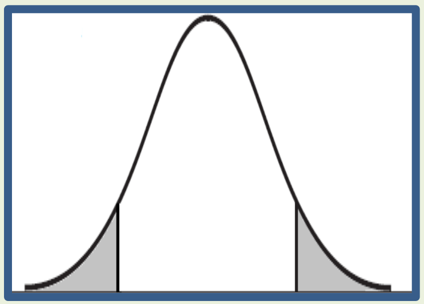
Métodos no paramétricos

Métodos estadísticos que requieren pocas o ninguna suposición acerca de la distribución de probabilidad de la población y acerca del nivel de medición. Estos métodos suelen usarse cuando se cuenta con datos nominales y ordinales.

Los análisis estadísticos se realizan con datos ya sea nominales, ordinales, de intervalo o de razón.

Métodos paramétricos



Cuando se conoce la forma de la distribución de probabilidad que sigue la variable aleatoria a estudiar en la población.

El problema consiste en determinar los diferentes parámetros de dicha distribución (ejemplo: media y varianza para la distribución normal)

Métodos no paramétricos

Cuando la distribución de la población es desconocida

El problema principal es encontrar la forma y características de la distribución.

Métodos no paramétricos

- ✓**Prueba de los signos**
- ✓**Prueba de los rangos con signo de Wilcoxon**
- ✓prueba de Mann-Whitney-Wilcoxon
- ✓Prueba de Kruskal-Wallis
- ✓Correlación de los rangos de Spearman.

Prueba de signos

La prueba de signo se utiliza para probar hipótesis sobre una mediana poblacional.

Al probar la hipótesis nula H_0 de que $\bar{\mu} = \bar{\mu}_0$ contra una alternativa apropiada, sobre la base de una muestra aleatoria de tamaño *n*, **reemplazamos cada valor de la muestra** que exceda $\bar{\mu}_0$ con un signo mas y cada valor de la muestra menor que $\bar{\mu}_0$ con un signo menos.

Hipótesis nula verdadera y distribución simétrica $\sum + \approx \sum - \approx$

Cuando un signo aparece con mas frecuencia de lo que debería, con base solo en el azar rechazamos la Hipótesis de que la mediana poblacional $\bar{\mu}$ igual a $\bar{\mu}_0$

Caso de muestras grandes $H_0 : p=0.5$

Media: $\mu = 0.5n$

Desviación estándar: $\sigma = \sqrt{0.25n}$

Prueba estadística no paramétrica para identificar diferencias entre dos poblaciones con base en el análisis de datos nominales.

En un sondeo realizado durante una campaña para elecciones presidenciales se pidió a 200 votantes registrados que evaluaran a los candidatos demócrata y republicano con relación a su política exterior. El resultado obtenido fue: 72 de los encuestados evaluaron mejor al candidato demócrata, 103 evaluaron mejor al republicano y 25 no encontraron diferencia entre los candidatos. ¿Con este sondeo puede observarse que exista una diferencia significativa, entre los candidatos, en términos de la opinión pública acerca de su política exterior?

