Unidad II

PRUEBAS ESTADÍSTICAS PARA EVALUAR UNA MUESTRA

“Cada uno de nosotros ha estado haciendo todas las estadísticas de su vida, en el sentido de que cada uno de nosotros ha querido llegar a conclusiones basadas en observaciones empíricas desde su nacimiento”

William Kruskal

Introducción

En algunas investigaciones, el objetivo es analizar el comportamiento de uno o varios parámetros de las variables de interés. Para realizar este análisis, se debe seleccionar una muestra aleatoria de tamaño n proveniente de la población de tamaño N.
Esta muestra se obtiene por lo general siguiendo un esquema de Muestreo Aleatorio Simple (MAS) sin reemplazo. En otras palabras, un supuesto importante para poder aplicar las pruebas que se desarrollarán en esta unidad es que los datos recolectados provienen de una muestra aleatoria (adicionalmente se presentarán supuestos propios para cada prueba). Esto desmiente lo que algunos investigadores consideran erróneamente que las pruebas no paramétricas no requieren del cumplimiento de supuestos.

De los elementos o unidades de muestreo que forman parte de una muestra aleatoria se miden u observan diferentes variables (cuantitativas o cualitativas), para analizar cada una de ellas se puede utilizar distintas pruebas estadísticas.
Una vez que se tienen los datos de los elementos que forman parte de la muestra, dos son los aspectos más importantes que se deben considerar para elegir la prueba estadística adecuada, ya sea paramétrica o no paramétrica: primero se debe tener claro el objetivo de la investigación; segundo, se debe identificar el tipo de variable que se desea analizar.

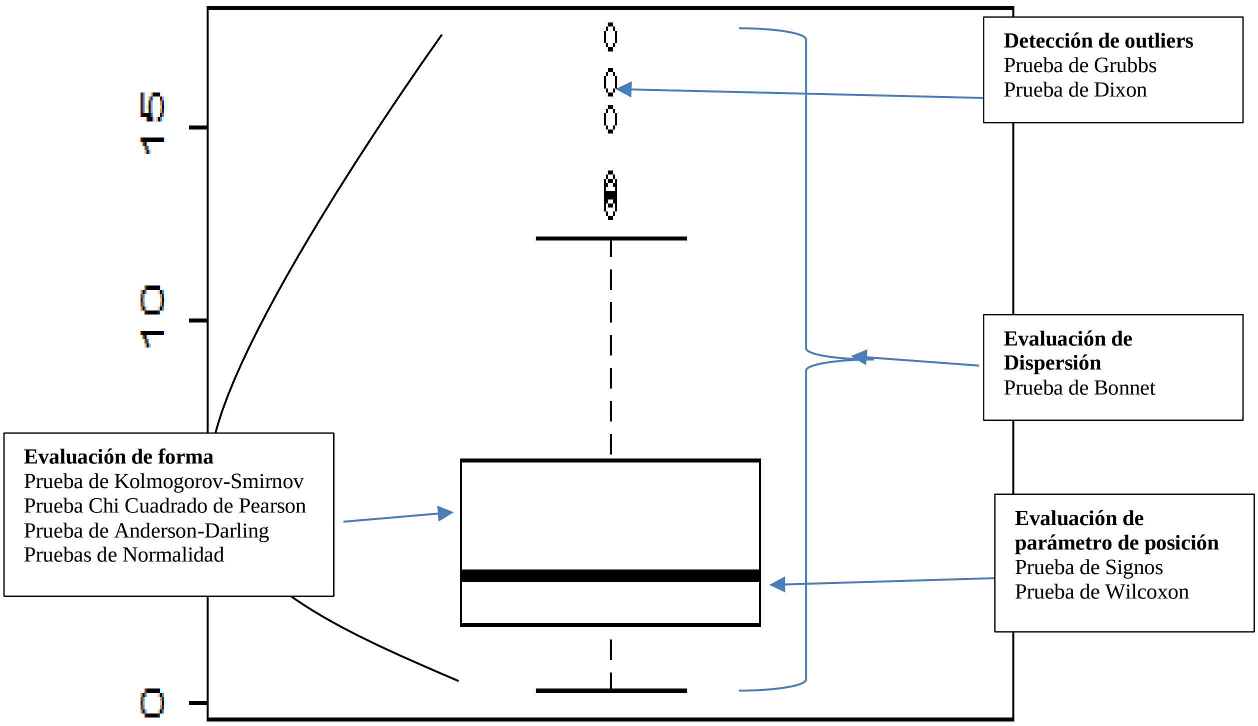
Por otro lado, a partir de los datos de una muestra aleatoria se puede tener como objetivo, por ejemplo, evaluar un parámetro de locación (como la mediana), el cual puede ser analizado por diferentes pruebas no paramétricas como la prueba de Signos o la prueba de Wilcoxon. Es decir, una interrogante que se debe despejar es determinar o elegir ¿Cuál de las pruebas es la más poderosa? Esto se puede lograr aplicando simulación, es decir generando r muestras aleatorias y en cada una de ellas se evalúan las pruebas para determinar cuál de ellas se equivoca menos (bajo las mismas condiciones).
Esta unidad tiene como objetivo desarrollar algunas pruebas (de las muchas pruebas no paramétricas que se pueden utilizar) que permiten analizar los datos de una variable (cualitativa o cuantitativa) provenientes de los elementos que forman parte de una muestra aleatoria.

Una inquietud adicional que se puede hacer el lector es ¿Cómo se puede clasificar a las pruebas no paramétricas para analizar a los datos provenientes de una muestra?

Nosotros lo haremos dando respuesta a ciertas inquietudes que puedan surgir en una investigación, como las siguientes: ¿Qué prueba permite el (la)

- Análisis de una variable cualitativa o cuantitativa?
- Evaluación de supuestos estadísticos (aleatoriedad, normalidad) que exigen los métodos paramétricos?
- Evaluación de la existencia de datos atípicos en la muestra?
- Análisis de un parámetro de posición o escala?.

Con la ayuda del siguiente gráfico de cajas se pueden resumir algunas de las pruebas que se desarrollarán en la presente unidad.



Finalmente, en esta unidad se brindará los primeros alcances sobre la comparación de diferentes pruebas no paramétricas mediante simulación.

1. Prueba para evaluar una variable dicotómica

A. Prueba Binomial

- **Aspectos Generales**

Esta prueba se emplea cuando se evalúan los datos provenientes de una variable X que se puede expresar en forma dicotómica. Los dos tipos de resultados se pueden clasificar en "éxito" o "fracaso" con lo que posteriormente se puede aplicar la función de probabilidad binomial para poder calcular la probabilidad de ocurrencia de la hipótesis nula formulada. Cada resultado pertenece a una de las clases, pero no a ambos.
El número de observaciones de la clase "éxito" es O_1 y el número de observaciones de la clase "fracaso" es O_2=n-O_1.

- **Supuestos**

- Los n elementos pertenecientes a la muestra son mutuamente independientes, es decir la selección de un elemento no afecta la selección de otro elemento.
- Cada elemento tiene probabilidad π de que resulte de la clase "éxito", siendo π el mismo para las *n* elementos.

- **Inferencia Estadística**

a) Prueba de Hipótesis

Procedimiento para el desarrollo de la Prueba de Hipótesis

- Contabilizar el número de éxitos O_1 de la prueba
- Calcular el p-valor según la hipótesis alterna, mediante los criterios que a continuación se presentarán

Sea π_0 es alguna constante especificada, ($0<\pi_0<1$), es decir el valor hipotético a evaluar.
Las hipótesis pueden tomar alguna de las siguientes formas:

Hipótesis

Bilateral	Unilateral	
Caso A	Caso B	Caso C
$H_0 : \pi = \pi_0$	$H_0 : \pi \leq \pi_0$	$H_0 : \pi \geq \pi_0$
$H_1 : \pi \neq \pi_0$	$H_1 : \pi > \pi_0$	$H_1 : \pi < \pi_0$

Prueba Estadística
Caso A

Suponga que α_1 y α_2 son los tamaños de la cola superior e inferior de la región crítica, respectivamente, tal que $\alpha = \alpha_1 + \alpha_2$ es el tamaño de la región crítica. Luego se tendría que hallar t_1 y t_2 tal que:

$$P(Y \leq t_1) = \alpha_1 \text{ y } P(Y \leq t_2) = 1-\alpha_2$$

Los valores de α_1 y α_2 podrían ser aproximadamente iguales.

Se rechaza H_0 si: $T \leq t_1$ ó $T > t_2$
No se rechaza H_0 si: $t_1 < T \leq t_2$

Caso B
Dado que un valor grande de *T* indica que H_0 es falsa la región crítica de tamaño α consiste en considerar todos los valores de *T* mayores que *t* tal que:

$$P(Y \leq t) = 1-\alpha$$

Se rechaza H_0 si: $T>t$

Caso C
Dado que un valor pequeño de *T* indica que H_0 es falsa la región crítica de tamaño α consiste en considerar todos los valores de *T* menores o iguales a *t* tal que:

$$P(Y \leq t) = \alpha$$

Se rechaza H_0 si: $T \leq t$

P-valor
Si se está interesado en la probabilidad de que el resultado sea de la clase "éxito", entonces el estadístico de prueba *T* está dado por el número de observaciones de la clase "éxito".

$$T=O_1$$

El cálculo del p-valor para cada uno de los casos mencionados anteriormente es de la siguiente manera:

Caso A
Calcular el p-valor para una prueba bilateral no es tan directo porque depende del número de éxitos obtenidos y del valor esperado de la prueba $E(X)=n\pi=m$ y del valor de probabilidad $d = P(X=O_1)$
Si $O_1 \leq m$

- Obtener una secuencia i desde m (redondeado por exceso) hasta n.
- Obtener "y" como la suma de probabilidades en i hasta que estas sean menores que d.

El p-valor es $P(X \leq O_1) + P(X>n-y)$

Si $O_1 > m$

- Obtener una secuencia desde 0 hasta m (redondeado por defecto).
- Obtener "y" como la suma de probabilidades en i hasta que estas sean menores que d.

El p-valor es $P(X \leq y-1) + P(X>O_1-1)$

Caso B

$$p-valor = P(X \geq O_1) = \sum_{i=O_1}^{n} \binom{n}{i} \pi^i (1-\pi)^{n-i}$$

Caso C

$$p-valor = P(X \leq O_1) = \sum_{i=0}^{O_1} \binom{n}{i} \pi^i (1-\pi)^{n-i}$$

Si el tamaño de muestra es grande se puede hacer uso de la aproximación a la normal, mediante:

$$Z = \frac{(O_1 \pm 0.5) - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} = \frac{(O_1/n \pm 0.5) - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Donde: $O_1+0.5$ si $x<n\pi_0$ y $O_1-0.5$ si $x>n\pi_0$

b) Intervalo de confianza para π

El intervalo exacto viene dado por la siguiente expresión:

$$LI(\pi) = \frac{O_1}{(n-O_1+1)F_{[1-\alpha/2,2(n-O_1+1),2O_1]} + O_1}$$

$$LS(\pi) = \frac{(O_1+1)F_{[1-\alpha/2,2(O_1+1),2(n-O_1)]}}{(n-O_1)+(O_1+1)F_{[1-\alpha/2,2(O_1+1),2(n-O_1)]}}$$

Cuando se tiene un tamaño de muestra grande, se puede hacer uso de la aproximación a la normal. Con lo que se obtendría:

$$LI(\pi) = \frac{O_1}{n} - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{O_1(n-O_1)}{n^3}}$$

$$LS(\pi) = \frac{O_1}{n} + Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{O_1(n-O_1)}{n^3}}$$

Puede deducirse que haciendo $p = O_1/n$, este es el clásico intervalo para una proporción poblacional π, el cual presenta muchas deficiencias.
A esta última expresión de intervalo se le conoce como el intervalo para una proporción de Wald.

- **Aplicación**

Ejemplo 1
A una muestra aleatoria de 26 degustadores igualmente entrenados se les dio a probar un nuevo sabor de refresco en tres vasos, los cuales les fueron proporcionados de manera aleatoria. Dos de los vasos tenían igual cantidad del saborizante (ingrediente base del refresco) y un vaso tenía una mayor cantidad de saborizante. A los degustadores se les pidió que identifiquen el vaso que contenía la mayor cantidad de saborizante. Dieciocho degustadores identificaron correctamente que vaso contenía mayor cantidad de saborizante. Pruebe si la proporción de degustadores que realizaron la correcta identificación es mayor a 1/3.
De sus conclusiones a un nivel de significación de 0.05.

Sea Y: Número de degustadores que identificaron correctamente que vaso contenía mayor cantidad de saborizante.

$$H_0 : \pi \le \frac{1}{3}$$

$$H_1 : \pi > \frac{1}{3}$$

α=0.05

Prueba Estadística
T=O_1=18

Asumiendo que $Y \sim B\left(26, \frac{1}{3}\right)$

$p-valor = P(Y \geq 18) = \sum_{y=18}^{26} \binom{26}{y} \left(\frac{1}{3}\right)^y \left(\frac{2}{3}\right)^{26-y} = 0.000197$ < α se rechaza H_0.

Conclusión
A un nivel de significación de 0.05, para rechazar H_0.
Por lo tanto, se puede afirmar que la proporción de degustadores que identifican correctamente el vaso que contenía mayor cantidad de saborizante es mayor a 1/3.

Ejemplo 1 (Continuación)
Halle e interprete un intervalo del 95% de confianza para el ejemplo anterior. El resultado exacto se obtiene de la siguiente manera:

$$LI(\pi) = \frac{18}{(26-18+1)F_{[0.025,2(26-18+1),2(18)]} + 18} = \frac{18}{(9)(2.148)+18} = 0.4821$$

$$LS(\pi) = \frac{(18+1)F_{[1-\alpha/2,2(18+1),2(26-18)]}}{(26-18)+(18+1)F_{[1-\alpha/2,2(18+1),2(26-18)]}} = \frac{19(2.518)}{8+19(2.518)} = 0.856$$

IC(π)=[0.4821;0.856]

El intervalo que va de 0.4821 a 0.856 brinda un 95% de confianza de contener a la proporción de degustadores que identificaron correctamente que vaso contenía mayor cantidad de saborizante.
El resultado aproximado es:

$$LI(\pi) = \frac{18}{26} - 1.96\sqrt{\frac{18(8)}{26^3}} = 0.5148$$

$$LS(\pi) = \frac{18}{26} + 1.96\sqrt{\frac{18(8)}{26^3}} = 0.8697$$

IC(π)=[0.5148;0.8697]

- **Funciones en R**

```
1-pbinom(17,26,1/3)
[1] 0.0001971944
#La opción lower.tail permite calcular P(X>x)
pbinom(17,26,1/3,lower.tail=FALSE)
[1] 0.0001971944
```

También existe la función binom.test

```
binom.test(O1,n,valor hipotético, alternative, conf.level)
```

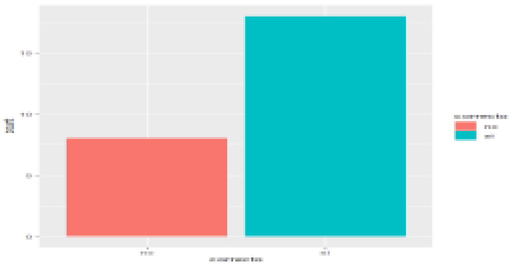
Para muestras grandes se puede hacer uso de la función prop.test con los mismos argumentos de la función binom.test. La función prop.test presenta el argumento `correct`, cuando se coloca TRUE se calcula el estadístico aplicando la corrección por continuidad.

- **Resultados con R**

```
datos1<-data.frame(c(rep("si",18),rep("no",8)))
fix(datos1)
```

Aquí aparecerá una ventana, si se da doble sobre el encabezado se puede cambiar el nombre a correcto

```
library(ggplot2)
ggplot(data=datos1)+geom_bar(mapping=aes(x=correcto,fill=correcto))
```



Para la prueba de hipótesis

```
binom.test(18,26,1/3,alternative="greater")
```

```
        Exact binomial test
data:  18 and 26
number of successes = 18, number of trials = 26, p-value =
0.0001972
alternative hypothesis: true probability of success is
greater than 0.3333333
```

Para el intervalo de confianza

```
binom.test(18,26,1/3,alternative="two.sided")
```

```
           Exact binomial test
data:  18 and 26
number of successes = 18, number of trials = 26, p-value =
0.0002236
alternative hypothesis: true probability of success is not
equal to 0.3333333
95 percent confidence interval:
 0.4821036 0.8567400
```

- **Algunas consideraciones de R**

- Permite solo trabajar con datos agrupados, es decir se debe conocer la cantidad de éxitos existentes.
- Permite analizar una hipótesis bilateral o unilateral.
- Si se tienen datos no agrupados en una tabla de frecuencia primero se debe utilizar la función *table* para poder obtener la frecuencia de la categoría de interés.

2. Prueba para evaluar supuestos

2.1 Pruebas para determinar la distribución de los datos (Pruebas de Bondad de Ajuste)

En general, con este tipo de pruebas se utiliza para verificar la siguiente hipótesis:

H_0: X se ajusta a una F(X)

H_1: X no se ajusta a una F(X)

A. Prueba de Kolmogorov-Smirnov

- **Aspectos Generales**

Esta prueba es utilizada para determinar si los datos de la muestra de tamaño n se ajusta a alguna distribución teórica.

La prueba consiste en hacer una comparación entre alguna función de distribución acumulada teórica, $F_T(x)$ y la función de distribución acumulada empírica de la muestra, $F_S(x)$.

Por medio de la función de distribución acumulada de la muestra $F_S(x)$, se puede determinar $P(X \leq x)$; si existe una estrecha concordancia entre las distribuciones acumuladas teórica y de la muestra, se apoya la hipótesis de que la muestra se extrajo de la población con la función de distribución acumulada que se especifica, $F_T(x)$. Sin embargo, si existe una discrepancia entre las funciones de distribución acumulada teórica y observada, demasiado grande como para ser atribuida unicamente al azar, se rechaza la hipótesis, cuando H_0 es verdadera.

Andrey Kolgomorov
(1903 -1987)

Nikolai Smirnov
(1900 – 1966)

- **Supuestos**
  - La distribución $F_T(x)$ establecida en la hipótesis es continua.

- **Inferencia Estadística**

Procedimiento para el desarrollo de la Prueba de Hipótesis

  - Obtener los valores de la función de distribución acumulada empírica de una muestra ($F_S(x)$).

- Calcular función de distribución acumulada y teórica ($F_T(x)$); convirtiendo cada valor observado de *x* a un valor de una variable normal unitaria.
  La diferencia entre la función de distribución, acumulada y teórica, $F_T(x)$, y la función de distribución acumulada de la muestra $F_S(x)$, se mide por medio de la estadística D, que es la distancia vertical máxima entre $F_S(x)$ y $F_T(x)$.

  La estadística es:

$$D = \sup_x \left| F_S(x) - F_T(x) \right|$$

  la cual se lee, "D es igual al supremo (máximo) sobre todos los *x*, del valor absoluto de la diferencia $F_S(x)$ menos $F_T(x)$.

  El valor del estadístico de prueba, no siempre se determina calculando y eligiendo los posibles valores de $\left| F_S(x) - F_T(x) \right|$. La distancia vertical máxima entre $F_S(x)$ y $F_T(x)$ puede no ser un valor observado, *x*, pero sí algún otro valor de X.

