PRUEBAS DE NORMALIDAD

MÉTODO DE KOLMOGOROV – SMIRNOV

Tal vez el método más recomendable para el caso en que *F(x)* es una distribución continua es el método para una muestra de Kolmogorov-Smirnov o (K-S). Consiste en una prueba de hipótesis en el que la hipótesis nula afirma que los datos sí se ajustan a la distribución *F(x)* y la hipótesis alterna establece que no se ajustan. El estadístico de prueba está dado por

$$D_c = Max\left\{\left|H_{i-1} - F_i\right|, \left|H_i - F_i\right|\right\}$$

este valor se compara con el valor crítico que se encuentra en una tabla. Se rechaza la hipótesis nula si *Dc* es mayor que el valor de tabla para el nivel de confianza y el tamaño de muestra que se estén considerando.

DURACIONES DE LAS BATERIAS DE UN AUTOMOVIL

2.2	4.1	3.5	4.5	3.2	3.7	3.0	2.6
3.4	1.6	3.1	3.3	3.8	3.1	4.7	3.7
2.5	4.3	3.4	3.6	2.9	3.3	3.9	3.1
3.3	3.1	3.7	4.4	3.2	4.1	1.9	3.4
4.7	3.8	3.2	2.6	3.9	3.0	4.2	3.5

Probar que los datos si se ajustan a una distribución normal con $\mu = 3.5$ y $\sigma = 0.7$

METODO DE KOLMOGOROV Y SMIRNOV (K-S)

OBSERVACION	F.REL.ACM	F(X)	$\lvert H_{i-1} - F_i \rvert$	$\lvert H_i - F_i \rvert$
1.60000	0.025000	0.0033	0.0033	0.0217
1.90000	0.050000	0.0111	0.0139	0.0389
2.20000	0.075000	0.0317	0.0183	0.0433
2.50000	0.100000	0.0766	0.0016	0.0234
2.60000	0.150000	0.0993	0.0007	0.0507
2.90000	0.175000	0.1957	0.0457	0.0207
3.00000	0.225000	0.2375	0.0625	0.0125
3.10000	0.325000	0.2839	0.0589	0.0411
3.20000	0.400000	0.3341	0.0091	0.0659
3.30000	0.475000	0.3875	0.0125	0.0875
3.40000	0.550000	0.4432	0.0318	**0.1068** *
3.50000	0.600000	0.5000	0.0500	0.1000
3.60000	0.625000	0.5568	0.0432	0.0682
3.70000	0.700000	0.6125	0.0125	0.0875
3.80000	0.750000	0.6659	0.0341	0.0841
3.90000	0.800000	0.7161	0.0339	0.0839
4.10000	0.850000	0.8043	0.0043	0.0457
4.20000	0.875000	0.8414	0.0086	0.0336
4.30000	0.900000	0.8735	0.0015	0.0265
4.40000	0.925000	0.9007	0.0007	0.0243
4.50000	0.950000	0.9234	0.0016	0.0266
4.70000	1.000000	0.9568	0.0068	0.0432

$D_c = 0.1068$, Es el mayor valor de las dos últimas columnas.
$D_T = 0.2150$, Para un nivel de significancia $\alpha = 0.05$ y n=40.
Como $D_c \prec D_T$, no se rechaza la hipótesis nula de que los datos se ajustan a una distribución normal con $\mu = 3.5$ y $\sigma = 0.7$

EL CONTRASTE DE SHAPIRO Y WILKS

Este contraste mide el ajuste de la muestra al dibujarla en papel probabilístico normal a una recta. Se rechaza la normalidad cuando el ajuste es malo, que corresponde a valores pequeños del estadístico. El estadístico es:

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

$$W = \frac{A^2}{n\,s^2}$$

Donde:

$$ns^2 = \sum_{i=1}^{i=n}\left(x_i - \overline{X}\right)^2 \qquad h = \frac{n}{2}$$ Si n es par. Si no, $h = \frac{n-1}{2}$

Los coeficientes $a_{j,n}$ están tabulados (tabla 10), y x_j es el valor ordenado en la muestra que ocupa el lugar j. La distribución de w está tabulada (tabla 11) y se rechaza la normalidad cuando el valor calculado es menor que el valor crítico dado en las tablas.

