Pruebas noparametricas para dos poblaciones

Estadistica no parametrica

Humberto Vaquera Huerta
2024-05-14

Contenido:

Comparacion de la localidad de dos poblaciones no relacionadas

- **Prueba U de Mann-Withney**

Comparacion de la variabilidad de dos poblaciones no relacionadas

- **Prueba de Ansari- Bradley**

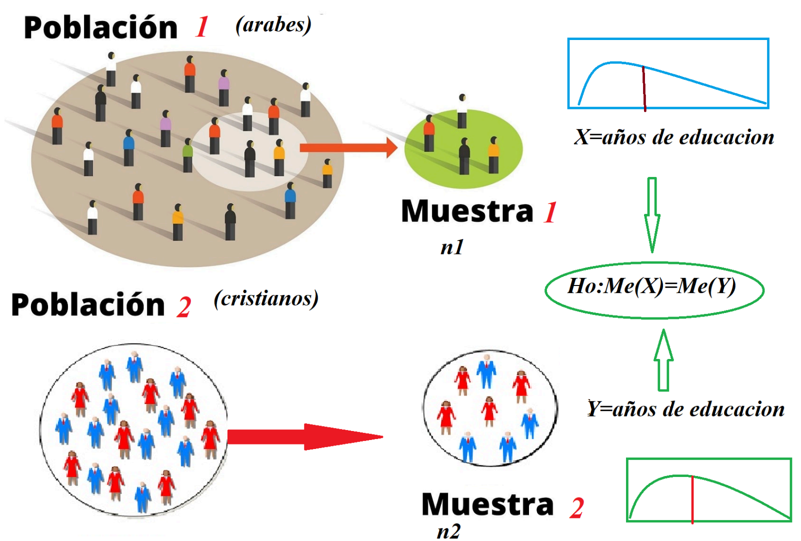
Comparacion de la distribución de dos poblaciones no relacionadas

- **Prueba de Kolmogorov-Smirnov para comparar dos poblaciones**

Comparacion de dos poblaciones no relacionadas

La investigación implica a menudo la comparación de dos o más muestras de diferentes poblaciones.

Con datos no pareados cuando obtienes un valor de una variable de cada individuo en dos grupos no relacionados. Los dos grupos se conocen como muestras independientes.



Comparacion de dos poblaciones independientes

Las muestras independientes se obtienen cuando se toman dos muestras de forma completamente independiente, o si se toma una muestra y luego se asignan aleatoriamente los miembros a los grupos. La aleatorización siempre te da muestras independientes.

Ejemplo: Diseño experimental y animales.

Se utilizó un diseño completamente al azar para estudiar las características de comportamiento en 24 cerdos m mestizos comerciales (Yorkshire x Landrace x Duroc), machos castrados y con un peso medio inicial de 26 kg. Los animales pertenecían al rebaño del CP. Estos cerdos se distribuyeron aleatoriamente en dos tratamientos consistentes en dietas con 0 y 30 % de harina de nuez de palma, en base seca.

Indicator	Royal palm nut meal, %		SE ±
	0	30	
Weight , kg			
Initial	26.5	26.7	0.46
Final	73.8	71.6	1.61
Intake, kg DM/day[1]	2.42	2.43	-
Gain, g/day	706	680	31
Conversion,			
kg DM/kg gain	3.43	3.57	0.18
kg protein/kg gain	0.501	0.435	0.060

Ejemplo: No bebas de la taza azul

En un artículo de noviembre de 2014 en la revista *Flavour*, investigadores de la Universidad de Oxford en Inglaterra investigaron si el aroma de una taza de café podría verse afectado por el color de la taza en la que lo beben. En el experimento, 12 personas fueron seleccionadas al azar para beber su café de una taza con funda azul y 12 fueron seleccionadas al azar para beber de una taza con funda blanca. Se pidió a los sujetos que calificaran subjetivamente el aroma del café en una escala de (0-100) puntos. El café de las tazas con funda blanca recibió una calificación promedio de 57.33 con una desviación estándar de 16.27, mientras que el café de las tazas con funda azul recibió una calificación promedio de 35,57 con una desviación estándar de 25,34.

Ejemplo: Calidad del agua

Datos históricos de calidad del agua subterránea para un acuífero poco profundo subyacente a tierras agrícolas muestra las siguientes concentraciones de nitrato (mg/L):

pre-1970				**post-1970**			
	1	2	4		1	5	14
	1	3	5		2	8	15
	1	3	5		2	10	18
	2	4	10		4	11	23

Pruebas para comparar dos poblaciones independientes

Pruebas de ubicacion.