  Puede determinarse algebraicamente el valor correcto de *D* al calcular, además de las diferencias $|F_S(x)-F_T(x)|$, las diferencias $|F_S(x_{i-1})-F_T(x_i)|$ para todos los valores de $i = 1, 2, \ldots, r+1$, donde r es el número de diferentes valores de *x* y $F_S(x_0) = 0$. El valor correcto de la estadística de prueba será entonces:

$$D = \underset{1 \le i \le r}{máximo} \left\{ máximo \left[\left| F_S(x_i) - F_T(x_i) \right|, \left| F_S(x_{i-1}) - F_T(x_i) \right| \right] \right\}$$

  Otra forma de obtener el estadístico de prueba y el p-valor es la siguiente:
- Calcular el punto medio $M=(X_{max}+X_{min})/2$ y la desviación estándar muestral (s) de los datos.
- Calcular las puntuaciones z y frecuencias relativas observadas para cada valor de la puntuación

$$z = \left| \frac{x_i - M}{s} \right|$$

- Calcular las probabilidades p_x de la siguiente manera $P(Z>z)$
- Obtener el valor de la frecuencia empírica $\hat{f}_{ri} = p_{xi} - \hat{f}_{ri-1}$.
  Es decir, ya se debe contar con: la frecuencia relativa observada f_{ri} la frecuencia empírica $\hat{f}_{ri}$, la frecuencia relativa acumulada observada F_{ri} y la frecuencia relativa acumulada empírica $\hat{F}_{ri}$.
- Estimar las divergencias $\tilde{D} = \left| \hat{F}_{xi} - F_{ri} \right|$ y $\tilde{D} = \left| \hat{F}_{xi} - F_{ri-1} \right|$
- Obtener el estadístico de prueba dado por:

$$Z = \sqrt{n} \max\left(|D|, |\tilde{D}| \right)$$

  El p-valor se puede determinar de la siguiente manera:
  Si $0 \le Z < 0.27$, entonces p-valor=1

Si 0.27≤Z<1, entonces $p-valor = 1 - \frac{2.506628}{Z}\left(Q + Q^9 + Q^{25}\right)$ con $Q = e^{-1.233701Z^{-2}}$

Si 1≤Z<3.1, entonces $p-valor = 2\left(Q - Q^4 + Q^9 - Q^{16}\right)$ con $Q = e^{-2Z^2}$

Si Z≥ 3.1, entonces p-valor=0

- **Aplicación**

Ejemplo

Las determinaciones de glucosa en la sangre en ayunas de 36 hombres adultos no obesos elegidos al azar, y aparentemente sanos, se muestran en la siguiente tabla.

75	84	80	77	68	87	92	77	92
86	78	76	80	81	72	77	92	80
80	77	77	92	68	87	84	75	78
80	80	77	72	81	76	78	81	86

Pruebe a un nivel de significación de 0.05 si los datos provienen de una distribución normal con media de 80 y desviación estándar de 6.

Solución

H_0: $F(x) = F_T(x)$, para (Las observacones se distribuyen normalmente con los parámetros especificados)

H_1: $F(x) \neq F_T(x)$, para al menos x (Las observacones no se distribuyen normalmente con los parámetros especificados)

α= 0.05.

Prueba Estadística

$$D = \underset{1 \le i \le r}{máximo}\left\{máximo\left[\left|F_S(x_i) - F_T(x_i)\right|, \left|F_S(x_{i-1}) - F_T(x_i)\right|\right]\right\}$$

Criterio de Decisión

No se rechaza si H_0 $D \le 0.1547$

Se rechaza H_0 si $D > 0.1547$

Desarrollo de la Prueba

Cálculo de $F_S(x)$

X	*Frecuencia*	*Frecuencia Acumulada*	$F_S(x)$
68	2	2	2/36 = 0.0556
72	2	4	4/36 = 0.1111
75	2	6	6/36 = 0.1667
76	2	8	0.2222
77	6	14	0.3889
78	3	17	0.4722
80	6	23	0.6389
81	3	26	0.7222

84	2	28	0.7778
86	2	30	0.8333
87	2	32	0.8889
92	4	36	1.0000
	36		

Cálculo de $F_T(x)$

X	***z=(x-80)/6***	***$F_T(x)$***
68	-2.00	0.0228
72	-1.33	0.0918
75	-0.83	0.2033
76	-0.67	0.2514
77	-0.50	0.3085
78	-0.33	0.3707
80	0.00	0.5000
81	0.17	0.5675
84	0.67	0.7486
86	1.00	0.8413
87	1.17	0.8790
92	2.00	0.9772

X	***$F_S(x)$***	***$F_T(x)$***	$\left\|F_S(x)-F_T(x)\right\|$
68	0.0556	0,0228	0.0328
72	0.1111	0,0918	0.0193
75	0.1667	0,2033	0.0366
76	0.2222	0,2514	0.0292
77	0.3889	0,3085	0.0804
78	0.4722	0,3707	0.1015
80	0.6389	0,5000	0.1389
81	0.7222	0,5675	**0.1547**
84	0.7778	0,7486	0.0292
86	0.8333	0,8413	0.0080
87	0.8889	0,8790	0.0099
92	1.000	0,9772	0.0228

Conclusiones
Existe suficiente evidencia estadística a un nivel de significación de 0.05 para no rechazar la H_0.
Por lo tanto, podemos afirmar que la muestra puede haber provenido de una distribución normal con los parámetros especificados.

- **Funciones en R**

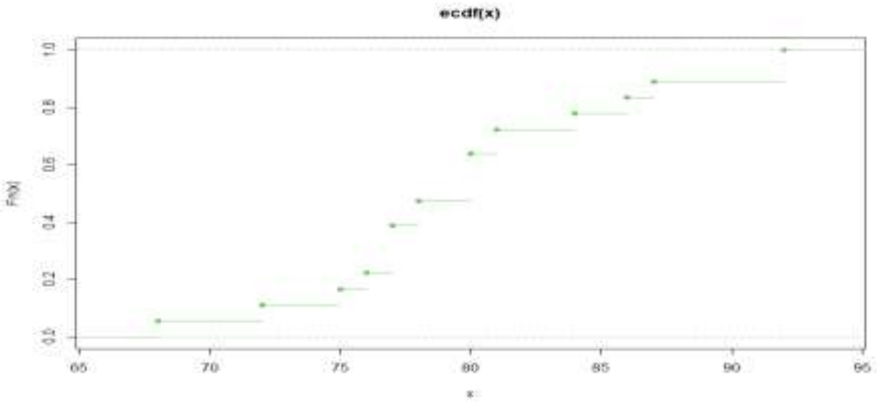
```
ks.test(vector de datos, distribución teórica, parámetros de
la distribución)
```

Existe la función ks.test.imp del paquete kolmim que presenta una rutina mejorada que ks,test pero de igual manera no presenta la corrección para empates.

➢ **Resultados con R**

```
datos<-c(75,84,80,77,68,87,92,77,92,86,78,76,80,81,72,77,92
,80,80,77,77,92,68,87,84,75,78,80,80,77, 72,81,76,78,81,86)

plot.ecdf(datos,col=3)
```




```
ks.test(datos,"pnorm",mean=80,sd=6)
```

```
       One-sample Kolmogorov-Smirnov test
data:  datos
D = 0.156, p-value = 0.3447
alternative hypothesis: two-sided
Mensajes de aviso perdidos
In ks.test(datos, "pnorm", 80, 6) :
  ties should not be present for the Kolmogorov-Smirnov test
```

```
library(kolmim)
ks.test.imp(datos,"pnorm",mean=80,sd=6)
```

```
     One-sample two-sided exact Kolmogorov-Smirnov test

data:  datos
D = 0.15604, p-value = 0.3115
alternative hypothesis: two-sided

Warning message:
In ks.test.imp(datos, "pnorm", mean = 80, sd = 6) :
  ties should not be present for the Kolmogorov-Smirnov test
```

➢ **Algunas consideraciones de R**

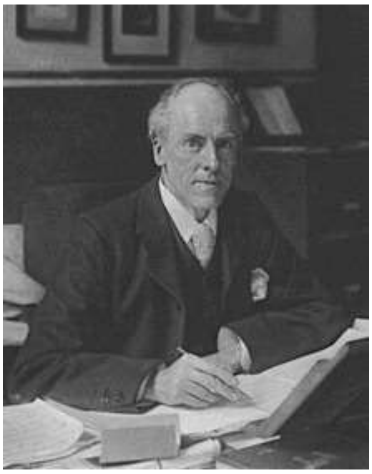
Se puede utilizar la función ks.test, la cual permite verificar si los datos se ajustan a las distintas funciones de distribución teóricas que ofrece el R.

B. Prueba Chi Cuadrado de Pearson

B.1 Ajuste a la Multinomial (Prueba de Frecuencias o Proporciones)

- **Aspectos Generales**

Esta prueba se utiliza cuando se desea verificar si al menos una de las probabilidades teóricas (π_i) de las categorías de una variable nominal es diferente en al menos una de las especificadas. Esto se puede verificar comparando si las frecuencias observadas (O_i) pertenecientes a la *i*-ésima categoría con su respectiva frecuencia teórica o frecuencia esperada (e_i) difieren significativamente.



Karl Pearson
(1857 – 1936)

- **Supuesto**
  - La variable de interés es de tipo cualitativa (nominal u ordinal).

- **Inferencia Estadística**

En esta prueba las probabilidades teóricas o hipotéticas (π_i) son establecidas por el investigador, de tal manera que:

$$\sum_{i=1}^{k} \pi_i = 1$$

- **Procedimiento para el Desarrollo de la Prueba de Hipótesis**

Prueba de Hipótesis

H_0: $\pi_i = \pi_{i0}$ $\forall$ i =1, …,k

H_1: Al menos una π_i es diferente a las especificadas.

- Agrupar los datos en una tabla de frecuencia de la siguiente manera:

N°	Nombre de la categoría	Frecuencia observada (O_i)
1	A_1	O_1
2	A_2	O_2
⋮	⋮	⋮
K	A_k	O_k
	Total	n

- Calcular las frecuencias esperadas (e_i) multiplicando el tamaño de la muestra *n* por la probabilidad teórica correspondiente (π_i), $e_i=n\pi_i$ de tal manera que $\sum_{i=1}^{k} e_i = n$

- Aplicar la prueba estadística

$$\sum_{i=1}^{k}\frac{(O_i-e_i)^2}{e_i} \sim \chi^2_{(1-\alpha,k-1)}$$

Si se tiene un solo grado de libertad para el valor crítico, el tamaño de la muestra es pequeño o si se tienen varios valores esperados menores a 5, se puede hacer uso de la Corrección por continuidad de Yates, el cual hace un ajuste al estadístico χ^2.

$$\chi^2_c = \sum_{i=1}^{k}\frac{\left(\left|o_i-e_i\right|-0.5\right)^2}{e_i} \sim \chi^2_{(1-\alpha,k-1)}$$

- **Aplicación**
Ejemplo
En una fábrica se cuenta con tres máquinas que producen el mismo producto. El jefe de producción desea determinar si las máquinas están produciendo en diferentes proporciones. Para despejar sus dudas selecciona al azar 135 artículos de la última semana de producción y los clasifica según la máquina que lo ha producido. A continuación, se presenta la tabla de frecuencia de las cantidades producidas por cada máquina:

Máquina A	Máquina B	Máquina C
43	53	39

Use nivel de significación 5% para probar si la cantidad producida no es la misma en las 3 máquinas.

Solución:
Planteamiento de la hipótesis.
H_0: $\pi_1=\pi_2=\pi_3=1/3$ (Las 3 máquinas producen en igual proporción)
H_1: Al menos un π_i es diferente a los especificados i=1,2,3. (Las 3 máquinas no producen en igual proporción)
$\alpha=0.05$
Prueba Estadística

$$\sum_{i=1}^{k}\frac{(O_i-e_i)^2}{e_i} \sim \chi^2_{(1-\alpha,k-1)}$$

Desarrollo de la Prueba

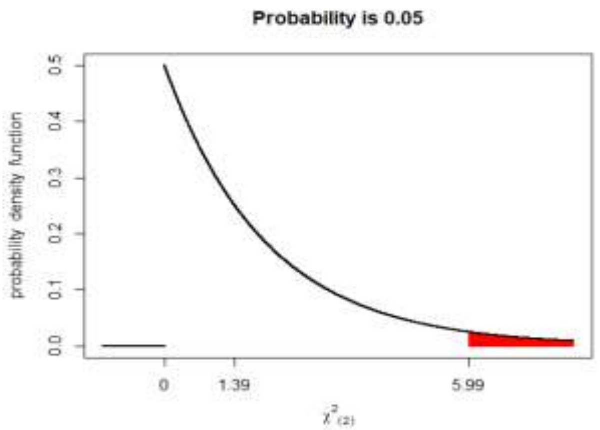
A continuación, se muestra la tabla que contiene las frecuencias observadas, las frecuencias esperadas entre otros valores que se requieren para esta prueba.

Máquina	O_i	π_i	$e_i=n\pi_i$	$(o_i-e_i)^2/e_i$
A	43	1/3	45	0.08888889
B	53	1/3	45	1.42222222
C	39	1/3	45	0.80000000
Total	135	1	135	2.31111111

$$\sum_{i=1}^{k}\frac{(O_i-e_i)^2}{e_i}=2.311$$

```
library(fastGraph)
shadeDist(qchisq(0.05,2,lower.tail=F),"dchisq",2,lower.tail
=F)
```

Criterio de decisión



No se rechaza H_0 si: $\chi^2_c \leq 5.9915$
Se rechaza H_0 si: $\chi^2_c > 5.9915$

Conclusión.
A un nivel de significación del 5% no se puede afirmar que las 3 máquinas no producen en igual proporción.

- **Funciones en R**

```
chisq.test(vector de datos, probabilidades teóricas)
```

El método exacto se consigue usando la distribución Multinomial que se encuentra en la función multinomial.test del paquete RVAideMemoire.

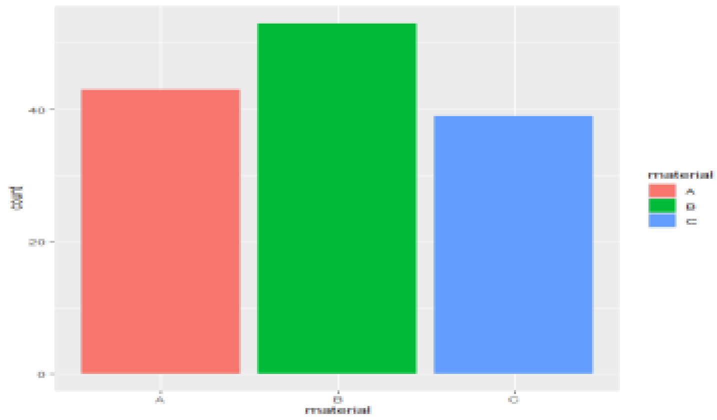
- **Resultados en R**

```
datos<-c(43,53,39)
prob<-c(1/3,1/3,1/3)

data<-data.frame(rep(c("A","B","C"),datos))
fix(data)
```

Cambie el nombre de la columna por el material.

```
library(ggplot2)
ggplot(data = data) +
  geom_bar(mapping = aes(x = material,fill=material))
```




```
chisq.test(datos,p=prob)
```

```
        Chi-squared test for given probabilities
data:  datos
X-squared = 2.3111, df = 2, p-value = 0.3149
```

```
library(RVAideMemoire)
datos<-c(43,53,39)
prob<-c(1/3,1/3,1/3)
multinomial.test(datos,prob)
```

```
     Exact multinomial test

data:  datos
p-value = 0.3189
```

Otra función que brinda el método exacto es la función xmulti del paquete Xnomial:

```
library(XNomial)
datos<-c(43,53,39)
prob<-c(1,1,1)
xmulti(datos,prob)
```

```
P value (LLR) = 0.3189
```

- **Algunas consideraciones con R**

- La función chisq.test permite solo trabajar con datos agrupados.
- Si se tienen datos no agrupados en una tabla de frecuencia primero se debe utilizar la función *table* para poder obtener la frecuencia de la categoría de interés.

B.2 Ajuste a una distribución teórica

- **Aspectos Generales**

Esta prueba se utiliza cuando se desea verificar si los datos recolectados se ajustan a una distribución teórica. Esto se puede verificar comparando las frecuencias observadas (O_i) perteneciente al *i*-ésimo valor de la variable (o i-ésimo intervalo de clase) difiere significativamente de su respectiva frecuencia teórica o frecuencia esperada (e_i).

En esta prueba las probabilidades teóricas o hipotéticas (π_i) son calculadas considerando la distribución teórica *f(x),* de tal manera que:

$$\sum_{i=1}^{k} \pi_i = 1$$

Recordemos que una *f(x)* puede tener uno o más parámetros. Por lo tanto, todos los parámetros pueden ser conocidos o algunos conocidos y otros desconocidos. El valor crítico depende de la cantidad "m" de parámetros a estimar a partir de la muestra.

- **Supuesto**

- La variable de interés es de tipo cuantitativa (discreta o continua).

- **Inferencia Estadística**

Prueba de Hipótesis

H_0: Los datos de la variable X se ajustan a la distribución teórica F(x)
H_1: Los datos de la variable X no se ajustan a la distribución teórica F(x)

Procedimiento para el Desarrollo de la Prueba de Hipótesis

- Agrupar los datos en una tabla de frecuencia de la siguiente manera:

N°	Valor de la variable(si es discreta) o Intervalos de clase (si es continua)	Frecuencia observada (O_i)
1	X_1 o $[LI_1; LS_1)$	O_1
2	X_2 o $[LI_2; LS_2)$	O_2
⋮	⋮	⋮
k	X_k o $[LI_k; LS_k]$	O_k
	Total	n

- Calcular las frecuencias esperadas (e_i) multiplicando el tamaño de la muestra *n* por la probabilidad teórica correspondiente (π_i), $e_i=n\pi_i$ de tal manera que $\sum_{i=1}^{k} e_i = n$
- Aplicar la prueba estadística

$$\sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(1-\alpha, k-m-1)}$$

Donde:
m: Es el número de parámetros a estimar de la distribución F(X)

Si se tienen valores esperados menores a 5, se pueden agrupar con el valor de la variable (o intervalo de clase) más cercano (superior o inferior) y de menor valor esperado.

- **Aplicación**

Se cree que el número de accidentes automovilísticos diarios en un cruce de dos avenidas de determinada ciudad tiene una distribución de Poisson. En una muestra aleatoria de 80 días del año pasado se obtuvieron los datos de la tabla adjunta. ¿Apoyan estos datos la hipótesis de que el número diario de accidentes tiene una distribución de Poisson? Use nivel de significación 0.05.

N° accidentes	O_i
0	34
1	25
2	11
3	7
4	3

Solución:
Procedimiento:
Planteamiento de la hipótesis.
H_0: Los datos provenientes del número de accidentes automovilísticos en el cruce de las avenidas de interés sigue una distribución de Poisson.
H_1: Los datos provenientes del número de accidentes automovilísticos en el cruce de las avenidas de interés no sigue una distribución de Poisson.
α=0.05
Prueba Estadística.

$$\sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} \sim \chi^2_{(1-\alpha, k-m-1)}$$

Desarrollo de la Prueba
Calculando la media (un parámetro a estimar)

N° accidentes (x_i)	O_i	x_iO_i
0	34	0
1	25	25
2	11	22
3	7	21
4	3	12
	80	80

$$\lambda = \frac{\sum_{i=1}^{k} x_i o_i}{n} = \frac{80}{80} = 1$$

A continuación, tenemos otros cálculos que nos permiten realizar la prueba y obtener los grados de libertad de la estadística de prueba.

$$\pi_1 = P(X = 0) = \frac{e^{-1}(1)^0}{0!} = 0.3679$$

:

π_5 = P(X≥4) = 1-P(X<4) =1-P(X≤3)=1-0.981=0.019

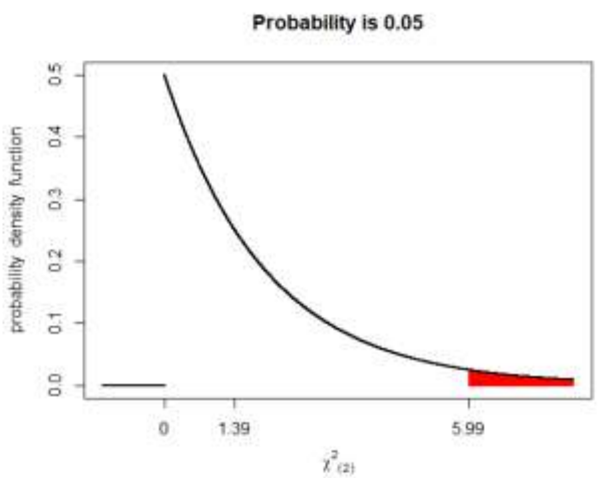
N° accidentes	π_i	$e_i=n\pi_i$
0	0.3679	29.43
1	0.3679	29.43
2	0.1839	14.72
3	0.0613	4.91
4 o más	0.0190	1.52
	1.0000	80.00

Observe que las dos últimas clases tienen frecuencias menores a cinco
Tenemos la siguiente tabla que resulta de unir las tres últimas clases.

