

# Aplicación de tests no paramétricos

Caso: Distribucion de notas de estudiantes de la academia Trilce en los ultimos 4 años

Alvarez, Sandro    Bautista, Walter  
Burga, Ever    Casanova, Italo    Cuyate, Brayan

Facultad de Ingenieria Industrial y de Sistemas  
Estadística Aplicada - PC2  
**Universidad Nacional de Ingenieria**

Diciembre 2022



# Objetivos del trabajo

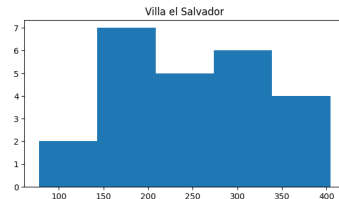
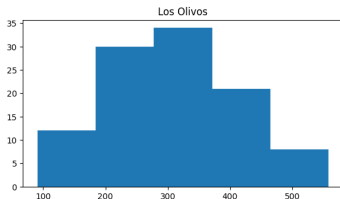
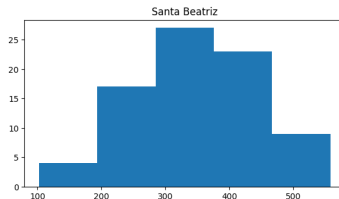
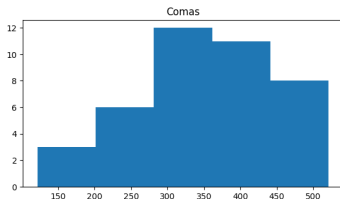
## General

Determinar si el ciclo de repaso ha sido útil para los postulantes a la Universidad Nacional de Ingeniería en la academia Trilce tomando en cuenta sus sedes y el efecto de la pandemia.

# Test de Kruskal-Wallis

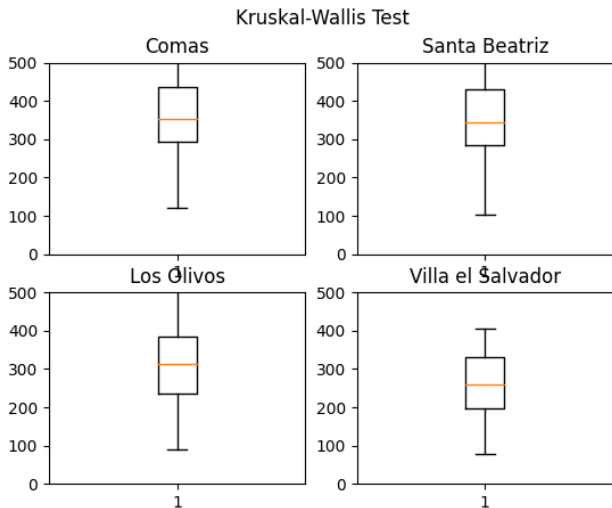
La distribución de las notas de los estudiantes de la academia Trilce de acuerdo a sede es la siguiente:

Kruskal-Wallis Test



# Test de Kruskal-Wallis

Con un gráfico de cajas se puede observar que la distribución de Villa el Salvador tiene una mediana muestral que es distinta a las demás.



# Test de Kruskal-Wallis

- Sea la hipótesis nula  $H_0$  es que las distribuciones de las notas de los estudiantes de la academia Trilce de acuerdo a sede son iguales.
- La hipótesis alternativa  $H_1$  es que al menos una de las distribuciones es distinta.
- La prueba de Kruskal-Wallis arroja los siguientes resultados:

```
Kruskal-Wallis Test
=====
Todos:
KruskalResult(statistic=17.971801375029507, pvalue=0.0004457791596021973)
Comas, Santa Beatriz, Los Olivos:
KruskalResult(statistic=7.967390787511344, pvalue=0.0186167159707278)
Comas, Santa Beatriz, Villa el Salvador:
KruskalResult(statistic=14.980319853682376, pvalue=0.0005585536256191514)
```

- Se rechaza  $H_0$  y se concluye que al menos alguna distribución de notas de los estudiantes de la academia Trilce de acuerdo a sede difiere con un nivel de significancia del 5 %

# Test de Kruskal-Wallis

Al comparar las cuatro distribuciones el p-value es 0.000445.

```
Kruskal-Wallis Test
=====
Todos:
KruskalResult(statistic=17.971801375029507, pvalue=0.0004457791596021973)
Comas, Santa Beatriz, Los Olivos:
KruskalResult(statistic=7.967390787511344, pvalue=0.0186167159707278)
Comas, Santa Beatriz, Villa el Salvador:
KruskalResult(statistic=14.980319853682376, pvalue=0.0005585536256191514)
```

Sin embargo, un detalle que resalta es que al excluir la distribución de Villa el Salvador, el p-value es 0.0186. Este valor sería significativo para un nivel de confianza del 1 % y señala una posible diferencia que consideramos que es susceptible de futura investigación.