Prueba U de Mann-Withney

La prueba U de Mann-Whitney se puede utilizar para comprobar si existe una diferencia entre dos grupos y no es necesario que los datos se distribuyan normalmente.

Para determinar esta diferencia, se utilizan las **sumas de rango de los dos grupos** en lugar de las medias como en la prueba t para muestras independientes.

Supuestos de la Prueba U de Mann-Whitney

- Tenemos dos muestras aleatorias independientes
- con al menos variables en escala ordinal.

Hipótesis nula: No hay diferencia entre los dos grupos de la población.

Hipótesis alternativa: Hay una diferencia entre los dos grupos de la población.

Prueba U de Mann-Withney

Supongamos que tenemos una muestra de n_1 observaciones $x_1, x_2, \ldots, x_{n_1}$ en un grupo (es decir, de una población) y una muestra de n_2 observaciones $y_1, y_2, \ldots y_{n_2}$ en otro grupo (es decir, de otra población). La prueba de Mann-Whitney se basa en una comparación de cada observación x_i en la primer muestra con cada observación y_j en la otra muestra. El número total de pares de comparaciones que se pueden hacer es $n_1 n_2$. Si las muestras tienen la misma mediana, entonces cada x_i tiene la misma oportunidad (es decir, probabilidad $1/2$) de ser mayor o menor que cada y_j.

- Entonces, bajo la hipótesis nula $H_0 : P(x_i > y_j) = 1/2$
- y bajo la hipótesis alternativa $H_1 : P(x_i > y_j) \neq 1/2$

Prueba U de Mann-Withney

- Prueba de suma de rango de dos muestras de Wilcoxon
  - Muestra combinada de $N = n_1 + n_2$.
  - Sea S_n el rango de Y_n en este orden conjunto.
  - Estadística de suma de rango de dos muestras de Wilcoxon $W = \sum_{j=1}^{n} S_j$.
- Bajo H_0, todas las asignaciones posibles de $\binom{N}{n}$ para los rangos Y son igualmente probables, cada una con probabilidad $\frac{1}{\binom{N}{n}}$.

Estadistico de Mann−Whitney

$U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi\left(X_i, Y_j\right)$, donde $\phi\left(X_i, Y_j\right) = 1\ if\ X_i < Y_j$.

- Pruebas con W y U son equivalentes.
  - $W = U + \dfrac{n_1(n_1+1)}{2}$.

Prueba de de Mann−Whitney

THE MANN-WHITNEY TEST

Sample 1	Sample 2
y_{11}	y_{21}
y_{12}	y_{22}
...	...
y_{1n1}	y_{1n2}

Sort the n = n1 + n2 values

Replace y_{ij} by its overall rank r_{ij}

Sample 1	Sample 2
3 (r_{11})	5 (r_{21})
1 (r_{12})	2 (r_{22})
...	...
6 (r_{1n1})	9 (r_{2n2})

$R_1 = \sum_{j=1}^{n_1} r_{1j}$ and $R_2 = \sum_{j=1}^{n_2} r_{2j}$ (sums of ranks)

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \qquad \left(\text{or } \; U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \, ; \;\; U = n_1 n_2 - U' \right)$$

Table A5.07: Critical Values for the Wilcoxon/Mann-Whitney Test (U)

Nondirectional α=.05 (Directional α=.025)

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	-	0	2	4	7	10	12	15	17	21	23	26	28	31	34	37	39	42	45	48
10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	-	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	-	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	-	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	-	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	-	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	-	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	-	2	8	14	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

Nondirectional α=.01 (Directional α=.005)

