Topic 6: NONPARAMETRIC HYPOTHESIS TESTING

So far the inference process has been aimed at giving values to the unknown population parameters and has been based in the normal distribution, apart from the fulfillment of certain assumptions arisen from the sampling technique used (s.r.s.). In particular, hypothesis tests elaborated in chapter 5 on these premises are called parametric hypothesis testing.

The nonparametric tests to be explained in this topic allow us to check the goodness of fit of a sample to a given probability distribution, specifically to a normal population. In addition this tool makes possible to study if the observations in the sample verify independence or if they have been obtained from an homogeneous population according to a s.r.s. The lasts two features will also be considered in relation to two variables not necessarily following a normal distribution.

Sometimes, a given hypothesis can be tested using both, parametric and nonparametric tools. In this regard the latter have the advantage of needing fewer assumptions, therefore having a greater robustness (their main results do not change when failing to comply with certain conjecture). However parametric tests are preferred provided that they are more powerful than the nonparametric ones.

6.1. DIAGNOSIS OF MODELS

When estimating the unknown parameters of a given probability distribution certain assumptions are made regarding:

- The probability distribution followed by the random variable under study (normal, binomial, poisson, uniform …)
- The sample obtained:
  - Considering the obervations are independent
  - All observations come from the same distribution (homogeneity)

As we have shown both the confidence intervals and the parametric hypothesis tests analysed in chapters 4 and 5 are based in the normal distribution, either directly or indirectly (Student’s t, Chi-square, Snedecor’s F).

Therefore, to verify that the sample obtained comes from a *normal distribution* becomes crucial for upholding the results achieved through the use of those techniques.

However both, intervals and parametric hypothesis tests regarding the population mean (or the means if two populations are considered) are robust in general. This means that if large samples are obtained the normality based on the central limit theory provides approximately correct inferences. In these situations when the normality assumption is false the inference is correct but not optimal (large confidence intervals or tests with low power).

At the same time intervals and hypothesis tests over the variances turn out to have low precision if the normality assumption is not accomplished.

By another side if the independence assumption is not verified the variances of the point estimators will be affected, then modifying both intervals and hypothesis tests built upon them.

Finally, the absence of homogeneity in the data collected may be due either to the variable under study being composed of different groups or to the existence of outliers.

Accordingly, in order to guarantee that the inference realized is optimal the fulfilling of the previous assumptions must be carried out using goodness of fit tests over the probability distribution used (particularly for normality), independence tests and homogeneity tests.

6.2. GOODNESS OF FIT

6.2.1. Pearson's Chi square

The null hypothesis being tested states that the sample comes from a given probability distribution.

The observed frequencies in the sample are compared with those corresponding to the theoretical distribution being tested.

Then an estimator incorporating those differences is calculated. That estimator follows a Pearson's Chi square as it can be shown.

After that the critical region is established on the basis of a certain significance level.

Finally the test estimate is obtained observing in which range is located, either in the critical region or in the acceptance region.

Characteristics of the test:

- It can be applied to any kind of random variable, either discrete or continuous.
- Two modalities can be considered, depending on 1) the variable being studied is completely determined under the null hypothesis or 2) the null hypothesis states the distribution followed by the variable but not its parameters being estimated by means of the sample.
- The power of the test increases when the sample size grows, the technique requiring at least $n \geq 25$ according to D.Peña or $n > 30$ conforming to other authors (100 or more in case of normality).
- The distribution studied under the null hypothesis is not specifically tested but all those distributions assigning to the specified classes the same probabilities established by the variable analysed.

<u>Steps</u>:

1. Hypothesis:
   - H_0: The sample comes from a given probability distribution
   - H_1: The sample DOES NOT come from a given distribution
2. Determination of the *d* test statistic:
   - First the *n* observations of the sample must be arranged in *k* classes selected to cover all data without overlapping

