula cuando es verdadera.
b) La probabilidad de rechazar la hipótesis nula cuando es falsa.
c) La probabilidad de aceptar la hipótesis nula cuando es falsa.

Estimación de Parámetros diferencia de medias (grupos pequeños y grandes, e independientes).

Ejemplo 3:

Grupo I: VETERANOS	Grupo II: NOVATOS
$X_1 = 7{,}5$	$X_2 = 5{,}2$
$s_1 = 1{,}2$	$s_2 = 1{,}3$
$N_1 = 12$	$N_2 = 12$

$$IC = (\bar{X}_1 - \bar{X}_2) \pm EM$$

$$EM = t_{(\alpha/2)} \cdot \sigma_{(\bar{X}_1 - \bar{X}_2)}$$

$NC = 99\%$ $\quad \alpha = 0{,}01$ (2 colas) $\quad \alpha/2 = 0{,}005$ (1 cola)

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\left(\frac{N_1 \cdot \sigma_1^2 + N_2 \cdot \sigma_2^2}{N_1 + N_2 - 2}\right) \cdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}$$

$$\sigma_{(\bar{X}_1 - \bar{X}_2)} = \sqrt{\left(\frac{12 \cdot (1{,}2)^2 + 12 \cdot (1{,}3)^2}{12 + 12 - 2}\right) \cdot \left(\frac{1}{12} + \frac{1}{12}\right)} = 0{,}53$$

$$EM = 2{,}82 \cdot 0{,}53 = 1{,}49$$

$$IC = (7{,}5 - 5{,}2) \pm 1{,}49 = 2{,}3 \pm 1{,}49 \left| \begin{array}{l} L_{inf} = +0{,}81 \\ L_{sup} = +3{,}79 \end{array} \right.$$

											Colas
α			0,200	0,100	0,050	0,025	0,010	$\boxed{0{,}005}$	0,001	0,0005	una
α	0,900	0,500	0,400	0,200	0,100	0,050	0,020	$\boxed{0{,}010}$	0,002	0,001	dos
n											n
1	0,158	1,000	1,376	3,078	6,314	12,706	31,821	63,657	318,289	636,619	1
2	0,142	0,816	1,061	1,886	2,920	4,303	6,965	9,925	22,328	31,598	2
3	0,137	0,765	0,978	1,638	2,353	3,182	4,541	5,841	10,214	12,924	3
4	0,134	0,741	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610	4
5	0,132	0,727	0,920	1,476	2,015	2,571	3,365	4,032	5,894	6,869	5
6	0,131	0,718	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959	6
7	0,130	0,711	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,408	7
8	0,130	0,706	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041	8
9	0,129	0,703	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781	9
10	0,129	0,700	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587	10
11	0,129	0,697	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437	11
12	0,128	0,695	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318	12
13	0,128	0,694	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221	13
14	0,128	0,692	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140	14
15	0,128	0,691	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073	15
16	0,128	0,690	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015	16
17	0,128	0,698	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965	17
18	0,127	0,688	0,862	1,330	1,734	2,101	2,552	2,878	3,610	3,922	18
19	0,127	0,688	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883	19
20	0,127	0,687	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850	20
21	0,127	0,686	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819	21
$\boxed{22}$	0,127	0,686	0,858	1,321	1,717	2,074	2,508	$\boxed{2{,}819}$	3,505	3,792	22
23	0,127	0,685	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767	23

Como en el intervalo no está la diferencia cero la diferencia de medias es estadísticamente significativa.

Podemos rechazar H_0

Contraste de Hipótesis diferencia de medias (grupos pequeños y grandes, e independientes).

Ejemplo 3:

$H_0 \rightarrow \mu_1 = \mu_2 \quad ; \quad \mu_1 - \mu_2 = 0$

$H_1 \rightarrow \mu_1 \neq \mu_2$

$$t = \frac{|\bar{X}_1 - \bar{X}_2| - 0}{\sqrt{\left(\frac{N_1 \cdot \sigma_1^2 + N_2 \cdot \sigma_1^2}{N_1 + N_2 - 2}\right) \cdot \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

$$t_{emp} = \frac{|7{,}5 - 5{,}2|}{0{,}53} = 4{,}33$$

H_0 se rechaza si: $t_{emp} > t_{crit} \rightarrow 4{,}33 > 2{,}82$

$$p(t_{emp}) \leq \alpha \rightarrow \begin{matrix} 0{,}001 < 0{,}01\ (\alpha) \\ 0{,}0005 < 0{,}005\ (\alpha/2) \end{matrix}$$

23) Si se formula la hipótesis: H_0: μ = 4 frente a H_1: μ < 4

a) La hipótesis es bilateral.

b) La hipótesis cubre el parámetro.

c) La hipótesis es unilateral izquierda.