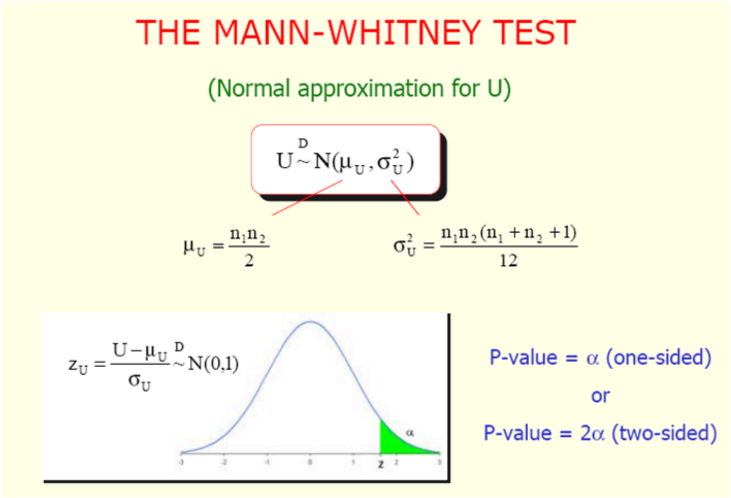
n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0
3	-	-	-	-	-	-	-	-	0	0	0	1	1	1	2	2	2	2	3	3
4	-	-	-			0	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	-	-	-		0	1	1	2	3	4	5	6	7	7	8	9	10	11	12	13
6	-	-	-	0	1	2	3	4	5	6	7	9	10	11	12	13	15	16	17	18
7	-	-	-	0	1	3	4	6	7	9	10	12	13	15	16	18	19	21	22	24
8	-	-	-	1	2	4	6	7	9	11	13	15	17	18	20	22	24	26	28	30
9	-	-	0	1	3	5	7	9	11	13	16	18	20	22	24	27	29	31	33	36
10	-	-	0	2	4	6	9	11	13	16	18	21	24	26	29	31	34	37	39	42
11	-	-	0	2	5	7	10	13	16	18	21	24	27	30	33	36	39	42	45	46
12	-	-	1	3	6	9	12	15	18	21	24	27	31	34	37	41	44	47	51	54
13	-	-	1	3	7	10	13	17	20	24	27	31	34	38	42	45	49	53	56	60
14	-	-	1	4	7	11	15	18	22	26	30	34	38	42	46	50	54	58	63	67
15	-	-	2	5	8	12	16	20	24	29	33	37	42	46	51	55	60	64	69	73
16	-	-	2	5	9	13	18	22	27	31	36	41	45	50	55	60	65	70	74	79
17	-	-	2	6	10	15	19	24	29	34	39	44	49	54	60	65	70	75	81	86
18	-	-	2	6	11	16	21	26	31	37	42	47	53	58	64	70	75	81	87	92
19	-	0	3	7	12	17	22	28	33	39	45	51	56	63	69	74	81	87	93	99
20	-	0	3	8	13	18	24	30	36	42	46	54	60	67	73	79	86	92	99	105

U_{obs} is the lesser of the two calculated test statistics (U_1 & U_2). If $U_{obs} \le U_{crit}$, reject H_0.
Dashes (-) indicate that the sample size is too small to reject the Null Hypothesis at the chosen α level.

If n > 20 this table cannot be used. A p can be computed for U_{obs}, using the normal distribution approximation:

$$z_U = \frac{\left|U_{obs} - \left(\frac{n_1 n_2}{2}\right)\right|}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Prueba de de Mann−Whitney aprox normal



Estadística de prueba: la versión estandarizada de W.

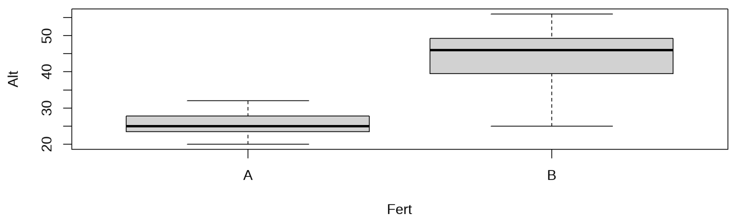
$$Z = \frac{W - \{n_2\,(n_2 + n_2 + 1)\,/2\}}{\{n_2 n_1\,(n_2 + n_1 + 1)\,/12\}^{1/2}} \sim \mathrm{N}\,(0, 1)$$

.

- Si hay empates
  - Utilizar la media de los rangos.
  - no es una prueba exacta.

Ejemplo: Prueba de Mann–Whitney

Se diseña un experimento en el que 15 semillas se asignan al azar a uno de dos fertilizantes y la altura de la planta resultante se mide después de dos semanas. Quiere saber si uno de los fertilizantes proporciona más crecimiento vertical que el otro.



Altura de plantas en cm.

Fert A	Fert B
20	25
23	46
32	56
24	45
25	46
28	51
27.5	34
	47.5

Hipótesis a probar

- H_0 : Mediana A=Mediana B.
- H_1 : Mediana A $\neq$ Mediana B

Calculos

Fert	Altura	Rango	rangoA	rangoB	n	n1	n2
A	20	1	*1*		15	7	8
A	23	2	*2*				
A	24	3	*3*				
A	25	4.5	*4.5*				
B	25	4.5		*4.5*			
A	27.5	6	*6*				
A	28	7	*7*				
A	32	8	*8*				
B	34	9		*9*			
B	45	10		*10*			
B	46	11.5		*11.5*			
B	46	11.5		*11.5*			
B	47.5	13		*13*			
B	51	14		*14*			
B	56	15		*15*			
		120	**31.5**	**88.5**			