- $k \geq 5$ is recommended
- Then the number of data included in each class O_i (observed frequency) is obtained. A minimum of 3 data in each class is suggested.
- After that the probability p_i assigned by the distribution under the null hypothesis to each category must be calculated, verifying that $\Sigma p_i = 1$
- Next the expected frequency in each class under the null hypothesis E_i is determined using this expression:

$$E_i = n \cdot p_i$$

- $E_i > 5$ in each class is recommended. Otherwise a redefinition of classes is advisable.
- Table 1 in the next page may be useful in order to inform about categories, frequencies and probabilities
- Finally the d test statistic is elaborated:

$$d = \sum_{1}^{k} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2 \quad If\ H_0\ is\ true$$

- Degrees of freedom:
    - *k-1* if the distribution is fully established under the null hypothesis
    - *k-r-1* if *r* parameters have been estimated through the sample

3. At this stage α is fixed in a low level and the rejection region is obtained, depending on the corresponding degrees of freedom:

$$d > \chi^2_{k-1,\alpha} \quad or \quad d > \chi^2_{k-r-1,\alpha}$$

4. At last the value d_0 of the statistic in the sample is calculated or, alternatively, the *p-value* is determined and a decision is made.

***Table 1**: classes, frequencies and probabilities*

Class	Observed frequency O_i	Probability p_i	Expected frequency E_i
1	O_1	p_1	$E_1 = np_1$
2	O_2	p_2	$E_2 = np_2$
…	…	…	…
k	O_k	p_k	$E_k = np_k$
Total	n	1	n

Pearson's Chi squared applied to **normality**

The most common rule is forming classes with equal probability p_i determining k so as to fulfil this requirement:

$$E_i > 3$$

If k categories are selected each class will have a probability of $1/k$.

Table 2 includes some quantiles of the Z distribution needed to establish categories with the same probability.

For example, if we want to use six classes with the same probability the corresponding intervals are the following:

$$(-\infty; \bar{x} - 0.97s_1); (\bar{x} - 0.97s_1; \bar{x} - 0.43s_1); (\bar{x} - 0.43s_1; \bar{x});$$
$$(\bar{x}; \bar{x} + 0.43s_1); (\bar{x} + 0.43s_1; \bar{x} + 0.97s_1); (\bar{x} + 0.97s_1; +\infty);$$

μ and σ will be estimated by the sample mean and the bias corrected sample standard deviation respectively using $k - 3$ degrees of freedom.

It is recommended this goodness of fit technique to be used when $n \geq 100$

For lower sample sizes the test will not reject normality in almost any symmetric distribution with a unique maximum (D. Peña).

Table 2: Quantiles of the N(0;1) distribution

P(Z ≤ a)	1/10	1/8	1/6	1/5	1/4	3/10	1/3	3/8	2/5
a	-1,28	-1,15	-0,97	-0,84	-0,67	-0,52	-0,43	-0,32	-0,25

6.2.2. Kolmogorov-Smirnov Test

This test compares the cumulative frequency distribution in a sample with the cumulative distribution function of the distribution studied in the null hypothesis.

It is only acceptable for continuous variables.

Stages:

1. Hypothesis:

   H_0: The sample comes from a given continuous distribution

   H_1: The sample DOES NOT come from a given continuous distribution

2. Determination of the *d* test statistic:

   - First data in the sample must be ordered from the least to greatest
   - Then the cumulative frequency distribution in the sample is calculated:

   $$F_{Oi}(x_i) = \frac{N_i}{n}$$

   $$n \ is \ the \ sample \ size$$

   $$N_i \ is \ the \ absolute \ cumulative \ frecquency \ of \ the \ observation \ x_i$$

   - Afterwards the cumulative distribution function over the sampling observations $F(x_i)$ is obtained
   - So is the difference in absolute terms d_{i-1} between the empirical distribution function in x_{i-1} and the theoretic cumulative distribution function in x_i:

   $$d_{i-1} = |F_{Oi}(x_{i-1}) - F(x_i)|$$

   - Besides that the difference in absolute terms d_i between the empirical distribution function in x_i and the theoretic cumulative distribution function in x_i is determined:

$$d_i = |F_{Oi}(x_i) - F(x_i)|$$

- Then the d test statistic is calculated as the maximum value of all the differences obtained:

$$d = \max(d_{i-1}, d_i)\ \forall i$$

- If the null hypothesis is true d behaves as a Kolmogorov-Smirnov statistic

3. Determination of the critical region once α is fixed:

   As this is a one tailed test to the right the null hypothesis will be rejected when the difference between both cumulative functions be large enough:

