Colegio de Postgraduados

Estadistica no parametrica

Introduccion

Humberto Vaquera Huerta
2024-05-14

Introducción

Si no se cumplen los supuestos de **normalidad** , linealidad o hay valores atípicos (extremos) , a veces es apropiado utilizar métodos no paramétricos libres de distribución, que no involucran inferencia estadística de parámetros. La mayoría de los métodos considerados en este capítulo involucran el uso de rangos. Los métodos no parametricos aún requieren la importante suposición de la independencia de las observaciones.

Modelos paramétricos y no paramétricos

Un **modelo estadístico paramétrico** es un modelo donde la distribución conjunta de las observaciones involucra varias constantes desconocidas llamadas *parámetros* La forma funcional de la distribución conjunta se asume por conocer y las únicas incógnitas en el modelo son los parámetros.

Dos modelos paramétricos comúnmente utilizados en experimentación agronomica.

1. El modelo de **Poisson** en el que suponemos que las observaciones son variables aleatorias de Poisson independientes con media común desconocida θ:
2. El modelo **normal** en el que las observaciones son independientes distribuidos con media μ desconocida y varianza desconocida σ^2 :

Estadística paramétrica

- En el caso de la prueba paramétrica, el proceso de realizar una prueba es relativamente simple. Una hipótesis se construye con base en la suposición de que una ley particular se puede aplicar a una población.
- Con base en la hipótesis, se realiza una prueba y los datos recopilados se analizan estadísticamente.
- Los datos recopilados generalmente se representan como un gráfico y la ley de hipótesis se considera el valor medio de esos datos en particular.
- Se utiliza una prueba paramétrica donde existen suposiciones claras sobre la población y hay mucha información disponible sobre ella.
- La prueba paramétrica es una prueba en la que los datos se recopilan mediante estadísticas paramétricas.

Estadística no paramétrica

- Un enfoque no paramétrico no hace suposiciones y la tendencia central se mide con el valor de la mediana.
- En caso de distribución no paramétrica de la población no se requiere que se especifiquen utilizando diferentes parámetros.
- No se basa en la hipótesis subyacente sino que se basa más en las diferencias de la mediana.
- El método de prueba utilizado en no paramétrico se conoce como prueba libre de distribución.
- Las variables de prueba se basan en el nivel ordinal o nominal.

La distinción básica entre paramétrico y no paramétrico es:

- Si su escala de medición es nominal u ordinal, entonces usa estadísticas no paramétricas
- Si está utilizando escalas de intervalo o de razón, utilice estadísticas paramétricas.

Historia de la no parametrica

- Comienzos de las estadística no paramétricas (Hotelling y Pabst 1936).
- Wilcoxon (1945) introdujo la prueba de suma de rangos de dos muestras para tamaños de muestra iguales, y Mann y Whitney (1947) la generalizaron Pitman (1948), Hodges y Lehmann (1956) y Chernoff y Savage (1958) mostraron propiedades de eficiencia deseables
- Jackknife, introducido por Quenouille (1949) como una técnica de reducción del sesgo y ampliado por Tukey (1958, 1962) para proporcionar pruebas de significancia aproximadas e intervalos de confianza
- Hodges y Lehmann (1963) mostraron cómo derivar estimadores a partir de pruebas de rango .
- Cox (1972) modelo y métodos para el análisis de supervivencia

Historia de la no parametrica

- Bootstrap de Efron (1979)
- Análisis robusto de modelos lineales (McKean, 2004)
- Estimación de densidad (Sheather, 2004)
- Modelado de datos a través de métodos de cuantiles (Parzen, 2004)
- Modelado de datos a través de métodos de cuantiles (Parzen, 2004)
- Suavizantes de nucleos (Schucany, 2004)
- Inferencia basada en permutaciones (Ernst, 2004)
- Pruebas multivariadas de rango con signo en problemas de series de tiempo (Hallin y Paindaveine, 2004)
- Generalizaciones para variedades no lineales (Patrangenaru y Ellingson 2015)

Tipos de escalas y niveles de medición

Variables discretas y continuas

- las variables discretas son variables en las que no hay valores intermedios posibles. Por ejemplo, la cantidad de llamadas telefónicas que recibe por día. No puede recibir 6.3 llamadas telefónicas.
- Las variables continuas son todo lo demás; cualquier variable que teóricamente puede tener valores entre puntos (por ejemplo, entre 153 y 154 libras, por ejemplo).
- ¡Comprender el nivel de medición de una variable (o escala o medida) es la primera y más importante distinción que uno debe hacer sobre una variable al hacer estadística!