$$R_1 = 31.5$$

$$R_2 = 88.5$$

$$n_1 = 7,\ n_2 = 8$$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= 7 \times 8 + \frac{7(7+1)}{2} - 31.5$$

$$= 52.5$$

$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$= 7 \times 8 + \frac{8(8+1)}{2} - 88.5$$

$$= 3.5$$

Ejemplo: Prueba de Mann–Whitney

Es estadistico de prueba es el valor mas pequeño de $W = Minimo(U, U')$ en este caso $U' = 3.5$ Valor de tablas de Mann Withney con $alpha = 0.05$ y $n_1 = 7,\ n_2 = 8$ es 10 Como $W = 3.5 < 10$ rechazo H_0

Prueba de de Mann−Whitney aprox normal

Cuando $n > 50$

$\mu_W = \frac{n_1 n_2}{2}$, $\sigma_W^2 = \frac{n_1 n_2 (n_1+n_2+1)}{12}$

$$Z_c = \frac{W - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}}$$

$$Z_c = \frac{3.5 - \frac{7\times 8}{2}}{\sqrt{\frac{7\times 8(7+8+1)}{12}}} = -2.835$$

Usando un $\alpha = 0.05$ el valor de las tablas de la normal estandar es N

```
Ztablas=qnorm(.975,0,1); Pvalue=2*pnorm(-2,835)
Ztablas
```

```
## [1] 1.959964
```

```
Pvalue
```

Prueba de de Mann−Whitney en R

```
Fert=c("A","A","A","A","A","A","A","B","B","B","B","B","B","B","B")
Alt=c(20,23,32,24,25,28,27.5,25,46,56,45,46,51,34,47.5)
wilcox.test (Alt~Fert)
```

```
##
##      Wilcoxon rank sum test with continuity correction
##
## data:  Alt by Fert
## W = 3.5, p-value = 0.005395
## alternative hypothesis: true location shift is not equal to 0
```

Prueba de de Mann–Whitney en SAS

```
DATA FERTI;
INPUT FERT $ ALT;
CARDS;
A 20
A 23
A 32
A 24
A 25
A 28.
A 27.5
B 25
B 46
B 56
B 45
B 46
B 51
B 34
B 47.5
;
/* Mann Whitney U test*/
proc npar1way data=FERTI wilcoxon;
    class FERT;
    var ALT;
run;
```

Prueba de de Mann–Whitney en SAS

Procedimineto NPAR1WAY

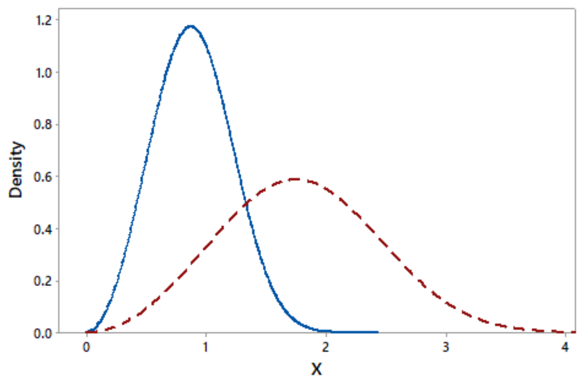
Puntuaciones de Wilcoxon (Sumas de rango) para variable ALT Clasificado por variable FERT					
FERT	N	Suma de puntuaciones	Esperado debajo de H0	Desv. est. debajo de H0	Puntuación media
A	7	31.50	56.0	8.625543	4.50000
B	8	88.50	64.0	8.625543	11.06250
Se utilizaron puntuaciones media para valores repetidos.					

Test de dos muestras de Wilcoxon					
				Aproximación t	
Estadístico	Z	Pr < Z	Pr > |Z|	Pr < Z	Pr > |Z|
31.5000	-2.7824	0.0027	0.0054	0.0073	0.0147
Z incluye una corrección de continuidad de 0.5.					

Test de Kruskal-Wallis		
Chi-cuadrado	DF	Pr > ChiSq
8.0679	1	0.0045

Pruebas de dispersion entre dos poblaciones

En algunos estudios es importante detectar diferencias en variabilidad o dispersión en lugar de ubicación. De hecho, la comparación de la variabilidad puede ser de interés en muchas áreas, incluidas las ciencias sociales, la biología, Agronomia, la ingeniería, la industria, etc. Además, a menudo se requieren pruebas para la igualdad de las varianzas como herramienta preliminar para el análisis de la varianza. (ANOVA), modelado dosis-respuesta, análisis discriminante, etc. La prueba de Bartlett y la prueba de Levene son pruebas parametricas de uso cotidiano.