$$\alpha = P(d > k_\alpha)$$

   k_α is a critical value obtained in the Kolmogorov-Smirnov table

4. Finally d_o is calculated and a decision is made:

Posición	x_i	$F_{oi}(x_{i-1})$	$F_{oi}(x_i)$	$F(x_i)$	d_{i-1}	d_i
1	x_1	0	N_1/n	$F(x_1)$	$\|0-F(x_1)\|$	$\|(N_1/n)-F(x_1)\|$
2	x_2	N_1/n	N_2/n	$F(x_2)$	$\|(N_1/n)-F(x_2)\|$	$\|(N_2/n)-F(x_2)\|$
3	x_3	N_2/n	N_3/n	$F(x_3)$	$\|(N_2/n)-F(x_3)\|$	$\|(N_3/n)-F(x_3)\|$
…	…	…	…	…	…	…
n-1	x_{n-1}	N_{n-2}/n	N_{n-1}/n	$F(x_{n-1})$	$\|(N_{n-2}/n)-F(x_{n-1})\|$	$\|(N_{n-1})/n)-F(x_{n-1})\|$
n	x_n	N_{n-1}/n	1	$F(x_n)$	$\|(N_{n-1})/n)-F(x_n)\|$	$\|1-F(x_n)\|$

$$d_0 = \max(d_{i-1}, d_i) \qquad \forall i = 1..n$$

$$If\ d_o \in RC \rightarrow Reject\ H_o and\ Accept\ H_1; \quad If\ d_o \in RA \rightarrow Accept\ H_o$$

Apart from that the p-value can be assessed and the corresponding decision been made.

The K-S table follows:

n\α	**0.20**	**0.10**	**0.05**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.90000	0.95000	0.97500	0.99000	0.99500	0.99750	0.99900	0.99950
2	0.68337	0.77639	0.84189	0.90000	0.92929	0.95000	0.96838	0.97764
3	0.56481	0.63604	0.70760	0.78456	0.82900	0.86428	0.90000	0.92065
4	0.49265	0.56522	0.62394	0.68887	0.73424	0.77639	0.82217	0.85047
5	0.44698	0.50945	0.56328	0.62718	0.66853	0.70543	0.75000	0.78137
6	0.41037	0.46799	0.51926	0.57741	0.61661	0.65287	0.69571	0.72479
7	0.38148	0.43607	0.48342	0.53844	0.57581	0.60975	0.65071	0.67930
8	0.35831	0.40962	0.45427	0.50654	0.54179	0.57429	0.61368	0.64098
9	0.33910	0.38746	0.43001	0.47960	0.51332	0.54443	0.58210	0.60846
10	0.32260	0.36866	0.40925	0.45562	0.48893	0.51872	0.55500	0.58042
11	0.30829	0.35242	0.39122	0.43670	0.46770	0.49539	0.53135	0-55588
12	0.29577	0.33815	0.37543	0.41918	0.44905	0.47672	0.51047	0.53422
13	0.28470	0.32549	0.36143	0.40362	0.43247	0.45921	0.49189	0.51490
14	0.27481	0.31417	0.34890	0.38970	0.41762	0.44352	0.47520	0.49753
15	0.26589	0.30397	0.33750	0.37713	0.40420	0.42934	0.45611	0.48182
16	0.25778	0.29472	0.32733	0.36571	0.39201	0.41644	0.44637	0.46750
17	0.25039	0.28627	0.31796	0.35528	0.38086	0.40464	0.43380	0.45540
18	0.24360	0.27851	0.30936	0.34569	0.37062	0.39380	0.42224	0.44234
19	0.23735	0.27136	0..30143	0.33685	0.36117	0.38379	0.41156	0.43119
20	0.23156	0.26473	0.29408	0.32866	0.35241	0.37451	0.40165	0.42085
21	0.22517	0.25858	0.28724	0.32104	0.34426	0.36588	0.39243	0.41122
22	0.221I5	0.25283	0.28087	0.31394	0.33666	0.35782	0.38382	0.40223
23	0.21646	0.24746	0.2749tl	0.30728	0.32954	0.35027	0.37575	0.39380
24	0.21205	0.24242	0.26931	0.30104	0.32286	0.34318	0.36787	0.38588
25	0.20790	0.23768	0.26404	0.29518	0.31657	0.33651	0.36104	0.37743
26	0.20399	0.23320	0.25908	0.28962	0.30963	0.33022	0.35431	0.37139
27	0.20030	0.22898	0.25438	0.28438	0.30502	0.32425	0.34794	0.36473
28	0.19680	0.22497	0.24993	0.27942	0.29971	0.31862	0.34190	0.35842
29	0.19348	0.22117	0.24571	0.27471	0.29466	0.31327	0.33617	0.35242
30	0.19032	0.21756	0.24170	0.27023	0.28986	0.30818	0.33072	0.34672
31	0.18732	0.21412	0.23788	0.26596	0.28529	0.30333	0.32553	0.34129
32	0.18445	0.21085	0.23424	0.26189	0.28094	0.29870	0.32058	0.33611
33	0.18171	0.20771	0.23076	0.25801	0.27577	0.29428	0.31584	0.33115
34	0.17909	0.21472	0.22743	0.25429	0.27271	0.29005	0.31131	0.32641
35	0.17659	0.20185	0.22425	0.25073	0.26897	0.28600	0.30597	0.32187
36	0.17418	0.19910	0.22119	0.24732	0.26532	0.28211	0.30281	0.31751
37	0.17188	0.19646	0.21826	0.24404	0.26180	0.27838	0.29882	0.31333
38	0.16966	0.19392	0.21544	0.24089	0.25843	0.27483	0.29498	0.30931
39	0.16753	0.19148	0.21273	0.23785	0.25518	0.27135	0.29125	0.30544
$n \ge 40$	$1,07/\sqrt{n}$	$1,22/\sqrt{n}$	$1,36/\sqrt{n}$	$1,52/\sqrt{n}$	$1,63/\sqrt{n}$	$1,73/\sqrt{n}$	$1,85/\sqrt{n}$	$1,95/\sqrt{n}$

