SÉPTIMA Y OCTAVA CLASE

Analytics AoZ

1. Inferencia Estadística – Estimación Puntual.
2. Estimación Interválica.
3. Pruebas de Hipótesis y tipos de errores.
4. Pruebas Chi-cuadrado
5. Test en R
   - Ejemplo de t-student una muestra, dos muestras y muestras parejas.
6. Diseño experimental – ANOVA
7. **Pruebas no Paramétricas**
   - **Wilcoxon**
   - **U de Mann-Whitney**
   - **K de Kruskall-Wallis**
8. Casos aplicados de Geología de los puntos 1 al 7 en Rstudio.

REPASO5 y REPASO6 : EJERCICIO PARA AFIANZAR LO APRENDIDO

GUIA PARA CALSIFICAR TEST DE HIPÓTESIS CON VARIABLES DE RESPUESTA CONTINUA

Analytics AoZ

Table 4.1. Guide to the classification of some hypothesis tests with continuous response variables.

[-, not applicable]

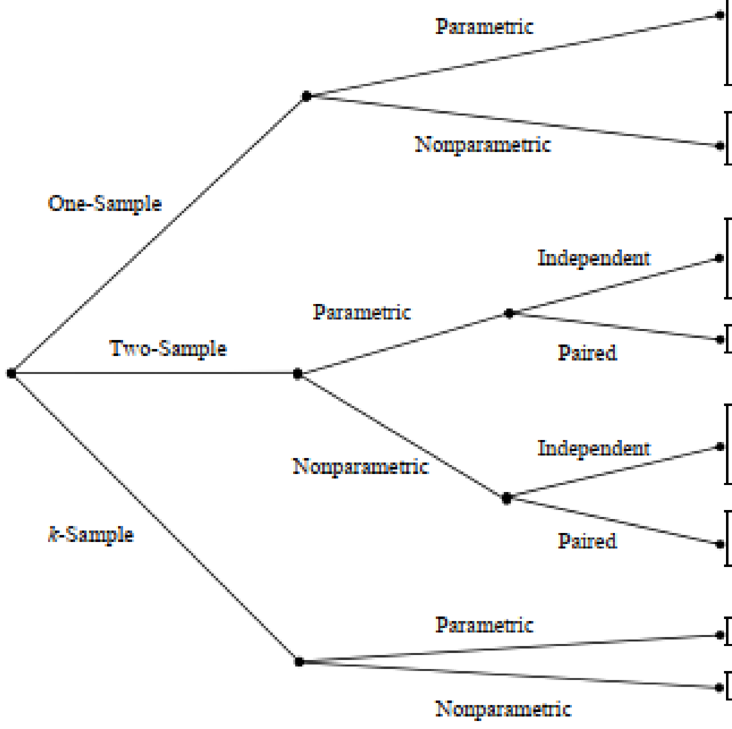
Parametric	Nonparametric	Permutation
	Two independent data groups (chap. 5)	
Two-sample *t*-test	Rank-sum test (two-sample Wilcoxon; Mann-Whitney test)	Two-sample permutation test
	Matched pairs of data (chap. 6)	
Paired *t*-test	Signed-rank test, sign test	Paired permutation test
	Three or more independent data groups (chap. 7)	
Analysis of variance	Kruskal-Wallis test	One-way permutation test
	Three or more dependent data groups (chap. 7)	
Analysis of variance without replication	Friedman test, aligned-rank test	-
	Two-factor group comparisons (chap. 7)	
Two-factor analysis of variance	Brunner-Dette-Munk (BDM) test	Two-factor permutation test
	Correlation between two continuous variables (chap. 8)	
Pearson's r (linear correlation)	Spearman's ρ or Kendall's τ (monotonic correlation)	Permutation test for Pearson's r
	Model of relation between two continuous variables (chaps. 9 and 10)	
Linear regression	Theil-Sen line	Bootstrap of linear regression

Referencia: Statical Methods in Water Resource 2020 , USGS

ÁRBOL DE DECISIÓN PARA EL USO DE TES ESTADÍSTICOS PARAMÉTRICOS Y NO PARAMÉTRICOS

Analytics AoZ

Decision Tree for Selecting the Specific Method



Test	Population Parameter	Distributional Assumption	Section
t-Test and CI	*Mean*	Normal	3.2.1.1
Stratified *t*-Test	*Mean*	Normal	3.2.1.3
Chen Test	*Mean*	Right-Skewed	3.2.1.4
Land's CI Method	*Mean*	Lognormal	3.2.1.5
Test for a Proportion and CI	*Proportion*		3.2.1.6
Sign Test	*Median*	None	3.2.2.1
Wilcoxon Signed Ranks Test	*Median /Mean*	Symmetric	3.2.2.2
t-Test and CI (equal variances)	*Diff in Means*	Normal	3.3.1.1.1
t-Test and CI (unequal variances)	*Diff in Means*	Normal	3.3.1.1.2
Test for Proportions and CI	*Diff in Props*		3.3.1.1.3
Paired *t*-Test	*Diff in Means*	Normal	3.3.1.2.1
Wilcoxon Rank Sum Test	*Diff in Means*	Same Variance	3.3.2.1.1
Quantile Test	Right-Tail		3.3.2.1.2
Slippage Test	Right-Tail		3.3.2.1.3
Sign Test	*Median*	None	3.3.2.2.1
Wilcoxon Signed Ranks Test	*Median /Mean*	Symmetric	3.3.2.2.2
Dunnett's Test	*Mean*		3.4.1.1
Kruskal-Wallis Test	*Mean*		3.4.1.2