Niveles de medición

Hay cuatro niveles básicos: nominal, ordinal, de intervalo y de razón.

Nominal

Una variable medida en una escala "nominal" es una variable que realmente no tiene ninguna distinción evaluativa. Un valor realmente no es mayor que otro. Un buen ejemplo de variable nominal es el sexo (o género).

Ordinal

Algo medido en una escala "ordinal" tiene una connotación evaluativa. Un valor es mayor o mayor o mejor que el otro. Ejemplo podría ser calificar su satisfacción laboral en una escala del 1 al 10, donde 10 representa la satisfacción total. Con escalas ordinales, solo sabemos que 2 es mejor que 1 o 10 es mejor que 9; no sabemos por cuánto. Puede variar. La distancia entre 1 y 2 puede ser más corta que entre 9 y 10.

Niveles de medición

Intervalo

Una variable medida en una escala de intervalo da información sobre más o mejor como lo hacen las escalas ordinales, pero las variables de intervalo tienen la misma distancia entre cada valor. La distancia entre 1 y 2 es igual a la distancia entre 9 y 10. La temperatura usando *Celsius* o *Fahrenheit* es un buen ejemplo, hay exactamente la misma diferencia entre 100 grados y 90 que entre 42 y 32.

Razon

Algo medido en una escala de razón tiene las mismas propiedades que una escala de intervalo excepto que, con una escala de razón, hay un punto cero absoluto. La temperatura medida en Kelvin es un ejemplo. No hay valor posible por debajo de 0 grados Kelvin, es el cero absoluto. El peso es otro

Operaciones con variables

Provides:	Nominal	Ordinal	Interval	Ratio
The "order" of values is known		✔	✔	✔
"Counts," aka "Frequency of Distribution"	✔	✔	✔	✔
Mode	✔	✔	✔	✔
Median		✔	✔	✔
Mean			✔	✔
Can quantify the difference between each value			✔	✔
Can add or subtract values			✔	✔
Can multiple and divide values				✔
Has "true zero"				✔

Métodos no paramétricos

Aplicaciones:

- ampliamente utilizados en datos en una escala ordinal.
- Los métodos se basan en menos suposiciones y, por lo tanto, su la aplicabilidad es mucho más amplia que la correspondiente métodos paramétricos.
- Son más fáciles de usar.
- El término no paramétrico fue utilizado por primera vez por Wolfowitz, 1942

Ventajas de los métodos no paramétricos

1. Los métodos no paramétricos se pueden aplicar a una amplia variedad de situaciones porque no tienen las más rígidas requisitos de los métodos paramétricos correspondientes. En particular, los métodos no paramétricos no requieren poblaciones normalmente distribuidas.
2. A diferencia de los métodos paramétricos, los métodos no paramétricos pueden a menudo se aplica a datos categóricos, como los géneros de los encuestados.
3. Los métodos no paramétricos por lo general involucran cálculos que el paramétrico correspondiente métodos y por lo tanto son más fáciles de entender y aplicar.

Desventajas de Métodos no paramétricos

1. Los métodos no paramétricos tienden a desperdiciar información porque los datos numéricos exactos a menudo se reducen a un forma cualitativa.
2. Las pruebas no paramétricas no son tan eficientes como las paramétricas pruebas, por lo que con una prueba no paramétrica generalmente necesitamos evidencia más fuerte (como una muestra más grande o mayor diferencias) antes de rechazar una hipótesis nula.

Definiciones

Rangos

Los datos se ordenan cuando se organizan de acuerdo con algún criterio, como el más pequeño al más grande o mejor al peor.