N° accidentes (x)	O_i	$e_i=n\pi_i$	$(o_i-e_i)^2/e_i$
0	34	29.43	0.7096
1	25	29.43	0.6668
2	11	14.72	0.9401
3 o más	10	6.42	1.9963
	80	80	4.3129

$$\sum_{i=1}^{k} \frac{(O_i - e_i)^2}{e_i} = 4.3129$$

Criterios de decisión.
Los g.l. para la distribución Chi- cuadrado de la prueba son: *k* – *m* - 1 = 4 – 1 - 1 = 2 grados de libertad.



No se rechaza H_0 si: $\chi^2_c \leq 5.9915$
Se rechaza H_0 si: $\chi^2_c > 5.9915$

Conclusión.
A un nivel de significación del 5% no podemos afirmar que la variable número de accidentes automovilísticos en el cruce de las avenidas de interés no sigue una distribución Poisson.

Funciones en R

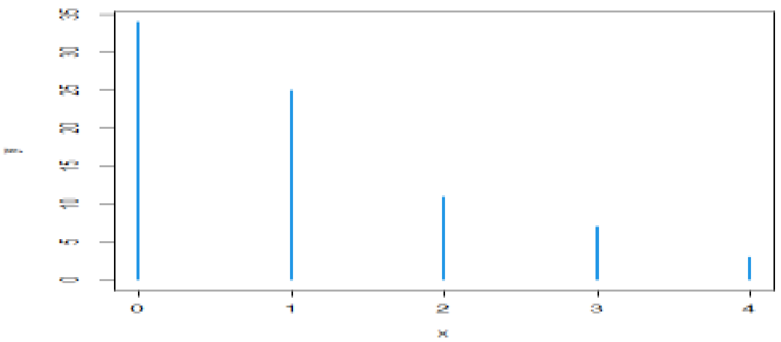
```
chisq.test(vector de datos, probabilidades teóricas)
library(vcd)
goodfit(tabla, distribución teórica,method="MinChisq")
```

Resultados en R

```
datos<-c(34,25,11,7,3)
prob<-c(dpois(0:3,1),1-ppois(3,1))
chisq.test(datos,p=prob)
```

```
        Chi-squared test for given probabilities
data:  datos
X-squared = 4.653, df = 4, p-value = 0.3248
Warning message:
In chisq.test(datos, p = prob) : Chi-squared approximation may be inco
rrect
```

```
x<-c(rep(0,34),rep(1,25),rep(2,11),rep(3,7),rep(4,3))
y<-table(x)
plot(y,type="h",col=4)
```



```
library(vcd)
res<-goodfit(y,type="poisson",method="MinChisq")
summary(res)
```

```
Goodness-of-fit test for poisson distribution
             X^2 df  P(> X^2)
Pearson 4.512272  3 0.2111983
Warning message:
In summary.goodfit(res) : Chi-squared approximation may be incorrect
```

Algunas consideraciones de R

- Las funciones chisq.test y goodfit (del paquete vcd) permiten solo trabajar con datos agrupados.
- Si se tienen datos no agrupados en una tabla de frecuencia primero se debe utilizar la función *table* para poder obtener la frecuencia de la categoría de interés.
- No considera el agrupamiento cuando hay valores esperados menores a 5.

C. Prueba de Cramer-von Mises

➢ Aspectos Generales

La~~s~~ prueba~~s~~ de Cramér-von Mises (CvM), al igual que la prueba de Anderson-Darling (AD) y Kolmogorov-Smirnov también se basa en una comparación de la función de distribución empírica, denotada por $F(x)$, y la función de distribución planteada en la hipótesis nula.
En estadística, el criterio de Cramer-von Mises es un criterio utilizado para juzgar la bondad de ajuste de una función de distribución acumulativa teórica en comparación con una función de distribución empírica dada $F(x)$ o también para comparar dos distribuciones empíricas.

Richard Edler von Mises
1883 - 1953

Harald Cramér
1893 - 1985

➢ Supuestos

- Los datos de la variable de interés deben ser de tipo cuantitativa continua.

➢ Inferencia estadística

- Para evaluar una sola muestra, el estadístico de prueba es:

$$CVM = W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - F_0(X_{(i)})\right)^2$$

- Criterio de rechazo de $H_0 : CVM > CVM_{crit}$
- Para decidir si el valor de W_n^2 es grande, y así rechazar H_0, se utilizan los valores críticos de la distribución de W_n^2 ,estadístico de prueba de Cramer-von Mises que se puede hallar con la qCvM del paquete goftest del programa R **,** bajo H_0.

➢ Aplicación

El tiempo que un trabajador de construcción civil utiliza durante su refrigerio para ir a la cafetería, almorzar y regresar a su puesto de trabajo está en el intervalo de 10 a 30 minutos.

El ingeniero civil a cargo de la construcción afirma que el tiempo que utiliza el trabajador durante su refrigerio no presenta un comportamiento uniforme para lo cual toma una muestra aleatoria de 30 trabajadores, a los cuales les mide el tiempo que utilizan para su refrigerio
A un nivel de significación de 0.03, la afirmación del ingeniero es correcta

H_0:El tiempo que se utiliza para refrigerio se ajusta a una distribución Uniforme
H_1:El tiempo que se utiliza para refrigerio no se ajusta a una distribución Uniforme

α = 0.05

$$CVM = W_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - F0(X(i))\right)^2$$

$$W_{30}^2 = \frac{1}{12(30)} + \left(\left(\frac{2(1)-1}{2(30)} - \frac{10-10}{(30-10)}\right)^2 + \left(\frac{2(2)-1}{2(30)} - \frac{10-10}{(30-10)}\right)^2\right.$$

$$\left. + \left(\frac{2(3)-1}{2(30)} - \frac{11-10}{(30-10)}\right)^2 + \ldots + \left(\frac{2(30)-1}{2(30)} - \frac{30-10}{(30-10)}\right)^2\right)$$

$$W_{30}^2 = 0.1208333$$

Hallando el estadístico Cramer von Mises crítico con n=30 y α = 0.03

$$W_{(30,0.97)}^2 = 0.5447941$$

Conclusión
A un nivel de significación de 0.03, dado que el Wcal es menor al Wcrit, existe suficiente evidencia estadística para no rechazar H_0.
Por lo tanto, podemos afirmar que el tiempo en minutos se ajusta a una distribución Uniforme.

- **Funciones en R**

```
Tiempo<-c(28, 10, 11, 17, 30, 17, 20, 21, 20, 16, 13, 15,
10, 22, 30, 14, 15, 16, 25, 11, 21, 18, 12, 18, 22, 16, 29,
29, 24, 27)
goftest::cvm.test(Tiempo,"punif",min(Tiempo),max(Tiempo),es
timated=FALSE)
qCvM(0.03,30,lower.tail = F)
```

- **Resultados con R**

```
Cramer-von Mises test of goodness-of-fit
     Null hypothesis: uniform distribution
     Parameters assumed to be fixed

data:  Tiempo
omega2 = 0.12083, p-value = 0.4945
```

D. Prueba de Anderson-Darling

- **Aspectos Generales**

Es una prueba que utiliza en su estadístico de prueba el logaritmo de la distribución acumulada teórica.
La prueba de Anderson-Darling otorga una mayor relevancia a los datos existentes en las colas de la distribución y se supone una modificación del test de Cramer von Mises, la cual se basa en la diferencia de cuadrados entre las distribuciones (Farrel & Rogers-Stewart, 2006).
Es una prueba de Bondad de Ajuste dado que no solo permite verificar la normalidad

Theodore Wilbur Anderson
(1918 – 2016)

Donald Allan Darling
(1915 – 2014)

- **Supuestos**

- La muestra ha sido elegida aleatoriamente.
- Los datos de la variable de interés deben ser de tipo cuantitativa continua.

- **Inferencia estadística**

- Utilizar el estadístico de prueba que está dado por $A^2 = -n - S$

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

Donde n es el número de observaciones, F(Y) es la distribución de probabilidad acumulada con parámetros especificados a partir de la muestra y Y_i son los datos obtenidos en la muestra, ordenados en forma ascendente.

- Definir la Región critica:

H_0 se rechaza con un nivel de significación α si A^2 es mayor que el valor crítico A^2_{crit}. A continuación, se presenta una tabla para algunos valores de α

α	0.1	0.05	0.025	0.01
A^2_{crit}	0.631	0.752	0.873	1.035

- **Aplicación**

Se desea analizar el tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción para el examen de admisión a una universidad. Los tiempos correspondientes a una muestra aleatoria de 40 postulantes se presentan a continuación:

Nº	Tiempo	Nº	Tiempo	Nº	Tiempo	Nº	Tiempo
1	10.9	11	9.1	21	12.1	31	10.1
2	12.0	12	14.2	22	11.2	32	9.2
3	11.7	13	8.9	23	9.4	33	11.6
4	13.7	14	11.9	24	10.3	34	12.6
5	14.1	15	9.7	25	15.7	35	14.0
6	10.0	16	11.5	26	10.6	36	12.3
7	13.0	17	13.3	27	14.2	37	10.4
8	12.7	18	11.1	28	12.0	38	10.6
9	13.7	19	11.3	29	10.6	39	13.2
10	12.9	20	13.0	30	12.4	40	12.5

Mediante la prueba de Anderson-Darling, pruebe si el conjunto de datos se distribuye normalmente. Use α=0.05

a) Hipótesis

H_0: El tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción para el examen de admisión se ajusta a una distribución normal

H_1: El tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción para el examen de admisión no se ajusta a una distribución normal

b) α=0.05.

c) Estadístico de contraste

El estadístico de prueba está dado por A^2

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

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

d) Desarrollo de la prueba

$\bar{x} = 11.843 \qquad s = 1.636$

$$F(x_i = 8.9) = P(X < 8.9) = P\left(\frac{X - \overline{X}}{s} < \frac{8.9 - 11.843}{1.636}\right) = 0.0360$$

Nº	$(2i-1)$	Y_i	$F(Y_i)$	$F(Y_{n+1-i})$	$\ln F(Y_i)$	$\ln\left(1-F(Y_{n+1-i})\right)$
1	1	8.9	0.0360	0.9908	-3.324	-4.689
2	3	9.1	0.0468	0.9252	-3.062	-2.593
3	5	9.2	0.0531	0.9252	-2.936	-2.593
4	7	9.4	0.0677	0.9063	-2.693	-2.368
5	9	9.7	0.0951	0.9161	-2.353	-2.478

6	11	10	0.1300	0.8719	-2.040	-2.055
7	13	10.1	0.1434	0.8719	-1.942	-2.055
8	15	10.3	0.1728	0.8135	-1.756	-1.679
9	17	10.4	0.1889	0.7966	-1.667	-1.593
10	19	10.6	0.2237	0.7603	-1.497	-1.428
11	21	10.6	0.2237	0.7603	-1.497	-1.428
12	23	10.6	0.2237	0.7409	-1.497	-1.351
13	25	10.9	0.2822	0.6999	-1.265	-1.204
14	27	11.1	0.3249	0.6783	-1.124	-1.134
15	29	11.2	0.3472	0.6561	-1.058	-1.067
16	31	11.3	0.3700	0.6333	-0.994	-1.003
17	33	11.5	0.4170	0.6101	-0.875	-0.942
18	35	11.6	0.4410	0.5625	-0.819	-0.827
19	37	11.7	0.4652	0.5383	-0.765	-0.773
20	39	11.9	0.5140	0.5383	-0.666	-0.773
21	41	12.0	0.5383	0.5140	-0.619	-0.722
22	43	12.0	0.5383	0.4652	-0.619	-0.626
23	45	12.1	0.5625	0.4410	-0.575	-0.582
24	47	12.3	0.6101	0.4170	-0.494	-0.540
25	49	12.4	0.6333	0.3700	-0.457	-0.462
26	51	12.5	0.6561	0.3472	-0.421	-0.426
27	53	12.6	0.6783	0.3249	-0.388	-0.393
28	55	12.7	0.6999	0.2822	-0.357	-0.332
29	57	12.9	0.7409	0.2237	-0.300	-0.253
30	59	13.0	0.7603	0.2237	-0.274	-0.253
31	61	13.0	0.7603	0.2237	-0.274	-0.253
32	63	13.2	0.7966	0.1889	-0.227	-0.209
33	65	13.3	0.8135	0.1728	-0.206	-0.190
34	67	13.7	0.8719	0.1434	-0.137	-0.155
35	69	13.7	0.8719	0.1300	-0.137	-0.139
36	71	14.0	0.9063	0.0951	-0.098	-0.100
37	73	14.1	0.9161	0.0677	-0.088	-0.070
38	75	14.2	0.9252	0.0531	-0.078	-0.055
39	77	14.2	0.9252	0.0468	-0.078	-0.048
40	79	15.7	0.9908	0.0360	-0.009	-0.037

$$A^2 = -40 + 40.16905 = 0.16905$$

e) Criterio de Decisión

$A^2 = 0.16905 < A^2_{crit} = 0.752$ no rechazo H_0

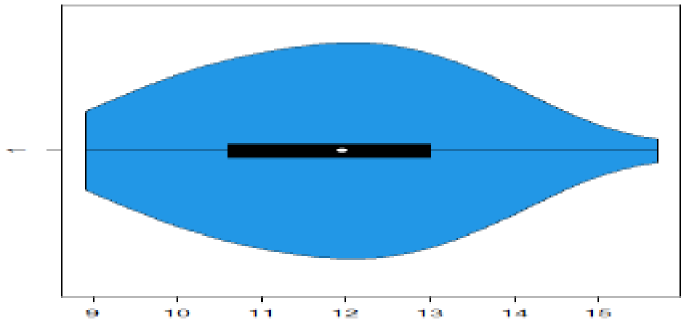
f) Conclusiones

No existe suficiente evidencia estadística para rechazar H_0 a un nivel de significación de 0.05.

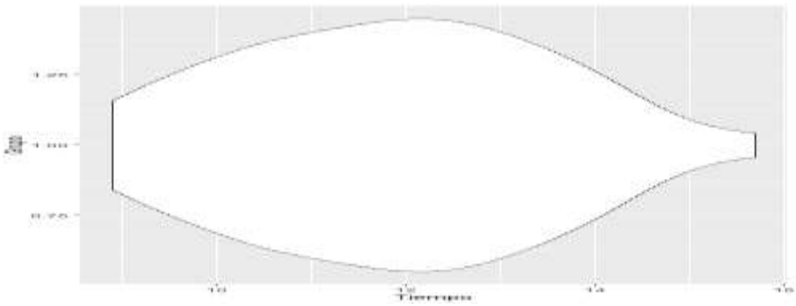
Por lo tanto, no se puede afirmar que el tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción para el examen de admisión no sigue una distribución normal.

- **Funciones en R**

```
Tiempo<-c(10.9, 12.0, 11.7, 13.7, 14.1, 10.0, 13.0, 12.7,
13.7, 12.9,  9.1, 14.2,  8.9, 11.9,  9.7, 11.5, 13.3, 11.1,
11.3, 13.0, 12.1, 11.2,  9.4, 10.3, 15.7, 10.6, 14.2, 12.0,
10.6, 12.4, 10.1,  9.2, 11.6, 12.6, 14.0, 12.3, 10.4, 10.6,
13.2,12.5)
library(vioplot)
vioplot(Tiempo,col=4,horizontal=T)
```



```
library(ggplot2)
Grupo<-rep(1,length(Tiempo))
datos2<-data.frame(Grupo,Tiempo)
g3<-ggplot(data=datos2,aes(x=Grupo,y=Tiempo))+
geom_violin()+coord_flip()
g3
```



```
library(goftest)
goftest::ad.test(Tiempo,"pnorm",mean(Tiempo),sd(Tiempo),
        estimated=F)
nortest::ad.test(Tiempo)
```

- **Resultados con R**

```
      Anderson-Darling test of goodness-of-fit
      Null hypothesis: Normal distribution
      Parameters assumed to be fixed

data:  as.matrix(tiempo)
An = 0.16905, p-value = 0.9967
```

```
      Anderson-Darling normality test
data:  Tiempo
A = 0.16905, p-value = 0.9292
```

2.2 Pruebas de Normalidad

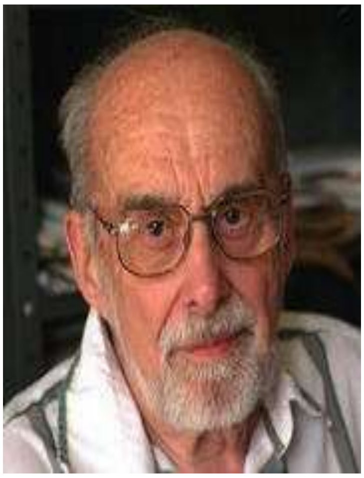
Muchos métodos paramétricos requieren el cumplimiento de la normalidad de sus datos o errores. Esto se puede verificar mediante diferentes pruebas que serán discutidas en esta sección.

Hipótesis
H_0: Los datos provienen de una distribución normal
H_1: Los datos no provienen de una distribución normal

A. Prueba de Shapiro Wilk

- **Aspectos Generales**

Utiliza en la construcción de su estadístico la dispersión de las estadísticas de orden.
Esta pruebase emplea para contrastar normalidad cuando el tamaño de la muestra es menor a 50 observaciones y en muestras grandes es equivalente al test de Kolmogórov-Smirnov. Sin embargo, en R se puede utilizar un vector de longitud de 3 a 5000 datos.
Esta prueba es considerada en muchas situaciones como una de las más poderosa cuando se compara con otras pruebas de normalidad.



Samuel Sanford Shapiro
(1930 – Actualidad)

Martín Bradbury Wilk
(1922 – 2013)

- **Supuestos**

- Los datos de la variable de interés deben estar medidos en al menos una escala intervalo.

- **Inferencia estadística**

Procedimiento para el desarrollo de la Prueba de Hipótesis

- Calcular el denominador D

$$D=\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$$

- Ordenar la muestra de menor a mayor:

$$X_{(1)}\leq X_{(2)}\leq\ldots\leq X_{(n)}$$

Donde $X_{(i)}$ el i-ésimo estadístico de orden.