Referencia: EPA QA/G-9S

ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

dicotómicas

politómicas

Chi cuadrado

Chi cuadrado

Prueba exacta de Fisher

ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

Cualitativa dicotómica

Cualitativa politómica

t student

Wilcoxon-Mann-Withney

Anova

Kruskal Wallis

ANALISIS BIVARIADO

Análisis Bivariado

dos Variables cualitativas

Cualitativa vs. Cuantitativa

Cuantitativa vs Cuantitativa

Gráficos

politómicas

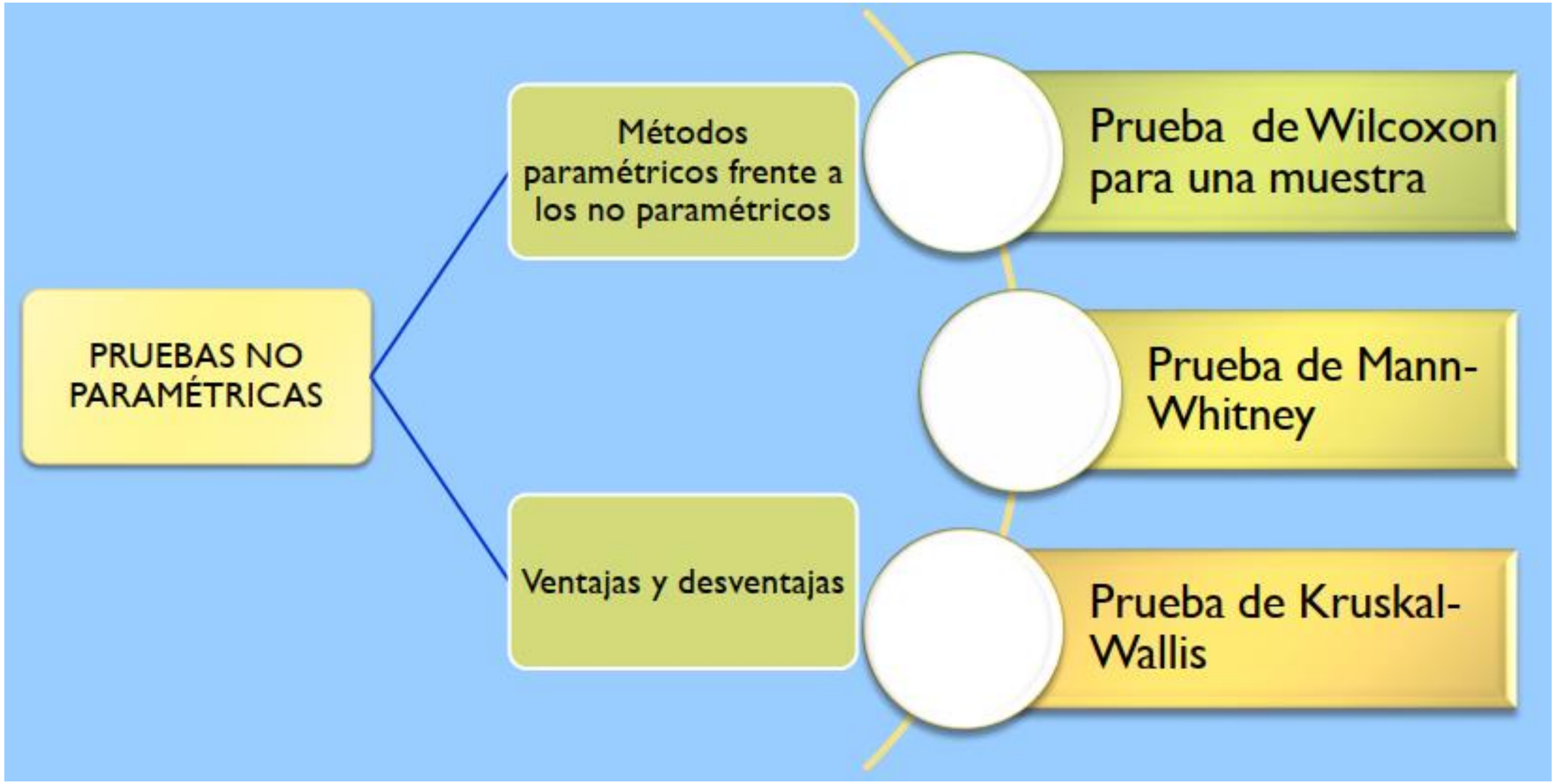
Grafico de dispersión

Correlación de Pearson

Regresión lineal

Sobrevida

PRUEBAS NO PARAMÉTRICAS



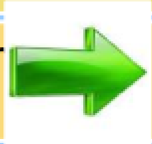
Uno de los problemas más difíciles en estadística, es decidir cuál es la prueba estadística más adecuada para analizar un conjunto de datos.