Comparison with the Pearson’s chi square test:

- When grouping information in Pearson’s tool some information is lost; this does not happen in K-S technique.
- The K-S test is suitable for small samples.
- In large samples the K-S test gives similar results to Pearson’s chi square.

Problem: when population parameters are estimated the test becomes conservative: it is prone to accept H_0.

Solution for the normality case: to apply this test with the Lilliefors’s correction: the K-S-L test.

Kolmogorov-Smirnov with Lilliefors's correction: the procedure is the same as that explained for the K-S test but the table used:

Distribución del estadístico de Kolmogorov-Smirnov-Lilliefors (D_n) para el contraste de normalidad.
Se tabula d tal que $P(D_n > d) = \alpha$.

n \ α	0'2	0'15	0'1	0'05	0'01	0'001
4	0'303	0'321	0'346	0'376	0'413	0'433
5	0'289	0'303	0'319	0'343	0'997	0'439
6	0'269	0'281	0'297	0'323	0'371	0'424
7	0'252	0'264	0'280	0'304	0'351	0'402
8	0'239	0'250	0'265	0'288	0'333	0'384
9	0'227	0'238	0'252	0'274	0'317	0'365
10	0'217	0'228	0'241	0'262	0'304	0'352
11	0'208	0'218	0'231	0'251	0'291	0'338
12	0'200	0'210	0'222	0'242	0'281	0'325
13	0'193	0'202	0'215	0'234	0'271	0'314
14	0'187	0'196	0'208	0'226	0'262	0'305
15	0'181	0'190	0'201	0'219	0'254	0'296
16	0'176	0'184	0'195	0'213	0'247	0'287
17	0'171	0'179	0'190	0'207	0'240	0'279
18	0'167	0'175	0'185	0'202	0'234	0'273
19	0'163	0'170	0'181	0'197	0'228	0'266
20	0'159	0'166	0'176	0'192	0'223	0; 260
25	0'143	0'150	0'159	0'173	0'201	0'236
30	0'131	0'138	0'146	0'159	0'185	0'217
> 30	$\frac{0'740}{\sqrt{n}}$	$\frac{0'770}{\sqrt{n}}$	$\frac{0'820}{\sqrt{n}}$	$\frac{0'890}{\sqrt{n}}$	$\frac{1'040}{\sqrt{n}}$	$\frac{1'220}{\sqrt{n}}$

6.2.3. The Shapiro-Wilks's test

It is only applied to test if a sample fits a normal distribution.