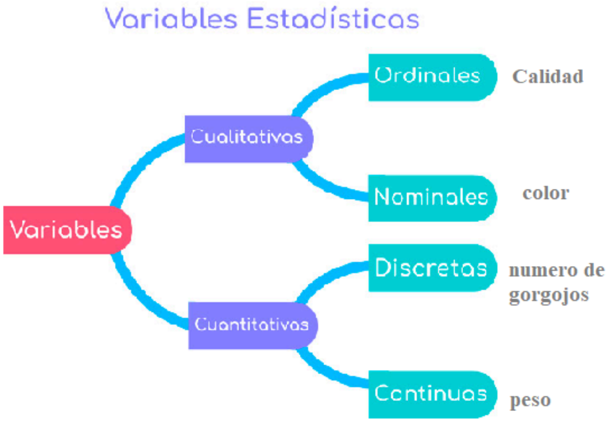
Un rango es un número asignado a un individuo elemento de muestra de acuerdo con su orden en el lista ordenada. Al primer elemento se le asigna un rango de 1, al segundo se le asigna un rango de 2, y así sucesivamente

Una población de estudio

Poblacion

Muestra

Medición



Analisis estadistico

Estadística inferencial

1 muestra poblacional:

- *-Contraste con valor de refencia*
  - *-t de Student (p)*
  - *- Prueba de signos (np)*

Variable:

- *-Estadística descriptiva (media, DE, ...)*
- *-Representación gráfica*
- *-Test de Normalidad :*
  - *-Paramétrico (p)*
  - *-No paramétrico (np)*

Ejemplo: Aroma de variedades de arroz

Aroma of new plant type rice lines based on leaf aroma method in four environments

> Angelita Puji L, Buang Abdullaha, Ahmad Junaedib, and Hajrial Aswidinnoor (2011). PERFORMANCE OF GRAIN QUALITY AND AROMA OF AROMATIC NEW PLANT TYPE PROMISING RICE LINES, Indonesian Journal of Agricultural Science 12(2), pag 84-93

Datos

Variety	Loc_A	Loc_A.1	Loc_A.2	Loc_A.3	category
IPB 113-F-1	0.7	0.5	0.6	0.7	Light
IPB 113-F-2	0.6	0.4	0.5	0.6	None
IPB 115-F-3-2	0.8	0.3	0.3	0.9	Light
IPB 115-F-11	0.6	0.6	0.4	0.7	Light
IPB 116-F-3-1	0.8	0.2	0.5	0.7	Light
IPB 116-F-44-1	0.5	0.4	0.7	0.8	Light
IPB 116-F-46-1	0.8	0.8	0.7	1.2	Light
IPB 117-F-1-3	0.4	0.5	0.7	1.1	Light
IPB 117-F-4-1	0.5	0.6	0.3	1.1	Light
IPB 117-F-6-1	0.9	0.6	0.3	0.7	Light
IPB 117-F-14-2	0.5	0.3	0.5	0.8	None
IPB 117-F-15-2	0.5	0.5	0.5	0.7	None
IPB 117-F-17-4	0.6	0.4	0.8	0.5	Light
IPB 117-F-17-5	0.6	0.4	0.7	0.9	Light
IPB 117-F-18-3	0.9	0.5	0.7	1.1	Light
IPB 117-F-45-2	0.7	0.6	0.6	0.9	Light
IPB 140-F-1-1	0.5	0.6	0.8	0.5	Light
IPB 140-F-2-1	0.5	0.3	0.6	0.7	None
IPB 140-F-3	0.9	0.7	0.8	0.9	Light
IPB 140-F-4	0.6	0.4	0.6	1.1	Light
IPB 140-F-5	0.5	0.3	0.6	0.8	Light
IPB 140-F-6	1.3	0.7	1.4	1.1	Aromatic
IPB 140-F-7	0.9	0.6	0.4	0.6	Light

Modelo de ubicación

$X_1, X_2, \ldots, X_n$ denota muestra aleatoria siguiendo el modelo $X_i = \theta + e_i,$, los errores aleatorios e_i se distribuyen de forma independiente e idéntica con una función de densidad de probabilidad continua $f(t)$ simétrica alrededor de 0.