- Con la ayuda de la tabla de coeficientes de la tabla de Shapiro Wilk, para el tamaño de muestra observado n, se obtienen los coeficientes $a_1,a_2,\ldots,a_k$, donde k es aproximadamente $n/2$.
- Aplicar el estadístico de prueba, dado por:

$$W=\frac{1}{D}\left[\sum_{i=1}^{k}a_i\left(X_{(n-i+1)}-X_{(i)}\right)\right]^2$$

- **Aplicación**

Se seleccionaron al azar a 50 empleados con altos cargos gerenciales de una empresa transnacional. Los sueldos mensuales (en cientos de dólares) se presentan a continuación:

23	23	24	27	29	31	32	33	33	35
36	37	40	42	43	43	44	45	48	48
54	54	56	57	57	58	58	58	58	59
61	61	62	63	64	65	66	68	68	70
73	73	74	75	77	81	87	89	93	97

Pruebe si los datos provenientes de los sueldos mensuales (en cientos de dólares) no se distribuyen normalmente a un α=0.05.

a) Hipótesis
H_0: Los sueldos mensuales se ajustan a una distribución normal
H_1: Los sueldos mensuales no se ajustan a una distribución normal

b) Nivel de significación α
c) Estadístico de contraste

$$W=\frac{1}{D}\left[\sum_{i=1}^{k}a_i\left(X_{(n-i+1)}-X_{(i)}\right)\right]^2$$

d) Desarrollo de la prueba
Los coeficientes obtenidos de la tabla de Shapiro Wilk y los estadísticos de orden $\left(X_{(n-i+1)}-X_{(i)}\right)$ son dados en la siguiente tabla:

i	a_i	$\left(X_{(n-i+1)}-X_{(i)}\right)$
1	0.3751	97-23
2	0.2574	93-23
3	0.2260	89-24

i	a_i	$\left(X_{(n-i+1)}-X_{(i)}\right)$
14	0.0846	66-42
15	0.0764	65-43
16	0.0685	64-43

4	0.2032	87-27	17	0.0608	63-44
5	0.1847	81-29	18	0.0532	62-45
6	0.1691	77-31	19	0.0459	61-48
7	0.1554	75-32	20	0.0386	61-48
8	0.1430	74-33	21	0.0314	59-54
9	0.1317	73-33	22	0.0244	58-54
10	0.1212	73-33	23	0.0174	58-56
11	0.1113	70-36	24	0.0104	58-57
12	0.1020	68-37	25	0.0035	58-57
13	0.0932	68-40			

El numerador del estadístico de prueba sería:

$$\left[\sum_{i=1}^{k} a_i\left(X_{(n-i+1)} - X_{(i)}\right)\right]^2 = \left[(0.3751)(97-23)+\ldots+(0.0035)(58-57)\right]^2 = [130.63]^2 = 17064$$

El denominador es dado por:

$$D = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = 17698$$

Por lo tanto, el estadístico de prueba es:

$$W = \frac{17064}{17698} = 0.9642$$

e) Criterio de Decisión
El cual se encuentra entre el percentil 0.10 y 0.50 de la distribución. Interpolando linealmente en la tabla de Shapiro-Wilk da $pvalor = \hat{\alpha} = 0.294$

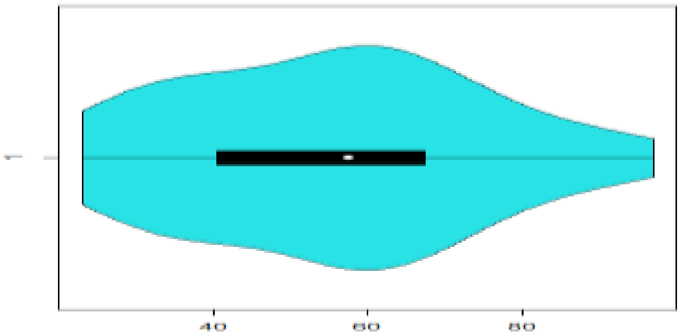
0.9550 ------ 0.10
0.9642 ------- x
0.9740 -------0.50

$$\frac{x-0.10}{0.9642-0.955} = \frac{0.5-0.1}{0.974-0.955} \Rightarrow x = 0.294$$

f) Conclusiones
No existe suficiente evidencia estadística a un nivel de significación de 0.05 para rechazar H_0.
Por lo tanto, no se puede afirmar que los sueldos mensuales de empleados con altos cargos gerenciales no sigan una distribución normal.

- **Funciones en R**

```
sueldos<-c(23, 23,  24,  27,  29,  31,  32,  33,  33,  35,
36,  37,  40,  42,  43,  43,  44,  45,  48,  48,  54,  54,
56,  57,  57,  58,  58,  58,  58,  59,  61,  61,  62,  63,
64,  65,  66,  68,  68,  70,  73,  73,  74,  75,  77,  81,
87,  89,  93,  97)
library(vioplot)
vioplot(sueldos,col=5,horizontal=T)
```




```
shapiro.test(vector de datos)
```

- **Resultados en R**

```
shapiro.test(sueldos)
```

```
      Shapiro-Wilk normality test

data:  as.matrix(sueldo)
W = 0.9744, p-value = 0.3473
```

La función shapiro_test del paquete rstatix también realiza la prueba de Shapiro Wilk

```
library(rstatix)
shapiro_test(sueldos)
```

```
# A tibble: 1 x 3
  variable statistic p.value
  <chr>        <dbl>   <dbl>
1 sueldos      0.974   0.347
```

La función shapiroTest del paquete fBasics también realiza la prueba de Shapiro Wilk

```
library(fBasics)
shapiroTest(sueldos)
```

```
Title:
 Shapiro - Wilk Normality Test

Test Results:
  STATISTIC:
    W: 0.9744
  P VALUE:
    0.3473
```

- **Algunas consideraciones de R**

- En el paquete como el nortest se ofrece la función sf.test de la prueba de Shapiro-Francia que se aproxima a la de Shapiro- Wilk cuando se analizan muestras grandes (n>50).

B. Prueba de D´Agostino

- **Aspectos Generales**

Es una variante de la prueba de Shapiro-Wilk, y es preferentemente utilizada cuando se tienen una muestra grande (n>50)

Ralph Benedict D' Agostino
(1940 – Actualidad)

- **Supuestos**

- Los datos de la variable de estudio deben estar medidos en al menos una escala intervalo.

- **Inferencia estadística**

- Utilizar el estadístico de D'Agostino es:

$$D = \sum_{i=1}^{n} \frac{\left(i - \frac{n+1}{2}\right) x_{(i)}}{n^2 s_n}$$

donde

Donde $s_n^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$ y $X_{(i)}$ son los valores ordenados de manera ascendente.

El valor esperado de este estadístico es aproximadamente $\frac{1}{2\sqrt{\pi}}$.

Para muestras pequeñas se dispone de un criterio de decisión que se puede obtener en la tabla de D'Agostino, para muestras grandes, la variable:

$$\sqrt{n}\,\frac{D - \frac{1}{2\sqrt{\pi}}}{\left(\frac{12\sqrt{3} - 27 + 2\pi}{24n\pi}\right)^{\frac{1}{2}}}$$

se puede aproximar por una variable normal estándar.

- **Aplicación**
  Utilice los datos de los tiempos que demoran los postulantes en inscribirse y verifique mediante la prueba de D´Agostino si los tiempos no se ajustan a una distribución Normal. Use α=0.05

a) H_0: El tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción tienen distribución normal.
   H_1: El tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción no tienen distribución normal.
b) $\alpha = 0.05$
c) Estadístico de prueba:

$$D = \sum_{i=1}^{n} \frac{(i - \frac{n+1}{2})x_{(i)}}{n^2 s_n} \qquad s_n^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

Para el cálculo del estadístico realizamos la siguiente tabla, la cual nos facilita los cálculos.

i	x_i	$x_{(i)}$	$(i - \frac{n+1}{2})$	$(i - \frac{n+1}{2})x_{(i)}$
1	10.90	8.90	-19.5	-173.55
2	12.00	9.10	-18.5	-168.35
3	11.70	9.20	-17.5	-161.00
4	13.70	9.40	-16.5	-155.10
5	14.10	9.70	-15.5	-150.35
6	10.00	10.00	-14.5	-145.00
7	13.00	10.10	-13.5	-136.35
8	12.70	10.30	-12.5	-128.75
9	13.70	10.40	-11.5	-119.60
10	12.90	10.60	-10.5	-111.30
11	9.10	10.60	-9.5	-100.70
12	14.2	10.60	-8.5	-90.10
13	8.9	10.90	-7.5	-81.75
14	11.9	11.10	-6.5	-72.15
15	9.7	11.20	-5.5	-61.6
16	11.5	11.30	-4.5	-50.85
17	13.3	11.50	-3.5	-40.25
18	11.1	11.60	-2.5	-29.00
19	11.3	11.70	-1.5	-17.55
20	13.0	11.90	-0.5	-5.95
21	12.1	12.00	0.5	6.00
22	11.2	12.00	1.5	18.00
23	9.4	12.10	2.5	30.25

24	10.3	12.30	3.5	43.05
25	15.7	12.40	4.5	55.80
26	10.6	12.50	5.5	68.75
27	14.2	12.60	6.5	81.90
28	12.0	12.70	7.5	95.25
29	10.6	12.90	8.5	109.65
30	12.4	13.00	9.5	123.50
31	10.1	13.00	10.5	136.50
32	9.2	13.20	11.5	151.80
33	11.6	13.30	12.5	166.25
34	12.6	13.70	13.5	184.95
35	14.0	13.70	14.5	198.65
36	12.3	14.00	15.5	217.00
37	10.4	14.10	16.5	232.65
38	10.6	14.20	17.5	248.5
39	13.2	14.20	18.5	262.7
40	12.5	15.70	19.5	306.15

d) Desarrollo de la prueba

$$D = \sum_{i=1}^{n} \frac{(i - \frac{n+1}{2})x_{(i)}}{n^2 s_n}$$

Para eso, primero hallamos el numerador sumando todos los elementos de la columna $(i - \frac{n+1}{2})x_i^*$ de la tabla:

$$\sum_{i=1}^{n}(i - \frac{n+1}{2})x_{(i)} = 738.05$$

$$s_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = 1.6153773$$

$$n^2 = 40^2 = 1600$$

Entonces:

$$n^2 s_n = 1600(1.6153773) = 2584.60364\ 5$$

Luego el Estadístico D'Agostino es:

$$D = \frac{738.05}{1600(1.6153773)} = \frac{738.05}{2584.603645} = 0.28555635$$

e) Decisión de la prueba:
Para decidir entre H_0 y H_1 nos fijamos en la tabla de D'Agostino para n =40. Como el estadístico de $D \in [0.2688; 0.2867]$, podemos concluir que con un α=0.05, no se rechaza la hipótesis nula, es decir, no se puede afirmar que los

datos del tiempo (en minutos) que les toma a los postulantes para completar el trámite de inscripción no tienen distribución normal.

- **Funciones con R**

En R existe la función statcompute del paquete PoweR que permite hacer muchas pruebas estadísticas. La número 24 realiza la prueba de D´Agostino, pero la última versión solo brinda el estadístico.

```
library(PoweR)
statcompute(24, data)
```

También existe la función dagoTest del paquete fBasics pero no brinda un único pvalor.

- **Resultados con R**

```
$statistic
[1] 0.5633672

library(fBasics)
dagoTest(data)
```

```
Title:
 D'Agostino Normality Test

Test Results:
  STATISTIC:
    Chi2 | Omnibus: 0.6755
    Z3  | Skewness: 0.258
    Z4  | Kurtosis: -0.7803
  P VALUE:
    Omnibus  Test: 0.7134
    Skewness Test: 0.7964
    Kurtosis Test: 0.4352
```

C. Prueba de Jarque-Bera

➢ Aspectos Generales

Fue propuesta por Carlos Jarque y Anil K. Bera.

Es una prueba de bondad de ajuste acerca de si una muestra de datos tiene la asimetría y la curtosis de una distribución Normal. Una distribución normal tiene asimetría cero (es decir, es perfectamente simétrica alrededor de la media) y una curtosis de 3. La curtosis mide el grado de concentración que presentan los datos en la región central de la distribución.

Esta prueba es usualmente utilizada para grandes conjuntos de datos, debido a que otras pruebas de normalidad no son confiables cuando n es grande (por ejemplo, Shapiro-Wilk no es confiable con un n mayor de 2000).

Carlos Manuel Jarque Uribe
(1954 – Actualidad)

Anil K. Bera
(1955 – Actualidad)

➢ Supuestos

- Los datos de la variable de estudio deben estar medidos en al menos una escala intervalo o debe ser cuantitativa continua.

➢ Inferencia estadística

El test estadístico JB se define como:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(C-3)^2\right)$$

Donde n es el tamaño de la muestra; S es la asimetría de la muestra, C la curtosis de la muestra.

La asimetría: Esta medida nos permite identificar si los datos se distribuyen uniformemente alrededor del punto central (media). Se calcula mediante la siguiente expresión:

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{3/2}}$$

Si el coeficiente es mayor a cero, la distribución es sesgada a la derecha, y en consecuencia presenta mayor número de observaciones a la izquierda.
La curtosis: Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución, cuando el coeficiente de curtosis es diferente de 3 tenemos los siguientes casos:
Platicúrtica si C < 3
Mesocúrtica si C = 3
Leptocúrtica si C > 3.

Y su expresión es:

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{2}}$$

Donde:

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \text{ y } \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$$

son las estimaciones de los momentos centrales tercer y cuarto, respectivamente, $\bar{x}$ es la media de la muestra y $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ es la estimación del segundo momento central, es decir la varianza.

El estadístico de Jarque-Bera con un n demasiado grande se distribuye asintóticamente como una distribución chi cuadrado con dos grados de libertad, el estadístico muestral de asimetría es:

$$\alpha_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{ns^3},$$

y si la hipótesis de normalidad es cierta, se sabe qué $E(\alpha_1) = 0$.

Para muestras grandes $(n \geq 50)$, se demuestra que $Var(\alpha_1) \cong 6/n$, lo que permite verificar la consistencia de la hipótesis de simetría con la evidencia que arrojan los datos obtenidos en la muestra. Por otro lado, el apuntamiento o curtosis en muestras de tamaño n, vendrá dado por la expresión $\alpha_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{ns^4}$. Se puede comprobar, asimismo, que este estadístico, para $n \geq 200$, sigue una distribución asintóticamente normal, con $E(\alpha_2) = 3$, el apuntamiento de una curva Normal, y $Var(\alpha_2) \cong 24/n$. Es posible, por tanto, verificar hipótesis acerca de si el apuntamiento en la población es el de una curva Normal.

Finalmente, si se combinan ambos estadísticos, puesto que se trata de la suma de los cuadrados de dos variables normales tipificadas, se tiene, como ya se sabe, que
$\left(\frac{\alpha_1-0}{\sqrt{6/n}}\right)^2+\left(\frac{\alpha_2-3}{\sqrt{24/n}}\right)^2 \sim\chi_2^2$, o lo que es igual $\frac{n}{6}\left(\alpha_1^2+\frac{(\alpha_2-3)^2}{4}\right)\sim\chi_2^2$ y puede usarse para probar la hipótesis nula de que los datos pertenecen a una distribución Normal. La hipótesis nula es una hipótesis conjunta de que la asimetría y el exceso de curtosis son nulos (asimetría =0 y curtosis =3).

- **Aplicación**

Considerando la información sobre ventas y gasto en publicidad de doce empresas elegidas al azar, verifique si los residuales resultantes del modelo siguen aproximadamente una distribución normal.

id	Y	X
1	69	9
2	76	12
3	52	6
4	56	10
5	57	9
6	77	10
7	58	7
8	55	8
9	67	12
10	53	6
11	72	11
12	64	8

H_0: $Los\ errores\ se\ distribuyen\ normalmente$
H_1: $Los\ errores\ \ no\ se\ distribuyen\ normalmente$

Modelo de regresión Lineal
Coeficiente

$$b_1=\frac{\sum X_iY_i-n\overline{X}\,\overline{Y}}{\sum X_i^2-n\overline{X}^2}=\frac{6960-12(9)(63)}{1020-12(9)^2}=3.25$$

Intercepto

$$b_0=\overline{Y}-b_1\overline{X}=63-(3.25)(9)=33.75$$

Estimando los residuales

$$res=Y-\left(X\times b_1+b_0\right)$$

id	Y	X	Residual
1	69	9	6.00
2	76	12	3.25
3	52	6	-1.25
4	56	10	-10.25
5	57	9	-6.00
6	77	10	10.75
7	58	7	1.50
8	55	8	-4.75
9	67	12	-5.75
10	53	6	-0.25
11	72	11	2.50
12	64	8	4.25

Estimando los momentos

id	Y	X	Residual	$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2$	$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^3$	$\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^4$
1	69	9	6.00	36.00	216.00	1296.00
2	76	12	3.25	10.56	34.33	111.57
3	52	6	-1.25	1.56	-1.95	2.44
4	56	10	-10.25	105.06	-1076.89	11038.13
5	57	9	-6.00	36.00	-216	1296
6	77	10	10.75	115.56	1242.30	13354.69
7	58	7	1.50	2.25	3.38	5.06
8	55	8	-4.75	22.56	-107.17	509.07
9	67	12	-5.75	33.06	-190.11	1093.13
10	53	6	-0.25	0.06	-0.02	0.00
11	72	11	2.50	6.25	15.63	39.06
12	64	8	4.25	18.06	76.77	326.25
Total			0	387.00	-3.75	29071.41

El segundo momento central es $\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\frac{387}{12}=32.25$

El tercer momento central es $\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^3=\frac{-3.75}{12}=-0.3125$

El cuarto momento central es $\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^4=\frac{29071.41}{12}=2422.617$

Hallando Asimetría

$$S = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right)^{3/2}} \qquad = \frac{-0.3125}{\left(32.25\right)^{3/2}} \qquad = -0.0017063$$

Calculando la curtosis

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right)^{2}} = \frac{2422.617}{\left(32.25\right)^{2}} \qquad = 2.329299$$

Hallando el coeficiente Jarque – Bera (JB)

$$JB = \frac{n-k+1}{6}\left(S^2 + \frac{1}{4}\left(C-3\right)^2\right)$$

$$JB = \frac{12-1+1}{6}\left((-0.0017063)^2 + \frac{1}{4}\left(2.329299-3\right)^2\right)$$

$$JB = 0.2249253$$

Estadístico de prueba

$\chi^2_{(2,0.05)} = 5.99$

Entonces JB=0.2249253 < $\chi^2_{(2,0.05)} = 5.99$

Por lo tanto, no se Rechaza H_0

Conclusión: No se rechaza la hipótesis nula, no se puede afirmar que los errores de la regresión lineal simple no se distribuyen normalmente. Por lo tanto, los residuales resultantes del modelo siguen aproximadamente una distribución normal.

- **Funciones en R**

```
y<-c(69,76,52,56,57,77,58,55,67,53,72,64)
x<-c(9,12,6,10,9,10,7,8,12,6,11,8)
res<-lm(y~x)$res
library(moments)
jarque.test(res)

library(tseries)
jarque.bera.test

library(normtest)
jb.norm.test(res)

library(normtest)
ajb.norm.test(res)
```

```
library(DescTools)
JarqueBeraTest(res)
JarqueBeraTest(res,robust=F)

library(Jmisc)
JBTest(res)

library(fBasics)
jarqueberaTest(res)
```

- **Resultados con R**

```
Jarque-Bera Normality Test
data:  res
JB = 0.22493, p-value = 0.8936
alternative hypothesis: greater
```

```
     Jarque Bera Test
data:  res
X-squared = 0.22493, df = 2, p-value = 0.8936
```

```
     Jarque-Bera test for normality

data:  res
JB = 0.22493, p-value = 0.875
```

Utiliza la simulación de Monte Carlo

```
     Adjusted Jarque-Bera test for normality
data:  res
AJB = 0.072747, p-value = 0.971
```

Prueba Robusta de Jarque Bera: utiliza el método #chi-cuadrado

```
     Robust Jarque Bera Test
data:  res
X-squared = 0.18829, df = 2, p-value = 0.9102
```

```
Jarque Bera Test
data:  res
X-squared = 0.22493, df = 2, p-value = 0.8936
```

```
[1] 0.8936307
```

```
Title:
 Jarque - Bera Normalality Test
Test Results:
  STATISTIC:
    X-squared: 0.2249
  P VALUE:
    Asymptotic p Value: 0.8936
```

D. Prueba de Lilliefors

Esta prueba se presenta como una modificación de la prueba de Kolmogorov-Smirnov, para poder resolver el problema de los parámetros desconocidos en el caso normal y presenta mejores resultados que la prueba de Kolmogorov-Smirnov.

Para su empleo, primero se estima la media y la varianza de la población en función de los datos, luego se encuentra la discrepancia máxima entre la función de distribución empírica y la función de distribución acumulativa de la distribución normal con la media estimada y la varianza estimada. Finalmente, evalúa si la discrepancia máxima es lo suficientemente grande para ser significativa, para rechazar la Hipótesis Nula (H_0).

Tiene un valor estadístico que coincide con el estadístico de Kolmogorov-Smirnov, cuya diferencia radica en la probabilidad asociada, ya que el estadístico de Lilliefors tiene una distribución muestral propia.

Hubert Lilliefors Whitman (1928 - 23 de febrero de 2008, Bethesda, Maryland).

- **Supuestos:**
  - Los datos provienen de una muestra aleatoria.
  - Los datos de la variable de interés deben ser de tipo cuantitativa continua.

- **Inferencia estadística**

El desarrollo o procedimiento teórico es el siguiente:

Paso 1

Sea $y_1, y_2, \dots, y_n$ una muestra aleatoria de las cuales se calcula la media y la varianza muestral si es que no se conoce alguno de ellos o ambos.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}\right]$$

Así, para calcular el valor normalizado:

$$z_i = \frac{y_i - \bar{y}}{s}, \qquad i = 1,2,\dots,n$$

Del cálculo del valor de z_i, la prueba está estructurada de manera análoga a la prueba de Kolmogorov-Smirnov.
Después ordenar los datos forma creciente.