La selección de la prueba estadística más adecuada para cada caso, depende de varios factores:

- La escala de medida de los datos que se analizarán.
- Si los datos corresponde a la medición de la variable o de las variables seleccionadas en el estudio.
- Si las variables son las características de interés de los sujetos u objetos que se estudian.

Indudablemente que lo mas importante para analizar un conjunto de datos es ***la escala de medida de los datos***.

A continuación se hará una breve explicación de las diferentes escalas con se pueden medir las variables.



Nominal

- Sólo permite asignar un nombre, etiqueta o valor al elemento sometido a medición.
- Los números que se puedan asignar a las propiedades de los elementos, se utilizan **sólo como etiquetas** con la finalidad de clasificarlos.
- Con esta escala no tiene sentido realizar operaciones aritméticas.

Ordinal

- Los datos son etiquetas y además el orden es significativo.
- Los datos se pueden ordenar en forma ascendente o descendente, de tal manera que puedan expresar grados de la característica medida.

Intervalo

- Además de asignar un nombre o **etiqueta** y establecer un **orden** entre los elementos, esta escala permite calcular **diferencias** entre los números asignados a las mediciones (el intervalo entre observaciones que se expresa en términos de una unidad fija de medida).
- Los datos son numéricos.

Razón

- Los datos tienen todas las propiedades de los datos de intervalo y el cociente de los dos valores es significativo.
- Tiene un punto **cero absoluto**, es decir, el cero indica la ausencia de la característica medida.
- Se puede realizar las operaciones aritméticas a los números asignados.

Los procedimientos estadísticos están diseñados para analizar variables cuantitativas y coinciden en las siguientes características:

- Permiten contrastar hipótesis referidas a algún parámetro, como es el caso de la media μ, la varianza σ^2 o la proporción π.
- Exigen el cumplimiento de determinados supuestos sobre las poblaciones originales de donde se extraen los datos (generalmente normalidad y homocedasticidad).
- Analizan datos obtenidos con una escala de medida de intervalo o razón.

Estas tres características combinadas permiten agrupar estos procedimientos estadísticos en la gran familia de las técnicas de análisis, denominadas *Pruebas Paramétricas*, pero su utilidad se ve reducida, fundamentalmente, por dos razones:

- Exigen el cumplimiento de algunos supuestos que en ocasiones pueden resultar demasiado exigentes;
- Obligan a trabajar con unos niveles de medida que no siempre resulta fácil alcanzar.

Analytics AoZ

PARAMÉTRICAS

Sin embargo, existen otras pruebas que:

- Permiten poner a prueba hipótesis no referidas a parámetros poblacionales.
- No necesitan establecer supuestos exigentes sobre las poblaciones de donde se extraen las muestras.
- No necesitan trabajar con datos obtenidos con una escala de medida de intervalo o razón.

NO PARAMÉTRICAS

Esta familia de contrastes se conoce con el nombre de *Pruebas No Paramétricas*.

- La expresión ***no paramétrico*** se usa para referirse a los contrastes que no plantean hipótesis sobre parámetros y, que no se limitan a analizar las propiedades nominales u ordinales de los datos y le añaden el término de distribución libre para referirse a los contrastes que no necesitan establecer supuestos sobre las poblaciones.

No tienen restricciones paramétricas que se usan adicionalmente para analizarla forma de las distribuciones donde se extrae la muestra.

- Permite poner a prueba hipótesis referida, no necesariamente a parámetros N.
- No requiere el cumplimiento de supuestos exigentes sobre la N donde se extrae n.
- No exige trabajar con datos obtenidos en escala de medida de intervalo o de razón.

VENTAJAS	DESVENTAJAS
No exige establecer supuestos relacionadas con que la distribución de la población deben ser normal o que debe tener una distribución específica.	Ignora cierta cantidad de información
Más fácil de llevar acabo y entender.	A menudo no son tan eficientes o exactas como las paramétricas.

WILCOXON (UNA MUESTRA)

Analytics AoZ

Se aplica a una muestra aleatoria independiente tomada de una población que tiene **una distribución continua y simétrica.** El objetivo es probar si la **mediana de la población** de donde proviene la muestra es igual o mayor que un valor propuesto.

Esta prueba es una excelente alternativa a la prueba t para contrastar el valor de la mediana de una población cuando:

- No se cumple los supuestos que se basa la prueba t, como normalidad.
- No es apropiado utilizar la prueba t porque la medida de los datos es ordinal.

PROCEDIMIENTO

1. Teniendo en cuenta que el valor propuesto de la media o mediana es μ_0, se obtiene las diferencias absolutas de las m observaciones de la muestra respecto μ_0, *no considerando las diferencia nulas*; Es decir se obtiene: $D_i = | x_i - \mu_0 |$, para toda $i = 1,\dots, n \leq m$.
2. Asignamos rangos R_i a las n diferencias absolutas; esto es:
   - 1 para la más pequeña,
   - 2 para la segunda más pequeña de las restantes, ..., y así sucesivamente hasta
   - n para la más grande;

   Resolviendo los empates asignando el rango promedio .