$\neq \rightarrow$ *bilateral*
$> \rightarrow$ *unilateral derecha*
$< \rightarrow$ *unilateral izquierda*

Errores frecuentes en la PSHN.

La PSHN (Prueba de significación de la hipótesis nula) se basa en la asunción de una serie de condiciones:

- Se ha realizado un muestreo aleatorio de N observaciones independientes. (Supuesto de independencia).
- Las puntuaciones de la variable dependiente son cuantitativas y su nivel de medida es al menos de intervalo.
- Es deseable que la distribución de la variable dependiente sea normal. (Supuesto de normalidad).
- El investigador debe seguir rigurosamente los pasos que hemos definido en el epígrafe anterior. (Contraste de hipótesis).
- El investigador debería realizar sólo uno o un número limitado de contrastes en el mismo estudio.

En la investigación educativa estas condiciones no se cumplen estrictamente en muchas ocasiones → se acude a la evidencia acumulativa por replicación de los estudios.

Aunque habitualmente se habla de rechazar o aceptar la hipótesis nula, en realidad debemos de hablar de rechazar o no rechazar la H_0

Esta distinción de matiz (entre "aceptar" y "no rechazar" la hipótesis nula) es importante porque puede haber muchas razones que expliquen que un resultado ha sido estadísticamente significativo, cuando en realidad sí lo es.

UN REDUCIDO TAMAÑO DE LA MUESTRA.

Otra explicación factible es que se eligieron o no se manipularon bien los niveles de la variable independiente…

…También debe considerarse la posibilidad de que nuestra variable independiente correlacione altamente con otra variable independiente no tenida en cuenta y que sea ésta última, en realidad, la que está produciendo los efectos sobre la variable dependiente que atribuimos a la primera.

El Tamaño del Efecto.

Una diferencia estadísticamente significativa **no** nos dice que tal diferencia sea importante; sino tan sólo que existen diferencias en la población de referencia, aunque dicha diferencia sea en la práctica irrelevante…

…por tanto, la significación estadística por sí sola **no** nos dice gran cosa como para tomar decisiones en la práctica; ni nos permite comparar resultados entre distintas investigaciones cuando la unidad de medida cambia.

Para solucionar este problema **se utiliza como indicador complementario a la significación estadística el tamaño del efecto, siendo la "d" de Cohen el índice más utilizado**.

Ejemplo 3:

Grupo I: VETERANOS	Grupo 2: NOVATOS
$X_1 = 7{,}5$	$X_2 = 5{,}2$
$s_1 = 1{,}2$	$s_2 = 1{,}3$
$N_1 = 12$	$N_2 = 12$

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{combinada}}$$

$$\sigma = \sqrt{\frac{(N_1 \cdot s_1^2) + (N_2 \cdot s_2^2)}{N_1 + N_2}}$$

$$\sigma = \sqrt{\frac{(12 \cdot (1{,}2)^2) + (12 \cdot (1{,}3)^2)}{12 + 12}} = 1{,}25$$

$$d = \frac{7{,}5 - 5{,}2}{1{,}25} = \frac{2{,}3}{1{,}25} = 1{,}84 \rightarrow efecto\ grnade$$

d ≤ .20 efecto *pequeño*
d ≈ .50 efecto *moderado*
d ≥ .80 efecto *grande*

(1) z Puntuación tipificada $\left(\frac{x}{\sigma}\right)$	(2) A Area desde la media a $\frac{x}{\sigma}$	(3) B Area de la parte mayor	(4) C Area de la parte menor
1.80	.4641	.9641	.0359
1.81	.4649	.9649	.0351
1.82	.4656	.9656	.0344
1.83	.4664	.9664	.0336
$\boxed{1.84}$	.4671	$\boxed{.9671}$	.0329
1.85	.4648	.9678	.0322
1.86	.4686	.9686	.0314
1.87	.4693	.9693	.0307
1.88	.4699	.9699	.0301

El sujeto medio del grupo de veteranos metido en el grupo de novatos sería un percentil 97

Supuestos de las Pruebas Paramétricas.