Paso 2
Se encuentra la máxima discrepancia entre la función de distribución empírica y la función de distribución acumulada de la distribución normal con la media 0 y varianza 1.

Siendo $S(X)$ la función de distribución empírica, la fracción de observaciones muestrales menor o igual al valor de X. $Y_1, Y_2, \dots, Y_n$ son las estadísticas de orden de la muestra aleatoria observada.

$$S(X) = \begin{cases} 0, & para\ \ X < Y_i \\ \frac{i}{n}, & para\ \ Y_i \le X < Y_{i+1}, \qquad i = 1,2,\dots,n-1 \\ 1, & para\ \ X \ge Y_n \end{cases}$$

En que $F(X) = \ \phi([z_{(i)}])$, donde ϕ es la función de distribución acumulativa de la distribución normal estándar.

$$F(X) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-1}{2} y^2\right\} dy$$

El estadístico de prueba es el máximo o supremo entre las diferencias absolutas de la función de distribución acumulativa empírica y la hipotética. Se puede calcular como:

$$D_0 = \max\{D^+, D^-\}$$

Con

$$D^+ = \max_{i=1,\dots,n} |\{S(X_i) - F(X_i)\}|, \quad D^- = \max_{i=1,\dots,n} |\{F(X_i) - S(X_{i-1})\}|$$

Paso 3
Ubicar en la tabla de Lilliefors, según el nivel de confianza y el número de observaciones

Table A22 Table of Critical Values for the Lilliefors Test for Normality

One-tailed Two-tailed	.20 .40	.15 .30	.10 .20	.05 .10	.01 .02
n = 4	.300	.319	.352	.381	.417
5	.285	.299	.315	.337	.405
6	.265	.277	.294	.319	.364
7	.247	.258	.276	.300	.348
8	.233	.244	.261	.285	.331
9	.223	.233	.249	.271	.311
10	.215	.224	.239	.258	.294
11	.206	.217	.230	.249	.284
12	.199	.212	.223	.242	.275
13	.190	.202	.214	.234	.268
14	.183	.194	.207	.227	.261
15	.177	.187	.201	.220	.257
16	.173	.182	.195	.213	.250
17	.169	.177	.189	.206	.245
18	.166	.173	.184	.200	.239
19	.163	.169	.179	.195	.235
20	.160	.166	.174	.190	.231
25	.142	.147	.158	.173	.200
30	.131	.136	.144	.161	.187
n > 30	$.736/\sqrt{n}$	$.768/\sqrt{n}$	$.805/\sqrt{n}$	$.886/\sqrt{n}$	$1.031/\sqrt{n}$

Paso 4
Comparar el Paso 4.2 (valor tabular) y Paso 4.3 (valor crítico) para poder tomar la decisión de aceptar o rechazar la hipótesis nula.

Otra forma de cálculo del pvalor
El p valor se calcula a partir de la fórmula de Dallal–Wilkinson (1986), que afirma que solo es confiable cuando el p valor es menor que 0.1.
La probabilidad de cola superior p de la distribución de D_{max} de tamaño de muestra n, entre 5 y 100 está dada aproximadamente por:

$$\mathrm{P}_{\text{valor}} = \exp\left(-7.01256D_{max}^{2}(n+2.78019) + 2.99587D_{max}(n+2.78019)^{\frac{1}{2}} - 0.122119 + \frac{0.974598}{n^{\frac{1}{2}}} + \frac{1.67997}{n}\right)$$

Para tamaños de muestra mayores que 100, se usa la misma expresión con D_{max} reemplazada por $D_{max}\ x\ (\frac{m}{100})^{0.49}$ y n reemplazada por 100.
Si el p valor de Dallal-Wilkinson (calculado en la fórmula anterior) resulta ser mayor a 0.10, entonces el p valor se calcula a partir del estadístico modificado $Z = D_{max} * (\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}})$. Simplemente, el estimador se obtiene mediante una simulación.
Para estadístico modificado sea menor e igual 0.5

$$P_{valor} = 2.76773 - 19.828315 * Z + 80.709644 * Z^2 - 138.55152 * Z^3 + 81.218052 * Z^4$$

Para estadístico modificado mayor a 0.5 y menor e igual a 0.9

$$P_{valor} = -4.901232 + 40.662806 * Z - 97.490286 * Z^2 + 94.029866 * Z^3 - 32.355711 * Z^4$$

Para estadístico modificado mayor a 0.9 y menor e igual a 1.31

$$P_{valor} = 6.198765 - 19.558097 * Z + 23.186922 * Z^2 - 12.234627 * Z^3 + 2.423045 * Z^4$$

- **Funciones en R**

En R existe la función lillie.test del paquete nortest y statcompute del paquete PoweR

```
library(nortest)
lillie.test(muestra_1$Ingreso)

library(PoweR)
statcompute(01,sueldos)
```

- **Resultados con R**

```
Lilliefors (Kolmogorov-Smirnov) normality test

data:  sueldos
D = 0.081071, p-value = 0.5662
```

```
$statistic
[1] 0.08107085

$pvalue
[1] 0.5661892
```

Otra pruebas de normalidad es:
Prueba de Shapiro-Francia que es una variante de la Prueba de Shapiro-Wilk.

Comparación de pruebas de normalidad mediante su potencia

A continuación, se presentan algunas líneas de comandos para comparar Pruebas de Normalidad

```
library(nortest)
set.seed(20)
r<-1000
sw<-c()
ad<-c()
da<-c()
jb<-c()
for (i in 1:r){
  datos<-rexp(50,1/10)
  sw[i]<-shapiro.test(datos)$p.value
  ad[i]<-ad.test(datos)$p.value
  da[i]<-agostino.test(datos)$p.value
  jb[i]<-jarque.test(datos)$p.value
}
psw<-mean(ifelse(sw<0.05,1,0))
psw
```

```
pad<-mean(ifelse(ad<0.05,1,0))
pad
pda<-mean(ifelse(da<0.05,1,0))
pda
pjb<-mean(ifelse(jb<0.05,1,0))
pjb
```

2.3 Prueba para evaluar aleatoriedad

A. Prueba de Aleatoriedad, Corridas o Rachas

- **Aspectos Generales**

La prueba de corridas se usa para probar la aleatoriedad de una serie de observaciones con respecto a un valor de referencia cuando cada observación puede ser asignada a una de dos categorías (menor igual a un valor de referencia o mayor a este valor).

Subhash Kak
(1947 – Actualidad)

Por rachas se entiende a una sucesión de símbolos idénticos que pueden estar separados o no por otro tipo de símbolos. Por ejemplo, sea una serie de mediciones de magnitudes dicotómicas identificadas con los símbolos de resultado positivo (+) o negativo (-) a juicio del investigador de acuerdo con cierto criterio profesional empleado.
Así se podría tener los resultados de 14 artículos producidos por una máquina en la cual clasificamos a los artículos no defectuosos con (+) y a los artículos defectuosos con (-):

Observaciones	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Resultado	+	+	-	-	-	+	-	-	-	-	+	+	-	+
Racha	1		2			3	4				5		6	7

La primera racha empieza con una serie de 2 símbolos positivos, la segunda racha con 3 negativos, la tercera con un positivo, y así sucesivamente hasta la séptima racha con un positivo. El número de rachas es r = 7.
El número total de rachas indica si una muestra es o no aleatoria. Si se da un número pequeño de rachas puede deberse a una falta de independencia o a una tendencia temporal. Mientras que, si por el contrario hay un número muy grande de rachas, las fluctuaciones cíclicas sistemáticas, en un período corto de tiempo, pueden causar influencia en los valores asignados por el investigador.
En otras palabras, mucho menos corridas o mucho más corridas que las que serían de esperar al azar, resultarían en el rechazo de la hipótesis nula (H_0) de que la secuencia de observaciones es una secuencia aleatoria.

Respecto a datos numéricos, un medio para obtener el esquema requerido de dos categorías es clasificar cada observación según si es superior o inferior a la media, mediana u otro valor de interés del grupo de observaciones.
Así, por ejemplo, si tenemos el número de artículos producidos en 10 días por una máquina, las rachas con respecto a la media son:

```
datos<-c(12,16,15,40,26,17,22,26,12,34)
mean(datos)
[1] 22
```

Día	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Producción	12	16	15	40	26	17	22	26	12	34
Comp. con Prom.	-	-	-	+	+	-	0	+	-	+
Rachas		1			2	3		4	5	6

- **Supuestos**
- Las observaciones deben ser susceptibles de transformación dicotómica.

- **Inferencia Estadística**

Prueba de Hipótesis

H_0: La secuencia de observaciones es aleatoria.
H_1: La secuencia de observaciones no es aleatoria.

Procedimiento para el Desarrollo de la Prueba de Hipótesis

- Hallar el número n_1 de elementos de una clase, identificadas por un símbolo (+) y n_2 la cantidad de elementos de la otra clase (-).
- Contabilizar el número *r* de rachas.

Para determinar el valor de la probabilidad o valor calculado se presentan dos casos según el tamaño de la muestra:
Muestras pequeñas: Si los tamaños muestrales de ambos resultados (n_1 o n_2) son menores o iguales que 20 se emplea una tabla especial para obtener valores de r_{min} y r_{max} que será comparado con el valor de r. Si el valor de r se encuentra fuera de los límites (r_{min} y r_{max}) se rechazará la H_o caso contrario no se podrá rechazar la H_o.

Muestras grandes: Si alguno de los tamaños muestrales (n_1 o n_2) es mayor a 20, se determina el valor del estadígrafo Z con una distribución aproximadamente normal:

$$Z_c = \frac{r - \mu_r}{\sigma_r} \sim \mathrm{N}(0,1)$$

donde: $$\mu_r = \frac{2n_1 n_2}{n_1 + n_2} + 1 \qquad \sigma_r = \sqrt{\frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}}$$

n_1: Número de signos positivos
n_2: Número de signos negativos

- **Aplicación**

Ejemplo 1

En un laboratorio de investigación se prueba un antinflamatorio nuevo. Los resultados son aceptables si al segundo día de aplicado al paciente se observa una reducción del 90% en la inflamación; se le asigna (+) a ese caso. Se quiere probar la hipótesis que la sucesión de signos positivos y negativos se produce al azar. La sucesión de los 24 casos analizados fue la siguiente:

Nº Paciente	% de Red. de Infl.	Signo	Racha	Nº Paciente	% de Red. de Infl.	Signo	Racha
1	92	+	1	13	62	-	6
2	82	-	2	14	77	-	
3	96	+	3	15	96	+	7
4	98	+		16	67	-	8
5	92	+		17	62	-	
6	96	+		18	93	+	9
7	85	-	4	19	94	+	
8	93	+	5	20	99	+	
9	94	+		21	76	-	10
10	95	+		22	74	-	
11	75	-	6	23	84	-	
12	72	-		24	82	-	

Pruebe si las observaciones siguen una secuencia aleatoria a un nivel de significación de 0.05.

Solución
H_0: La secuencia de observaciones es aleatoria.
H_1: La secuencia de observaciones no es aleatoria.
$\alpha = 0.05$.
Cálculos previos
El tamaño de cada muestra (n_1 y n_2) es 12 y $r = 10$.

Criterio de Decisión
Usando las se determina la zona de rechazo de la hipótesis nula (H_0);
Zona de no rechazo: $7 < r < 19$
Como $r = 10$ cae dentro de esta zona, no se rechaza (H_0).

Conclusión
A un nivel de significación de 0.05 existe suficiente evidencia estadística para no rechazar la H_0.
Se concluye que se deben suponer aleatorias a las muestras tomadas.

Ejemplo 2

En un hospital se forma todas las mañanas temprano una cola de pacientes esperando su turno para la extracción de sangre. La bioquímica a cargo decide verificar si la colocación de hombres (H) y mujeres (M) es al azar. Anota el sexo de cada uno de los primeros 50 pacientes que entraron al laboratorio. Los resultados fueron:

Nº de Paciente	Sexo	Racha
1	H	1
2	H	
3	M	2
4	H	3
5	M	4
6	H	5
7	H	
8	H	
9	M	6
10	M	
11	H	7
12	M	8
13	M	
14	H	9
15	M	10
16	H	11
17	H	
18	M	12
19	M	
20	M	
21	H	13
22	H	
23	M	14
24	M	
25	H	15

Nº de Paciente	Sexo	Racha
26	H	15
27	M	16
28	M	
29	H	17
30	M	18
31	H	19
32	M	20
33	H	21
34	M	22
35	M	
36	H	23
37	M	24
38	H	25
39	H	
40	M	26
41	H	27
42	H	
43	M	28
44	H	29
45	M	30
46	H	31
47	M	32
48	H	33
49	M	34
50	M	

Pruebe a un nivel de significación de 0.05 si la ubicación de hombres y mujeres es aleatoria.

Solución:

H_0: La ubicación de hombres y mujeres en la fila para esperar su turno es aleatoria

H_1: La ubicación de hombres y mujeres en la fila para esperar su turno no es aleatoria

$\alpha = 0.05$.

Prueba Estadística

$$Z_c = \frac{r-\mu_r}{\sigma_r} \sim \mathrm{N}(0,1)$$

Desarrollo de la Prueba

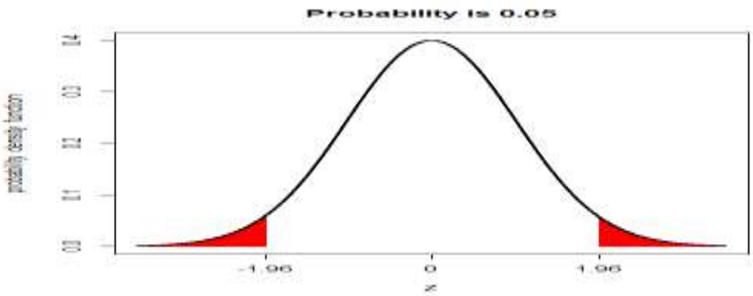
Como r = 34 y $n_1 = n_2 = 25$

$$\mu_r = \frac{2(25)(25)}{25+25} + 1 = 26 \quad \sigma_r = \sqrt{\frac{2(25)(25)[2(25)(25)-25-25]}{(25+25)^2(25+25-1)}} = 3.5$$

$$Z_c = \frac{34-26}{3.5} = 2.286$$

Criterio de Decisión

```
library(fastGraph)
shadeDist(qnorm(c(0.025,0.975),lower.tail = F),"dnorm")
```



Como $Z_c = 2.286$ cae fuera de la zona de no rechazo, se rechaza H_0.

Conclusión

A un nivel de significación de 0.05 existe suficiente evidencia estadística para rechazar la H_0.

Por lo tanto, se puede afirmar que los sexos no guardan un orden aleatorio al formar la fila.

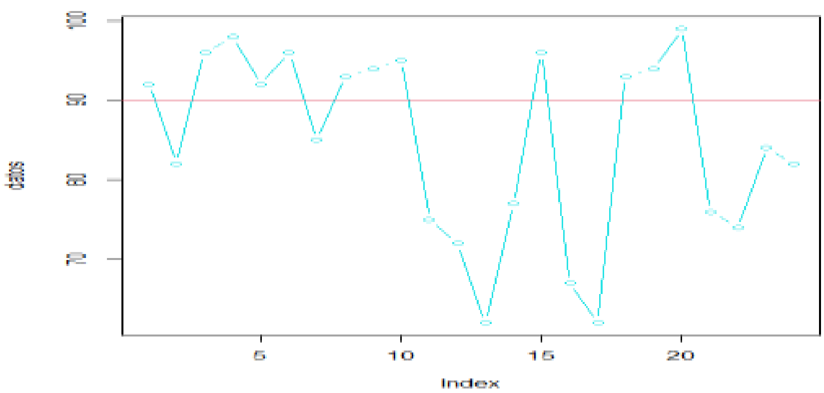
- **Funciones en R**

Se puede utilizar la función runs.test del paquete lawstat, runs.test del paquete randtests o la función del mismo nombre del paquete tseries pero la variable debe estar dicotomizada

```
runs.test(vector de datos)
```

- **Resultados con R**

```
datos<-
c(92,82,96,98,92,96,85,93,94,95,75,72,62,77,96,67,62,93,94,
99,76,74,84,82)
plot(datos,col=5,type="b")
abline(h=90,col=2)
```



```
library(lawstat)
runs.test(datos)
```

```
        Runs Test - Two sided
data:  datos
Standardized Runs Statistic = -1.2523, p-value = 0.2105
```

```
library(randtests)
runs.test(datos)
```

```
Runs Test

data:  datos
statistic = -1.2523, runs = 10, n1 = 12, n2 =
12, n = 24, p-value = 0.2105
alternative hypothesis: nonrandomness
```

```
library(tseries)
runs.test(as.factor(datos>90))
```

```
Runs Test
data:  as.factor(datos > 90)
Standard Normal = -1.2523, p-value = 0.2105
alternative hypothesis: two.sided
```

```
datos<-c(1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1,
0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0,
1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0)
runs.test(as.factor(datos>0.5))
```

```
Runs Test

data:  as.factor(datos > 0.5)
Standard Normal = 2.2862, p-value = 0.02224
alternative hypothesis: two.sided
```

- **Algunas consideraciones en R**
  - Se debe tener cuidado con la función runs.test del paquete lawstat o del paquete randtests porque no permite indicar un valor de referencia para definir las rachas.

2.4 Prueba para evaluar simetría

En algunas situaciones se puede estar interesado en la forma de la distribución de donde provienen los datos.
Tal es el caso de la prueba de Wilcoxon la cual tiene como supuesto que las observaciones provienen de una distribución simétrica.

Prueba de Hipótesis

Estas pruebas permiten determinar si las observaciones provienen de una distribución simétrica, asimétrica (positiva o negativa); de estos casos es quizás más importante probar si los datos provienen o no de una distribución simétrica
H_0: Las observaciones provienen de una distribución simétrica. As=0
H_1: Las observaciones no provienen de una distribución simétrica. As≠0

A. Prueba de Triadas

- **Aspectos Generales**

Esta prueba se basa en la comparación de la media de todas las posibles submuestras de tamaño 3 (C_3^n) con la mediana de estas 3 observaciones. Si la media supera a la mediana tiene un comportamiento asimétrico positiva, caso contrario tendrá un comportamiento asimétrico negativo. Si la media coincide con la media tendría un comportamiento simétrico.

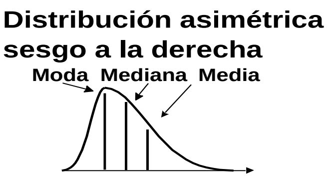
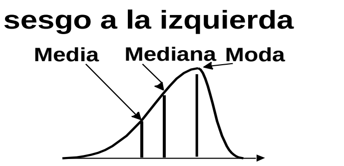
- **Supuestos**

- La muestra debe ser mayor a 20.
- La variable de interés debe estar medida en una escala al menos intervalo.

- **Inferencia Estadística**

Procedimiento para el Desarrollo de la Prueba de Hipótesis

- Se debe agrupar las observaciones en todas las posibles triadas que se puedan generar (grupo de 3 observaciones).
- Clasificar cada una de las n(n-1)(n-2)/6 posibles triadas de la siguiente manera:

Triada derecha  	$(X_i+X_j+X_k)/3 > med(X_i+X_j+X_k)$
Triada izquierda  	$(X_i+X_j+X_k)/3 < med(X_i+X_j+X_k)$
Ninguna	$(X_i+X_j+X_k)/3 = med(X_i+X_j+X_k)$

- Calcular el estadístico de interés:

T = # triadas derechas - # triadas izquierdas

Si se desea realizar la aproximación asintótica se debe adicionalmente calcular lo siguiente:
B_i = # triadas derechas que incluyen a X_i - # triadas izquierdas que incluyen a X_i
B_{jk} = # triadas derechas que incluyen tanto a X_i como X_k - # triadas izquierdas que incluyen a X_i como X_k.