Evaluación de la hipótesis:

$$H_0 : \theta = 0 \text{ versus } H_A : \theta > 0$$

Una hipótesis nula está asociada con una contradicción a una teoría que uno quisiera probar

Una hipótesis alternativa está asociada con la teoría real que uno quisiera probar

Prueba del signo

Es una prueba no paramétrica en datos de muestras pareadas , en datos nominales con dos categorías, o en hipotesis sobre la mediana de la población frente a un valor hipotético k.

A cada prueba estadística como la prueba de signos, asociamos una estadística de prueba

Un estadístico de prueba es una función de los datos

La prueba de signos tiene estadística de prueba:

$$S = \sum_{i=1}^{n} \text{signo}(X_i)$$

$$\text{signo}(t) = \begin{cases} -1 & \text{si } t < 0 \\ 0 & \text{si } t = 0 \\ 1 & \text{si } t > 0 \end{cases}$$

Prueba del signo

Mirando solo las observaciones positivas:

$$S^+ = \#_i\{X_i > 0\}$$

- 0 se ignoran las observaciones y se reduce el tamaño de la muestra
- Si observando un −1 o 1 para cada X_i entonces es un lanzamiento de moneda
- S^+ sigue una distribución binomial con n intentos y probabilidad de éxito $1/2$
- Esta es la distribución nula del estadístico de prueba.
- No depende de la distribución de errores.
- Esta propiedad se llama distribución libre.
- Compare la distribución nula con la estadística de prueba observada

Ejemplo de la prueba del signo

En datos nominales

- Comparación de las marcas de Cafe A y B Teoría: se prefiere la marca A a la B
- Hipótesis nula: no hay diferencia entre A y B Alternativa: se prefiere A a B
- Catador con los ojos vendados da preferencia a un cafe sobre el otro Experimento con 12 catadores

  ```
  10 catadores prefieren la marca A a la marca B
  1 catador prefiere la marca B a la A
  1 sin preferencia
  ```

  Evidencia bastante convincente a favor de la marca A

Pero, ¿qué tan probable es que tal resultado se deba al azar si la hipótesis nula es verdadera (sin preferencia en las marcas)?

Ejemplo de la prueba del signo

Para nuestra prueba de signos, el estadístico de prueba S^+

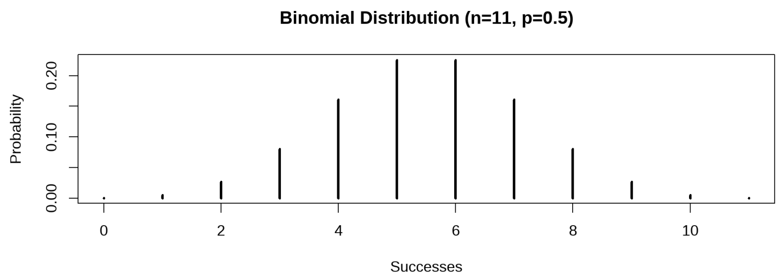
es el numero signos positivos Evaluado en nuestros datos, la estadística de prueba observada es $S^+ = 10$ Comparamos el estadístico de prueba observado con la distribución binomial con $n = 11$ Pruebas y probabilidad de éxito $1/2$

El p-value es $P_{H_0}(S^+ \geq s^+) = 1 - F_B(s^+ - 1, n, \frac{1}{2})$

Ejemplo de la prueba del signo

```
Pvalue=binom.test(10, 11, 0.5, alternative="greater")
Pvalue
```

```
##
##      Exact binomial test
##
## data:  10 and 11
## number of successes = 10, number of trials = 11, p-value = 0.005859
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.6356405 1.0000000
## sample estimates:
## probability of success
##              0.9090909
```



Ejemplo de prueba del signo

Nivel de significación: la probabilidad de rechazar la nula cuando es verdadera

Podemos rechazar la hipótesis nula al nivel de significancia $\alpha = 0.05$ a favor de la alternativa de que la Marca A es más sabrosa que la Marca B.

