



Ingeniería en Sistemas
Computacionales
Software

PROBABILIDAD Y ESTADÍSTICA

Ing. Iván García S., PhD.

idgarcia@utn.edu.ec

www.ivangarciasantillan.com

Abril 2024

CONTENIDO GENERAL

Cap. 1 Análisis exploratorio de datos

Cap. 2 Análisis de Regresión y correlación

Cap. 3 Muestreo estadístico

Cap. 4 Validación de instrumentos de investigación

Cap. 5 Pruebas paramétricas

Cap. 6 Pruebas no paramétricas



Estadística Inferencial

- La estadística inferencial es una rama de las matemáticas que se enfoca en sacar conclusiones (inferir) sobre una población a partir de una muestra de datos.
- En la ciencia de datos, la estadística inferencial nos permite tomar decisiones informadas y podrían ayudar a establecer relaciones causales entre variables. Por ejemplo, ayudar a determinar si hay una relación entre el consumo de tabaco y el cáncer de pulmón en una población específica.
- La estadística inferencial nos permite generalizar nuestros hallazgos, ayudar a establecer relaciones causales y evaluar la significancia estadística de los resultados. Sin ella, las conclusiones que extraemos de los datos serían menos precisas y confiables.

Pruebas de hipótesis y causalidad

- Las **pruebas de hipótesis** por sí mismas no pueden establecer **causalidad** entre variables. Estas pruebas se utilizan para evaluar la evidencia en contra o a favor de una afirmación o hipótesis estadística sobre la relación entre variables, pero no pueden probar causalidad por sí solas. Las pruebas de hipótesis estadísticas pueden ayudar a proporcionar evidencia en apoyo o en contra de una relación causal propuesta, pero por sí solas no son suficientes para establecer la causalidad.
- La **causalidad** implica una relación causal directa entre una variable y otra, donde los cambios en una variable causan cambios en la otra variable. Establecer la causalidad generalmente requiere un diseño de investigación cuidadoso, que puede incluir métodos, tales como: experimentos controlados aleatorizados, análisis de series temporales, estudios longitudinales, análisis de causalidad de Granger, modelos de ecuaciones estructurales, entre otros.

Conceptos importantes

- **Muestreo:** Proceso de seleccionar un subconjunto de datos de una población con el objetivo de inferir características de la población completa.
- **Estimación:** determinación del valor de un parámetro desconocido de una población a partir de una muestra de datos.
- **Pruebas de hipótesis:** Proceso estadístico para determinar si hay suficiente evidencia para rechazar una afirmación (o hipótesis) sobre un parámetro desconocido de una población.
- **P-valor:** Medida de la probabilidad de obtener un resultado tan extremo o más extremo que el observado, bajo la suposición de que la hipótesis nula (H_0) es cierta.
 - Un p-valor bajo (< 0.05) sugiere que es poco probable que la diferencia observada (significativa) sea debido al azar, y por lo tanto, se puede rechazar la hipótesis nula H_0 .

Aplicaciones de las Pruebas de Hipótesis

Medicina: Los investigadores utilizan técnicas inferenciales para determinar si un tratamiento es eficaz en comparación con un placebo o con otro tratamiento. Por ejemplo, se puede utilizar una prueba de hipótesis para determinar si una nueva droga tiene un efecto significativamente diferente en la presión arterial en comparación con un grupo de control que recibe un placebo.

Agricultura: determinar si un tratamiento es eficaz en el control de plagas (o rendimiento del cultivo) en comparación con otro tratamiento.

Software: determinar si un proceso automatizado es mejor (precisión, tiempo, etc.) en comparación con el proceso manual.

Principales pruebas de hipótesis: paramétricas y no paramétricas

- Las pruebas no paramétricas tienen como denominador común la ausencia de asunciones acerca de la ley de probabilidad que sigue la población de la que ha sido extraída la muestra: sobre todo cuando la **normalidad** de las distribuciones de la variable en estudio esté en duda y el **tamaño de la muestra** sea menor a 30 casos (pequeñas).
- Muestras Relacionadas:** que los mismos sujetos sean evaluados antes y después de un tratamiento/intervención/capacitación.