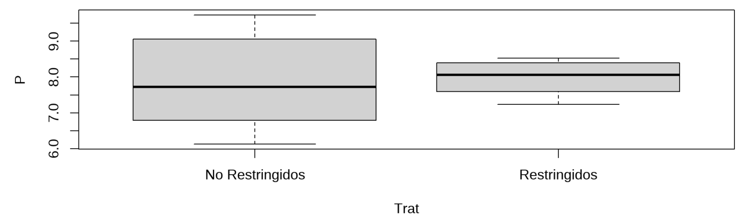


Ejemplo: Metabolitos de Fosforo en Sangre

A los fines de determinar los efectos de la restricción alimentaria en la químicasanguínea de vacunos se midieron los metabolitos Fósforo (P) en sangre. El experimento se realizó tomando un lote de novillos de 180 kilos de peso promedio. De ellos, se eligieron aleatoriamente 10 para constituir el lote control (no restringidos) que eran alimentados con centeno a voluntad. El otro lote (restringidos) se conformó por los 10 animales restantes. La restricción consistió de dejar los novillos en pastoreo por 3 horas por día y luego pasarlos a corral pelado.

No Restringidos	Restringidos
8.69	7.24
6.13	7.46
6.79	7.59
6.79	7.73
6.93	7.86
7.59	8.26
7.86	8.39
9.06	8.39
9.59	8.53
9.73	8.53

Julio Alejandro Di Rienzo et al. 2009, Estadística para la ciencias agropecuarias, 7a ed.,372 p ISBN 978-987-591-112-3



Hipotesis de interes

- H_0 : Varibilidad NR=Variabilidad R.
- H_1 : Varibilidad NR $\neq$ Varibilidad R

Prueba de Ansari- Bradley

Supuestos

> Esta prueba requiere que las muestras tengan medianas iguales.
>
> las distribuciones de las muestras son continuas e idénticas, la prueba es independiente de las distribuciones.

Si las muestras no tienen las mismas medianas, los resultados pueden ser engañosos. Si desea realizar las pruebas restando las medianas, debe restar las medianas de x e y antes de usar Ansari- Bradley.

Hipotesis

- hipótesis nula H_0 : Los datos en los vectores x e y provienen de la misma distribución.
- La hipótesis alternativa H_1: Los datos en x e y provienen de distribuciones con la misma mediana y forma pero diferentes dispersiones (por ejemplo, varianzas).

Estadística de prueba de Ansari-Bardley

Calculo de Rangos de Ansari-Bardley

$$R_{AB_i} = \frac{n+1}{2} - |R_i - \frac{n+1}{2}|$$

La estadística de Ansari-Bradley C se define como:

$$C = \sum_{i=1}^{n_2} R_{AB_i}$$

La suma de rangos de Ansari del grupo con la menor suma.