Entonces la prueba estadística Z es: T/σ_T donde:

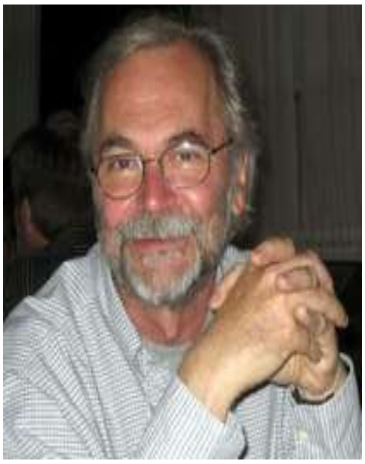
$$\sigma_T^2 = \frac{(n-3)(n-4)}{(n-1)(n-2)}\sum_{i=1}^{n} B_i^2 + \frac{(n-3)}{(n-4)}\sum_{i<j<k<n} B_{jk}^2 + \frac{n(n-1)(n-2)}{6} - \left[1 - \frac{(n-3)(n-4)(n-5)}{n(n-1)(n-2)}\right]T^2$$

Lamentablemente esta prueba no se encuentra implementada en R debido a su alto costo computacional para muestras grandes.

B. Prueba de Cabilio-Masaro

- **Aspectos Generales**

Esta prueba fue propuesta por Cabilio y Masaro en el año 1996.
Se utiliza para probar la simetría de una función de distribución sobre una mediana poblacional desconocida.



Paul Cabilio Joe Masaro

- **Supuestos**

- La muestra debe ser mayor a 20
- La variable de interés debe estar medida en una escala al menos intervalo.
- Los datos deben provenir de una distribución continua con mediana desconocida**.**

- **Inferencia Estadística**

La prueba de simetría Cabilio-Masaro considera el estadístico S_k , definido de la siguiente manera:

$$S_k = \frac{\sqrt{n}(\bar{X} - Me)}{s}$$

Donde:
$\bar{X}$: $Media\ Muestral$
Me: $Mediana\ Muestral$
s: $Desviación\ Estánda\ \ Muestral$
n: $Tamaño\ \ Muestral$

Se ha demostrado que S_k se distribuye de manera normal asintóticamente. Sin embargo, la varianza asintótica de S_k dependerá de la distribución subyacente de la población de donde se toma la muestra. Por lo tanto, la región de rechazo implícita del estadístico de prueba es:

$$|S_K| \geq P_{1-\frac{\alpha}{2}}$$

Donde $P_{1-\frac{\alpha}{2}}$ es el percentil $(1-\frac{\alpha}{2})$ de la muestra.
Por lo general la varianza asintótica en la distribución de la población es desconocida, frente a este problema Paul Cabilio y Joe Masaro sugieren para propósitos prácticos el uso de la varianza asintótica de 0.570796. Este valor se obtiene de la distribución empírica de S_k, simulando muestras de distinto tamaño provenientes de una distribución normal. Se halló que cuando el tamaño de la muestra tiende al infinito, $n \to \infty$ $\Phi^{-1}(\pi)\sqrt{0.570796}$ es el percentil 100π de la distribución asintótica de S_k cuando se toma de una normal.

La cual se deriva del supuesto de que la distribución poblacional tiene una distribución normal estándar donde el efecto de cualquier especificación errónea en la varianza asintótica es menor. Por lo tanto, usamos el estadístico *C*:

$$C = \frac{\bar{X} - Me}{s}$$

Entonces la región de rechazo queda definida de así:

$$|C| \geq \frac{\phi_{(\alpha/2)}\sqrt{0.570796}}{\sqrt{n}}$$

$$Rechazo\ H_1\ si\ \frac{|C|\sqrt{n}}{\sqrt{0.570796}} \geq \phi_{(\alpha/2)}$$

- **Aplicación**

Para evaluar el rendimiento de cierto medicamento contra la depresión y evitar una eventual dependencia a este, se analizan los niveles de dopamina producidos por un paciente medicado. Se desea probar si los datos de los niveles de dopamina medidos a una muestra de pacientes elegidos al azar no se distribuyen de manera simétrica.

Dopamina				
0.227	0.432	0.379	0.597	1.161
0.043	0.868	0.201	0.086	0.361
0.193	0.194	0.227	0.870	0.393
0.231	1.001	0.022	0.500	0.014
0.519	0.666	0.292	0.484	0.187
0.697	0.406	0.246	0.299	0.453
0.181	0.740	0.461	0.411	0.299
0.512	0.390	0.347	0.268	0.720

Solución:
$H_0: \quad A_s = 0$
$H_1: \ A_s \neq 0$

$\alpha = 0.05$

$\bar{X} = 0.41445$
$Me = P_{50} = 0.3845$

$s = 0.2651475$

$$C = \frac{\bar{X} - Me}{s} = \frac{0.41445 - 0.3845}{0.26514756} = 0.11295597$$

Entonces la región de rechazo queda definida de así:

$$Rechazo\ H_1\ si \qquad \frac{|0.11295597|\sqrt{40}}{\sqrt{0.570796}} \geq \phi_{(0.05/2)}$$

$$0.94558098 \geq 0.509972 \quad \Rightarrow Se\ rechaza\ la\ Hipotesis\ Alterna$$

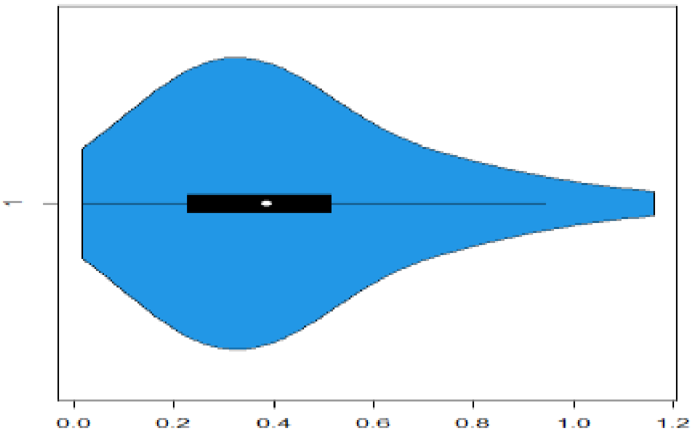
$$\mathrm{Z}_{0.94558098} = 0.1721812$$

$\text{Por ser valor simetrico} \quad 2 * \mathrm{Z}(0.94558098) = 0.344362 > 0.05 \quad No\ se\ rechaza\ H0$

Aun nivel de significación del 5%, se puede afirmar que los niveles de dopamina provienen de una distribución simétrica por lo tanto se concluye que el medicamento es efectivo.

- **Funciones en R**

```
dopamina<-c(0.227,  0.432,     0.379,     0.597,     1.161,
0.043,     0.868,     0.201,     0.086,     0.361,
0.193,     0.194,     0.227,     0.870,     0.393,
0.231,     1.001,     0.022,     0.500,     0.014,
0.519,     0.666,     0.292,     0.484,     0.187,
0.697,     0.406,     0.246,     0.299,     0.453,
0.181,     0.740,     0.461,     0.411,     0.299,
0.512,     0.390,     0.347,     0.268,     0.720)
vioplot(dopamina,col=4,horizontal=T)
```




```
library(lawstat)
symmetry.test(dompamina,option = "CM",boot = F)
```

- **Resultados con R**

```
     Symmetry test by Cabilio and Masaro (1996)

data:  dop[, "dop"]
Test statistic = 0.94558, p-value = 0.3444
alternative hypothesis: the distribution is asymmetric.
```

C. Prueba de Miao, Gel, y Gastwirth (MGG)

- **Aspectos Generales**

La prueba MGG tiene mayor poder que otras pruebas existentes para detectar la simetría de los datos normales contaminados sesgados, así como la mayoría de las distribuciones sesgadas pertenecientes a la familia lambda (que pertenecen a la familia exponencial).



Weiwen Miao | Yulia R. Gel | Joseph L. Gastwirth

- **Supuestos**

- Los datos provienen de una distribución continua con media, mediana y desviación estándar poblacional desconocida.
- Para propósitos prácticos, Cabilio y Masaro (1996) se usa la varianza asintótica de 0.5708 que se deriva del supuesto de que F tiene una distribución normal estándar.

- **Inferencia Estadística**

Procedimiento para el Desarrollo de la Prueba de Hipótesis

Se supone que se obtiene una muestra de observaciones independientes X_1, X_2 ,…., X_n de una distribución continua F con la función de densidad de probabilidad f, media desconocida μ, mediana v y desviación estándar σ. El objetivo es comprobar si F es simétrica respecto a v. Por lo tanto, la prueba de la hipótesis es dada por:

$$H_o: f(v-x) = f(v+x)$$
$$H_1: f(v-x) \neq f(v+x)$$

Existen tres pruebas de simetría sobre la mediana desconocida v que servirán de base para construir la prueba robusta de MGG. El primer procedimiento, propuesto por Cabilio y Masaro (1996), se basa en el estadístico de prueba.

$$C = \frac{\bar{X} - Me}{S_n},$$

donde $\bar{X}$, Me y S_n son la media muestral, la mediana y la desviación estándar, respectivamente. La estadística $\sqrt{n}C$ se distribuye de manera asintótica normalmente. Sin embargo, la varianza asintótica de C depende de la distribución subyacente F que generalmente se desconoce. Cabilio y Masaro

(1996) proponen utilizar la varianza asintótica de 0.5708 que se deriva de que F es una distribución normal estándar, y argumentan que el efecto de cualquier especificación errónea en la varianza asintótica es menor.
Miao, Gel y Gastwirth (2006) modifican el procedimiento de Cabilio-Masaro y proponen una estadística de prueba que puede considerarse como una versión robusta de C, es decir:

$$T = \frac{\overline{(X} - Me) * \sqrt{n} * \sqrt{0.5708}}{J_n}, \qquad J_n = \frac{c}{n}\sum_{i=1}^{n} |X_n - Me|, \qquad c = \sqrt{\frac{\pi}{2}}$$

- **Funciones en R**

En el paquete lawstat se encuentra la función symmetry.test que permite realizar hasta tres pruebas para evaluar la simetría de un conjunto de datos con respecto a la mediana.

- **Resultados con R**

```
datos<-
c(92,82,96,98,92,96,85,93,94,95,75,72,62,77,96,67,62,93,94,
99,76,74,84,82)

hist(datos)
library(e1071)
skewness(datos)
library(moments)
skewness(datos)
```

La función skewness permite calcular el coeficiente de asimetría.

```
library(lawstat)
symmetry.test(datos,option ="MGG",boot=F)
```

```
Symmetry test by Miao, Gel, and Gastwirth (2006)


data:  datos
Test statistic = -1.897, p-value = 0.05782
alternative hypothesis: the distribution is asymmetric.
```

Otra prueba es la prueba de Mira que también se encuentra en la función symmetry.test del paquete lawstat.

3. Pruebas para evaluar un parámetro de locación (o posición)

Las pruebas que se discutirán a continuación son alternativas no paramétricas a la prueba de hipótesis referente al valor de la media poblacional como son la prueba Z o t.

Hipótesis

Bilateral	Unilateral	
Caso A	Caso B	Caso C
$H_0 : Me = Me_0$	$H_0 : Me \leq Me_0$	$H_0 : Me \geq Me_0$
$H_1 : Me \neq Me_0$	$H_1 : Me > Me_0$	$H_1 : Me < Me_0$

A. Prueba de Signos

➢ **Aspectos Generales**

La Prueba de Signos es un caso particular de la prueba Binomial cuando π=0.5. Por esta razón, esta prueba puede utilizarse para probar una hipótesis nula referente al valor de la mediana de la población.

➢ **Supuestos**

- Los n elementos pertenecientes a la muestra son mutuamente independientes.
- Las observaciones de la variable de interés deben expresarse al menos en una escala de intervalos o la variable de interés debe ser de tipo cuantitativa.

➢ **Inferencia Estadística**

a) Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Se aplica un signo más (+) a cada valor muestral observado mayor que el valor hipotético de la mediana y un signo menos (-) a cada valor menor que el valor hipotético de la mediana. Si un valor muestral es exactamente igual al de la mediana hipotética, no se aplica ningún signo, con lo que el tamaño de muestra efectivo se reduce.

  Por ejemplo, a continuación, se presentan las notas del último examen parcial del curso de Estadística General de una muestra 10 estudiantes que fueron elegidos al azar:

```
notas<-c(9,12,8,8,15,11,8,11,13,15)
median(notas)
[1] 11
```

Alumno	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Nota	9	12	8	8	15	11	8	11	13	15
Signo	-	+	-	-	+	0	-	0	+	+

- Determinar el tamaño de la muestra (*n*) definitivo excluyendo aquellos valores que son similares a la mediana.
- Contabilizar "c" el número de signos (+) de acuerdo a la hipótesis que se desa probar.

- Hallar la probabilidad de ocurrencia (p-valor) de la hipótesis nula formulada con ayuda de la función de distribución Binomial y compararlo con el nivel de significación (α) adoptado por el investigador para el estudio.

Si la hipótesis nula (H_0) sobre el valor de la mediana es cierta, el número de signos más (+) debería ser aproximadamente igual al número de signos menos (-). Es decir la proporción de signos más (o de signos menos) debe ser alrededor de 0.5.

Caso A
Se toma el menor número de signos c´ ya sea (+) o (-)

$$p-valor = 2P(X \le c') = 2\sum_{x=0}^{c'}\binom{n}{x}2^{-n}$$

Caso B

$$p-valor = P\left[B(n,0.5) \ge c\right] = \sum_{i=c}^{n}\binom{n}{i}2^{-n}$$

Caso C

$$p-valor = P\left[B(n,0.5) \le c\right] = \sum_{i=0}^{c}\binom{n}{i}2^{-n}$$

Cuando *n* es suficientemente grande, se usa la aproximación de la distribución Normal a la distribución Binomial con $\pi = 1-\pi = 0.5$, para calcular el p-valor.

$$Z = \frac{x-\mu_p}{\sigma_p} = \frac{x-n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{x-0.5n}{0.50\sqrt{n}}$$

donde: *x*: es el número de signos positivos

b) Intervalo de Confianza
Se inicia ordenando las observaciones, es decir obteniendo las estadísticas de orden que son denotadas como $X_{(1)}<X_{(2)}<\cdots<X_{(n)}$.
Se desea obtener un intervalo de la forma:

$$X_{(a)} < Me < X_{(b)}$$

Tal que

$$P(X_{(a)} < Me < X_{(b)}) = 1-\alpha$$

Donde $1-\alpha$ es la probabilidad deseada que el intervalo capture a la mediana (Me).
Con el fin de tener $X_{(a)}$ < Me, al menos "a" de las observaciones deben ser menores a la Me y con el fin de tener Me < $X_{(b)}$, al menos b-1 de las observaciones podrían ser menores o iguales que Me.
Dado que las observaciones son independientes, la probabilidad que al menos *a* y al menos un b-1 de las observaciones sean menores a Me, es obtenido por la distribución Binomial con π=0.5

$$\sum_{i=a}^{b-1}\binom{n}{i}2^{-n}$$

Para la construcción de un intervalo con $100(1-\alpha)\%$ para la mediana se debe elegir a y b de tal manera que esta suma sea $1-\alpha$.

Debido a que la distribución Binomial es discreta no necesariamente se calcularán límites con el nivel de confianza deseado.
Para muestras grandes, valores aproximados de a y b pueden ser hallados usando la aproximación a la normal de la distribución Binomial. Si se usa la corrección por continuidad para esta aproximación, se podría obtener a y b resolviendo las siguientes ecuaciones y redondeando al entero más cercano:

$$\frac{a-np}{\sqrt{np(1-p)}} = -z_{(1-\alpha/2)} \qquad \frac{b-1-np}{\sqrt{np(1-p)}} = z_{(1-\alpha/2)}$$

En este caso particular p=0.5

- **Aplicación**

Ejemplo 1
Se afirma que el número de unidades ensambladas con un nuevo sistema rediseñado será mayor que el número de unidades ensambladas con el antiguo sistema, cuya mediana poblacional es de 80 unidades por turno laboral.
Pruebe si dicha afirmación es cierta al nivel de significación del 5%.
Los datos muestrales de 12 turnos elegidos azar sobre el número de unidades ensambladas con el nuevo sistema se reportan en la siguiente tabla:

Turno Laboral Muestreado	Unid. Ensam. (X)	Signo Dif (X - 80)
1	75	-
2	85	+
3	92	+
4	80	0
5	94	+
6	90	+

Turno Laboral Muestreado	Unid. Ensam. (X)	Signo Dif (X - 80)
7	91	+
8	76	-
9	88	+
10	82	+
11	96	+
12	83	+

Solución
H_0: Me ≤ 80 H_0: $\pi \le 0.50$
H_1: Me > 80 H_1: $\pi > 0.50$
α=0.05

$$p-valor = \sum_{i=c}^{n} \binom{n}{i} 2^{-n}$$

Desarrollo de la Prueba
El valor inicial de *n* es 12 pero debido a la existencia de un valor igual a la mediana a probar, ahora *n* sería 11.
Según los criterios se determina la probabilidad de observar 9 o más signos (+) de 11 observaciones.

$$p-valor = P(X \ge 9 / n = 11, \pi = 0.50) = \sum_{i=9}^{11} \binom{11}{i} 2^{-11} = 0.0328$$

Conclusión
A un nivel de significación de 0.05 existe suficiente evidencia estadística para rechazar la hipótesis nula.
Por lo tanto, podemos afirmar que la mediana de la producción por turno laboral con el nuevo sistema de ensamblaje es mayor de 80 unidades.

Ejemplo 2

El jefe de un laboratorio afirma que un nuevo medicamento que es aplicado a perros logra un tiempo de reacción mediano inferior a los 8 minutos. Para probar si dicha afirmación es verdadera se eligen al azar 33 animales experimentales y se les mide el tiempo de reacción (en minutos) luego de aplicarles una dosis del nuevo medicamento.

N°	Tiempo	Signo
1	6	-
2	9	+
3	9	+
4	5	-
5	7	-
6	7	-
7	9	+
8	7	-
9	7	-
10	7	-
11	10	+

N°	Tiempo	Signo
12	8	0
13	5	-
14	9	+
15	10	+
16	10	+
17	7	-
18	8	0
19	5	-
20	5	-
21	7	-
22	9	+

N°	Tiempo	Signo
23	10	+
24	7	-
25	7	-
26	9	+
27	7	-
28	7	-
29	6	-
30	7	-
31	7	-
32	5	-
33	11	+

Prueba si la afirmación del jefe de laboratorio es cierta. Use α=0.05.

Solución:
H_0: Me $\geq$ 8
H_1: Me < 8

α=0.05

Prueba Estadística

$$Z_c = \frac{x - 0.5n}{0.50\sqrt{n}} \sim \mathrm{N}(0,1) \text{ ó}$$

$$p\text{-value} = P(Z<Z_c)$$

Desarrollo de la Prueba
x = 11; n = 31

$$Z_c = \frac{11-0.5(31)}{0.5\sqrt{31}} = \frac{-4.5}{2.784} = -1.616$$

ó

p-valor = $P(Z<Z_c) = P(Z<-1.616) = 0.053 > \alpha$ no se rechaza H_0

Conclusión

A un nivel de significación de 0.05 no existe suficiente evidencia estadística para rechazar la H_0.

Por lo que se puede afirmar que el tiempo de reacción mediana no es inferior a los 8 minutos. Por lo tanto, no es correcta la afirmación del jefe del laboratorio.

Ejemplo 3
En base a los datos del Ejemplo1, calcular un intervalo del 95% de confianza.

$$\frac{a-0.5(12)}{\sqrt{0.25(12)}}=-1.96 \Rightarrow a=2.6\approx 3 \qquad \frac{b-1-0.5(12)}{\sqrt{0.25(12)}}=1.96 \Rightarrow b=10.39\approx 10$$

Entonces $X_{(3)}$=80 y $X_{(10)}$=92, con lo cual IC(Me) = [80;92]
El intervalo que va de 80 a 92 brinda un 95% de confianza de contener al número mediano de unidades ensambladas con el nuevo sistema.