6.2. Elección de la prueba estadística adecuada.

Una vez que está claro el tipo de diseño, hay que elegir la prueba de contraste más adecuada. Existen dos tipos de pruebas: las paramétricas y las no paramétricas. Las primeras son las más potentes para rechazar H_0, por lo que son las preferibles. Pero para utilizarlas, se deben cumplir los siguientes supuestos (partiendo del supuesto de ***independencia de las observaciones*** que incumplimos casi siempre en Ciencias de la Educación y que se refiere a que se realizó muestreo aleatorio en la selección de las muestras, como ya vimos al hablar de la PSHN):

1. ***Nivel de medida de la Variable Dependiente***: debe ser de **Intervalo** o **Razón**.
   En Ciencias Sociales se acepta también el nivel *cuasi-intervalo* (por ejemplo, las escalas tipo Likert que indican el grado de acuerdo con respecto a una afirmación: *"Valore de 1 a 6 su grado de acuerdo con las siguientes afirmaciones…").*
2. ***Normalidad de la distribución en la población:*** La distribución de la Variable Dependiente en la población se distribuye según ***la curva normal.*** Se comprueba con el test de χ^2 o el de Kolmogorov-Smirnov.
3. ***Homocedasticidad de las varianzas poblacionales:*** Igualdad en las varianzas, es decir, los dos grupos que comparamos tienen varianzas estadísticamente iguales o pertenecen a la mima población de varianza σ^2. Se prueba con **el test F de Snedecor o con la F de Levene**.

25) Una de las condiciones fundamentales exigidas para aplicar la prueba paramétrica t, es:

a) La linealidad.

b) La homocedasticidad.

c) La potencia.

Supuestos de las Pruebas Paramétricas: NORMALIDAD.

Prueba de Kolmogorov-Smirnov para una muestra

		VAR00003
N		15
Parámetros normales[a,b]	Media	67,3333
	Desviación típica	12,22799
Diferencias más extremas	Absoluta	,110
	Positiva	,110
	Negativa	-,091
Z de Kolmogorov-Smirnov		,426
Sig. asintót. (bilateral)		,993

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Grupo experimental.

H_0 → Nuestra distribución se ajusta a la distribución normal.

Como p ($Z_{K\text{-}S}$) = 0,993 > α (0,05) → no se rechaza H_0, y aceptamos el supuesto de normalidad de la distribución.

Prueba de Kolmogorov-Smirnov para una muestra

		VAR00004
N		15
Parámetros normales[a,b]	Media	55,0000
	Desviación típica	13,62770
Diferencias más extremas	Absoluta	,176
	Positiva	,102
	Negativa	-,176
Z de Kolmogorov-Smirnov		,684
Sig. asintót. (bilateral)		,738

a. La distribución de contraste es la Normal.

b. Se han calculado a partir de los datos.

Grupo control.

H_0 → Nuestra distribución se ajusta a la distribución normal.

Como p ($Z_{K\text{-}S}$) = 0,738 > α (0,05) → no se rechaza H_0, y aceptamos el supuesto de normalidad de la distribución.

<u>Supuestos de las Pruebas Paramétricas: HOMOCEDASTICIDAD.</u>

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
VAR00001	Se han asumido varianzas iguales	,363	,551	2,609	28	,014	12,333	4,727	2,649	22,017
	No se han asumido varianzas iguales			2,609	27,677	,014	12,333	4,727	2,644	22,022

H_0 → Ambos grupos tienen igualdad de varianza.

Como p (F) = 0,551 > α (0,05) → no se rechaza H_0, y aceptamos el supuesto de homocedasticidad.

	A		
1	Prueba F para varianzas de dos muestras		
2			
3		*Grupo Experimental*	*Grupo Control*
4	Media	67,33333333	55
5	Varianza	149,5238095	185,7142857
6	Observaciones	15	15
7	Grados de libertad	14	14
8	F	0,805128205	
9	P(F<=f) una cola	0,345331125	
10	Valor crítico para F (una cola)	0,402620943	
11			
12			

H_0 → Ambos grupos tienen igualdad de varianza.

Como p (F) = 0,345 > α (0,05) → no se rechaza H_0, y aceptamos el supuesto de homocedasticidad.