UNIT 11 .- NON-PARAMETRIC TESTS.

UNIT 11. GOALS

- Solve non parametric statistical testing problems.

Runs Test

•H_0: *Data comes from a random sample*

- Data is classified into two categories, defining a dummy variable

- If the nature of data is quantitative, the dummy variable is then defined as: 0, if x < Me and 1, if x > Me

- The number of runs R is calculated

  A run is a sequence of consecutive observations of the same category

If n is large, the test statistic is the following:

$$d_R = \frac{R - 2np(1-p)}{2\sqrt{np}(1-p)} \approx N(0,1)$$

p= % of "1" in the sample

The decision rule is based on the observed number of runs (r*), and leads to reject the null hypothesis if r* is either very large or very small (two-sided test)

Example

H_0: *Data comes from a random sample*

Example

Sample:

X	2	1	2,5	3	1,5	3	(Me = 2,25)
Dummy variable	0	0	1	1	0	1	

Nº of runs: r*= 4

$$d_R{}^* = \frac{r^* - 2np(1-p)}{2\sqrt{np(1-p)}} = \frac{4 - 2(6)(0,5)(0,5)}{2\sqrt{6}(0,5)(0,5)} = 0,89$$

$$p = P\left(\left|N(0,1)\right| > 0,89\right) = 0,37$$

We fail to reject H_0

Jarque -Bera Test

Hypothesis: Normality

This test compares the skewness and kurtosis of the data with those of the normal distribution

	Skewness (Asymmetry)	**Kurtosis**
	$g_1 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^3}{S^3}$	$g_2 = \dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^4}{S^4} - 3$
g=0	Symmetric distribution	Mesokurtic distribution
g>0	Positive asymmetry	Leptokurtic distribution
g<0	Negative asymmetry	Platykurtic distribution

Test statistic

$$d_{JB} = \frac{n}{6}\left(g_1^2 + \frac{1}{4}g_2^2\right) \qquad d_{JB} \approx \chi_2^2$$

P-value

$$\mathbf{p = P(d_{JB} > d_{JB}^{*}/H_0)}$$

χ^2 Test for independence

X is a nominal variable with "r" categories $A_1, A_2, \ldots, A_r$

GENDER

Male

Female

Y is a nominal variable with "s" categories $B_1, B_2, \ldots, B_s$

SMOKING

Smoker

Non-smoker

Is there any relationship between X and Y?

Observed Frequencies

X / Y	A_1	A_2	...	A_r
B_1	n_{11}	n_{21}	...	n_{r1}
B_2	n_{12}	n_{22}	...	n_{r2}
...	...	...	...	...
B_s	n_{1s}	n_{2s}	...	n_{rs}

$$n_{.j} = \sum_{i=1}^{r} n_{ij}$$

$$\boxed{\sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n}$$

$$n_{i.} = \sum_{j=1}^{s} n_{ij}$$

χ^2 Test for independence

H_0: X and Y are two random independent variables

- X and Y have r and s categories
- Observed and theoretical frecuencies are compared

Observed Frequency

$$n_{ij}$$

Theoretical Frequency

$$\frac{n_{i.}n_{.j}}{n}$$

$$d_{IND} = \sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\left(n_{ij} - \dfrac{n_{i.}n_{.j}}{n}\right)^2}{\dfrac{n_{i.}n_{.j}}{n}} \rightarrow \chi^2_{(r-1)(s-1)}$$

$$\frac{n_{i.}n_{.j}}{n} \geq 5$$

P-value

$$p = P(d_{IND} > d^*_{IND}/H_0)$$

χ^2 Test for independence

- H_0: Gender and smoking are independent

Observed Frequencies

Gender / Smoking	Male	Female	$n_{.j}$
Smoker	20	19	39
Non-smoker	10	16	26
$n_{i.}$	30	35	65

Marginal frequencies

Test Statistic

$$d^*_{IND} = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} = 1{,}03$$

Theoretical Frequencies

Gender / Smoking	Male	Female
Smoker	18	21
Non-smoker	12	14

P-value

$$p = P(\chi^2_1 > 1{,}03) = 0{,}31$$

We failed to reject H_0