Tabla 2. Pruebas paramétricas y su alternativa no paramétrica

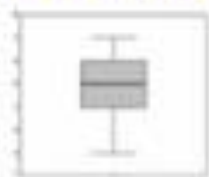
Muestra	Prueba paramétrica	Prueba no paramétrica
<i>Muestras relacionadas</i>		
2 muestras	t-Student	Wilcoxon
> 2 muestras	ANOVA	Friedman
<i>Muestras independientes</i>		
2 muestras	t-Student	U de Mann-Whitney
> 2 muestras	ANOVA	Kruskal-Wallis

¿Cómo distinguir la diferencia entre muestras dependientes e independientes?

Fuente: Lind et al., (2015) en Cap. 11, pp. 325-327.

- Las muestras **dependientes** son mediciones pareadas de un conjunto de elementos (las mismas imágenes, transcripciones, personas, etc.). Las muestras **independientes** son mediciones realizadas en dos conjuntos de elementos distintos.
 - Si los valores de una muestra afectan los valores de la otra muestra, entonces las muestras son dependientes.
 - Si los valores de una muestra no revelan información sobre los valores de la otra muestra, entonces las muestras son independientes.
- Una evaluación **manual** (experto) y una evaluación **automática** (software, modelo) pueden ser tratadas como muestras **dependientes** si se realizan sobre el mismo conjunto de datos (imágenes en nuestro caso), e independientes si se realizan en 2 conjuntos de datos diferentes.

CONTRASTE PARA LA IGUALDAD DE TENDENCIA CENTRAL



NORMALES

$$H_0: \mu_1 = \mu_2$$

medias

$$H_a: \begin{cases} \mu_1 \neq \mu_2 & \text{bilateral} \\ \mu_1 > \mu_2 \\ \mu_1 < \mu_2 \end{cases} \quad \text{unilateral}$$

1

TIPO DE DATOS

2

NO NORMALES

$$H_0: Me_1 = Me_2$$

medianas

$$H_a: \begin{cases} Me_1 \neq Me_2 & \text{bilateral} \\ Me_1 > Me_2 \\ Me_1 < Me_2 \end{cases} \quad \text{unilateral}$$

diferencia

$$H_0: Me_d = 0$$

$$H_a: Me_d \neq 0$$



también para tamaños muestrales muy pequeños

Hablando de pruebas de hipótesis:



Criterios para seleccionar una prueba de hipótesis

- 1. Normalidad de los datos.**
- 2. Tipo de dato: cuantitativo, ordinal, nominal.**
- 3. Muestras relacionadas o independientes**

Probando la normalidad de datos

- En muestras pareadas, donde se comparan dos variables **relacionadas**, no se prueban por separado la normalidad de cada variable. En cambio, se evalúa la normalidad de las **diferencias** entre las dos variables relacionadas (después - antes).
- La lógica detrás de esto es que, en una muestra pareada, las observaciones de las dos variables están vinculadas entre sí, y lo que realmente importa es si las diferencias entre las observaciones siguen una distribución normal. Por lo tanto, la normalidad se evalúa en las diferencias entre las dos variables relacionadas y no en las variables por separado.
- Para muestras **independientes**, donde se comparan dos variables independientes, se puede evaluar la normalidad en cada variable **por separado**. Esto se debe a que en este caso las dos muestras son independientes entre sí y no existe una relación directa entre las observaciones de una variable y las observaciones de la otra.