It is based on the idea that a graph analysing the fit of a sample to a normal variable would resemble that one of a linear regression model, in the sense that the nearer the observations to the line the better the fit and, vice versa, the greater the distance between the point cloud and the line the worse the fit.

It is suitable for small samples ($n < 30$) due to its power, although some authors consider its utility also in larger samples ($n < 50$). In fact, 50 is the maximum value for n included in Shapiro-Wilks tables.

- *Hypothesis*:

  H_0: the sample comes from a normal distribution

  H_1: the sample DOES NOT come from a normal distribution

- *SW test statistic*:

$$d \equiv SW = \frac{\left[\sum_{j=1}^{h} a_{j,n}(x_{n-j+1} - x_j)\right]^2}{ns^2}$$

  where:

  $h = \frac{n}{2} \quad if\ n\ is\ even$

  $h = \frac{n-1}{2} \quad if\ n\ is\ odd$

  coefficients $a_{j,n}$ are tabulated (see virtual campus)

  x_j is the observation occupying the position *j* in the ordered sample (from the least to the greatest)

- *Critical Region* once α has been defined:

  SW plays the role of the coefficient of determination R^2 in a linear regression analysis, provided that it measures the degree of fit between the sample and the normal distribution. Therefore high values of d would induce us to accept H_0. On the contrary low values of d would lead us to reject H_0:

$$d < SW_\alpha$$

  Values for the SW statistic are also tabulated (see virtual tables).

6.3. TESTS FOR INDEPENDENCE AND RANDOMNESS

6.3.1 Wald-Wolfowitz runs test

It is employed to test the hypothesis of randomness in the order the observations appear in the sample, or in other words, that the observations are mutually independent.

This tool is applied to both quantitative and qualitative variables.

Let's start with the former.

We call run to a sequence of observations of the same kind. For instance, imagine that we want to know if a sample has been randomly obtained regarding sex. Consider this sample:

X_0:	M	M	H	M	H	H	H	M	M	M

There are 5 runs:

Run 1:	M	M	
Run 2:	H		
Run 3:	M		
Run 4:	H	H	H
Run 5:	M	M	M

This result would make us think that the sample is random in relation to sex.

Now observe these samples:

X_1:	M	M	M	M	M	H	H	H	H	H
X_2:	M	H	M	H	M	H	M	H	M	H

In situations like these with either a low number of runs (2 in X_1) or a high one (9 in X_2) we would believe the observations being dependent regarding the sex appearance.

General procedure: let a sample of size n from a dichotomous random variable with events A and A^c having obtained n_1 observation of A and n_2 of A^c being $n = n_1 + n_2$

Hypothesis:

H_0: the sample is random

H_1: the sample is NOT random

Case 1: Small sizes for n_1 and n_2 ($n_1 \leq 20$ y $n_2 \leq 20$)

Test statistic: $d = r = number\ of\ runs$

Rejection region, once a significance level has been established:

$$d > r_{\alpha/2}\ \ o\ \ d < r_{1-\alpha/2}$$

The critical values are tabulated for different values of n_1 and n_2 (a table is available at virtual campus for $\alpha = 5\%$).

Case 2: Large sizes for n_1 and/or n_2 ($n_1 > 20$ and/or $n_2 > 20$)

Then *r* follows approximately as a Normal distribution:

$$r \sim N\left(\frac{2n_1n_2}{n} + 1;\ \sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}}\right)$$

According to this the test statistic is:

$$d = \frac{r - \left(\frac{2n_1n_2}{n} + 1\right)}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}}} \sim Z \quad if\ H_o\ is\ true$$

Critical region once α is fixed: $|d| > z_{\alpha/2}$

In quantitative variables the procedure is similar having previously dichotomized the variable being studied. This is achieved by fixing the median, the mean or the mode (usually the median) as the reference, then classifying the observations in two groups depending on the data being located above or below that reference.