El valor C se compara con el valor de tablas de Ansari-Bradley

Tablas para Ansari-Bradley

TABLE 1

Lower and Upper Significance Levels of W

Sample Sizes		Significance Levels							
m	n	.995	.99	.975	.95	.05	.025	.01	.005
2	5	—	—	—	2	—	—	—	—
2	6	—	—	—	2	8	—	—	—
2	7	—	—	—	2	9	—	—	—
2	8	—	—	2	2	10	10	—	—
2	9	—	—	2	2	11	11	—	—
2	10	—	—	2	2	12	12	—	—
2	11	—	—	2	2	13	13	—	—
2	12	—	—	2	2	14	14	—	—
2	13	—	2	2	3	14	15	—	—
2	14	—	2	2	3	15	16	16	—
2	15	—	2	2	3	16	17	—	—
2	16	—	2	2	3	17	17	18	—
2	17	—	2	2	3	18	19	—	—
2	18	—	2	2	3	19	19	20	—
3	5	—	—	—	4	11	—	—	—
3	6	—	—	4	4	13	13	—	—
3	7	—	—	4	5	13	14	—	—
3	8	—	—	4	5	15	16	16	—
3	9	—	4	4	5	16	17	17	—
3	10	—	4	5	5	17	18	18	19
3	11	—	4	5	6	18	19	20	—
3	12	4	4	5	6	20	21	22	22
3	13	4	4	5	6	21	22	23	23
3	14	4	5	6	7	22	23	24	25
3	15	4	5	6	7	23	24	25	26
3	16	4	5	6	7	24	25	27	28
3	17	4	5	6	8	25	26	28	29
4	4	—	—	6	6	14	14	—	—
4	5	—	6	6	7	14	16	—	—
4	6	6	6	7	7	17	17	18	18
4	7	6	6	7	8	19	19	20	—
4	8	6	6	7	8	20	21	22	22
4	9	6	7	8	9	21	22	23	24
4	10	7	7	8	9	23	24	25	25
4	11	7	7	9	10	24	26	27	27
4	12	7	8	9	10	26	27	28	29
4	13	7	8	9	11	27	29	30	31
4	14	8	9	10	11	29	30	31	32
4	15	8	9	10	12	30	32	33	34
4	16	8	9	11	12	32	33	35	36
5	5	—	9	10	10	20	20	21	—
5	6	9	9	10	11	22	23	24	24
5	7	9	10	11	11	24	24	25	26
5	8	10	10	11	12	26	26	28	29
5	9	10	11	12	13	27	28	29	30
5	10	10	11	12	14	29	30	32	32
5	11	11	12	13	14	31	32	33	34
5	12	11	12	14	15	33	34	36	37
5	13	12	13	14	16	34	36	37	38
5	14	12	13	15	16	36	38	40	41
5	15	12	14	15	17	38	40	41	43
6	6	12	13	14	15	27	28	29	30
6	7	13	14	15	16	29	30	32	32
6	8	14	14	16	17	31	32	34	34
6	9	14	15	16	18	34	35	36	37
6	10	15	16	17	18	36	37	38	39
6	11	15	16	18	19	38	40	41	42
6	12	16	17	19	20	40	41	43	44
6	13	16	18	19	21	42	44	46	47
6	14	17	18	20	22	44	46	48	49
7	7	17	18	19	21	35	37	38	39
7	8	18	19	20	22	38	39	41	42
7	9	19	20	21	23	40	42	43	44
7	10	20	21	22	24	43	44	46	47
7	11	20	22	23	25	45	47	48	50
8	8	23	24	26	27	45	46	48	49
8	9	24	25	27	29	48	49	51	52
8	10	25	26	28	30	50	52	54	55
8	11	26	27	29	31	53	55	57	58
8	12	27	28	30	32	56	58	60	61
9	9	30	31	33	35	55	57	59	60
9	10	31	32	34	36	58	60	62	64
9	11	32	34	36	38	61	63	65	67
10	10	38	39	41	43	67	69	71	72

A. R. Ansari. R. A. Bradley. "Rank-Sum Tests for Dispersions." Ann. Math. Statist. 31 (4) 1174 - 1189, December, 1960. https://doi.org/10.1214/aoms/1177705688

Aplicacion de Prueba de Ansari-Bradley

P	Trat	Ri	RABi	n	Suma RABi NR	Suma RABi R
6.13	No Restringidos	1	1	20	73	38
6.79	No Restringidos	2.5	2.5			
6.79	No Restringidos	2.5	2.5			
6.93	No Restringidos	4	4			
7.24	Restringidos	5	5			
7.46	Restringidos	6	6			
7.59	No Restringidos	7.5	7.5			
7.59	Restringidos	7.5	7.5			
7.73	Restringidos	9	9			
7.86	No Restringidos	10.5	10.5			
7.86	Restringidos	10.5	10.5			
8.26	Restringidos	12	9			
8.39	Restringidos	13.5	7.5			
8.39	Restringidos	13.5	7.5			
8.53	Restringidos	15.5	5.5			
8.53	Restringidos	15.5	5.5			
8.69	No Restringidos	17	4			
9.06	No Restringidos	18	3			
9.59	No Restringidos	19	2			
9.73	No Restringidos	20	1			

$R_{AB_i} = \frac{20+1}{2} - |R_i - \frac{20+1}{2}|$ con $n = 20$. $C = \sum_{i=1}^{n_1} R_{AB_i} = 38$

De tablas de Ansari y Bradley con $\alpha = 95, n_1 = 10, n_2 = 10$ el valor es 43 esto implica qie Rechazamos H_0 las variablidades no son iguales

Aplicacion de Prueba de Ansari-Bradley a datos con R

$$H_0 : \delta = \frac{var(NR)}{var(R)} = 1,\ vs\ H_1 : \delta = \frac{var(NR)}{var(R)} \neq 1$$

```
ansari.test(P~Trat,data=fosforo)
```

```
##
##      Ansari-Bradley test
##
## data:  P by Trat
## AB = 38, p-value = 0.01443
## alternative hypothesis: true ratio of scales is not equal to 1
```