Ejemplo de la prueba del signo

Sobre la ubicacion de la mediana

- Prueba de cola izquierda:

$$H_0 : mediana = k \; y \; H_1 : mediana < k$$

- Prueba de cola derecha:

$$H_0 : mediana = k \; y \; H_1 : mediana > k$$

- Prueba de dos colas:

$$H_0 : mediana = k \; y \; H_1 : mediana \neq k$$

Para usar la prueba de signos para una mediana de población, el valor está por debajo de la mediana,asignarle un signo −; si está por encima de la mediana, asígne un signo +; y si el entrada es igual a la mediana, asígnele un 0. Luego compare el número de + y −. (Los 0 se ignoran). Si hay una gran diferencia entre el número de +

Ejemplo: Datos de banco

Un gerente de banco afirma que la cantidad mediana de clientes por día ya no es de 750. Un cajero duda de la exactitud de esta afirmación. el numero de banco clientes por día durante 16 días seleccionados al azar se enumeran a continuación. A un $\alpha = 0.05$ nivel de significancia, ¿puede el cajero rechazar la afirmación del gerente del banco?

775	**739**	**756**	**765**	**751**	**760**	**801**	**745**	**782**	**742**	**750**	**789**	**754**	**777**	**753**	**769**

$H_0 : mediana = 750\ vs\ H_1 : mediana \neq 750$ Estadísticos $S^+ = 12$, $S^- = 3$

```
binom.test(12, 15, 0.5, alternative="two.sided")
```

```
##
##      Exact binomial test
##
## data:  12 and 15
## number of successes = 12, number of trials = 15, p-value = 0.03516
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5191089 0.9566880
## sample estimates:
```

Calculo de p-value usando distribucion Binomial

$$
\begin{aligned}
pvalue = 2 * P(X \leq 3) &= 2 \sum_{i=0}^{3} \binom{15}{i} (0.5)^{15-i} (0.5)^{i} && (1) \\
&= 2 \sum_{i=0}^{3} \binom{15}{i} (0.5)^{15} && (2) \\
&\approx 0.03516 && (3)
\end{aligned}
$$

Ejemplo en SAS

Codigo SAS

```
data bancos;
input numero @@;
cards;
775 739 756 765 751 760 801 745 782 742 750 789 754 777 753 769
;
footnote1 'Ejemplo bancos';
proc univariate location=750;
var numero;
run;
```

Tests para posición: Mu0=750				
Test	**Estadístico**		**P valor**	
T de Student	**t**	2.920918	**Pr > |t|**	0.0105
Signo	**M**	4.5	**Pr >= |M|**	0.0352
Rango con signo	**S**	42	**Pr >= |S|**	0.0151

Inferencia para la mediana de la población

Hasta ahora, los métodos que aprendimos eran para la media de la población. La media es una buena medida del centro cuando los datos tienen forma de campana, pero es sensible a valores atípicos y extremos. Sin embargo, cuando los datos están sesgados, una mejor medida del centro sería la mediana. La mediana, como recordará, es una medida resistente. Presentamos un ejemplo a continuación que demuestra por qué podríamos considerar un método alternativo al presentado hasta ahora. En otras palabras, podemos querer considerar una prueba para la mediana y no para la media.

Prueba de t tradicional para una poblacion

La prueba t es una prueba estadística parametrica para la media μ de una población. Supuestos:

- El tamaño de la muestra $n < 30$,
- se desconoce la desviación estándar de la población σ, y
- se sabe que la población se distribuye normalmente
- Independencia entre observaciones

Juego de Hipótesis

- Para la prueba de cola izquierda:

$$H_0 : \mu \geq k, \ vs \ H_1 : \mu < k$$

- Para la prueba de cola derecha:

$$H_0 : \mu \leq k, \ vs \ H_1 : \mu > k$$

- Para la prueba de dos colas:

Prueba de t tradicional para una poblacion

$$H_0 : \mu = k, \; vs \; H_1 : \mu \neq k$$

- la prueba depende de la fdp de la población $f(t)$ Por lo tanto, no es de distribución libre.