6.3.2 Independence test

Sometimes we have observations in a sample that can be classified under two or more categories, then a multidimensional variable arising. In this scenario we may be interested in testing the independence hypothesis between the different outcomes.

The most common situation is a bidimensional variable. Then the sampling observations are included in a contingency table with the values of the variables appearing in the first column and in the first row and the cells incorporating the joint frequencies, either the absolute or the relative ones:

Contingency table with absolute frequencies

X \ Y	y_1	y_2	...	y_j	...	y_k	$n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1k}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2k}	$n_{2.}$
...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ik}	$n_{i.}$
...	...	...	...	...	...	...	...
x_h	n_{h1}	n_{h2}	...	n_{hj}	...	n_{hk}	$n_{h.}$
$n_{.j}$	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.k}$	n

Where *n* is the sample size: $n = \sum_{i=1}^{h}\sum_{j=1}^{k} n_{ij} = \sum_{i=1}^{h} n_{i.} = \sum_{j=1}^{k} n_{.j}$

Hypothesis:

H_0: the variable Y is independent from the variable X

H_1: the variable Y is NOT independent from the variable X

Test statistic *d*:

If the null hypothesis is true any relative joint frequency can be expressed as a product of the respective marginal relative frequencies:

$$\frac{n_{ij}^{H0}}{n} = \frac{n_{i.}}{n}\frac{n_{.j}}{n} \quad \forall\, i = 1..h \quad \forall\, j = 1..k$$

There we can solve for the expected frequencies E_{ij} under the null hypothesis:

$$n_{ij}^{H0} \equiv E_{ij} = n \frac{n_{i\cdot}}{n} \frac{n_{\cdot j}}{n} = \frac{n_{i\cdot} n_{\cdot j}}{n} \quad \forall\, i = 1..h \quad \forall\, j = 1..k$$

Where $E_{ij} > 5$ according to the restrictions given in the goodness of fit test.

Then compare the observed frequencies n_{ij} with the expected ones under H_0:

$$d = \sum_{i=1}^{h} \sum_{j=1}^{k} \frac{\left(n_{ij} - E_{ij}\right)^2}{E_{ij}} \sim \chi^2_{(h-1)(k-1)} \quad If\ H_0\ is\ true$$

Critical region after fixing α:

$$d > \chi^2_{(h-1)(k-1),\alpha}$$

This technique can also be applied when samples obtained from different populations are available to test if a given characteristic (variable) behaves homogenously among those populations (Esteban García J. et al.). This situation is referred to as a homogeneity test (and it could have been included in the next section):

H_0: the variable under study is homogeneous among different populations

H_1: the variable under study is NOT homogeneous among those populations

6.4. HOMOGENEITY TESTS

Sometimes we have a sample presenting high dispersion with low curtosis, hence giving the clue for the existence of an heterogeneous population. To test it we would split the sample in two independent homogeneous subsamples and compare them (D. Peña).

Other times two independent samples coming from two populations are available, wishing to know if there are significant differences between them.

Yet a comparison between two dependent samples may be the case.

In all these situations when variables are normal the respective parameter tests analysed in chapter 5 must be run. However if the variables do not fit a normal distribution comparisons should be carried out using non parametric tests. Two of them are shown next.

6.4.1 Mann-Whitney U test for two independent samples

Let two independent samples be obtained from two populations (measured at least in an ordinal scale) or a given sample split in two independent samples coming from two populations (either ordinal or quantitative):

$$(x_1, x_2 \dots x_i \dots x_{n1})\ is\ a\ s.r.s. from\ \xi_x$$

$$(y_1, y_2 \dots y_j \dots y_{n2})\ is\ a\ s.r.s. from\ \xi_y$$

Hipothesis:

H_0: the populations are identical

H_1: the populations are NOT identical

It is assumed that if the populations are not identical it is only due to their measures of central tendency, in particular, to their median.