3. Luego consideramos como W_+ a la suma de los rangos R_i asignados a las diferencias multiplicado por ψ_i.

$$\psi_i = \begin{cases} 0 & \text{si} \quad D_i < 0 \\ 1 & \text{si} \quad D_i > 0 \end{cases} \Rightarrow \quad W_+ = \sum_{i=1}^{n} \psi_i R_i$$

4. Finalmente se obtiene el estadístico de prueba **Z**, se ajusta a distribución normal estándar, que tiene la siguiente forma:

$$Z = \frac{W_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Analytics AoZ

Dependiendo del tipo de prueba, las hipótesis nula y alternativa son las siguientes:

Hipótesis	Prueba bilateral (Prueba de dos colas)	Prueba unilateral a la derecha (Prueba de cola derecha)	Prueba unilateral a la izquierda (Prueba de cola izquierda)
H_0 :	$m = m_0$	$m \leq m_0$	$m \geq m_0$
H_1 :	$m \neq m_0$	$m > m_0$	$m < m_0$
donde m = es la mediana de la población, m_0=un valor especificado			

Para el caso de una cola, la prueba se especifica de la siguiente forma:

1. Formular las Hipótesis.

   H_0: $\mu \leq \mu_0$ (La mediana de la población no es mayor que μ_0)

   H_1: $\mu > \mu_0$ (La mediana de la población es mayor que μ_0)

2. Fijar el nivel de significación: $0 \leq \alpha \leq 1$
3. Calcular el estadístico de Prueba:

$$Z_c = \frac{W_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

4. Calcular el valor crítico: $z_{1-\alpha}$
5. Tomar la decisión de acuerdo a la siguiente regla de decisión: H_0 se rechaza si: $z_c > z_{1-\alpha}$, en caso contrario no se rechaza H_0.

Factor de Corrección por Empates

- Si el número de empates es grande se tiene que utilizar el factor de corrección dado por:

$$\frac{\sum t^3 - \sum t}{48}$$

- Luego de lo cual el valor de Z_c sería:

$$Z_c = \frac{W - n(n+1)/4}{\sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^3 - \sum t}{48}}}$$

Para el caso de dos colas, la prueba se especifica de la siguiente forma:

1. Formular las Hipótesis.

   H_0: $\mu = \mu_0$ (La mediana de la población es igual que μ_0)

   H_1: $\mu \neq \mu_0$ (La mediana de la población no es igual que μ_0)

2. Fijar el nivel de significación: $0 \leq \alpha \leq 1$
3. Calcular el estadístico de Prueba:

$$Z_c = \frac{W_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

4. Calcular el valor crítico: $z_{1-\alpha/2}$
5. Tomar la decisión de acuerdo a la siguiente regla de decisión: H_0 se rechaza si: $|z_c| > z_{1-\alpha/2}$, en caso contrario no se rechaza H_0.

Ejercicio 1:

Una empresa defensora del medio ambiente cree que el agua potable de su comunidad contiene por lo menos el límite recomendado por los funcionarios de salud de 40 partes por millón (ppm) de cierto metal. En respuesta a su afirmación, el departamento de salud obtiene muestras y analiza el agua potable de 11 hogares en la comunidad. Los resultados son los niveles residuales en ppm:

Hogar	1	2	3	4	5	6	7	8	9	10	11
Niveles residuales	39	20.2	40	32.2	30.5	26.5	42.1	45.6	42.1	29.9	40.9

¿Podemos concluir que el agua de la comunidad excede el límite recomendado de 40 partes por millón (ppm)? Use **α=0.05.**

1. Hipótesis

H_0: $m \leq 40$ (El agua de la comunidad no excede el límite recomendado de 40 partes por millón)

H_1: $m > 40$ (El agua de la comunidad excede el límite recomendado de 40 partes por millón)

Hogar	1	2	3	4	5	6	7	8	9	10	11
Niveles residuales (xi)	39	20.2	40	32.2	30.5	26.5	42.1	45.6	42.1	29.9	40.9
di = xi-40	-1	-19.8	0	-7.8	-9.5	-13.5	2.1	5.6	2.1	-10.1	0.9

Analytics AoZ

2. Estadístico de prueba

Analytics AoZ

Ordenado según Idil

Hogar	Niveles residual	di = xi-40	Idil	Rango	Rangof	R +	R-
3	40	0	0	-	-		
11	40.9	0.9	0.9	1	1	1	
1	39	-1	1	2	2		2
7	42.1	2.1	2.1	3	3.5	3.5	
9	42.1	2.1	2.1	4	3.5	3.5	
8	45.6	5.6	5.6	5	5	5	
4	32.2	-7.8	7.8	6	6		6
5	30.5	-9.5	9.5	7	7		7
10	29.9	-10.1	10.1	8	8		8
6	26.5	-13.5	13.5	9	9		9
2	20.2	-19.8	19.8	10	10		10
						13	42

(3+4)/2 =3.5

Empates

W = 13

3. Regla de decisión

Wilcoxon Signed Rank Test: Niveles

Test of median = 40.00 versus median > 40.00

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Niveles	11	10	13.0	0.937	35.85

W= 13

P-valor = 0.937 > alpha=0.05 *No rechazamos Ho.*

Conclusión: El agua de la comunidad no excede el límite recomendado de 40 partes por millón, a un nivel de significancia del 5%.