El estadístico de prueba es:

$$t = \frac{\bar{X}}{s/\sqrt{n}}$$

donde $\bar{X}$ es la media y s es la desviación estándar de la muestra Usualmente s no se conoce y debe estimarse, entonces si la población es normal t tiene una distribución t de Student con $n-1$ grado de libertad y el p-value es $P_{H_0}(t \geq t_0) = 1 - F_T(t_0, n-1)$

Ejemplo prueba de t datos banco

$$H_0 : \mu = 750, \; vs \; H_1 : \mu \neq 750$$

Ejemplo de Bancos en R

```
x=c(775, 739, 756, 765, 751, 760, 801, 745,
    782, 742, 750, 789, 754, 777, 753, 769)
t.test(x, mu = 750, alternative = "two.sided")
```

```
##
##      One Sample t-test
##
## data:  x
## t = 2.9209, df = 15, p-value = 0.01054
## alternative hypothesis: true mean is not equal to 750
## 95 percent confidence interval:
##  753.5137 772.4863
## sample estimates:
## mean of x
##       763
```

La Prueba de Rangos con signos de Wilcoxon

La prueba de hipótesis no paramétrica de Wilcoxon de 1 muestra es una de las pruebas no paramétricas mas populares.

La prueba de **rangos con signo de Wilcoxon** es una prueba no parametrica equivalente a la *prueba t* y se puede usar cuando la variable dependiente no tiene una distribución normal. La prueba con signos de Wilcoxon está diseñada para probar hipótesis sobre la ubicación (mediana) de una población.

Juego de Hipótesis

- Para la prueba de cola izquierda:

$$H_0 : M_e \geq k \; vs \; H_1 : M_e < k$$

- Para la prueba de cola derecha:

$$H_0 : M_e \leq k \; vs \; H_1 : M_e > k$$

- Para la prueba de dos colas:

$$H_0 : M_e = k \; vs \; H_1 : M_e \neq k$$

Supuestos de la prueba de Wilcoxon de una muestra

Las diferencias entre el valor de los datos y la mediana hipotética son continuas. Los datos siguen una distribución simétrica. Las observaciones son mutuamente independientes entre sí. La escala de medición es al menos el intervalo

Procedimiento para ejecutar una prueba de hipótesis no paramétrica de Wicoxon de una muestra

- Identificar la diferencia entre cada valor individual y la mediana.
- Si la diferencia del valor individual y la mediana es cero, ignórela.
- Ignore los signos de los valores de diferencia y asigne el rango más bajo al valor de diferencia más pequeño. Si los valores han empatado, considere el valor medio.
- Calcule la suma de rangos de valores de diferencia positivos y valores de diferencia negativos (W+ y W-)
- Si los valores son (>20), la aproximación normal sería adecuada.

$$W' = \frac{\sum_{i=1}^{n} Z_i R_i - \dfrac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim n(0,1)$$

con Z_i es 1 si el rango es positivo, cero de otro modo

Ejemplo de Bancos con Wilcoxon

$$H_0 : M_e = 750 \; vs \; H_1 : M_e \neq 750$$

X	775	739	756	765	751	760	801	745	782	742	750	789	754	777
X-750	25	-11	6	15	1	10	51	-5	32	-8	0	39	4	27
Rangos	11	1	7	9	4	8	15	3	13	2		14	6	12
positivos	11		7	9	4	8	15		13			14	6	12
negativos		1						3		2				

- Suma de rangos + =114
- Suma de rangos − =6
- Valor mas pequeño entre 114 & 6 =6
- Estadistico de Prueba =6

Tablas de Wilcoxon Signed-Rank

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
5	--	--	--	--	--	0	2
6	--	--	--	--	0	2	3
7	--	--	--	0	2	3	5
8	--	--	0	2	3	5	8
9	--	0	1	3	5	8	10
10	--	1	3	5	8	10	14
11	0	3	5	8	10	13	17
12	1	5	7	10	13	17	21
13	2	7	9	13	17	21	26
14	4	9	12	17	21	25	31
15	6	12	15	20	25	30	36
16	8	15	19	25	29	35	42
17	11	19	23	29	34	41	48
18	14	23	27	34	40	47	55
19	18	27	32	39	46	53	62
20	21	32	37	45	52	60	69
21	25	37	42	51	58	67	77
22	30	42	48	57	65	75	86
23	35	48	54	64	73	83	94
24	40	54	61	72	81	91	104
25	45	60	68	79	89	100	113
26	51	67	75	87	98	110	124
27	57	74	83	96	107	119	134