# Prueba de aleatoriedad de rachas

Se realiza una prueba de rachas para determinar si los rankings de estudiantes de acuerdo a género son aleatorios

Mérito	1	2	2	3	4	5	6	7	8	9	10	11	11	12	13	14	15		
Genero	M	F	F	M	M	M	M	M	F	M	M	M	M	M	M	M	M		
	16	17	18	18	19	20	20	21	22	23	23	24	25	26	27	28	29	30	
	M	M	M	M	F	M	M	M	M	M	M	F	M	M	F	M	M	M	
	31	32	33	34	34	35	35	35	36	36	37	37	38	39	40	41	41	42	
	M	F	M	F	M	M	F	M	M	F	F	M	M	M	M	M	M	M	
	43	44	45	45	45	45	46	47	48	48	49	50	50	51	52	53	54	55	56
	M	M	F	M	M	M	F	M	M	M	M	M	M	M	M	M	F	M	M



# Prueba de aleatoriedad de rachas

$H_0$ = El orden de merito por genero ha ocurrido de forma aleatoria

$H_1$ = El orden de merito por genero han ocurrido de forma no aleatoria

$n_1=58$

$n_2=14$

$R=25$

La distribución de probabilidad para  $R$  tiende hacia la normalidad cuando  $n_1$  y  $n_2$  se hacen grandes. La aproximación es buena cuando  $n_1$  y  $n_2$  sean ambos mayores que 10. Por esta razón, podemos usar el estadístico  $Z$  como estadístico de prueba con muestras grandes, donde

$$z = \frac{R - E(R)}{\sqrt{V(R)}}$$
$$E(R) = \frac{2n_1n_2}{n_1 + n_2} + 1 = 23.56$$
$$V(R) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)} = 6.85$$
$$Z = \frac{25 - 23.56}{\sqrt{6.85}} = 0.55$$

La region de rechazo para una prueba de dos colas, con un nivel de suficiencia del 5%.

$|z| > 1.96$ .

$Z$  no pertenece a la región de rechazo, entonces se acepta la hipotesis nula.

# Test U de Mann-Whitney

Se ha comparado los resultados pre-pandemia y post-pandemia de los estudiantes de la academia Trilce.

## TEST U MANN-WHITNEY

Ejemplo: Se han recolectado las muestras de los puntos obtenidos por alumnos en simulacros tipo examen de admisión UNI de la academia TRILCE en los años 2018 y 2022. Se quiere saber si los resultados de puntos obtenidos en simulacros en 2018 (pre-pandemia) y en 2022 (post-pandemia) son diferentes.

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Estadístico de Prueba

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

$$U = \min\{U_1, U_2\}$$

Donde:

$n_1$ : tamaño de la muestra de 2018

$n_2$ : tamaño de la muestra de 2022

$R_1$ : sumatoria de los rangos de 2018

$R_2$ : sumatoria de los rangos de 2022

2018	$n_1 =$	92	$R_1 =$	9315	$U_1 =$	943
2022	$n_2 =$	65	$R_2 =$	3088	$U_2 =$	5037
$U =$ 943						

$$Z(\text{cal}) = \frac{-7.295}{0.05}$$

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0, 1)$$

NOTA: para muestras mayores a 10 se puede aproximar a una distribución

Valor crítico:

$$Z_{1-\alpha/2} = 1.960$$

**Decisión:** Se Rechaza la hipótesis nula ( $H_0$ )

**Conclusión:** Con un NS de 0,05 sobre los datos, existen evidencias suficiente para decir que los resultados en los simulacros post-pandemia y pre-pandemia son diferentes.

# Prueba de signos

Se realiza una prueba de signos para evaluar el desempeño de los estudiantes post-repaso.