Prueba de **Wilcoxon**

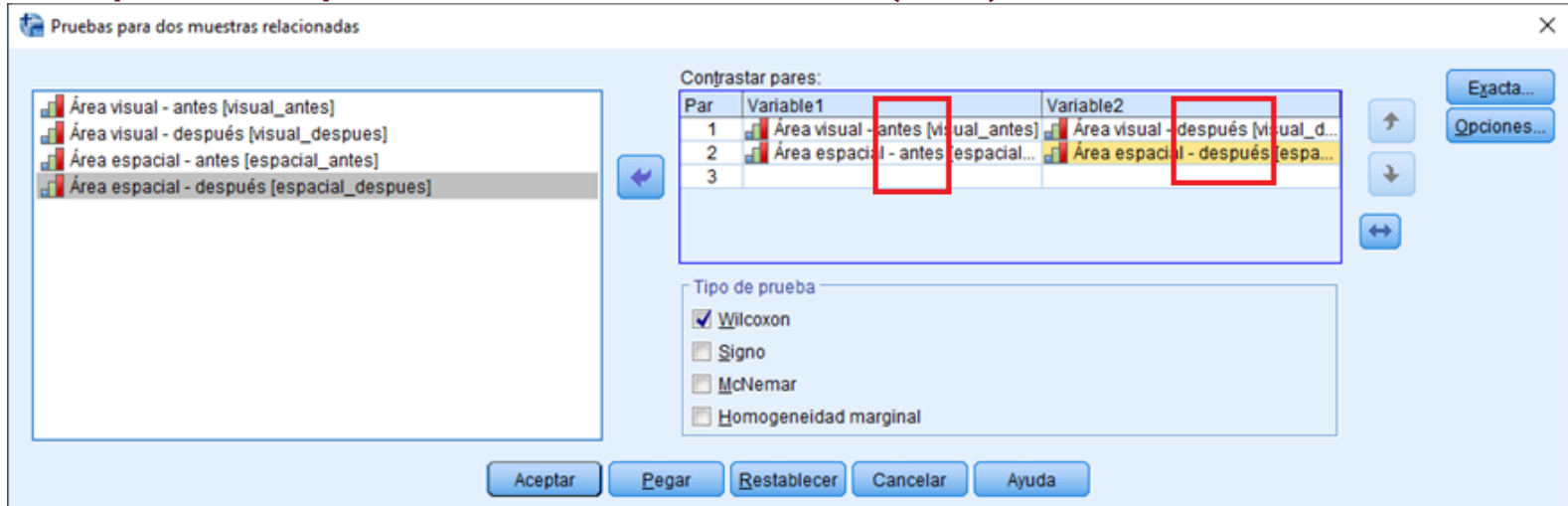
- Dado que es un test no paramétrico, no asume una distribución específica de los datos.
- En la prueba de Wilcoxon, no es necesario probar la normalidad de las variables relacionadas antes de realizar el análisis, ya que es un test no paramétrico que no depende de la distribución de los datos.
- A diferencia de las pruebas paramétricas, como la prueba t de Student, la prueba de Wilcoxon se basa en los rangos de los datos y no en sus valores exactos. Por lo tanto, la prueba de Wilcoxon es robusta a las violaciones de la suposición de normalidad.

Prueba de Wilcoxon

- Permite contrastar la hipótesis de igualdad entre dos **mediana**s poblacionales de grupos relacionados.
- Es la homóloga no paramétrica de la prueba paramétrica t para muestras relacionadas (pareadas).
- Puede ser usada sobre variables de tipo **ordinal o cuantitativo**.
- Para probar la no normalidad de los datos, se trabaja sobre la variable **diferencia** = después - antes. No sobre cada variable separadamente.

SPSS: Transformar → calcular variable

- **SPSS**: Analizar → pruebas no paramétricas → cuadro de diálogos antiguos → 2 muestras relacionadas.
- Interpretación: pág. 8 de Berlanga & Rubio (2012)



Ejercicio:

- Se desea estudiar la efectividad de cierta dieta en mujeres.
- Pregunta: ¿Es efectiva la dieta en la reducción del peso (lb)?
- H_0 : No hay diferencias significativas (medianas iguales) en el peso de las mujeres antes y después de la dieta.
- H_1 : Si hay diferencias significativas en el peso de las mujeres antes y después.
- **Regla:** Si $p\text{-value} > 0.05$ se acepta H_0 , sino se rechaza H_0 (acepta H_1)

Paciente	Peso-Antes	Peso-Después
1	186	160
2	147	148
3	128	125
4	176	178
5	212	203
6	158	158
7	204	197
8	157	160
9	189	181
10	149	151
11	191	187
12	200	195

