

<p style="text-align: center;">UOC ESTUDIOS DE INFORMÁTICA Estadística Primavera 2014</p>

PEC – 1. Grado

Fecha de propuesta: 2-3-2014. Fecha de entrega: 23-3-2014.

Instrucciones

- **El nombre del fichero que nos adjuntéis con la respuesta tiene que seguir la regla siguiente:** 3 primeras letras primer apellido + 3 primeras letras segundo apellido + “1”. Por ejemplo, un estudiante que se llame **Jorge Gratacos Pellicer** tendría que enviar un fichero con el nombre: **GRAPEL1.doc** (o la extensión que corresponda).
- Sed breves. No es necesario escribir mucho.
- **No deis sólo el resultado de los ejercicios**, hace falta explicar el motivo de los razonamientos que utilizáis.
- Poned vuestro nombre completo dentro de la prueba.
- Tenéis que enviar la solución al buzón de **"Entrega de actividades"**.
- Tiempo previsto: cinco horas para "realizar" la PEC y dos horas para pasarla al ordenador.
- Dad los resultados numéricos redondeando a dos decimales si no se indica lo contrario..

Apellidos y Nombre

LA PAC CONSTA DE 4 CUESTIONES Y DOS PROBLEMAS QUE REQUIEREN EL USO DE R.

CUESTIÓN 1 (15%)

Tenemos los siguientes datos correspondientes al tiempo de CPU (en segundos) que ha consumido una máquina con Windows8 para ejecutar el llamado algoritmo de Spiess en una red de transporte de tamaño grande:

14.83, 11.03, 20.02, 155.11, 121.32, 16.59, 69.39, 121.32

- Calculad su media.
- Calculad su desviación típica y la varianza.
- Calculad la mediana y los cuartiles.
- Haced a mano algún gráfico que sirva para ilustrar estos datos.
- Calculad en cuánto aumentan la media, la desviación y la mediana si sustituimos el valor más grande por 3000 y justificad el resultado obtenido.

Indicaciones: Todos los apartados valen lo mismo y se valorará la claridad en los cálculos y en el uso de las fórmulas. Las justificaciones tienen que estar argumentadas y apoyadas por los resultados numéricos correspondientes. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con un 0.

CUESTIÓN 2 (15%)

Disponemos de los siguientes datos correspondientes al número a veces que se han ejecutado unos ciertos programas y el número a veces que se llama al algoritmo de Spiess en cada

ejecución:

Programa	Ejecuciones	Algoritmo de Spiess
RedGrande	1000	100
RedMedia	1200	75
RedPequeña	23	34

Queremos estudiar ahora la variable que recoge el número a veces que se llama el algoritmo de Spiess cuando se ejecutan los programas:

- Calculad su media.
- Calculad su desviación típica y la varianza.
- Calculad la moda, la mediana y los cuartiles.
- Haced a mano algún gráfico que sirva para ilustrar estos datos.
- Calculad en cuánto aumentan la media, la desviación típica, la mediana, los cuartiles y la moda si suponemos ahora que en las ejecuciones sobre la red pequeña el algoritmo de Spiess se llama 3000 veces; justificad los resultados obtenidos.

Indicaciones: Todos los apartados valen lo mismo y se valorará la claridad en los cálculos y en el uso de las fórmulas. Las justificaciones tienen que estar argumentadas y apoyadas por los resultados numéricos correspondientes. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con un 0.

CUESTIÓN 3 (10%)

Indicad razonadamente de qué manera podríamos obtener una muestra de 1000 individuos para determinar la marca y el sistema operativo de los teléfonos móviles entre los estudiantes de

- Una universidad presencial para la que disponemos de la lista de sus estudiantes, distribuidos por cursos.
- Una universidad presencial para la que no disponemos de la lista de sus estudiantes pero de la que sabemos que tiene tres campus, y que en el primer campus hay el 25% de los estudiantes y en el segundo el 60% de los estudiantes.
- Una universidad no presencial para la que tenemos los correos electrónicos y la indicación del semestre en la que están los estudiantes y dónde además sabemos que hay 1000 estudiantes en su primer semestre, 1100 en su segundo semestre y 3000 en semestres posteriores.

Indicaciones: Todos los apartados valen el mismo y se valorará la identificación razonada de cada tipo de muestreo y la indicación del número de individuos de cada tipo que tienen que componer la muestra. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con un 0.