Aplicacion de Prueba de Ansari-Bradley a datos con SAS

```
DATA FOSFORO;
INPUT P TRAT$ @@;
CARDS;
6.13 No_Restringidos 7.24 Restringidos
6.79 No_Restringidos 7.46 Restringidos
6.79 No_Restringidos 7.59 Restringidos
6.93 No_Restringidos 7.73 Restringidos
7.59 No_Restringidos 7.86 Restringidos
7.86 No_Restringidos 8.26 Restringidos
8.69 No_Restringidos 8.39 Restringidos
9.06 No_Restringidos 8.39 Restringidos
9.59 No_Restringidos 8.53 Restringidos
9.73 No_Restringidos 8.53 Restringidos
;
/* Ansari Bradley*/
proc npar1way data=FOSFORO ab;
class TRAT;
var P;
exact;
run;
```

Aplicacion de Prueba de Ansari-Bradley a datos con SAS

Puntuaciones de Ansari-Bradley para variable P Clasificado por variable TRAT					
TRAT	N	Suma de puntuaciones	Esperado debajo de H0	Desv. est. debajo de H0	Puntuación media
No_Restr	10	37.50	55.0	6.549407	3.750
Restring	10	72.50	55.0	6.549407	7.250
Se utilizaron puntuaciones media para valores repetidos.					

Test de dos muestras de Ansari-Bradley					
				Exacto	
Statistic (S)	Z	Pr < Z	Pr > |Z|	Pr <= S	Pr >= |S-media|
37.5000	-2.6720	0.0038	0.0075	0.0032	0.0064

Análisis de un factor de Ansari-Bradley		
Chi-cuadrado	DF	Pr > ChiSq
7.1396	1	0.0075

Prueba de Kolmogorov-Smirnov para comparar dos poblaciones

La prueba de Kolmogorov-Smirnov es un procedimiento de prueba de hipótesis para determinar si dos muestras de datos pertenecen a la misma distribución. La prueba en no paramétrica y es completamente independiente de lo que realmente es esta distribución.

Prueba

- Hipótesis nula, $H_0 : F(x) = G(x)$, de que las dos muestras que estamos probando provienen de la misma distribución.
- Hipótesis nula $H_1 : F(x) \neq G(x)$, de que las dos muestras son de diferentes distribuciones.

Supuestos;

- Supongamos que tenemos una muestra de n_1 observaciones $x_1, x_2, \ldots, x_{n_1}$ en un grupo (es decir, de una población) y una muestra de n_2 observaciones $y_1, y_2, \ldots y_{n_2}$ en otro grupo (es decir, de otra

calcular el estadístico Z de la prueba bilateral de Kolmogorov-Smirnov.

- Encuentre las funciones de distribución empíricas $F_{n_1}(t)$ y $G_{n_2}(t)$ para las muestras de X y Y, respectivamente. Combine y clasifique ambos conjuntos de valores. Por cada número real t,

$$F_{n_1}(t) = \frac{\text{numero de X obs } < t}{n_1}$$

$$G_{n_2}(t) = \frac{\text{numero de Y obs } < t}{n_2}$$

Encuentre cada valor absoluto de divergencia D entre el Funciones de distribuciones empíricas:

$$D = |F_{n_1}(t) - G_{n_2}(t)|$$

Use la mayor divergencia $Dmax$ esel estadistico de prueba de Kolmogorov−Smirnov:

$$Dmax = Max|F_{n_1}(t) - G_{n_2}(t)|$$

Table gives critical D-values for $\alpha = 0.05$ (upper value) and $\alpha = 0.01$ (lower value) for various sample sizes. * means you cannot reject H_0 regardless of observed D.

$n_2 \backslash n_1$	3	4	5	6	7	8	9	10	11	12
1	* *	* *	* *	* *	* *	* *	* *	* *	* *	* *
2	* *	* *	* *	* *	* *	16/16 *	18/18 *	20/20 *	22/22 *	24/24 *
3	* *	* *	15/15 *	18/18 *	21/21 *	21/24 24/24	24/27 27/27	27/30 30/30	30/33 33/33	30/36 36/36
4		16/16 *	20/20 *	20/24 24/24	24/28 28/28	28/32 32/32	28/36 32/36	30/40 36/40	33/44 40/44	36/48 44/48
5			* *	24/30 30/30	30/35 35/35	30/40 35/40	35/45 40/45	40/50 45/50	39/55 45/55	43/60 50/60
6				30/36 36/36	30/42 36/42	34/48 40/48	39/54 45/54	40/60 48/60	43/66 54/66	48/72 60/72
7					42/49 42/49	40/56 48/56	42/63 49/63	46/70 53/70	48/77 59/77	53/84 60/84
8						48/64 56/64	46/72 55/72	48/80 60/80	53/88 64/88	60/96 68/96
9							54/81 63/81	53/90 70/90	59/99 70/99	63/108 75/108
10								70/100 80/100	60/110 77/110	66/120 80/120
11									77/121 88/121	72/132 86/132
12										96/144 84/144

For larger sample sizes, the approximate critical value D_α is given by the equation

$$D_\alpha = c(\alpha)\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

where the coefficient is given by the table below.