n= 200 -25= 175

$$\mu = 0.5n$$

$$\mu = 0.5 \times 175 \Rightarrow \mu = 87.5$$

$$\sigma = \sqrt{0.25n}$$

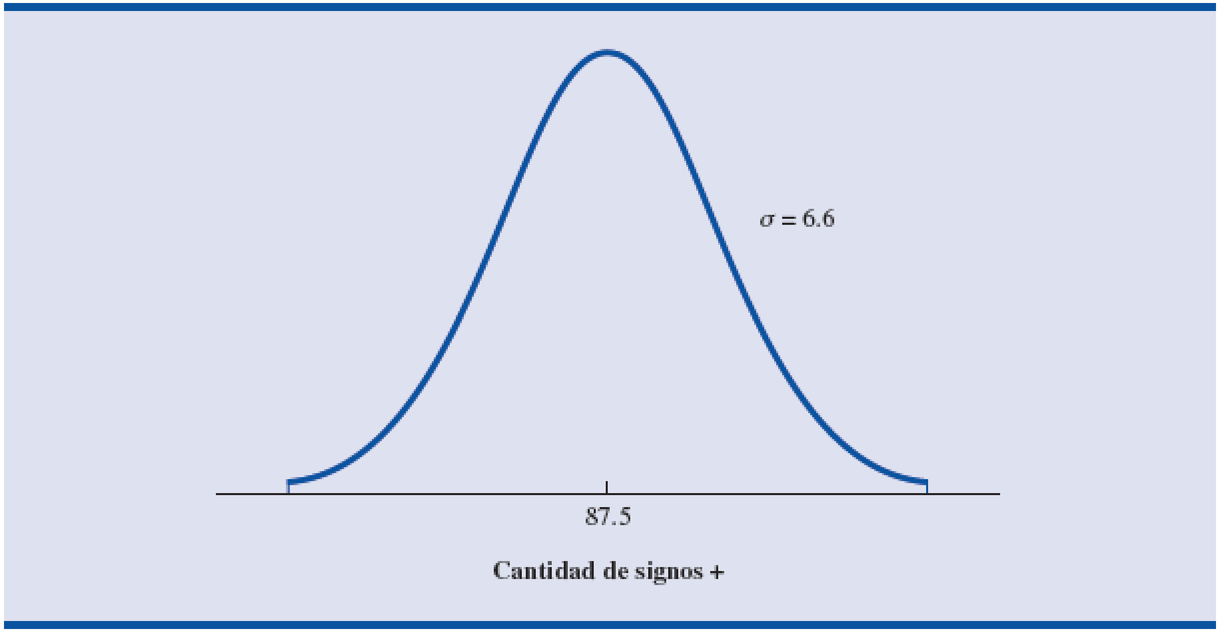
$$\sigma = \sqrt{0.25 \times 175} \Rightarrow \sigma = 6.6$$

Con base en el número de signos más (*x=72) que corresponden* al número de personas que evaluaron como mejor la política exterior del candidato demócrata

estadístico de prueba

$$z = \frac{x - \mu_T}{\sigma_T} \Rightarrow z = \frac{72 - 87.5}{6.6} \Rightarrow z = -2.35$$

DISTRIBUCIÓN DE PROBABILIDAD PARA EL NÚMERO DE SIGNOS MÁS EN UNA PRUEBA DE LOS SIGNOS EN LA QUE ***n =175***

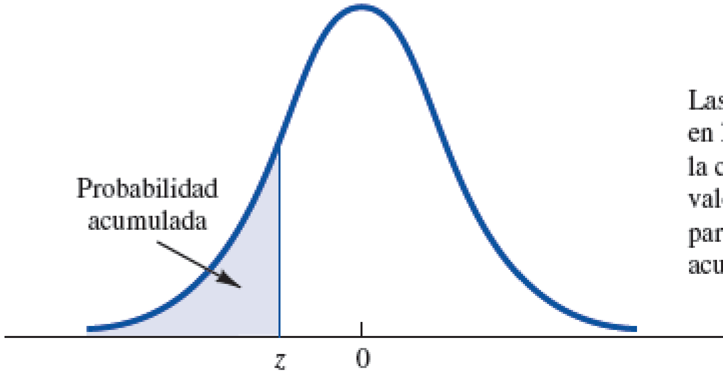


En las tablas de probabilidad normal estándar se encuentra que el **área en la cola, a la izquierda** de $z = -2.35$ *es 0.0094.*

Como se trata de una prueba de dos colas, el *valor-p = 2(0.0094)* = 0.0188.

valor-$p < \alpha$ $0.0188 < 0.05$	se rechaza H_0

Como resultado de este estudio se encuentra que los candidatos difieren en términos de la opinión pública acerca de su política exterior.



Las entradas que aparecen en la tabla dan el área bajo la curva y a la izquierda del valor de z. Por ejemplo, para $z = -0.85$, la probabilidad acumulada es 0.1977.

2.3+0.05=2.35

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

En cualquier libro de estadística, estas tablas se encuentran al final de los mismos. Para nuestros ejemplos he tomado una fotografía de la región de interés.

Los siguientes datos representan el numero de horas que un compensador opera antes de requerir una recarga: 1.5 2.2,0.9,1.3,2.0,1.6,1.8,1.5,2.0, 1.2, y 1.7. Utilice la prueba de signos para probar la hipótesis en el nivel de significancia de 0.05 que este compensador particular opera con una mediana de 1.8 horas antes de requerir una recarga.