EJEMPLO

Contrastar la hipótesis de que los datos siguientes provienen de una distribución normal: (20, 22, 24, 30, 31, 32, 38). Para aplicar el test calcularemos los valores $a_{j,n}$ directamente en la tabla 10, entonces:

a_{17}= 0.6233
a_{27}= 0.3031
a_{37}= 0.1401

Por lo tanto, A será:

$$A = a_{17}(x_7 - x_1) + a_{27}(x_6 - x_2) + a_{37}(x_5 - x_2)$$

= 0.6233. (18) + 0.3031 (10) + 0.1401(7)=15.2311

Como:

s^2 = 34.9796, ns^2 = 244.8571

A^2 = 231.9864

El estadístico resultante será:

$$\omega = \frac{231.9864}{244.8571} = 0.9474$$

El valor de ω para n=7 y un nivel de significación de 0.05 es, 0.803, menor que el obtenido, por lo que aceptamos la hipótesis de normalidad.

PRUEBA DE GEARY

OBSERVACION	X_i	X_i^2	$(X_i - \bar{X})$	$\lvert X_i - \bar{X} \rvert$	$(X_i - \bar{X})^2$
1	-0,6653	0,4426	-0,6653	0,6653	0,4426
2	-2,5205	6,3529	-2,5205	2,5205	6,3529
3	-4,3756	19,1459	-4,3756	4,3756	19,1459
4	-0,0860	0,0074	-0,0860	0,0860	0,0074
5	-0,0136	0,0002	-0,0136	0,0136	0,0002
6	5,1899	26,9351	5,1899	5,1899	26,9351
7	-0,4480	0,2007	-0,4480	0,4480	0,2007
8	-2,7963	7,8193	-2,7963	2,7963	7,8193
9	0,7830	0,6131	0,7830	0,7830	0,6131
10	0,6381	0,4072	0,6381	0,6381	0,4072
11	3,1313	9,8050	3,1313	3,1313	9,8050
12	0,4933	0,2433	0,4933	0,4933	0,2433
13	0,3485	0,1215	0,3485	0,3485	0,1215
14	2,5520	6,5127	2,5520	2,5520	6,5127
15	-2,2308	4,9765	-2,2308	2,2308	4,9765
$\sum$	**0,0000**	**83,5833**	**0,0000**	**26,2722**	**83,5833**

$$\mu = \frac{\sqrt{\frac{\pi}{2}}\left(\frac{\sum_{i=1}^{i=15}\lvert X_i - \bar{X}\rvert}{n}\right)}{\sqrt{\frac{\sum_{i=1}^{i=15}(X_i - \bar{X})^2}{n}}} = \frac{(1.253314)\left(\frac{26.2722}{15}\right)}{\sqrt{\frac{83.5833}{15}}} = \frac{2.19568}{2.36056}$$

$$\mu = 0.9302$$

$$Z = \frac{\mu - 1}{\frac{0.2661}{\sqrt{15}}} = \frac{0.9302 - 1}{\frac{0.2661}{3.873}} = \frac{-0.0698}{0.0687} = -1.02$$

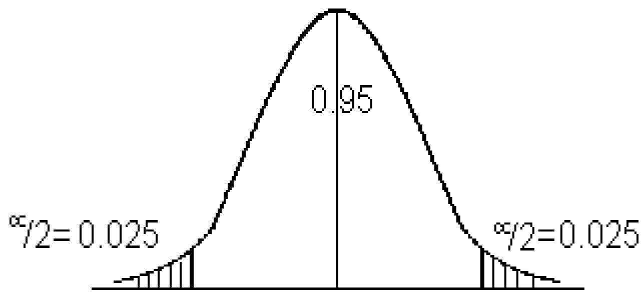
$$P = 2p\left(z = -1.02\right) = 2\left[1 - p\left(z = 1.02\right)\right] = 2\left[1 - F\left(Z = 1.02\right)\right]$$

$$P = 2\left(1 - 0.8461\right) = 2\left(0.1539\right) = 0.3078$$

1. $H_0 =$ Las observaciones se ajustan a una distribución normal
2. $H_a =$ *Las observaciones no se ajustan a una distribución normal*
3. $\alpha = 0.05$
4. *Estadístico de prueba:*

$$\mu = \frac{\sqrt{\frac{\pi}{2}}\left(\frac{\sum_{i=1}^{i=15}\left|X_i - \bar{X}\right|}{n}\right)}{\sqrt{\frac{\sum_{i=1}^{i=15}\left(X_i - \bar{X}\right)^2}{n}}}$$

5. Criterio de decisión: Si P < α se rechaza H_O
6. *Gráfica:*



7. *Decisión: como P > α, no se rechaza H_O*

8. Conclusión: Las observaciones se distribuyen en forma normal

PRUEBA DE ANDERSON – DARLING

Esta prueba es aplicada para evaluar el ajuste a cualquier distribución de probabilidades. Se basa en al comparación de la distribución de probabilidades acumulada empírica (resultado de los datos) con la distribución de probabilidades acumulada teórica (definida por H0).