OBSERVACIÓN

Cuando el número de observaciones para los cuales di ≠ 0 es n ≥ 20, la prueba Z será una aproximación cercana a la prueba de rango con signo de Wilcoxon. Esto es posible porque la distribución W se aproxima a una curva Normal a medida que n se hace más grande.

El estadístico de prueba es:

$$Z_c = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Donde W: suma de los rangos R+

n: el número de observaciones para las cuales di≠ 0.

R function to compute Wilcoxon test

To perform one-sample Wilcoxon-test, the R function **wilcox.test**() can be used as follow:

```
wilcox.test(x, mu = 0, alternative = "two.sided")
```

- **x**: a numeric vector containing your data values
- **mu**: the theoretical mean/median value. Default is 0 but you can change it.
- **alternative**: the alternative hypothesis. Allowed value is one of “two.sided” (default), “greater” or “less”.

Analytics AoZ

“LO QUE ESCUCHO LO OLVIDO.

LO QUE VEO LO RECUERDO.

PERO LO QUE HAGO, LO ENTIENDO.”

VAMOS A RESOLVER UN EJERCICIO

U MANN-WHITNEY

Se tiene ***dos poblaciones***, de cada población se extrae una muestra aleatoria. Ambas muestras son ***independientes***. Se quiere probar la hipótesis de que las ***medianas*** de ambas poblaciones son diferentes, una mediana es mayor que otra, o una mediana es inferior que la otra.

Objetivo probar si las muestras **proceden de una misma población continua o de poblaciones diferentes con características similares**.

Esta prueba es una excelente alternativa a la prueba t para la comparación de dos medias poblacionales, cuando:

- No se cumple los supuestos que se basa la prueba t, como normalidad.
- No es apropiado utilizar la prueba t porque la medida de los datos es ordinal.

PROCEDIMIENTO

1. Consideremos dos muestras independientes, de tamaño n_1 y n_2, extraídas de la misma población o de dos poblaciones idénticas.
2. Si mezclamos las $n_1 + n_2 = n$ observaciones y, como si se tratara de una sola muestra, asignamos rangos R_i a las n puntuaciones; esto es, 1 para la más pequeña, 2 para la más pequeña de las restantes, ..., n para la más grande; resolviendo los empates asignando el rango promedio.
3. Así tendremos rangos R_{i1} (los n_1 rangos correspondientes a las observaciones de la primera muestra) y los rangos R_{i2} (los n_2 rangos correspondientes a las observaciones de la segunda muestra).
4. Luego se calculan los estadísticos S_1 (sumando los rangos de la primera muestra) y S_2 (sumando los rangos de la segunda muestra).
5. Posteriormente se calcula:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - S_1 \quad \text{y} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - S_2$$

7. Dado que suponemos que las dos muestras se han extraído de dos poblaciones idénticas, cabe esperar que U_1 y U_2 sean iguales; si fuesen distintas, existirá la evidencia de que ambos promedios poblacionales son iguales, siempre y cuando U_1 (o U_2) sea demasiado grande o demasiado pequeño. Entonces, para determinar esto último, podemos basar nuestra decisión en:

$$U = U_1, \quad \text{si} \quad U_1 < \frac{n_1 n_2}{2} \qquad U = U_2, \quad \text{si} \quad U_1 > \frac{n_1 n_2}{2}$$

8. Se debe estandarizar (tipificar) el valor de U de modo que se distribuya aproximadamente como una normal estándar. Esto es:

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

Analytics AoZ

La prueba se especifica de la siguiente forma:

1. Formular las Hipótesis.

   H_0: Las poblaciones son similares ó tienen promedios iguales

   H_1: Las poblaciones no son similares ó no tienen promedios iguales

2. Fijar el nivel de significación: $0 \leq \alpha \leq 1$

3. Calcular el estadístico de Prueba:

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

4. Calcular el valor crítico: $z_{1-\alpha/2}$

5. Tomar la decisión de acuerdo a la siguiente regla de decisión: H_0 se rechaza si: $|z_c| > z_{1-\alpha}$, en caso contrario no se rechaza H_0.

Si n_1 y n_2 son **mayores o iguales que 10**, se puede usar la aproximación normal siguiente:

$$Z_c = \frac{U - \mu_U}{\sigma_U}$$

donde:

$$\mu_U = \frac{n_1 n_2}{2} \quad \text{y} \quad \sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$$

Analytics AoZ

Ejercicio 2:

La siguiente información corresponde a los puntajes que obtuvieron en una prueba estudiantes de dos especialidades de una universidad.

Especia-lidad 1	100	110	80	75	130	95	105	125	140	85
Especia-lidad 2	92	112	83	136	65	75	89	160	90	114

Pruebe que las puntuaciones medianas en ambas especialidades es la misma. Use $\alpha=0.10$.