- **Funciones en R**

Se puede utilizar la función binom.test con valor hipotético igual a 0.5
`binom.test(O1,n,valor hipotético, alternative, conf.level)`
En R, también existe la función SIGN.test del paquete BSDA.
En el caso de esta última función se debe ingresar el vector de datos.
`SIGN.test(vector de datos, valor hipotético, alternativa, conf.level)`
También existe la función signTest del paquete EnvStats
`signTest(vector de datos, valor hipotético, alternativa, conf.level)`
Otras funciones son: signtest del paquete nonpar y SignTest del paquete DescTools

- **Resultados en R**

```
ens<-c(75,76,80,82,83,85,88,90,91,92,94,96)

binom.test(9,11,0.5,alternative="greater")
```

```
        Exact binomial test
data:  9 and 11
number of successes = 9, number of trials = 11, p-value =
0.03271
alternative hypothesis: true probability of success is
greater than 0.5
95 percent confidence interval:
 0.5299132 1.0000000
sample estimates:
probability of success
             0.8181818
```

```
Library(BSDA)
SIGN.test(ens,md=80,alternative="g")
```

```
        One-sample Sign-Test
data:  ens
s = 9, p-value = 0.03271
alternative hypothesis: true median is greater than 80
95 percent confidence interval:
 81.14364      Inf
sample estimates:
median of x
```

```
       86.5
                  Conf.Level  L.E.pt U.E.pt
Lower Achieved CI     0.9270 82.0000    Inf
Interpolated CI       0.9500 81.1436    Inf
Upper Achieved CI     0.9807 80.0000    Inf
```

```
signTest(ens,alternative="g",mu=80)
```

```
Results of Hypothesis Test
--------------------------

Null Hypothesis:                 median = 80
Alternative Hypothesis:          True median is greater tha
n 80
Test Name:                       Sign test
Estimated Parameter(s):          median = 88

Data:                            ens

Test Statistic:                  # Obs > median = 9

P-value:                         0.03271484

Confidence Interval for:         median

Confidence Interval Method:      exact

Confidence Interval Type:        lower

Confidence Level:                88.67188%

Confidence Limit Rank(s):        4

Confidence Interval:             LCL =  83
                                 UCL = Inf
```

```
library(DescTools)
SignTest(ens,alternative="g",mu=80)
```

```
      One-sample Sign-Test

data:  ens
S = 9, number of differences = 11, p-value =
0.03271
alternative hypothesis: true median is greater than 80
98.1 percent confidence interval:
  80 Inf
sample estimates:
median of the differences
                     86.5
```

```
binom.test(11,31,0.5,alternative="less")
```

```
        Exact binomial test
data:  11 and 31
number of successes = 11, number of trials = 31, p-value =
0.07481
alternative hypothesis: true probability of success is less
than 0.5
95 percent confidence interval:
 0.0000000 0.5182493
sample estimates:
probability of success
             0.3548387
```

Resultados con R para intervalos de confianza

```
library(BSDA)
SIGN.test(ens,conf.level=0.95)
```

```
                  Conf.Level  L.E.pt  U.E.pt

Lower Achieved CI     0.8540 82.0000 91.0000
Interpolated CI       0.9500 80.2127 91.8936
Upper Achieved CI     0.9614 80.0000 92.0000
```

```
library(EnvStats)
signTest(ens,alternative="g",mu=80)

library(DescTools)
SignTest(ens,alternative="g",mu=80)
```

- **Algunas consideraciones en R**

  La función SIGN.test:

- Permite analizar una hipótesis bilateral o unilateral.
- Permite obtener intervalos de confianza.

B. Prueba de Rango de Wilcoxon

- **Aspectos Generales**

La Prueba de Rango Wilcoxon al igual que la Prueba de Signos, puede usarse para probar una hipótesis nula referente al valor de la mediana de la población.
Pero dado que la prueba de Wilcoxon considera la magnitud de la diferencia entre el valor muestral y el valor hipotético de la mediana, es una prueba más sensible que la Prueba de Signos.
Por otra parte, puesto que se determinan las diferencias D_i (entre el valor observado y el valor hipotético), los valores deben estar al menos en la escala de intervalo, es decir, la variable de interés debe ser cuantitativa.
Probar si la diferencia D_i o los datos originales provienen de una distribución simétrica es equivalente dado que restar un valor constante a las observaciones solo hace que estas se desplacen en el eje horizontal.

Frank Wilcoxon
(1892 – 1965)

- **Supuestos**

- Las observaciones de la variable de interés deben expresarse al menos en una escala de intervalos.
- Las observaciones provienen de una distribución simétrica.

- **Inferencia estadística**

a) Prueba de Hipótesis

Procedimiento para el Desarrollo de la Prueba

- Determinar la diferencia entre cada valor observado y el valor hipotético de la mediana, diferencia que, con el signo aritmético que le corresponda, se designa como D_i = (X_i - Med_0). Si alguna diferencia es igual a cero, la observación asociada se excluye del análisis y el tamaño de muestra efectivo se reduce.
- Se calcula los rangos de las diferencias sin tomar en cuenta el signo de las mismas (en valor absoluto). En caso de haber empate se asigna un rango promedio a todas las diferencias empatadas.
- Se obtiene la suma de los rangos en forma separada para las diferencias positivas y para las negativas; la suma de los rangos positivos (S^+) será utilizado para determinar el valor de *p-valor* y compararlo con el nivel de significación α.

Caso Bilateral
La hipótesis nula (H_0) se rechaza al nivel de significación α si se cumple una de las dos desigualdades

$$S^+ \leq w_{(n,\alpha/2)}\,;\ S^+ \geq \frac{n(n+1)}{2} - w_{(n,\alpha/2)}$$

Caso Unilateral derecho
La hipótesis nula (H_0) se rechaza al nivel de significación α si se cumple la desigualdad

$$S^+ \geq \frac{n(n+1)}{2} - w_{(n,\alpha)}$$

Caso Unilateral izquierdo
La hipótesis nula (H_0) se rechaza al nivel de significación α si se cumple la desigualdad

$$S^+ \leq w_{(n,\alpha)}$$

Siendo $w_{(n,\alpha/2)}$ y $w_{(n,\alpha)}$ los valores críticos de la distribución de Wilcoxon correspondiente al tamaño de muestra n.
En R, los valores críticos se pueden obtener mediante la función qsignrank(nivel de significación, n).

Cuando el número de los valores de la muestra que son distintos al valor de la mediana que aparece en la hipótesis nula es grande se puede aproximar el estadístico de Wilcoxon a una distribución Normal
El estadístico *w* de Wilcoxon será la suma de los rangos correspondientes a las diferencias positivas.

$$Z_c = \frac{S^+ - \mu_w}{\sigma_w} \sim \mathrm{N}(0,1)$$

donde: $\mu_{S^+} = \dfrac{n(n+1)}{4}$ $\qquad \sigma_{S^+} = \sqrt{\dfrac{n(n+1)(2n+1)}{24}}$

Si hubiera empates, la varianza sufre una ligera modificación y se aplica:

$$\sigma_{S^+} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \sum_{i=1}^{g} \frac{\left(t_i^3 - t_i\right)}{2}}$$

donde, *g* es el número de grupos empatados y t_i es el tamaño del i-ésimo grupo empatado.

b) Intervalo de Confianza
Dada una muestra aleatoria $X_1, X_2, \ldots, X_n$ para obtener un intervalo de confianza, primero se deben calcular los n(n-1)/2 promedios de Walsh definidos como:

$$\frac{X_i + X_j}{2} \qquad \forall i \leq j$$

A la mediana de estos promedios se le conoce como el estimador de Hodges-Lehmann.
Si $X_{(1)}, X_{(2)},\ldots,X_{(m)}$ son las estadísticas de orden de los m=n(n-1)/2 promedios de Walsh y a es tal que $P(S^+ \leq a) = P(S^+ \geq m - a) = \alpha/2$, entonces:

$[X_{(a+1)}, X_{(m-a+1)}]$ es un intervalo de confianza de nivel $1-\alpha$, basado en S^+.

En R la función walsh del paquete Rfit permite obtener los promedios de Walsh.

Si n es suficientemente grande, "a" se puede obtener a partir de la distribución Normal con corrección por continuidad

$$a \cong \frac{n(n+1)}{4} - 0.5 - Z_{(1-\alpha/2)}\sqrt{\frac{n(n+1)(2n+1)}{24}}$$

- **Aplicación**

Ejemplo
A una muestra de 19 alumnos elegidos al azar se les evaluó su aprendizaje de un nuevo sistema computacional obteniendo los siguientes resultados:

Alumno	Puntaje (X)	Dif (X - 90)	Rango de |d|	Rango con Signo	
				(+)	(-)
1	76	-14	17.5		17.5
2	96	6	10.0	10	
3	92	2	3.5	3.5	
4	80	-10	14.5		14.5
5	84	-6	10.0		10
6	80	-10	14.5		14.5
7	91	1	1.0	1	
8	76	-14	17.5		17.5
9	88	-2	3.5		3.5
10	92	2	3.5	3.5	
11	86	-4	8.0		8
12	83	-7	12.0		12
13	88	-2	3.5		3.5
14	84	-6	10.0		10
15	99	9	13.0	13	
16	87	-3	6.5		6.5
17	68	-22	19.0		19
18	87	-3	6.5		6.5
19	78	-12	16.0		16
Total				31	159

Se considera que el programa de estudios ha sido provechoso si el puntaje mediano es superior a los 90 puntos. Pruebe si el programa ha sido provechoso a un nivel de significación de 0.05.

Primero se debe evaluar que los datos provienen de una distribución simétrica
H_0: As = 0
H_1: As ≠ 0
α=0.05
Pvalor= 0.4307
Conclusión
A un nivel de significación de 0.05, no se puede afirmar que los datos del puntaje provienen de una distribución asimétrica.
Por lo tanto, se puede proceder a realizar la prueba de Wilcoxon.

Solución (Si analiza como muestra pequeña)
H_0: Me ≤ 90
H_1: Me > 90
α=0.05

Prueba Estadística

$$S^+=31$$

Criterio de Decisión
Se rechaza H_0 si

$$S^+ \geq \frac{n(n+1)}{2} - w_{(n,\alpha)} = \frac{19(19+1)}{2} - 54 = 136$$

```
qsignrank(0.05,19)
[1] 54
```

Conclusión
A un nivel de significación de 0.05 no existe suficiente evidencia estadística para rechazar la hipótesis nula.
Por lo tanto, no se puede afirmar que el programa de estudios haya sido provechoso

Solución (Si analiza como muestra grande)
H_0: Me ≤ 90
H_1: Me > 90
α=0.05

Prueba Estadística

$$Z_c = \frac{S^+ - \mu_w}{\sigma_w} \sim \text{N}(0,1) \text{ ó}$$

$$p\text{-valor} = P(Z>Z_c) = 1 - P(Z<Z_c)$$

Desarrollo de la Prueba

$$\mu_w = \frac{19(19+1)}{4} = 95$$

$$\sigma_w = \sqrt{\frac{19(19+1)[2(19)+1)]}{24} - \left[\frac{4^3-4}{2} + \frac{2^3-2}{2} + \frac{3^3-3}{2} + \frac{2^3-2}{2} + \frac{2^3-2}{2}\right]} = 23.80$$

$$Z_c = \frac{31-95}{23.80} = -2.71$$

p-value = P(Z>-2.71) =1- P(Z<-2.71)=1-0.0034=0.9966

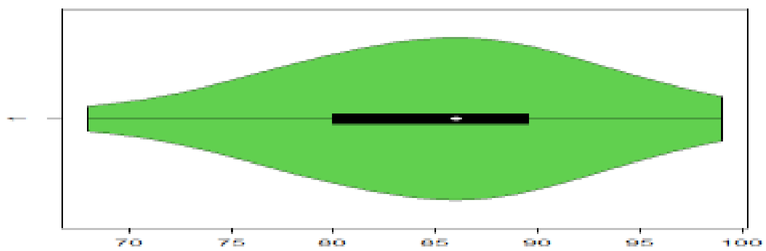
Criterio de Decisión

No se rechaza H_0 si $Z_c \le 1.65$. ó Si p-value $\ge \alpha$ no se rechaza la H_0.
Se rechaza la H_0 si $Z_c > 1.65$. Si p-value $< \alpha$ se rechaza la H_0.

Conclusión
A un nivel de significación de 0.05 existe suficiente evidencia estadística para no rechazar la hipótesis nula.
Por lo tanto, no se puede afirmar que el programa de estudios haya sido provechoso.

Ejemplo 2
Calcule un intervalo del 95% de confianza para el puntaje mediano obtenido utilizando el nuevo sistema computacional

```
puntaje<-c(76,96,92,80,84,80,91,76,88,92,86,83,88,84,99,87,
68,87,78)
library(vioplot)
vioplot(puntaje,horizontal = T,col=3)
```




```
n<-length(puntaje)
m<-n*(n+1)/2
alfa<-0.05
library(Rfit)
prom<-sort(walsh(puntaje))

a<-qsignrank(alfa/2,n)
limites<-c(prom[a+1],prom[m-a+1])
```

- **Funciones en R**

La función qsignrank(nivel de significación, n) permite obtener los cuantiles de los valores tabulares para la distribución de Wilcoxon.

```
wilcox.test(vector de datos, valor hipotético, alternative,
conf.level)
```

También se puede hacer uso de la función wilcoxE.test del paquete PASWR.
```
wilcoxE.test(vector de datos, valor hipotético,
alternative, conf.level)
```

La función wilcox.exact del paquete exactRankTests permite obtener los resultados con el ajuste para empates
```
wilcox.exact(vector de datos, valor hipotético,
alternative, conf.level)
```

- **Resultados con R**

```
symmetry.test(puntaje,option="MGG",boot=F)
```

```
      Symmetry test by Miao, Gel, and Gastwirth (2006)

data:  puntaje
Test statistic = -0.78796, p-value = 0.4307
alternative hypothesis: the distribution is asymmetric.
```

```
wilcox.test(puntaje,alternative="greater",mu=90)
```

```
        Wilcoxon signed rank test with continuity correction

data:  puntaje
V = 31, p-value = 0.9953
alternative hypothesis: true location is greater than 90

Mensajes de aviso perdidos
In wilcox.test.default(puntaje, alternative = "greater", mu
= 90) :
  cannot compute exact p-value with ties
```

```
wilcox.test(puntaje,alternative="two.sided",mu=90,conf.int=
TRUE,conf.level=0.95)
```

```
Wilcoxon signed rank test with continuity correction
data:  puntaje
V = 31, p-value = 0.01047
alternative hypothesis: true location is not equal to 90
95 percent confidence interval:
 81.49999 88.99992
sample estimates:
(pseudo)median
      85.49997
```

```
wilcoxE.test(puntaje, mu=90, alternative="g")
```

```
     Wilcoxon Signed Rank Test

data:  puntaje
```

```
t+ = 31, p-value = 0.9964
alternative hypothesis: true median is greater than 90
95.13798 percent confidence interval:
  82 Inf
sample estimates:
(pseudo)median
          85.5
```

```
wilcox.exact(puntaje,alternative="greater",mu=90)
```

```
        Exact Wilcoxon signed rank test

data:  puntaje
V = 31, p-value = 0.9964
alternative hypothesis: true mu is greater than 90
```

```
        Exact Wilcoxon signed rank test

data:  puntaje
V = 31, p-value = 0.007847
alternative hypothesis: true mu is not equal to 90
95 percent confidence interval:
 81.5 89.0
sample estimates:
(pseudo)median
         85.25
```

- **Algunas consideraciones** en R
  La función wilcox.test:
  - Permite analizar una hipótesis bilateral o unilateral.
  - Permite obtener intervalos de confianza.

4. Pruebas para detectar datos atípicos

Usualmente cuando realizamos una prueba estadística o procedimiento de estimación, asumimos que los datos son observaciones de variables aleatorias iid, frecuentemente de una distribución normal. Sin embargo, algunas veces se puede notar que algunas observaciones se encuentran alejadas de la mayoría. Estas observaciones son comúnmente llamadas datos atípicos (“outliers”).
Los datos atípicos se deben a varias fuentes como:

- Aleatoriedad: En muestras normales se espera que una de cada 20 observaciones puede ser mayor a dos desviaciones estándar de la media y uno de cada 100 pueda ser 2.5 desviaciones estándar de la media. Por lo tanto, la presencia de “datos atípicos” no necesariamente significa que es un problema.
- Falla en el proceso de la generación de los datos. En este caso, la identificación del dato atípico puede ser muy importante.
- Una persona realiza mediciones que no son homogéneas con la de otra persona.
- Falla del instrumento de medición sea mecánico o humano.
- Error en la unidad de medida al ingreso de datos en una computadora.

Una vez detectados los datos discordantes, existen muchas alternativas para saber qué hacer con ellos entre las cuales tenemos:

- Eliminar los datos discordantes y continuar con el análisis.
- Reemplazar los datos atípicos con los valores no atípicos más cercanos.
- Reemplazar los datos atípicos con nuevas observaciones.

Ciertas pruebas de detección de datos atípicos, asumen que el número de posibles datos atípicos en la muestra es conocido a priori, adicionalmente algunas de estas pruebas asumen el conocimiento de la dirección de la observación discordante.

Cuando dos o más datos atípicos ($k > 1$) aparecen en una muestra, una prueba para un único dato atípico podría no detectar ese dato atípico debido a que los otros datos atípicos inflan la varianza, de esta manera enmascaran la detección de otros datos atípicos. Por esta razón, es necesario considerar pruebas para más de un dato atípico.
Por otro lado, si se considera que hay $k > 1$ datos atípicos y hay menos que la cantidad k, podría ocurrir lo siguiente:

- La prueba podría rechazar la hipótesis nula de la no existencia de datos atípicos debido a la gran influencia que los verdaderos datos atípicos han tenido en la prueba estadística, es decir, se identifica más datos atípicos de los que realmente hay. Este es un efecto conocido como empantanamiento.
- Los verdaderos valores extremos podrían no tener suficiente influencia para lograr rechazar la hipótesis nula, es decir, no se identifica que hay datos atípicos. Esto es conocido como el enmascaramiento.

Hipótesis
Bilateral
H_0: No existe valores atípicos
H_1: El valor más pequeño o más grande de los datos es un valor atípico

Unilateral
H_0: No existe valores atípicos
H_1: El valor más pequeño de los datos es un valor atípico

H_0: No existe valores atípicos
H_1: El valor más grande de los datos es un valor atípico

A. Prueba de Grubbs

- **Aspectos Generales**

La prueba de Grubbs (1950) es quizás la más conocida prueba para identificar un dato atípico en una muestra de n observaciones.
También conocida como prueba residual normalizada máxima o prueba de desviación extrema estudiantil, es una prueba utilizada para detectar valores atípicos en un conjunto de datos univariados
La prueba se basa en la máxima desviación con respecto a la media muestral.

Frank E. Grubbs
(1913 – 2000)

- **Supuestos**

- Los datos de la variable de interés deben estar medidos en al menos una escala intervalo.

- **Inferencia estadística**

Procedimiento para el Desarrollo de la Prueba de Hipótesis
Calcular
$T = \max\left(T_1, T_n\right)$
Donde

$$T_1 = \frac{\left(\bar{x} - x_{(1)}\right)}{s} \text{ y } T_n = \frac{\left(x_{(n)} - \bar{x}\right)}{s}$$

En el caso de que la dirección del posible outlier es conocido, T_1 podría ser usado para identificar un extremo inferior y T_n podría ser usado para identificar un extremo superior. En cada caso, la prueba estadística es comparada con un percentil superior de la distribución. Valores grandes de la prueba indican la presencia de un outlier.

Grubbs mostró que T_n es idéntico a la razón de la suma de las diferencias al cuadrado con y sin el outlier.

$$\frac{S_n^2}{S^2} = \frac{\sum_{i=1}^{n-1}\left(x_{(i)} - \bar{x}_n\right)^2}{\sum_{i=1}^{n}\left(x_{(i)} - \bar{x}\right)^2} = 1 - \frac{T_n^2}{n-1}$$

Donde $\bar{x}_n$ es la media de la muestra excluyendo a $x_{(n)}$. Algo parecido para probar el estadístico de prueba con un outlier inferior.

$$\frac{S_1^2}{S^2} = \frac{\sum_{i=2}^{n}\left(x_{(i)} - \bar{x}_1\right)^2}{\sum_{i=1}^{n}\left(x_{(i)} - \bar{x}\right)^2} = 1 - \frac{T_1^2}{n-1}$$

- **Aplicación**

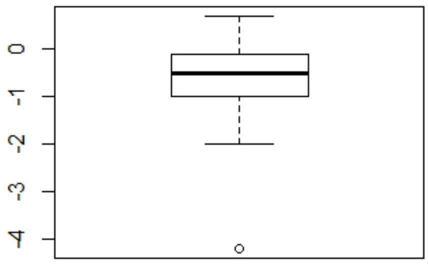
Trece promedios anuales de razón de erosión (m/año) elegidos al azar fueron estimados para la Costa Este de US. Los datos negativos indican erosión y los positivos indican acreción

Maine	-0.4
New Hampshire	-0.5
Massachusetts	-0.9
Rhode Island	-0.5
New York	0.1
New Jersey	-1.0
Delaware	0.1

Maryland	-1.5
Virginia	-4.2
North Carolina	-0.6
South Carolina	-2.0
Georgia	0.7
Florida	-0.1

Pruebe a un nivel de significación de 0.05, si existe un outlier.

