

Diseño experimental y estadística aplicada para ciencias ambientales

Diego Mota de Echeandía y Rocío López Flores
Escuela Politécnica Superior - Universidad de Zaragoza



Diseño experimental y estadística aplicada para ciencias ambientales by Diego Mota de Echeandía & Rocío López-Flores is licensed under a Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional License.

ESTADÍSTICOS DESCRIPTIVOS

Para obtener los estadísticos descriptivos de nuestros datos, utilizaremos la función “*summary statistics*” que encontraremos en el menú “*Univariate*”.

En primer lugar, seleccionaremos los datos de los que queremos obtener dichos estadísticos. Podemos seleccionar cada columna individualmente haciendo click sobre el nombre de la variable y obteniendo así los estadísticos de dicha variable, o seleccionar el conjunto de variables y obtenerlos de todas:

The screenshot shows the 'Univariate' menu open, with 'Summary statistics' highlighted. The menu options include: One-sample tests (t, Wilcoxon, single-case), Two-sample tests, ANOVA etc. (several samples), Correlation, Intraclass correlation, Normality tests, Contingency table (chi^2 etc.), Mantel-Cochran-Haenszel test, Risk/odds, Single proportion test, Multiple proportion CIs, Ratios of counts CI, Survival analysis, and Combine errors. The background data table is as follows:

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7	H	I	J	K
A	3,4				8,7	6,5	4,5				
B	2,6				6,2	5,2	4,3				
C	2,8				6,8	5,6	4,7				
D	2,5				6,0	5,0	4,2				
E	2,3				5,6	4,7	3,9				
F	3,1				7,3	6,1	5,1				
G	3,2				7,6	6,4	5,3				
H	2,8				6,6	5,5	4,6				
I	2,6				6,2	5,2	4,3				
J	2,6	1,8	3,0	2,7	6,2	5,2	4,3				
K	2,5	1,7	2,9	2,6	5,9	4,9	4,1				
L	2,2	1,5	2,5	2,3	5,2	4,3	3,6				

The 'Univariate statistics' window displays the following data:

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7
N	13	13	13	13	13	13	13
Min	2.1	1.5	2.4	2.2	5	4.2	3.5
Max	3.4	2.4	3.7	3.5	8.7	6.5	5.3
Sum	34.7	24.1	39.3	36	83.3	68.8	56.4
Mean	2.669231	1.853846	3.023077	2.769231	6.407692	5.292308	4.338462
Std. error	0.1070494	0.07391557	0.1020017	0.1070494	0.2809009	0.2004679	0.1443546
Variance	0.1489744	0.07102564	0.1352564	0.1489744	1.025769	0.5224359	0.2708974
Stand. dev	0.385972	0.2665064	0.3677722	0.385972	1.012803	0.7227973	0.5204781
Median	2.6	1.8	3	2.7	6.2	5.2	4.3
25 prcntil	2.4	1.65	2.8	2.5	5.75	4.8	4
75 prcntil	2.95	2.05	3.25	3.05	7.05	5.85	4.65
Skewness	0.4716941	0.5894379	0.05637305	0.4716941	0.8735488	0.2894129	0.2145572
Kurtosis	-0.3654047	-0.00345632	-0.1100665	-0.3654047	0.9042119	-0.5848028	-0.08363322
Geom. mean	2.644036	1.836694	3.002223	2.74494	6.337629	5.247113	4.309682
Coeff. var	14.46004	14.37586	12.16549	13.93788	15.80604	13.65751	11.99684

Additional options in the window include: Bootstrap, Bootstrap type: Simple, Bootstrap N: 9999, and a Recompute button.

NORMALIDAD Y HOMOGENEIDAD

Normalidad: en estadística, el supuesto de normalidad se refiere a la idea de que una variable aleatoria sigue una distribución normal o gaussiana. Una distribución normal es una distribución de probabilidad continua con una curva en forma de campana, donde la mayoría de los valores se encuentran cerca de la media y menos valores se encuentran a medida que nos alejamos de la media.

Existen diferentes métodos para evaluar la normalidad, el método que desarrollaremos será el de **contraste de hipótesis**.

-Se considera como **hipótesis nula** que los datos proceden de una distribución normal y como **hipótesis alternativa** que no lo hacen.

-En el caso de que los datos no se distribuyan normalmente, existen dos opciones:

- **Transformar** los datos para que se distribuyan de manera normal
- Realizar una **prueba no paramétrica** (se explicará más adelante). Los equivalentes no paramétricos de algunas de las pruebas paramétricas más usadas son:

Prueba paramétrica	Equivalente no paramétrico
Prueba t para una muestra	Prueba de rango con signo de Wilcoxon de una muestra
Prueba t de dos muestras	Prueba U de Mann-Whitney
Prueba t de muestras pareadas	Prueba de rango con signo de Wilcoxon de dos muestras
ANOVA unidireccional	Prueba de Kruskal-Wallis

Existen varios test para probar la normalidad, en este caso nos centraremos principalmente en dos de ellos:

-**Test de Shapiro-Wilk:** cuando el tamaño muestral es menor de 50

-**Test de Kolmogorov-Smirnov y modificación de Lilliefors:** cuando el tamaño muestral es mayor de 50

¿Cómo realizar una prueba de normalidad en PAST?

En primer lugar, seleccionaremos las variables de las cuales queremos probar su normalidad (de una a todas las variables).