1. Hipótesis y Procedimiento

Analytics AoZ

H_0: $m_1 = m_2$ (Las puntuaciones medianas en ambas especialidades es la misma)

H_1: $m_1 \neq m_2$ (Las puntuaciones medianas en ambas especialidades no es la misma)

- $S_1 = 110,5$
- $n_1 = 10$
- $n_2 = 10$

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - S_1$$

$$U = 10 \times 10 + \frac{10 \times 11}{2} - 110,5$$

$$U = 44,5$$

$$\mu_U = \frac{n_1 n_2}{2} = \frac{10 \times 10}{2} = 50$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(10 \times 10)(10 + 10 + 1)}{12}}$$

$$\sigma_U = 13,2288$$

$$Z_c = \frac{44,5 - 50}{13,288} = -0,4158$$

3. Regla de decisión

A un nivel de significación del 0,10 tenemos que:

Como $Z_{1-\alpha/2} = Z_{0.95} = 1,645$ no se rechaza H_0.

4. Conclusión

Las puntuaciones medianas de ambas especialidades son las mismas, a un nivel de significación del 5%.

R function to compute MANN-WHITNEY test

Analytics AoZ

To perform two-samples Wilcoxon test comparing the means of two independent samples (x & y), the R function **wilcox.test**() can be used as follow:

```
wilcox.test(x, y, alternative = "two.sided")
```

- **x,y**: numeric vectors
- **alternative**: the alternative hypothesis. Allowed value is one of “two.sided” (default), “greater” or “less”.

Analytics AoZ

“LO QUE ESCUCHO LO OLVIDO. LO QUE VEO LO RECUERDO. PERO LO QUE HAGO, LO ENTIENDO.”

VAMOS A RESOLVER UN EJERCICIO

KRUSKAL-WALLIS

Analytics AoZ

Esta prueba es una extensión de la Prueba U de Mann-Whitney propuesta por Kruskal y Wallis (1952)

Se resuelve de forma similar al Diseño de Experimento de un factor completamente aleatorio de k muestras aleatorias e independientes extraídas de k poblaciones (ANOVA), con el objetivo de averiguar si las k poblaciones son idénticas o alguna de ellas posee un promedio diferente a las otras.

Las ventajas fundamentales de esta prueba frente a la ANOVA son:

- No necesita establecer supuestos sobre las poblaciones originales tan exigentes como los de normalidad y homocedasticidad.
- Permite trabajar con datos ordinales.

PROCEDIMIENTO

1. Consideremos k muestras aleatorias e independientes de tamaños $n_1, n_2, \ldots, n_k$, extraídas de k poblaciones idénticas. Llamemos n como el conjunto total de observaciones: $n = n_1 + n_2 + \ldots + n_k$.
2. Asignemos rangos desde 1 hasta n a ese conjunto de n observaciones como si se tratara de una sola muestra (si existen empates se asigna el promedio de los rangos empatados).
3. Llamemos R_{ij} a los rangos asignados a las observaciones i de la muestra j y R_j a la suma de los rangos asignados a las n_j observaciones de la muestra j.
4. Así tendremos:

$$R_j = \sum_{j}^{n_k} R_{ij} \quad \text{y} \quad \overline{R}_j = \frac{R_j}{n_j}$$

5. ***Si la hipótesis nula**, **que indica que las k poblaciones son idénticas***, es verdadera, los R_j de las k muestras serán parecidos. Por tanto, es posible obtener, tomando como punto de partida la suma de los rangos de cada muestra, un estadístico capaz de proporcionarnos información sobre el parecido existente entre las k poblaciones.
6. De esta manera el estadístico H estará dado por la siguiente expresión:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

La prueba se especifica de la siguiente forma: :

1. Formular las Hipótesis.

   H_0: Las k poblaciones son similares ó tienen promedios iguales

   H_1: Las k poblaciones no son similares ó no tienen promedios iguales

2. Fijar el nivel de significación: $0 \leq \alpha \leq 1$

3. Calcular el estadístico de Prueba:

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

4. **Calcular el valor crítico:** $\chi^2_{(k-1;\ 1-\alpha)}$ siendo k el número de poblaciones.

5. **Tomar la decisión** de acuerdo a la siguiente regla de decisión: H_0 se rechaza si: $H > \chi^2_{(k-1;\ 1-\alpha)}$, en caso contrario no se rechaza H_0.

Ejercicio 3:

Se quiere saber si las calificaciones de los promedios ponderados de los estudiantes de Negocios, Ingeniería, Humanidades y Estudios Generales son iguales, para lo cual se ha registrado las calificaciones de una muestra de estas cuatro escuelas. Use alpha=0.05

NEGOCIOS	INGENIERIA	HUMANIDADES	EST.GENERALES
10	9.2	8.6	10.2
11	11.2	13.4	9.9
8	8.3	12.0	11.7
7.5	13.6	11.5	6.6
13.0	6.5	7.6	12.3
9.5	7.5	15.5	8.7
10.5	8.9	15.7	11.0
12.5	16.0	9.4	8.4
14.0	9.0	12.0	6.6
8.5	11.4		12.3
	10.2		

Analytics AoZ

KRUSKAL-WALLIS
COMPARACIONES MULTIPLES

Analytics AoZ

Cuando se aplica la prueba de Kruskal-Wallis y se rechaza H_0, se concluye que al menos uno de los k grupos es diferente del resto. Esto no indica cuál o cuáles son los grupos diferentes. Ahora corresponde realizar las pruebas de hipótesis para comparar las medianas por pares.