Lo más recomendable es realizar un gráfico que ayude a detectar la presencia de valores atípicos.



Se puede observar un valor atípico en la parte inferior.

H_0: No existe un valor atípico
H_1: El valor más pequeño de los datos es un valor atípico
$\alpha = 0.05$.

Estadístico de prueba

$$T_1 = \frac{(-0.83 - (-4.2))}{1.23} = 2.74$$

El p-valor asociado a la prueba es 0.004, por lo tanto, se rechaza H_0.

Conclusión
A un nivel de significación de 0.05, se puede afirmar que existe un valor atípico.

- **Funciones e**n R

```
library(outliers)
grubbs.test(datos,two.sided=TRUE)
```

- **Resultados con R**

```
datos<-c(-0.4,-0.5,-0.9,-0.5,0.1,-1.0,0.1,-1.5,-4.2,-0.6,-
2.0,0.7,-0.1)
grubbs.test(datos,two.sided=TRUE)
```

```
Grubbs test for one outlier
data:  datos
G = 2.73070, U = 0.32681, p-value = 0.007669
alternative hypothesis: lowest value -4.2 is an outlier
```

```
grubbs.test(datos,two.sided=FALSE)
```

```
      Grubbs test for one outlier

data:  datos
G = 2.73070, U = 0.32681, p-value = 0.003835
alternative hypothesis: lowest value -4.2 is an outlier
```

- **Algunas consideraciones de los programas estadísticos**
Se debe tener cuidado en identificar previamente si el outlier es superior, inferior o si se ubica en ambos lados. Esto se puede hacer con la ayuda de un diagrama de cajas.

B. Prueba de Dixon

- **Aspectos Generales**

Es una variante de la prueba de Grubbs. Es una prueba que puede ser aplicada a muestras pequeñas (menores a 30).

Wilfrid Joseph Dixon
(1915 – 2008)

- **Supuestos**

- Los datos de la variable de interés deben estar medidos en al menos una escala intervalo.

- **Inferencia estadística**

Procedimiento para el Desarrollo de la Prueba de Hipótesis

Para el caso de outliers superiores, calcular el ratio de la distancia entre las dos observaciones más grandes y el rango.

$$r_{10} = \frac{\left(x_{(n)} - x_{(n-1)}\right)}{\left(x_{(n)} - x_{(1)}\right)}$$

Para el caso de outliers inferiores, calcular:

$$r'_{10} = \frac{\left(x_{(2)} - x_{(1)}\right)}{\left(x_{(n)} - x_{(1)}\right)}$$

Una prueba para un simple outlier cuando la dirección es desconocida es obtenido por r=max(r_{10},r'_{10})

Dixon propuso una prueba alternativa para evitar el efecto de enmascaramiento debido a otras observaciones.

$$r_{jk} = \frac{\left(x_{(n)} - x_{(n-j)}\right)}{\left(x_{(n)} - x_{(k+1)}\right)}$$

La cual elimina la influencia de las j-1 otras observaciones grandes y k observaciones extremas pequeñas. Dixon sugirió usar j=1,2 y k=0,1,2. Para evitar el efecto de enmascaramiento de dos outliers superiores se podría usar j=2 y k=1 o 2, podría ser considerado como más robusto. Similarmente r'_{jk} son usados para simples outliers inferiores.

$$r'_{jk} = \frac{\left(x_{(j+1)} - x_{(1)}\right)}{\left(x_{(n)} - x_{(k+1)}\right)}$$

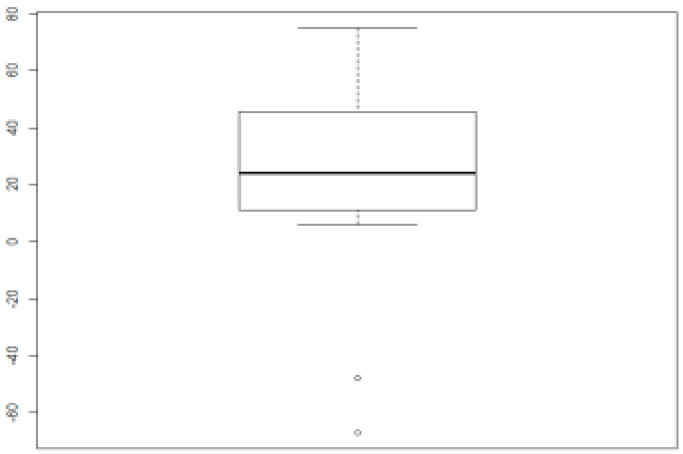
En general Dixon sugirió usar r_{jk} de acuerdo al tamaño de la muestra (n).
r_{10} para muestras muy pequeñas (n ≤ 7),
r_{11} para un tamaño de muestra de 8 a 10
r_{21} para un tamaño de muestra de 11 a 13
r_{22} para muestras mayores o iguales a 14.

- **Aplicación**

Las diferencias del consumo anual (en decenas de soles) de combustible en los años 2014 y 2015 de 15 familias seleccionadas al azar son:

50	-67	8	16	6	23	28	41
14	29	56	24	75	60	-48	

Es bueno realizar un análisis descriptivo previo como la elaboración de un diagrama para poder visualizar si realmente existen datos discordantes y si hay uno o varios para evitar el efecto de enmascaramiento.



Se puede visualizar la existencia de dos valores atípicos.
Con fines prácticos primero se realizará una prueba de dos colas, considerando que no se ha realizado previamente el diagrama de cajas.

Pruebe a un nivel de significación de 0.05, si existe un outlier.

H_0: No existen datos atípicos
H_1: El valor más pequeño o más grande de los datos es un valor atípico

α=0.05

Estadístico de prueba

$$r'_{10} = \frac{\left(-48-(-67)\right)}{75-\left(-67\right)} = 0.134$$

El p-valor asociado es 0.424.
Conclusión
A un nivel de significación de 0.05, no se puede afirmar que existe un valor atípico.

Este resultado no concuerda con lo observado en el diagrama de cajas. Como se puede apreciar r'_{10} no identificó el outlier debido al efecto de enmascaramiento de $x_{(2)}$.

Si se utiliza r'_{20} podría identificar el outlier.

$$r'_{20} = \frac{\left(6-(-67)\right)}{75-\left(-67\right)} = 0.51$$

El p-valor asociado es 0.012 con lo cual se rechaza H_0.

Si se utiliza r'_{22} de acuerdo a lo considerado por Dixon el pvalor sería 0.01507 con lo cual también se rechaza H_0.

- **Funciones en R**

```
library(outliers)
dixon.test(datos,type=20)
```

- **Resultados con** R

```
datos<-c(50,-67,8,16,6,23,28,41,14,29,56,24,75,60,-48)
dixon.test(datos,type=20,two.sided = FALSE)
```

```
        Dixon test for outliers

data:  datos
Q = 0.51408, p-value = 0.01176
alternative hypothesis: lowest value -67 is an outlier
```

```
        Dixon test for outliers

data:  datos
Q = 0.5935, p-value = 0.01507
alternative hypothesis: lowest value -67 is an outlier
```

- **Algunas consideraciones de R**

  Se debe tener en consideración la cantidad de posibles valores outliers para evitar el problema de enmascaramiento.

  Existe la función chisq.out.test del paquete outliers que realiza la prueba Chi Cuadrado para detección de outliers.

C. Prueba de Rosner

➢ Aspectos Generales

No necesariamente los valores atípicos son medidas incorrectas; por tal motivo, se recomienda que posteriormente se realice un análisis profundo de la existencia de tal valor atípico.
Existen muchos métodos para detectar la presencia de algún valor atípico, diferenciándose principalmente en que algunos métodos tienen condiciones de tamaño de muestra. Algunos de los métodos más recomendados para un único valor atípico son: La prueba de Grubbs, la prueba de Dixon.
la Prueba de Rosner o Generalized Extreme Studentized Deviate (1983).
La prueba de Rosner se utiliza para múltiples valores atípicos, exactamente, para detectar hasta 10 valores atípicos inferiores y superiores entre los valores de datos seleccionados de una misma muestra. Esta prueba detectará valores atípicos que sean mucho más pequeños o mucho más grandes que el resto de los datos. El enfoque de Rosner está diseñado para evitar el problema del enmascaramiento, que ocurre cuando un valor atípico (valor desapercibido) no se detecta porque tiene un valor cercano a otro valor atípico.

➢ Supuestos

- El tamaño de la muestra debe ser mayor o igual a 25 (n ≥ 25).
- Dentro del conjunto de datos puede existir hasta 10 valores atípicos inferiores y superiores.

➢ Inferencia estadística

Antes de realizar la prueba se necesita un número M de outliers potenciales presentes en el grupo de observaciones (conjunto de datos). La hipótesis nula es que todo el grupo de observaciones representa una muestra tomada de una distribución normal. La hipótesis alternativa es que hay M outliers, o $M-1$ outliers, o $M-2$ outliers, o $M-(algún\ número\ menor\ a\ M)$ outliers, o 1 outlier. (Gilbert, 1987)

Procedimiento para el Desarrollo de la Prueba de Hipótesis

- Los datos deben estar ordenados de manera ascendente (de menor a mayor).
- Se realiza una exploración de los datos (estadística descriptiva). Mediante el gráfico de cajas se podrá visualizar la existencia de algún outlier.
- Se determina la media muestral y la desviación estándar muestral.
- $R_s = \frac{|x_j - \bar{x}|}{s_X}$ es la fórmula del estadístico de la Prueba de Rosner.

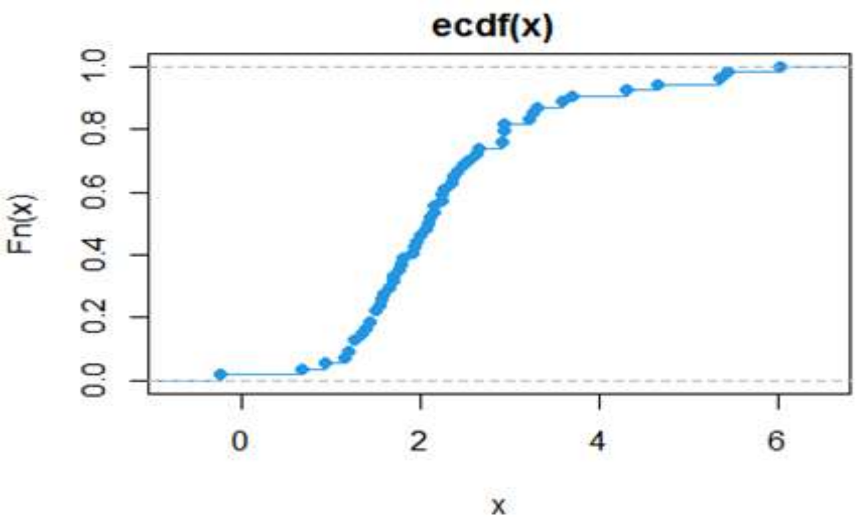
- El rango de los outliers es: i = 1, 2, 3, …, s. El cual indica el número de outliers que se irán probando en el estadístico R. Siendo i = 1 la prueba del primer outlier más distante a la media muestral (de entre los valores más distantes), y i = s el último outlier menos distante a la media muestral (de entre los valores más distantes). Por lo tanto, s representa el máximo número de outliers presentes en el conjunto de datos.
- $R_{i=1,2,3,\ldots,s}$ representa el estadístico de Rosner con un límite inferior fijo de 1 outlier presente en el conjunto de datos hasta un límite superior fijo de s outliers presentes en el conjunto de datos. Tal estadístico es comparado con el valor crítico.
- x_j es el valor del outlier más distante de la media (de entre los valores más distantes).
- Para realizar el procedimiento del cálculo del estadístico R_s , se tiene que hallar el valor de la media $\bar{X}$ y de la desviación estándar S sin considerar en el conjunto de datos al valor del outlier x_j más distante de la media muestral (de entre los valores más distantes).

> **Aplicación**

Se tienen los siguientes datos de ingresos de personas

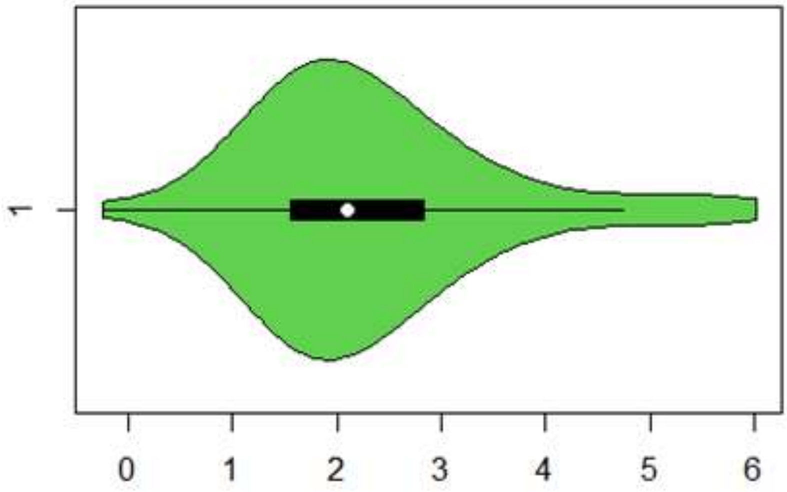
```
datos<-c(-0.25,0.68,0.94,1.15,1.2,1.26,1.26,1.34,1.38,1.43,1.49,1.49,
1.55,1.56,
         1.58,1.65,1.69,1.7,1.76,1.77,1.81,1.91,1.94,1.96,1.99,2.06,2.
09,2.1,
         2.14,2.15,2.23,2.24,2.26,2.35,2.37,2.4,2.47,2.54,2.62,2.64,2.
9,2.92,
         2.92,2.93,3.21,3.26,3.3,3.59,3.68,4.3,4.64,5.34,5.42,6.01)

plot.ecdf(datos,col=4)
```

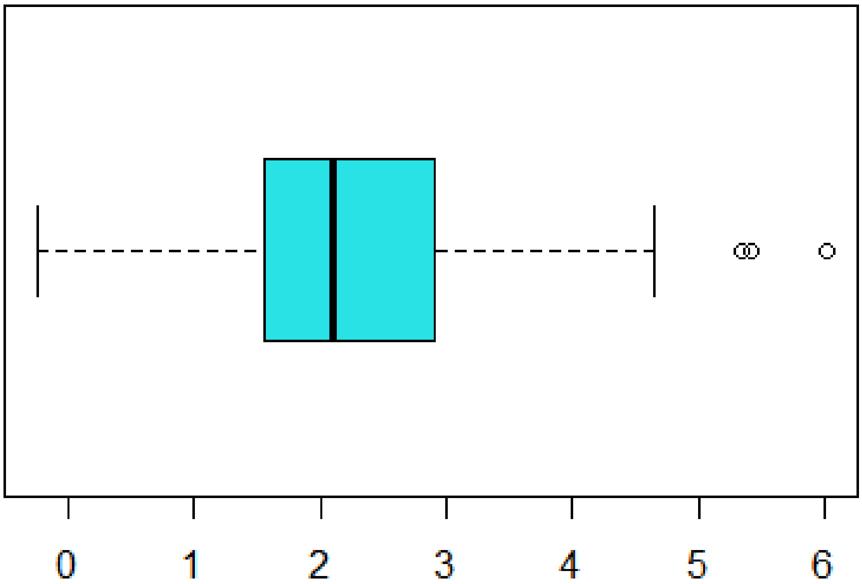



```
library(vioplot)

vioplot(datos,col=3,horizontal=T)
```




```
boxplot(datos,col=5,horizontal = TRUE)
```



- **Funciones en R**
  La función rosnerTest del paquete EnvStats permite obtener los resultados de los outliers de la muestra

- **Resultados con** R

```
## $distribution
## [1] "Normal"
##
## $statistic
##      R.1      R.2      R.3
## 3.118906 2.942973 3.179424
##
## $sample.size
## [1] 54
##
## $parameters
## k
## 3
##
## $alpha
## [1] 0.05
##
## $crit.value
## lambda.1 lambda.2 lambda.3
## 3.158794 3.151430 3.143890
##
## $n.outliers
## [1] 3
##
## $alternative
## [1] "Up to 3 observations are not\n      from the same Distribution."
##
## $method
## [1] "Rosner's Test for Outliers"
##
## $data
##  [1] -0.25  0.68  0.94  1.15  1.20  1.26  1.26  1.34  1.38  1.43  1.49
  1.49
## [13]  1.55  1.56  1.58  1.65  1.69  1.70  1.76  1.77  1.81  1.91  1.94
  1.96
## [25]  1.99  2.06  2.09  2.10  2.14  2.15  2.23  2.24  2.26  2.35  2.37
  2.40
## [37]  2.47  2.54  2.62  2.64  2.90  2.92  2.92  2.93  3.21  3.26  3.30
  3.59
## [49]  3.68  4.30  4.64  5.34  5.42  6.01
##
## $data.name
## [1] "datos"
##
## $bad.obs
## [1] 0
##
## $all.stats
##   i   Mean.i     SD.i Value Obs.Num    R.i+1 lambda.i+1 Outlier
## 1 0 2.320741 1.182870  6.01      54 3.118906   3.158794    TRUE
## 2 1 2.251132 1.076757  5.42      53 2.942973   3.151430    TRUE
## 3 2 2.190192 0.990685  5.34      52 3.179424   3.143890    TRUE
##
```

5. Comparación de pruebas mediante Potencia de Prueba

Si se desea comparar las pruebas de Signos y la de Wilcoxon esto se puede evaluar mediante su potencia de prueba.

La potencia de prueba se puede llevar a cabo mediante Simulación de Montecarlo realizando los siguientes pasos para una hipótesis

- Definir la hipótesis a evaluar, como por ejemplo H_0: $Me \leq Me_0$ versus H_1: $Me > Me_0$
- Simular r muestras de tamaño n proveniente de una distribución teórica F(X),por ejemplo la Normal.
- Para cada una de las muestras generadas calcular el estadístico de prueba (o p-valor).
- Para cada uno de los p-valores obtenidos en las pruebas compararlo con el nivel de significación α.
- Contabilizar el número de veces que se rechaza H_0 para cada prueba.

```
library(BSDA)
```

```
potencia<-function(n,mu0,mu,des,r,alfa){
  signos<-rep(0,r)
  wilcox<-rep(0,r)
  for (i in 1:r){
    datos<-rnorm(n,mu,des)
    signos[i]<- SIGN.test(datos, md = mu0, alternative =
"two.sided")$"p.value"
    wilcox[i]<-
wilcox.test(datos,mu=mu0,alternative="two.sided")$p.value
  }
  s<-sum(ifelse(signos<alfa, 1, 0))/r
  w<-sum(ifelse(wilcox<alfa, 1, 0))/r
  return(list(s=s,w=w))
}
```

```
potencia(50,16.5,17,1,100,0.05)
```

Como se puede apreciar para poder obtener la potencia de prueba se pueden definir distintos escenarios, los cuales podrían depender para esta comparación de Wilcoxon y de Signos de los siguientes aspectos:

- El tamaño de la muestra.
- La variabilidad de los datos.
- La distribución teórica utilizada para generar los datos pseudoaleatorios.

Otra prueba interesante es la prueba de Chen que es una prueba para datos asimétricos, lo puede encontrar en la función chenTTest del paquete EnvStat

La función statcompute del paquete PoweR ofrece varias pruebas no paramétricas.

En el paquete randtest existe varias pruebas sobre aleatoriedad.