Hipótesis nula: H_0 $\quad \bar{\mu} = 1.8$

Hipótesis alternativa: H_1 $\quad \bar{\mu} \neq 1.8$

Nivel de confianza: α $\quad 0.05$

Estadística de la Prueba: Variable binomial *X* con *p=1/2*

	1.5-1.8 -0.3	**2.2-1.8 0.4**	**0.9-1.8 -0.9**	**1.3-1.8 -0.5**	**2.0-1.8 0.2**	**1.6-1.8 -0.2**	**1.5-1.8 -0.3**	**2.0-1.8 0.2**	**1.2-1.8 -0.6**	**1.7-1.8 -0.1**
Secuencia de signos	-	+	-	-	+	-	-	+	-	-

$$n = 10 \qquad x = 3$$ **(Signos positivos)**

$$P\left(X \leq 3, p = \frac{1}{2}\right) = 2\sum_{x=0}^{3} b\left(x;10;\frac{1}{2}\right) = 2 \times 0.1719 = 0.3438$$

$$\sum_{x=0}^{3} b\left(x;10;\frac{1}{2}\right) = 0.1719$$

$$P = 0.3438 > 0.05$$

No rechazar la hipótesis nula **(H_0)** y concluir que el tiempo mediano de operación no es significativamente diferente de 1.8 horas

PROBABILIDADES BINOMIALES (*continuación*)

		p								
n	*x*	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
9	0	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.3874	0.3679	0.3020	0.2253	0.1556	0.1004	0.0605	0.0339	0.0176
	2	0.1722	0.2597	0.3020	0.3003	0.2668	0.2162	0.1612	0.1110	0.0703
	3	0.0446	0.1069	0.1762	0.2336	0.2668	0.2716	0.2508	0.2119	0.1641
	4	0.0074	0.0283	0.0661	0.1168	0.1715	0.2194	0.2508	0.2600	0.2461
	5	0.0008	0.0050	0.0165	0.0389	0.0735	0.1181	0.1672	0.2128	0.2461
	6	0.0001	0.0006	0.0028	0.0087	0.0210	0.0424	0.0743	0.1160	0.1641
	7	0.0000	0.0000	0.0003	0.0012	0.0039	0.0098	0.0212	0.0407	0.0703
	8	0.0000	0.0000	0.0000	0.0001	0.0004	0.0013	0.0035	0.0083	0.0176
	9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0008	0.0020
10	0	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.3874	0.3474	0.2684	0.1877	0.1211	0.0725	0.0403	0.0207	0.0098
	2	0.1937	0.2759	0.3020	0.2816	0.2335	0.1757	0.1209	0.0763	0.0439
	3	0.0574	0.1298	0.2013	0.2503	0.2668	0.2522	0.2150	0.1665	0.1172
	4	0.0112	0.0401	0.0881	0.1460	0.2001	0.2377	0.2508	0.2384	0.2051
	5	0.0015	0.0085	0.0264	0.0584	0.1029	0.1536	0.2007	0.2340	0.2461
	6	0.0001	0.0012	0.0055	0.0162	0.0368	0.0689	0.1115	0.1596	0.2051
	7	0.0000	0.0001	0.0008	0.0031	0.0090	0.0212	0.0425	0.0746	0.1172
	8	0.0000	0.0000	0.0001	0.0004	0.0014	0.0043	0.0106	0.0229	0.0439
	9	0.0000	0.0000	0.0000	0.0000	0.0001	0.0005	0.0016	0.0042	0.0098
	10	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0010

$\sum$ 0.1719

En cualquier libro de estadística, estas tablas se encuentran al final de los mismos. Para nuestros ejemplos he tomado una fotografía de la región de interés.

Prueba de los rangos con signo de Wilcoxon

Una prueba que utiliza dirección y magnitud

La prueba de rango de signo de Wilcoxon se aplica en el caso de una distribución continua simétrica. Se Debe probar la hipótesis nula $\mu \cong \mu_0$

1. Restamos de cada valor muestral y descartamos todas las diferencias iguales a cero.
2. Se clasifican las diferencias restantes sin importar el signo.
3. Se asigna un rango de **1** a la diferencia absoluta mas pequeña (es decir, sin signo), un rango de 2 a la siguiente mas pequeña, y así sucesivamente.
4. Cuando el valor absoluto de dos o mas diferencias es el mismo, se asigna a cada uno el promedio de los rangos que se asignarían si las diferencias se distinguieran.