Peso →
variable
cuantitativa

Mediana

181

169

Ejercicio: verificando la normalidad

Pruebas de normalidad

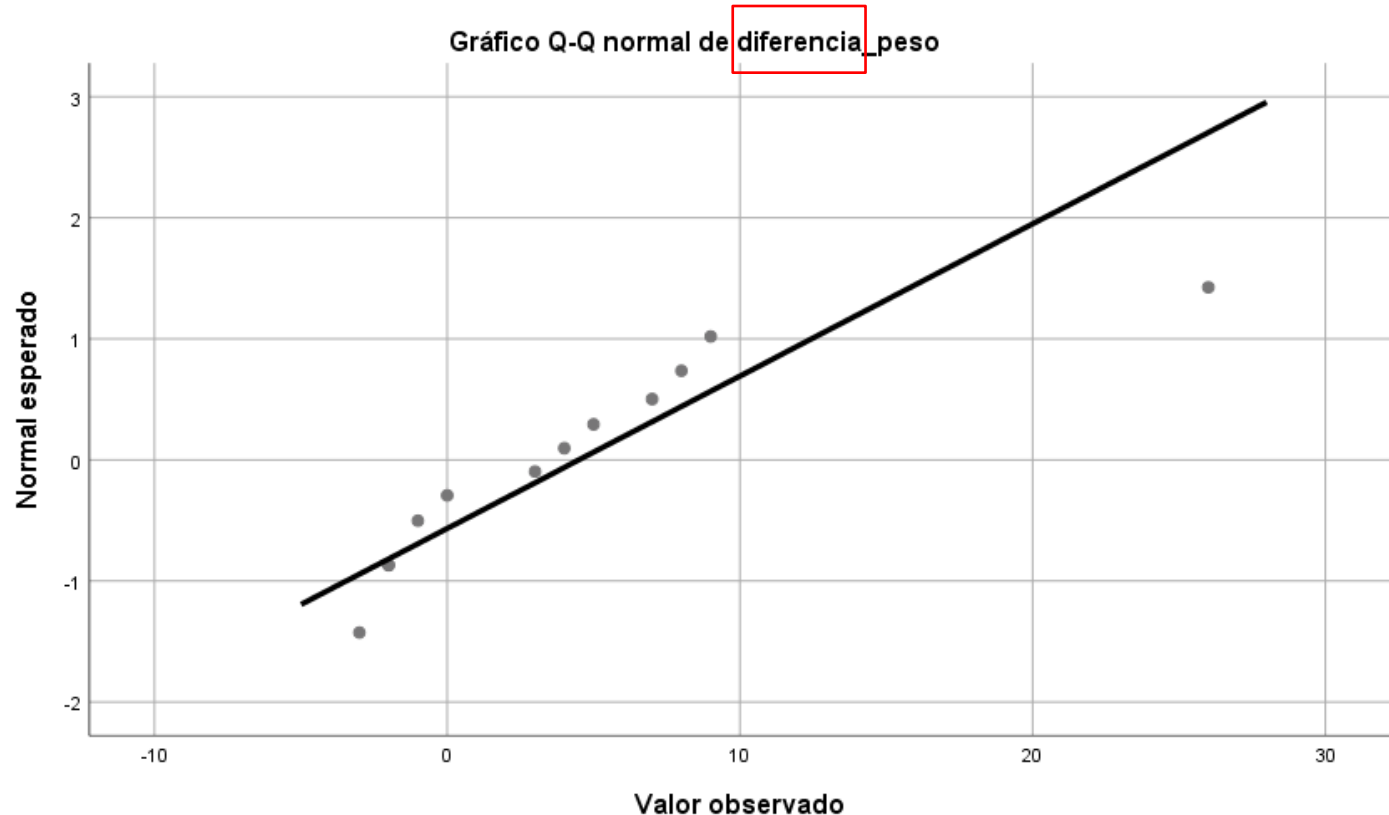
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
diferencia_peso	,202	12	,188	,804	12	,010

a. Corrección de significación de Lilliefors

Como $p\text{-value } 0.010 < 0.05$ se rechaza H_0 , es decir, se concluye que los datos no siguen una distribución normal.