- Las hipótesis son:

  $H_0: M_i = M_j$

  $H_1: M_i \neq M_j$

- Procedimiento:

1. Se calculan $\left|\bar{R}_i - \bar{R}_j\right|$ para todos los pares de grupos.
2. Si n (total de datos) es grande las diferencias se distribuyen aproximadamente normal, pero como las diferencias no son independientes, se debe hacer un ajuste.

Luego se utiliza la siguiente regla de decisión:

Se rechaza H_0, si:

$$\left|\bar{R}_i - \bar{R}_j\right| \geq Z_{\left(1-\frac{\alpha}{k(k-1)}\right)} \sqrt{\frac{n(n+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

donde $Z_{\left(1-\frac{\alpha}{k(k-1)}\right)}$ es tal que :

$$P\left(Z < Z_{\left(1-\frac{\alpha}{k(k-1)}\right)}\right) = 1 - \frac{\alpha}{k(k-1)}$$

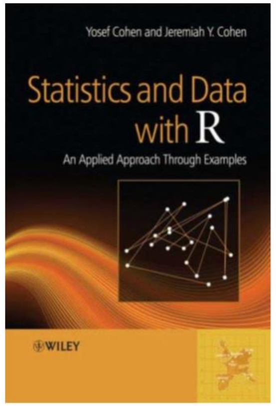
“LO QUE ESCUCHO LO OLVIDO. LO QUE VEO LO RECUERDO. PERO LO QUE HAGO, LO ENTIENDO.”

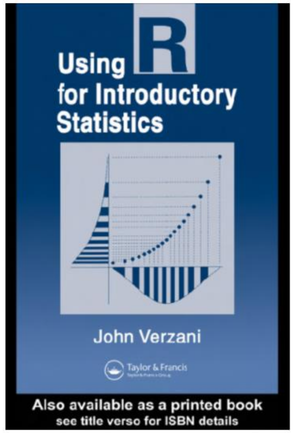
VAMOS A RESOLVER UN EJERCICIO

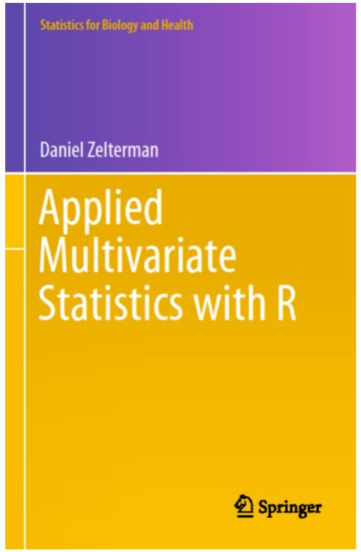
BIBLIOGRAFIA ESPECÍFICA

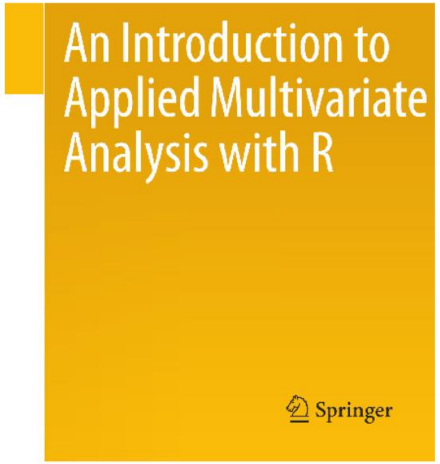
- http://www.sthda.com/english/wiki/one-sample-wilcoxon-signed-rank-test-in-r
- https://www.datanovia.com/en/lessons/wilcoxon-test-in-r/
- http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r
- https://rpubs.com/Joaquin_AR/218456
- https://www.sheffield.ac.uk/polopoly_fs/1.885207!/file/99_Mann_Whitney_U_Test.pdf
- http://www.r-tutor.com/elementary-statistics/non-parametric-methods/mann-whitney-wilcoxon-test
- http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r
- http://www.r-tutor.com/elementary-statistics/non-parametric-methods/kruskal-wallis-test
- https://rpubs.com/Joaquin_AR/219504
- https://rcompanion.org/rcompanion/d_06.html
- https://www.sheffield.ac.uk/polopoly_fs/1.714570!/file/stcp-karadimitriou-KW.pdf
- http://www.sthda.com/english/wiki/kruskal-wallis-test-in-r