DISTRIBUCIÓN MUESTRAL DE *T PARA POBLACIONES IDÉNTICAS*

Media: $\mu_T = 0$

Desviación estándar: $\sigma_T = \sqrt{\dfrac{n(n+1)(2n+1)}{6}}$

Prueba estadística no paramétrica para identificar diferencias entre dos poblaciones con base en el análisis de dos muestras pareadas.

Para probar la hipótesis nula de que se muestrean dos poblaciones simétricas para el caso de una muestra pareada, clasificamos las diferencias de las observaciones pareadas sin importar el signo, y procedemos como en el caso de una sola muestra. Los diversos procedimientos de prueba para los casos de una sola muestra y de una muestra pareada se resumen a continuación:

$$\mu = \mu_0 \quad \begin{cases} \mu < \mu_0 & w_+ \\ \mu > \mu_0 & w_- \\ \mu \neq \mu_0 & w \end{cases}$$

$$\mu_1 = \mu_2 \quad \begin{cases} \mu_1 < \mu_2 & w_+ \\ \mu_1 > \mu_2 & w_- \\ \mu_1 \neq \mu_2 & w \end{cases}$$

La hipótesis nula se rechaza si el valor **calculado**, W_+ , W_- o W **es menor o igual que** el valor **tabulado** apropiado (TABLAS).

Valores críticos para la prueba de rangos con signo

n	Unilateral $\alpha = 0.01$ Bilateral $\alpha = 0.02$	Unilateral $\alpha = 0.025$ Bilateral $\alpha = 0.05$	Unilateral $\alpha = 0.05$ Bilateral $\alpha = 0.10$
5			1
6		1	2
7	0	2	4
8	2	4	6
9	3	6	8
10	5	8	11
11	7	11	14
12	10	14	17
13	13	17	21
14	16	21	26
15	20	25	30
16	24	30	36
17	28	35	41
18	33	40	47
19	38	46	54
20	43	52	60
21	49	59	68
22	56	66	75
23	62	73	83
24	69	81	92
25	77	90	101
26	85	98	110
27	93	107	120
28	102	117	130
29	111	127	141
30	120	137	152

En cualquier libro de estadística, estas tablas se encuentran al final de los mismos. Para nuestros ejemplos he tomado una fotografía de la región de interés.

Los siguientes datos representan el numero de horas que un compensador opera antes de requerir una recarga: 1.5 2.2,0.9,1.3,2.0,1.6,1.8,1.5,2.0, 1.2, y 1.7. Utilice la prueba de signos para probar la hipótesis en el nivel de significancia de 0.05 que este compensador particular **opera con una mediana de 1.8 horas antes de requerir una recarga**.

Hipótesis nula: H_0 $\quad \bar{\mu} = 1.8$

n=10 (Se descarta 1.8)

Hipótesis alternativa: H_1 $\quad \bar{\mu} \neq 1.8$

Nivel de confianza: α $\quad 0.05$

$$P(X \geq r) = \sum_{x=r}^{n} b(x; n; p)$$

$$\bar{\mu} = 1.8$$

1.5-1.8 -0.3	2.2-1.8 +0.4	0.9-1.8 -0.9	1.3-1.8 -0.5	2.0-1.8 +0.2	1.6-1.8 -0.2	1.5-1.8 -0.3	2.0-1.8 +0.2	1.2-1.8 -0.6	1.7-1.8 -0.1
5.5	7	**10**	8	3	3	5.5	3	9	**1**

Ordenar las diferencias sin importar el signo

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0.1	0.2	0.2	0.2	0.3	0.3	0.4	0.5	0.6	0.9

$w_+ = 13$ y $w_- = 42$ $\qquad w = 13$ Mas pequeño entre los w_+ y w_-

$$\boxed{w_{cal} = 13 > 11}$$

No rechazar la hipótesis nula (**H_0**) y concluir que el tiempo promedio de operación no es significativamente diferente de 1.8 horas