CUESTIÓN 4 (15%)

Queremos estudiar la relación entre la versión del sistema operativo Android y la marca de teléfono móvil que utilizan los estudiantes de una universidad virtual, y hemos obtenido la siguiente información

	Sumum	Nokus	Syno
Petazeta	1000	123	200
Caramelo	233	223	344
CotCat	123	134	134

Tomando en los cálculos 4 decimales, calculad

- La probabilidad de que un estudiante tenga un teléfono de la marca Nokus.

- La probabilidad de que un estudiante tenga instalado el sistema CocCat.
- La probabilidad de que un estudiante tenga un teléfono Nokus con sistema Caramelo.
- La probabilidad de que un estudiante que tenga un teléfono Nokus lo tenga con con sistema Caramelo.
- La probabilidad de que un estudiante que usa el sistema CotCat no tenga un teléfono de la marca Sumum.
- Indicad si alguna de las combinaciones (Syno, Petazeta) o (Syno, Caramelo) o (Syno,CotCat) corresponde a una pareja de sucesos independientes.

Indicaciones: Todos los apartados valen lo mismo y se valorará que se indique mediante las correspondientes fórmulas qué probabilidad se calcula. Se valorará especialmente la distinción entre la probabilidad de la intersección y la probabilidad condicionada. Los resultados sobre independencia tienen que ir apoyados con los cálculos correspondientes. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con un 0.

PROBLEMA 1 – CON R. (25%)

En el fichero **Pac1p_1314.csv** tenéis datos publicados el 26/2/2103 por diario “El País” sobre las nuevas previsiones económicas de la Comisión Europea para los años 2013, 14 y 15, datos que mostramos a continuación:

NUEVAS PREVISIONES ECONÓMICAS DE LA COMISIÓN EUROPEA

	PIB (%)			INFLACIÓN (%)			DESEMPLEO (% de la pobl. activa)			DÉFICIT PÚBLICO (% del PIB)		
	2013	2014	2015	2013	2014	2015	2013	2014	2015	2013	2014	2015
Bélgica	0,2	1,4	1,7	1,2	0,9	1,4	8,4	8,5	8,2	-2,7	-2,6	-2,7
Alemania	0,4	1,8	2,0	1,6	1,4	1,4	5,3	5,2	5,1	-0,1	0,0	0,0
Estonia	0,7	2,3	3,6	3,2	1,8	2,8	8,8	8,3	7,7	-0,4	-0,4	-0,4
Irlanda	0,3	1,8	2,9	0,5	0,8	1,1	13,1	11,9	11,2	-7,2	-4,8	-4,3
Grecia	-3,7	0,6	2,9	-0,9	-0,6	0,2	27,3	26,0	24,0	-13,1	-2,2	-1,0
España	-1,2	1,0	1,7	1,5	0,3	0,9	26,4	25,7	24,6	-7,2	-5,8	-6,5
Francia	0,3	1,0	1,7	1,0	1,2	1,2	10,8	11,0	11,0	-4,2	-4,0	-3,9
Italia	-1,9	0,6	1,2	1,3	0,9	1,3	12,2	12,6	12,4	-3,0	-2,6	-2,2
Chipre	-6,0	-4,8	0,9	0,4	0,4	1,4	16,0	19,2	18,4	-5,5	-5,8	-6,1
Letonia	4,0	4,2	4,3	0,0	1,9	2,1	11,9	10,5	9,2	-1,3	-1,0	-1,0
Luxemburgo	2,1	2,2	2,5	1,7	1,5	1,7	5,9	6,0	5,9	-0,2	-0,5	-2,4
Malta	2,0	2,1	2,1	1,0	1,2	1,9	6,5	6,4	6,4	-3,0	-2,7	-2,7
Holanda	-0,8	1,0	1,3	2,6	1,1	1,3	6,7	7,4	7,2	-3,1	-3,2	-2,9
Austria	0,3	1,5	1,8	2,1	1,8	1,8	4,9	4,8	4,7	-1,7	-2,1	-1,8
Portugal	-1,6	0,8	1,5	0,4	0,8	1,2	16,5	16,8	16,5	-5,9	-4,0	-2,5
Eslovenia	-1,6	-0,1	1,3	1,9	0,8	1,3	10,2	10,8	10,7	-14,9	-3,9	-3,3
Eslovaquia	0,8	2,3	3,2	1,5	0,7	1,6	14,2	13,9	13,4	-2,5	-3,3	-3,4
Finlandia	-1,5	0,2	1,3	2,2	1,7	1,6	8,2	8,3	8,1	-2,4	-2,5	-2,3
Zona Euro	-0,4	1,2	1,8	1,4	1,0	1,3	12,1	12,0	11,7	-3,1	-2,6	-2,5
Suecia	0,9	2,5	3,3	0,4	0,9	1,8	8,0	7,7	7,3	-1,1	-1,5	-0,8
Reino Unido	1,9	2,5	2,4	2,6	2,0	2,0	7,6	6,8	6,5	-6,3	-5,2	-4,2
Unión Europea	0,1	1,5	2,0	1,5	1,2	1,5	10,9	10,7	10,4	-3,5	-2,7	-2,7