Seguidamente, en el menú **"Univariate"** haremos click sobre **"Normality tests"** y se desplegará la siguiente ventana:

Tests for normal distribution							
	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6	Var 7
N	13	13	13	13	13	13	13
Shapiro-Wilk W	0.9544	0.9457	0.9751	0.9544	0.9448	0.9518	0.9687
p(normal)	0.6655	0.5346	0.9464	0.6655	0.5213	0.6262	0.8788
Anderson-Darling A	0.2993	0.3149	0.2087	0.2993	0.3247	0.2659	0.219
p(normal)	0.5316	0.5045	0.8258	0.5316	0.4784	0.629	0.7936
p(Monte Carlo)	0.5579	0.5173	0.8417	0.5678	0.494	0.6699	0.818
Lilliefors L	0.1866	0.1954	0.1404	0.1866	0.1966	0.1662	0.1448
p(normal)	0.242	0.1856	0.6824	0.242	0.179	0.4116	0.6347
p(Monte Carlo)	0.2359	0.1872	0.6817	0.2372	0.1838	0.4076	0.6303
Jarque-Bera JB	0.6155	0.6844	0.1411	0.6155	1.297	0.4974	0.2039
p(normal)	0.7351	0.7102	0.9319	0.7351	0.5229	0.7798	0.9031
p(Monte Carlo)	0.6079	0.5549	0.9379	0.6126	0.2006	0.7003	0.9026

Copy Print Monte Carlo N: Recompute
Close Help

Como puede observarse, se muestran los resultados de todos los test de normalidad para todas las muestras que hayamos seleccionadas. Las pruebas que nos interesan son las marcadas en rojo. En primer lugar, aparece el valor del estadístico de normalidad aplicado y debajo, su *p-valor*.

-Si el *p-valor* es menor que el nivel de significancia (0,05) se puede rechazar la hipótesis nula y confirmar que los datos no se distribuyen normalmente

-En el caso contrario, cuando el *p-valor* supera al nivel de significancia, no se puede rechazar la hipótesis nula y se acepta el supuesto de normalidad.

Cabe destacar, que cuanto mayor sea el tamaño muestral, menos sensibles serán los métodos paramétricos a la falta de normalidad. Por ello, es importante no basar las conclusiones únicamente en el *p-valor*, sino también considerar la representación gráfica y el tamaño muestral.

Para ver si una muestra es normal, se pueden utilizar gráficos como el histograma y la normalidad Q-Q plot. El histograma es un gráfico de barras que muestra la frecuencia de cada valor en la muestra, mientras que la normalidad Q-Q plot compara los cuantiles de la muestra con los cuantiles esperados de una distribución normal.

Homogeneidad: En estadística, el supuesto de homogeneidad se refiere a la idea de que las muestras o grupos que se están comparando tienen varianzas o distribuciones similares. En otras palabras, se asume que las diferencias observadas entre las muestras o grupos son debidas solo al azar y no a diferencias reales en la población.

Este supuesto es esencial para muchas pruebas estadísticas, como la prueba t de Student y la prueba ANOVA, ya que estas pruebas se basan en la asunción de que las varianzas de las muestras son similares. Si este supuesto no se cumple, los resultados de estas pruebas pueden ser erróneos y no se pueden confiar en las conclusiones obtenidas.

Existen varias pruebas estadísticas que se pueden utilizar para conocer la homogeneidad de los datos. Algunas de las pruebas más comunes incluyen:

Prueba de Bartlett: Esta prueba compara la varianza entre las muestras para determinar si son homogéneas, es una variante de la prueba de Levene

Prueba de Kruskal-Wallis: Esta prueba se utiliza para comparar la distribución de una variable continua entre dos o más grupos para ver si son homogéneas.

- **Prueba de Levene:** es una prueba no paramétrica que compara la varianza entre las muestras o grupos para determinar si son homogéneas. Se puede utilizar una variante de dicha prueba conocida como **Prueba de Bartlett**, la cual es más sensible a los datos atípicos (outliers). *La prueba de Bartlett se explicará en el Apartado de R Commander*

Para calcular dicho test, en el menú **“Univariate”** haremos click sobre **“ANOVA etc. (several simples)”** y seguidamente seleccionaremos **“Several sample tests (ANOVA, Kruskal-Wallis)”**, se desplegará la siguiente ventana:

Several-sample tests

One-way ANOVA | Residuals | Tukey's pairwise | **Kruskal-Wallis** | Mann-Whitney pairwise | Dunn's post hoc

Test for equal means

	Sum of sqrs	df	Mean square	F	p (same)
Between groups:	208,527	6	34,7546	104,7	7,553E-37
Within groups:	27,88	84	0,331905		Permutation p (n=99999)
Total:	236,407	90			1E-05

Components of variance (only for random effects):

Var(group):	2,6479	Var(error):	0,331905	ICC:	0,888615
--------------------	--------	--------------------	----------	-------------	----------

omega²: 0,8724

Levene's test for homogeneity of variance, from means	p (same):	0,001324
Levene's test, from medians	p (same):	0,0139

Welch F test in the case of unequal variances: F=92,14, df=36,85, p=1,139E-20

Close Copy Print Help

Como puede observarse en el recuadro en rojo, se muestran los resultados del *p-valor* obtenidos por el test:

-Si el *p-valor* es menor que el nivel de significancia (0,05) se puede rechazar la hipótesis nula y suponemos que las varianzas entre las muestras son distintas.

-En el caso contrario, cuando el *p-valor* supera al nivel de significancia, no se puede rechazar la hipótesis nula y se acepta el supuesto de homogeneidad de varianzas entre muestras.

De manera complementaria se pueden realizar representaciones gráficas para corroborar la homogeneidad obtenida con los tests anteriores. Para ver si una muestra es homogénea, se pueden utilizar gráficos como el gráfico de caja y el gráfico de dispersión. El gráfico de caja muestra la distribución de los valores de la muestra y si hay valores atípicos, mientras que el gráfico de dispersión muestra la relación entre dos variables y puede mostrar si hay un patrón en los datos.