Case of small samples:

Test statistic:

The two samples are merged and the observations arranged from the least to the greatest, assigning them ranks r_i where tied data are assigned the average of the next corresponding ranks. Then the sum of the ranks is calculated for each of the former samples:

$$R_1 = \sum_{i=1}^{n1} r_{i1} \qquad R_2 = \sum_{j=1}^{n2} r_{j2}$$

So are the statistics U_1 and U_2:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Then the U statistic is the minimum of them:

$$d \equiv U = mín(U_1; U_2)$$

Rejection region, after fixing α:

If both samples come from the same population it would seem reasonable the ranks being randomly scattered between the samples, hence R_1 and R_2 becoming similar so being U_1 and U_2. Therefore low values of the test statistic (those below a certain critical value) would lead to reject H_0 (tables are not shown for this case):

$$d < U_{\alpha/2} \quad for\ a\ two\ tailed\ test$$

$$d < U_{\alpha} \quad for\ a\ one\ sided\ test$$

Case of large samples: $n_1 \geq 10$ and $n_2 \geq 10$

Test statistic:

In this case the *U* statistic approaches a normal distribution with parameters:

$$E(U) = \mu_U = \frac{n_1 n_2}{2}$$

$$V(U) = \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

The test statistic resulting:

$$d = \frac{U - \mu_U}{\sigma_\mu} \sim Z \quad if\ H_0\ is\ true$$

Critical region, once α is fixed:

$$d < -z_{\frac{\alpha}{2}} \qquad for\ a\ two\ sided\ test$$

$$d < -z_{\alpha} \qquad for\ a\ one\ sided\ test$$

6.4.2 Wilcoxon sign ranked test for paired samples

Let two dependent samples be obtained from two populations (measured at least in an ordinal scale) on which we want to know if there are significant differences between their respective medians:

$$(x_1, x_2 \dots x_i \dots x_m)\ is\ a\ s.r.s. from\ \xi_x$$

$$(y_1, y_2 \dots y_i \dots y_m)\ is\ a\ s.r.s. from\ \xi_y$$

This test is seen as the nonparametric alternative to the Student's t test for two dependent samples when the normality assumption cannot be held or when data are measured in an ordinal scale.

For instance, consider the opinions given by *m* subjects regarding their preferences in relation with two cola drinks, using an ordinal scale.

Hypothesis:

H_0: $Me_X = Me_Y$

H_1: $Me_X \neq Me_Y$

Case of small samples:

Test statistic:

First the differences d_i between the corresponding paired observations in the samples are calculated. It is assumed that those differences are mutually independent and follow a continuous distribution with symmetry centered at 0 (median null):

$$d_i = x_i - y_i$$

Next the null differences ($d_i = 0$) are not considered and the rest n ($n < m$) are arranged in ascending order on the basis of their absolute value. Then a rank r_i is assigned (tied data are assigned the average of the next corresponding ranks).

Then the sums of ranks, both positive and negative, are obtained:

$$R^+ = \sum r_i^+ \qquad R^- = \sum r_i^-$$

Now the W test statistic is the minimum of them:

$$d \equiv W = mín(R^+; R^-)$$

Rejection region, after α has been established:

The arrangement of the differences in a symmetric way around cero, under H_0, would prompt to think that R+ and R- would be similar. Hence the more separated these totals become the stronger the reasons to reject H_0. Therefore low values of the test statistic (those below a certain critical value) would make to reject H_0 (see Wilcoxon table at virtual campus):

$$d < W_{\alpha/2} \quad for\ a\ two\ sided\ test$$

$$d < W_{\alpha} \quad for\ a\ one\ sided\ test$$

Case of large samples (the number of differences d_i not being cero is large):

$$n > 20$$

Test statistic:

In this situation the *W* statistic approaches a normal distribution with parameters:

$$E(W) = \mu_W = \frac{n(n+1)}{4}$$

$$V(W) = \sigma_W^2 = \frac{n(n+1)(2n+1)}{24}$$

And the corresponding test statistic is:

$$d = \frac{W - \mu_W}{\sigma_W} \sim Z \quad if\ H_0\ is\ true$$

Rejection region, after fixing α:

$$d < -z_{\frac{\alpha}{2}} \qquad for\ a\ two\ tailed\ test$$

$$d < -z_{\alpha} \qquad for\ a\ one\ tailed\ test$$