Se pide que hagáis un análisis descriptivo de estos datos, en particular teniendo en cuenta estos aspectos:

- Resumid con los estadísticos habituales las previsiones de incremento del PIB correspondientes a 2013 y 2014 y comentad los resultados.
- Sobre estas mismas variables, haced unos gráficos (histogramas y box-plots) que muestren de la mejor manera posible el comportamiento de los dos años considerados.
- Cread una columna nueva que muestre la variable DIF, correspondiente a la diferencia entre la inflación prevista el 2015 y el 2014 y analizad esta diferencia.
- Haced un gráfico que muestre de manera clara el nombre del país y el valor de la variable DIF (indicación: investigad sobre la función *dotchart*, indicando la variable numérica y después la variable que contiene los nombres de los países; recordad que

para indicar las variables hay que poner el nombre del conjunto de datos, seguido de “\$” y después el nombre de la variable).

- e) Resumid los resultados obtenidos en los apartados anteriores sobre las variables consideradas.

Para importar el fichero hay que ir a “Datos--importar datos--desde archivo de texto...” y hay que tener en cuenta que el separador de campos ha de ser el “;” y que el carácter decimal es la “.”. Poned un nombre al “Conjunto de Datos ” que os sea relevante para poder hacer los cálculos después.

Indicaciones: Todos los apartados valen lo mismo y se valorará que se indique qué herramientas de R se han usado en el análisis y por qué se han seleccionado unos determinados indicadores o gráficos. Se valorará especialmente la adecuación, la creatividad, y la originalidad de los comentarios. Hay que indicar en la respuesta de la PAC las instrucciones usadas (o generadas por el R-Commander) y los resultados obtenidos en R. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con 0.

PROBLEMA 2 – CON R. (20%)

A partir de los datos del fichero anterior, recodificaremos los valores de las variables PIB del 2014 e Inflación del 2014, segmentándolos en 4 grupos del mismo número de elementos, para poder hacer después una tabla de contingencia de estas dos nuevas variables. Denominaremos PIB e INF a las nuevas variables y utilizaremos PIB1, PIB2, PIB3, PIB4 para indicar los cuatro posibles segmentos de la variable PIB, e INF1, INF2, INF3 e INF4 para los referentes a la inflación.

Para hacer la recodificación hay que ir a *Datos->Modifica variables de la tabla de datos activa -> Segmenta una variable numérica* y hay que seleccionar 4 segmentos con el mismo número de elementos.

- Explicad qué modificaciones se han producido en nuestro conjunto de datos, es decir, qué columnas aparecen nuevas y por qué. Concretamente indicad cuántos países hay en cada segmento de las nuevas variables e indicad los resúmenes estadísticos por grupo (*Estadísticos->Resúmenes->Distribuciones de frecuencias y Datos->Resúmenes numéricos->Resúmenes por grupo*).
- Obtened la tabla de contingencia que os proporciona el software y incorporadla al documento. Per hacer la tabla de contingencia hay que ir a *Estadísticos->Tabla de contingencia->Tabla de doble entrada*. A partir de esta información haced manualmente una nueva tabla de contingencia completa añadiendo los totales marginales.
- Calculad la probabilidad de que un país esté en el grupo de los países con menos PIB y en el grupo de los países con más inflación. (Usamos “país” en sentido amplio puesto que se incluyen “Zona Euro” y “Unión Europea”).
- Calculad la probabilidad de que un país que está en el grupo de los países con menos PIB esté en INF2.
- ¿Son independientes los acontecimientos “estar en el grupo de países con más inflación” y “estar en el grupo de países con mayor PIB”?

Indicaciones: Todos los apartados valen lo mismo y se valorará que se indique qué herramientas de R se han usado en el análisis y por qué se han seleccionado unos determinados indicadores o gráficos. Se valorará especialmente la adecuación, la creatividad, y la originalidad de los comentarios. Hay que indicar en la respuesta de la PAC las instrucciones usadas (o generadas por el R-Commander) y los resultados obtenidos en R. Resultados sin indicación de cálculo o falta de justificación se pueden valorar con 0.