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Examples: (1) At $\alpha = 0.05$ and samples sizes 5 and 8, $D_\alpha = 30/40 = 0.75$.

(2) At $\alpha = 0.01$ and samples sizes 15 and 28, $D_\alpha = 1.63\sqrt{\frac{15+28}{15 \cdot 28}} = 0.522$.

```r
A=c(20,23,32,24,25,28,27.5)
B=c(25,46,56,45,46,51,34,47.5)
ks.test(A, B, alternative="two.sided")
```

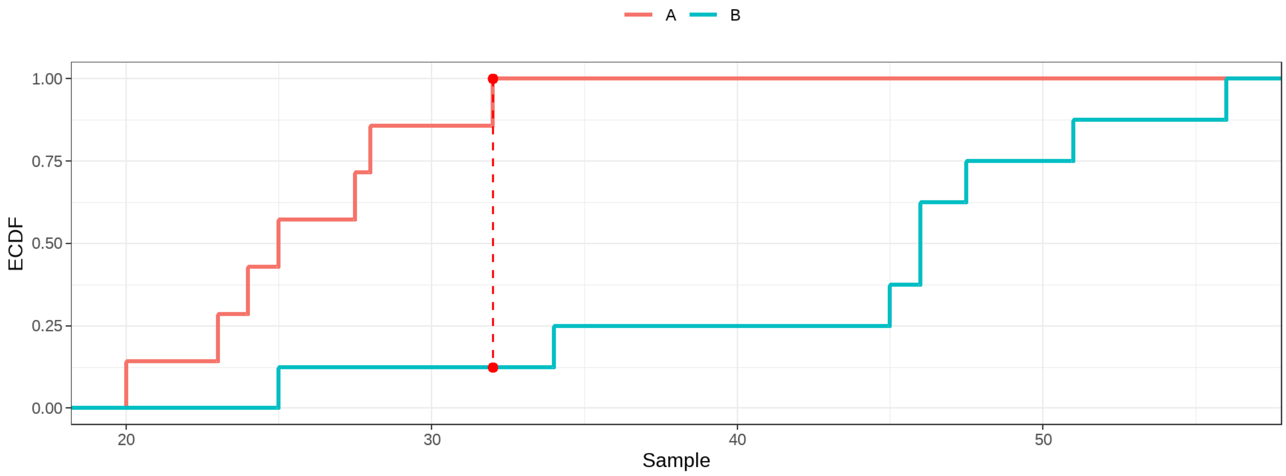
```
##
##      Exact two-sample Kolmogorov-Smirnov test
##
## data:  A and B
## D = 0.875, p-value = 0.002486
## alternative hypothesis: two-sided
```

```r
library(fBasics)
ks2Test(A,B)
```

```
##
## Title:
##  Kolmogorov-Smirnov Two Sample Test
##
## Test Results:
##   STATISTIC:
##     D | Two Sided: 0.875
##        D^- | Less: 0
##     D^+ | Greater: 0.875
##   P VALUE:
##     Alternative       Two-Sided: 0.002486
```

Prueba de Kolmogorov-Smirnov para comparar dos poblaciones

Grafica de funciones empiricas y Dmax



Prueba de Kolmogorov-Smirnov SAS

```
DATA FERTI;
INPUT FERT $ ALT;
CARDS;
A 20
A 23
A 32
A 24
A 25
A 28.
A 27.5
B 25
B 46
B 56
B 45
B 46
B 51
B 34
B 47.5
;
/* Kolmogorov Smirnov*/
proc npar1way data=FERTI edf plots=edfplot;
    class FERT;
    var ALT;
run;
```

Prueba de Kolmogorov-Smirnov SAS

Procedimiento NPAR1WAY

Test Kolmogorov-Smirnov para variable ALT Clasificado por variable FERT			
FERT	N	EDF al máximo	Desviación de la media al máximo
A	7	1.000000	1.234664
B	8	0.125000	-1.154941
Total	15	0.533333	
Desviación máxima producida en la observación 3			
Valor de ALT en Máximo = 32.0			

Test de dos muestras Kolmogorov-Smirnov (Asintótico)			
KS	0.436527	D	0.875000
KSa	1.690681	Pr > KSa	0.0066