	alpha values						
n	0.001	0.005	0.01	0.025	0.05	0.10	0.20
28	64	82	91	105	116	130	145
29	71	90	100	114	126	140	157
30	78	98	109	124	137	151	169
31	86	107	118	134	147	163	181
32	94	116	128	144	159	175	194
33	102	126	138	155	170	187	207
34	111	136	148	167	182	200	221
35	120	146	159	178	195	213	235
36	130	157	171	191	208	227	250
37	140	168	182	203	221	241	265
38	150	180	194	216	235	256	281
39	161	192	207	230	249	271	297
40	172	204	220	244	264	286	313
41	183	217	233	258	279	302	330
42	195	230	247	273	294	319	348
43	207	244	261	288	310	336	365
44	220	258	276	303	327	353	384
45	233	272	291	319	343	371	402
46	246	287	307	336	361	389	422
47	260	302	322	353	378	407	441
48	274	318	339	370	396	426	462
49	289	334	355	388	415	446	482
50	304	350	373	406	434	466	503

Con $n = 15$.

Valor de tablas 25

Note que Estadistico de Prueba =6 es menor que 25

Rechazo H_0.

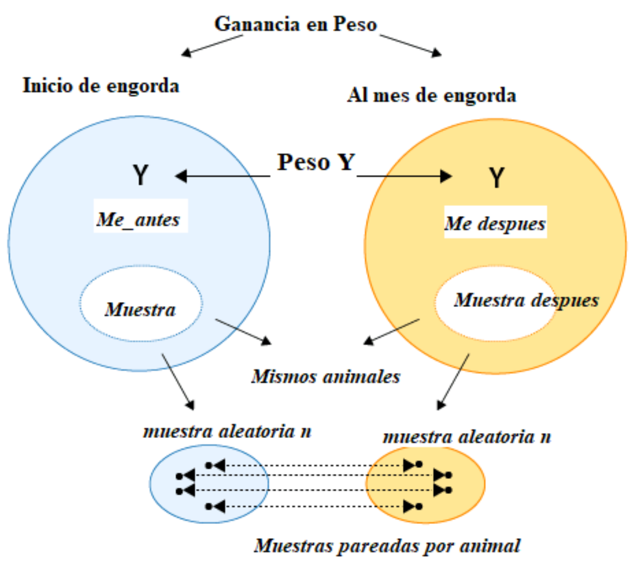
Ejemplo de Bancos en R

```r
x=c(775, 739, 756, 765, 751, 760, 801, 745,
    782, 742, 750, 789, 754, 777, 753, 769)
wilcox.test(x, mu = 750, alternative = "two.sided")
```

```
##
##      Wilcoxon signed rank test with continuity correction
##
## data:  x
## V = 102, p-value = 0.01842
## alternative hypothesis: true location is not equal to 750
```

Muestras dependientes pareadas

Para muestras dependientes, los valores medidos están disponibles en pares. Los pares resultan de repeticiones de medidas emparejadas. Este puede ser el caso, por ejemplo, en estudios longitudinales con varios puntos de medición en el tiempo (análisis de series temporales) o en estudios de intervención con diseños experimentales (medidas antes y después).



Ejemplo prueba del signo en dos muestras pareadas

The iron contents of fruits before and after applying farm yard manure were observed as follows.

Fruit No:	1	2	3	4	5	6	7	8	9	10
Before Applying	7.7	8.5	7.2	6.3	8.1	5.2	6.5	9.4	8.3	7.5
After Applying	8.1	8.9	7	6.1	8.2	8	5.8	8.9	8.7	8
sign	+	+	-	-	+	+	-	-	+	+

Is there any Significant difference between the mean iron contents in the fruits before & after the
farm yarn manure?