Hipótesis:

Ho: La variable sigue una distribución Normal $\left(\mu-\sigma^2\right)$

H1 : La variable no sigue una distribución Normal $\left(\mu-\sigma^2\right)$

Estadístico de Prueba:

$$A^2=-n-S$$

$$S=\frac{1}{n}\sum_{i=1}^{n}\left(2i-1\right)\left[\ln F\left(Yi\right)+\ln\left(1-F\left(Y_{n+1-i}\right)\right)\right]$$

Donde **n** es el número de observaciones, **F(Y)** es la distribución de probabilidades acumulada normal con media y varianza especificadas a partir de la muestra y **Yi** son los datos obtenidos en la muestra, ordenados de menor a mayor.

Regla de Decisión:

La hipótesis nula se rechaza con un nivel de significancia α si A^2 es mayor que el valor critico A_T^2. Aunque la prueba de Anderson – Darling puede ser aplicada a cualquier distribución, no se dispone de tablas para todos los casos. A continuación se presenta una tabla para la prueba a la distribución normal.

α	0.1	0.05	0.025	0.01
A_T^2	0.631	0.752	0.873	1.035

Ejemplo:

Pruebe si los siguientes datos se distribuyen o no en forma normal

0.25	0.00	0.70	-0.14	-0.15
0.50	0.10	0.26	0.55	-0.10
-0.40	-0.04	-0.65	0.20	
-0.64	-0.05	-0.40	-0.10	
0.05	-0.20	-0.30	0.56	

Ho: Los datos siguen una distribución Normal $\left(0,\sigma^2\right)$

H1: Los datos no siguen una distribución Normal $\left(0,\sigma^2\right)$

$$\overline{X} = \frac{\sum_{i=1}^{22} x_i}{n} = 0 \qquad S = \sqrt{\frac{\sum_{i=1}^{22} x_i^2 - \frac{\left(\sum_{i=1}^{22} x_i\right)^2}{n}}{n-1}} = 0.3708$$

(2i-1)	Yi	Yn+1-i	F(Yi)	F(Yn+1-i)	1-F(Yn+1-i)	ln(F(Yi))	ln(1-F(Yn+1-i))	Si
1	-0.65	0.70	0.0398	0.9705	0.0295	-3.224	-3.522	-0.307
3	-0.64	0.56	0.0422	0.9345	0.0655	-3.166	-2.726	-0.803
5	-0.40	0.55	0.1403	0.9310	0.0690	-1.964	-2.674	-1.054
7	-0.40	0.50	0.1403	0.9112	0.0888	-1.964	-2.422	-1.395
9	-0.30	0.26	0.2092	0.7584	0.2416	-1.564	-1.420	-1.221
11	-0.20	0.25	0.2948	0.7499	0.2501	-1.221	-1.386	-1.304
13	-0.15	0.20	0.3429	0.7052	0.2948	-1.070	-1.221	-1.354
15	-0.14	0.10	0.3529	0.6063	0.3937	-1.042	-0.932	-1.346
17	-0.10	0.05	0.3937	0.5536	0.4464	-0.932	-0.807	-1.344
19	-0.10	0.00	0.3937	0.5000	0.5000	-0.932	-0.693	-1.404
21	-0.05	-0.04	0.4464	0.4570	0.5430	-0.807	-0.611	-1.353
23	-0.04	-0.05	0.4570	0.4464	0.5536	-0.783	-0.591	-1.437
25	0.00	-0.10	0.5000	0.3937	0.6063	-0.693	-0.500	-1.356
27	0.05	-0.10	0.5536	0.3937	0.6063	-0.591	-0.500	-1.340
29	0.10	-0.14	0.6063	0.3529	0.6471	-0.500	-0.435	-1.233
31	0.20	-0.15	0.7052	0.3429	0.6571	-0.349	-0.420	-1.084
33	0.25	-0.20	0.7499	0.2948	0.7052	-0.288	-0.349	-0.956
35	0.26	-0.30	0.7584	0.2092	0.7908	-0.277	-0.235	-0.813
37	0.50	-0.40	0.9112	0.1403	0.8597	-0.093	-0.151	-0.411
39	0.55	-0.40	0.9310	0.1403	0.8597	-0.071	-0.151	-0.395
41	0.56	-0.64	0.9345	0.0422	0.9578	-0.068	-0.043	-0.207
43	0.70	-0.65	0.9705	0.0398	0.9602	-0.030	-0.041	-0.138
TOTAL								**-22.253**

$$A^2 = -22 + 22.253 = 0.2532$$

Este valor es menor inclusive al valor crítico correspondiente a $\alpha = 0.1$. Por lo tanto se acepta el supuesto de normalidad de los datos