PRUEBA DE SIGNOS			
CÓDIGO	NOTA SEMESTRAL	NOTA REPASO	DIFERENCIA
60776567	26.00	134.00	-108.00
70281430	73.00	57.00	16.00
70362935	-11.00	37.00	-48.00
70379032	184.00	166.00	18.00
70399298	109.00	120.00	-11.00
70418454	58.00	59.00	-1.00
70572968	146.00	135.00	11.00
70800481	7.00	-1.00	8.00
70844684	312.00	246.00	66.00
70995029	153.00	64.00	89.00
71121137	237.00	171.00	66.00
71245348	-21.00	43.00	-64.00
71253302	238.00	262.00	-24.00
71282841	185.00	115.00	70.00
71329406	314.00	223.00	91.00
71387944	51.00	68.00	-17.00
71425572	44.00	20.00	24.00
71455383	93.00	240.00	-147.00
71530184	48.00	175.00	-127.00
71535597	232.00	283.00	-51.00
71535634	86.00	53.00	33.00
71608886	218.00	228.00	-10.00
71660259	153.00	207.00	-54.00
71834363	198.00	259.00	-61.00
71900361	133.00	199.00	-66.00
71961932	71.00	5.00	66.00
72048579	350.00	302.00	48.00

## 1. HIPOTESIS

H0: No hay cambio en las notas de ambos exámenes,  $p = 0.5$

H1: El repaso no ayudó a mejorar las notas,  $p < 0.5$

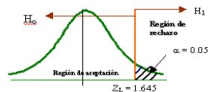
## 2. ESTADISTICO

Prueba de signos/muestras relacionadas

Negativos	40
Positivos	69
Total N	109

$$Z = \frac{r^+ - 0.5n}{0.5\sqrt{n}}$$

$$z = \frac{69 - 0.5(109)}{0.5\sqrt{109}} = 2.777$$



## 3. CONCLUSION

Como 2.777 es mayor que 1.645 se rechaza H0, y se concluye con un nivel de significancia de 0.05 que el repaso no ayudó a mejorar los conocimientos.

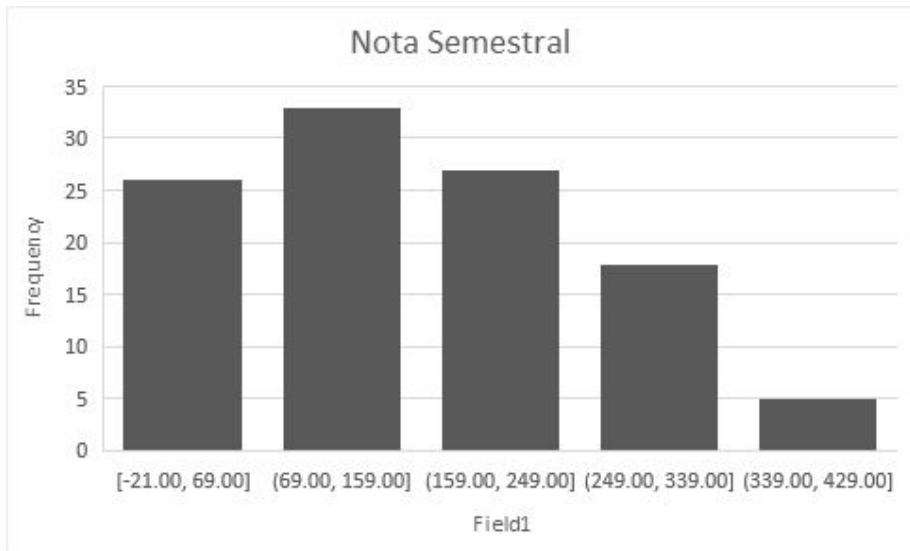
Si bien, el test de signos puede cumplir la misma función que el de **Wilcoxon**, este último tiene mayor potencia al momento de detectar diferencia de medias.

# Prueba de Wilcoxon

Se realiza una prueba de Wilcoxon para 2 muestras **relacionadas**



# Prueba de Wilcoxon



$H_0$  : Las distribuciones son semejantes

$H_1$  : Las distribuciones se encuentran desplazadas

Dado que el estadístico de prueba:

$$Z = 2,991$$

es mayor al

$$Z_{crit} = 1,645$$

se rechaza la hipótesis nula, por lo que se no se puede afirmar que ambas muestras sean idénticas.

**El repaso no ayudó a mejorar los conocimientos**

# Prueba de Wilcoxon

estadístico de la prueba	2008
tamaño	109
T+ y T- suman 5995, como el estadístico de prueba	
2008 no es 2997.7, entonces si rechaza la hipótesis nula en favor a la hipótesis alternativa	



- El ciclo de repaso no ha mejorado las notas de los estudiantes de la academia Trilce. Sin embargo esto podría deberse a que el examen después de repaso fue más complicado que el examen antes de repaso, cosa que no tenemos por seguro.
- La distribución de estudiantes de Villa el Salvador es diferente a la de los demás distritos como mínimo.
- Las mujeres y hombres tienen un desempeño similar en el examen de repaso.
- Existe diferencia en el desempeño pre-pandemia y post-pandemia de los estudiantes de la academia Trilce.