$S^+ = 6$, $S^- = 4$ $n = 10$

Hipotesis:

$$H_0 : Me_{antes} = Med_{despues} \; y \; H_1 : Me_{antes} \neq Med_{despues}$$

Ejemplo prueba del signo en dos muestras pareadas en R

```
binom.test(6, 10, 0.5, alternative="two.sided")
```

```
## 
##  Exact binomial test
## 
## data:  6 and 10
## number of successes = 6, number of trials = 10, p-value = 0.7539
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2623781 0.8784477
## sample estimates:
## probability of success 
##                    0.6
```

Ejemplo prueba del signo con rangos de Wilcoxon en dos muestras pareadas

Hipotesis:

$$H_0 : Me_{antes} = Med_{despues} \; y \; H_1 : Me_{antes} \neq Med_{despues}$$

Fruit No:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**suma**
Before Applying	7.7	8.5	7.2	6.3	8.1	5.2	6.5	9.4	8.3	7.5	
After Applying	8.1	8.9	7	6.1	8.2	8	5.8	8.9	8.7	8	
dif	0.4	0.4	-0.2	-0.2	0.1	2.8	-0.7	-0.5	0.4	0.5	
rangos de dif	7	7	3.5	3.5	5	10	1	2	7	9	
rangos +	7	7			5	10			7	9	45
rangos -			3.5	3.5			1	2			10

Ejemplo prueba del signo con rangos de Wilcoxon en dos muestras pareadas

- Suma de rangos + =45
- Suma de rangos − =10
- Valor mas pequeño entre 45 & 10 =10
- Estadístico de Prueba =10

Con $n = 10$.

Valor de tablas de Wilcoxon = 8

Note que Estadistico de Prueba =10 es mayor que 8

No Rechazo H_0.

Ejemplo prueba del signo con rangos de Wilcoxon en dos muestras pareadas en R.

```
antes=c(7.7,    8.5,    7.2,    6.3,    8.1,    5.2,    6.5,    9.4,    8
desp=c(8.1,    8.9,    7,    6.1,    8.2,    8,    5.8,    8.9,    8.7,
wilcox.test(antes, desp, paired = TRUE, alternative = "two.sided")
```

```
##
##      Wilcoxon signed rank test with continuity correction
##
## data:  antes and desp
## V = 21.5, p-value = 0.5746
## alternative hypothesis: true location shift is not equal to 0
```

Ejemplo prueba de t en dos muestras pareadas en SAS.

Subject	Before	After
1	54	56
2	61	65
3	50	52
4	74	73
5	80	82
6	62	61
7	52	59
8	60	64

```
data paired;
input subj before after;
cards;
1 54 56
2 61 65
3 50 52
4 74 73
5 80 82
6 62 61
7 52 59
8 60 64
;

proc print data=paired noobs;
title 'Two related samples';
run;

proc ttest data=paired alpha=.01;
paired before*after;
run;
```

Ejercicio 1

El nivel de amoníaco (en ppm) se estudia cerca de una rampa de salida de un túnel de la carretera en una zona urbana. La concentración diaria de amoníaco en ocho días elegidos al azar durante el tiempo de conducción por la tarde son:

$$1.37, 1.41, 1.42, 1.48, 1.50, 1.50, 1.51, 1.53, 1.55$$

¿Proporcionan los datos suficiente evidencia de que la concentración diaria **mediana** de amoníaco excede 1.5 ppm.

a) Para responder a esta pregunta, realice una prueba de signos, use $\alpha = 0.05$. b) Realice la prueba de rangos con signo de Wilcoxon c) Use la prueba de T (verifique supuestos).

Ejercicio 2 (pareada)

Desempeño fotosintético de diez plantas en dos ambientes en un invernadero (sombreado y soleado). El emparejamiento de datos se realiza para reducir la posible variación entre las mediciones con el objetivo de detectar mejor las diferencias entre los grupos.

Plant	Shady	Sunny
1	0.687	0.884
2	0.651	0.871
3	0.635	0.802
4	0.611	0.86
5	0.673	0.836
6	0.646	0.813
7	0.652	0.845
8	0.668	0.882
9	0.672	0.821
10	0.671	0.837