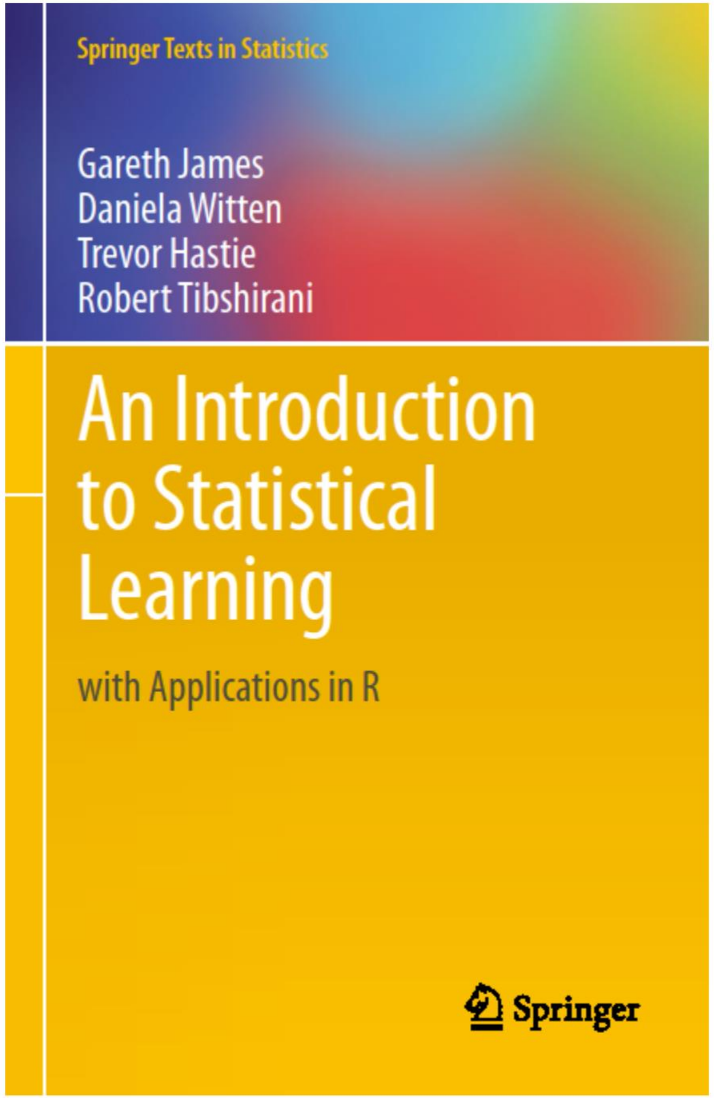
BIBLIOGRAFÍA:

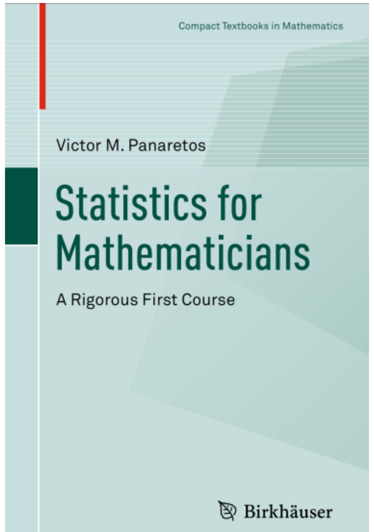


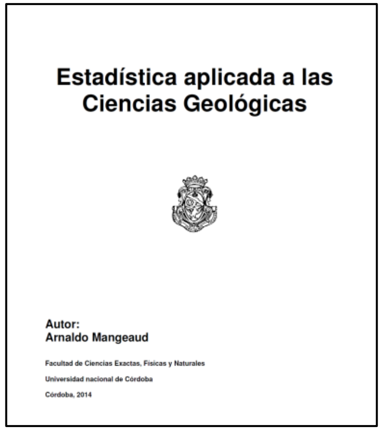


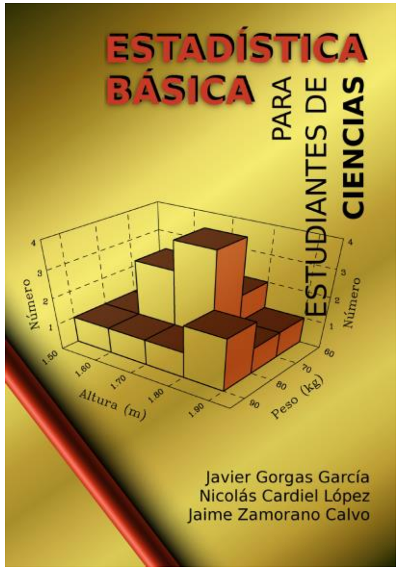












BIBLIOGRAFÍA ESTADÍSTICA APLICADA A GEOLOGÍA EXTRA:

Practical Methods for Data Analysis (US EPA QA/G-9, 2000)

Helsel, D. R., & Hirsch, R. M. (2002). *Statistical methods in water resources* (Vol. 323). Reston, VA: US Geological Survey.

Salvador Figu
eras, M y Gargallo, P. (2003): "Análisis Exploratorio de Datos, 5campus.com, Estadística <http://www.5campus.com/leccion/aed>

Ramalle-Gómara, E., & De Llano, J. A. (2003). Utilización de métodos robustos en la estadística inferencial. *Atención Primaria*, *32*(3), 177-182.

Verzani, J. (2005). *Using R for introductory statistics*. CRC press.

Cohen, Y., & Cohen, J. Y. (2008). *Statistics and Data with R: An applied approach through examples*. John Wiley & Sons.

Arnaldo Mangeaud (2014). *Estadística aplicada a las Ciencias Geológicas*. Universidad nacional de Córdova.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J., 2020, *Statistical methods in water resources: U.S. Geological Survey Techniques and Methods*, book 4, chapter A3, 458 p.