SPSS:

Analizar →
estadísticos
descriptivos →
explorar (lista de
dependientes;
gráficos: gráficos
de normalidad con
pruebas)



Ejercicio: resultados de Wilcoxon

- Prueba de rangos con signo de Wilcoxon

Rangos		N	Rango promedio	Suma de rangos
peso_después - peso_antes	Rangos negativos	7 ^a	7,93	55,50
	Rangos positivos	4 ^b	2,63	10,50
	Empates	1 ^c		
	Total	12		

a. peso_después < peso_antes

b. peso_después > peso_antes

c. peso_después = peso_antes

Estadísticos de prueba^a

	peso_después - peso_antes
Z	-2,002 ^b
Sig. asintótica(bilateral)	,045

a. Prueba de rangos con signo de Wilcoxon

b. Se basa en rangos positivos.

Como $p\text{-value } 0.045 < 0.05$ se rechaza H_0 , es decir, se concluye que si hay diferencias significativas en el peso de mujeres antes y después.

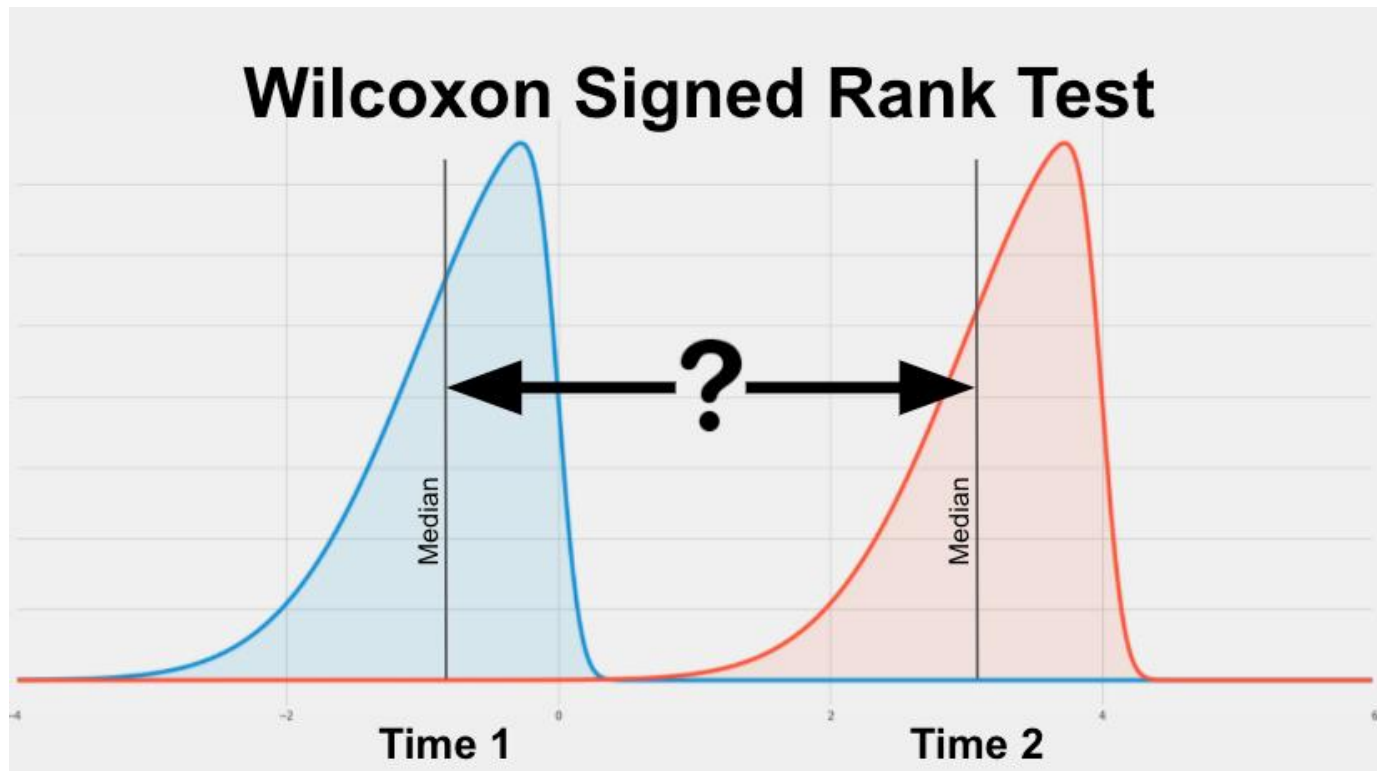
Basado en los rangos negativos (7) y positivos (4), aquí se concluye que, la dieta es efectiva en la **reducción** del peso (peso_después < peso_antes en 7 casos).

Taller:

- Utilice los datos de un examen de inteligencia visual (antes y después) para ver si un cierto tratamiento tuvo efectividad.
- ¿Con qué tipo de variable se está trabajando?
- ¿Es posible aplicar Wilcoxon?
- ¿Qué puede concluir acerca de la efectividad de la intervención/tratamiento?

Deber:

- Seleccione y descargue un conjunto de datos de Internet y aplique la prueba de Wilcoxon.



Prueba T-student

- Permite contrastar la hipótesis de igualdad entre dos **medias** poblacionales de grupos, tanto independientes como dependientes.
- En muestras relacionadas (dependientes, pareadas) es la homóloga **paramétrica** de la prueba no paramétrica de Wilcoxon.
- Puede ser usada sobre variable (grupo) de tipo **cuantitativo**.
- Para probar la normalidad de los datos, se trabaja sobre la variable **diferencia** = después - antes. No sobre cada variable separadamente.

SPSS: Transformar → calcular variable

SPSS:

- Analizar → comparar medias → Prueba T para muestras relacionadas.
- Analizar → comparar medias → Prueba T para muestras independientes.

T-student: resumen

La distribución t de Student se aplica idealmente a muestras tomadas de una población que sigue una distribución normal. Se aplica en el análisis de datos especialmente cuando el tamaño de la muestra es pequeño ($n < 30$) y la desviación estándar de la población es desconocida.

- Escenario 1: Dos muestras **relacionadas**
- Escenario 2: Dos muestras **independientes** con varianzas **iguales** (homocedasticidad).
- Escenario 3: Dos muestras **independientes** con varianzas **diferentes** (heterocedasticidad).

Las **varianzas** son una medida de dispersión, es decir, qué tan dispersos están los datos con respecto a la media.

Ejercicio escenario 1: muestras dependientes

- Se desea estudiar la efectividad de implantar un sistema informático.
- Pregunta: ¿El tiempo promedio de un proceso disminuye significativamente después de implantar el sistema?
- H_0 : No hay diferencias significativas antes y después de implantar el sistema.
- H_1 : Si hay diferencias significativas antes y después de implantar el sistema.
- **Regla:** Si $p\text{-value} > 0.05$ se acepta H_0 , sino se rechaza H_0 (se acepta H_1)

Tiempo promedio de un proceso (min)		
Num	Antes del Sistema	Después del sistema
1	15	15
2	20	14
3	10	11
4	18	12
5	17	10
6	14	20
7	16	15
8	12	11
9	11	10
10	19	12
11	15	13
12	17	10
13	13	13
14	19	14
15	12	14
Prom.	15.20	12.93



Se observa una disminución en el promedio del tiempo (15.20 vs 12.93 min).

¿Es esta diferencia estadísticamente significativa o es casualidad?

Ejercicio escenario 1: resultados

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
diferencia_sistema	,163	15	,200 [*]	,912	15	,146

*. Esto es un límite inferior de la significación verdadera.

a. Corrección de significación de Lilliefors

Como *p-value* 0.146 > 0.05 se acepta H_0 , es decir, se concluye que los datos siguen una distribución normal.

Estadísticas de muestras emparejadas

		Media	N	Desv. Desviación	Desv. Error promedio
Par 1	antes de implantar el sistema	15,20	15	3,144	,812
	después de implantar el sistema	12,93	15	2,631	,679

Como *p-value* 0.042 < 0.05 se rechaza H_0 , es decir, se concluye que si hay diferencias significativas en el promedio de tiempo antes y después. Por lo tanto, la implantación del sistema es efectiva (reducción del tiempo promedio).

Prueba de muestras emparejadas

		Diferencias emparejadas					gl = n-1		
		Media	Desv. Desviación	Desv. Error promedio	95% de intervalo de confianza de la diferencia		t	gl	Sig. (bilateral)
					Inferior	Superior			
Par 1	antes de implantar el sistema - después de implantar el sistema	2,267	3,918	1,012	,097	4,436	2,241	14	,042

Ejercicio escenario 2 y 3: muestras independientes

- Pregunta: ¿Existe diferencia significativa en el promedio del peso (kg) en cuanto al género (hombres/mujeres)?
- H_0 : No hay diferencias significativas en el promedio del peso.
- H_1 : Si hay diferencias significativas en el promedio del peso.
- **Regla:** Si $p\text{-value} > 0.05$ se acepta H_0 , sino se rechaza H_0 (acepta H_1)

	Nombre	Tipo
1	genero	Númérico
2	altura	Coma
3	peso	Coma
4	IMC	Coma
5	edad	Coma
6	ingreso_mensual	Coma
7	nota_promedio	Coma
8	velocidad_internet	Coma

Definir grupos

☒ Utilizar valores especificados

Grupo 1:

Grupo 2:

☐ Punto de corte:

Prueba T para muestras independientes

Variables de prueba:

Variable de agrupación:

Los valores de la variable género (1: masculino; 2 femenino)

Ejercicio 2: resultados

Primeramente, probamos la **normalidad** de los 2 grupos (independientes) **por separado**.

SPSS: Datos → Seleccionar casos (hombres/mujeres)
Analizar → Estadísticos descriptivos → Explorar

Estadísticas de grupo

	genero	N	Media	Desv. Desviación	Desv. Error promedio
en kg	Masculino	29	68.3586	9.98933	1.85497
	Femenino	19	61.7105	7.58895	1.74102

Ambos grupos independientes deben ser aproximadamente normales para garantizar resultados válidos.

$$gl = n_1 + n_2 - 2$$

Prueba de muestras independientes

		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias						95% de intervalo de confianza de la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar		Inferior	Superior
en kg	Se asumen varianzas iguales	1,879	,177	2,468	46	,017	6.64809	2.69342		1.22652	12.06967
	No se asumen varianzas iguales			2,613	44,882	,012	6.64809	2.54403		1.52378	11.77241

La **prueba F** sirve para ver si las **varianzas** son iguales (homocedasticidad) o no (heterocedasticidad).

Ho: Las varianzas son iguales

H1: Las varianzas son distintas

Como *p-value* 0.177 > 0.05 se acepta Ho, es decir, se concluye que las varianzas son iguales entre los 2 grupos independientes.

Regla: Si *p-value* > 0.05 se acepta Ho, sino se rechaza Ho (acepta H1).

Ejercicio 2: resultados (cont.)

Prueba de muestras independientes									
		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias					
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	95% de intervalo de confianza de la diferencia
									Inferior Superior
en kg	Se asumen varianzas iguales	1,879	,177	2,468	46	,017	6.64809	2.69342	1.22652 12.06967
	No se asumen varianzas iguales			2,613	44,882	,012	6.64809	2.54403	1.52378 11.77241

Como antes ya se probó la igualdad de varianzas (homocedasticidad), se trabajará con el $p\text{-value} = 0.017$ en este caso.

Como $p\text{-value} 0.017 < 0.05$ se rechaza H_0 , es decir, se concluye que si hay diferencias significativas en el promedio del peso entre hombres y mujeres.

*Nota: Si alguna de las muestras independientes no sigue una distribución normal, se pueden considerar alternativas al test t, como la prueba **U de Mann-Whitney**, que no asume normalidad en los datos.*

Taller:

- Usando el mismo dataset, se podría usar t-student para verificar si existe una diferencia estadísticamente significativa en los promedios de las variables: *altura*, *nota*, *velocidad-Internet* con respecto al género? ¿Justifique?

Bibliografía

- Lind et al., (2015). Estadística aplicada a los negocios y la economía. 16° ed. McGraw-Hill: México.
- Lind D., Marchal W., Wathen S. (2020). Statistical Techniques in Business and Economics. 18° ed. McGraw-Hill Education: New York.
- Lind D., Marchal W., Wathen S. (2022). Basic statistics for business and economics. 10° ed. McGraw-Hill LLC: New York.
- <http://www.spssfree.com/>
- Berlanga, V., Rubio, M. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. REIRE, Vol. 5, núm. 2. doi: 10.1344/reire2012.5.2528
- Curso de Estadística.
<https://www.geocities.ws/estadistica/>

