Capítulo 9: Comparación de medias

En capítulos anteriores estudiamos pruebas de hipótesis e intervalos de confianza para la proporción o la media de una población. Estos métodos pueden ser extendidos para otras situaciones. Lo bueno es que a pesar de que las situaciones sean diferentes (distintos parámetros de interés) la forma de tomar las decisiones sigue la misma lógica. En este capítulo revisaremos cómo usar los resultados de dos conjuntos de datos para decidir acerca de dos poblaciones, comparando medias de dos poblaciones μ_1 y μ_2. Antes debemos distinguir si los datos provienen de un diseño pareado (dependiente) o es un diseño independiente.

Diseño de muestras pareadas versus muestras independientes

En la comparación de medias vamos a distinguir dos tipos de diseños, pareado e independiente. Es necesario tomar en cuenta el diseño porque cada tipo de diseño dará origen a un test estadístico distinto. El tipo de diseño depende del objetivo y de la población en estudio. Cuando comparamos una variable en dos poblaciones sería ideal que las dos poblaciones sean similares en todos los aspectos menos en el factor o tratamiento que se está estudiando. Suponga que estamos estudiando los sueldos de hombres y mujeres y tomamos una muestra aleatoria de 4 matrimonios.

Diseño pareado 1

Juanita Pérez	↔	Pedro Moya
María Cárcamo	↔	Juan Cerda
Rosa Campos	↔	Diego Maroto
Julia del Valle	↔	Pablo Mora

Esta muestra se dice pareada y el tamaño muestral corresponde al número de pares, no al número de individuos. Cuando los datos son pareados, nos interesa comparar las respuestas de cada par. Entonces, analizaremos las *diferencias* entre las respuestas de cada par.

Datos pareados = sueldo mensual en miles de $

Salario mujer	Salario hombre	Diferencias
150	200	-50
450	310	140
500	500	0
160	300	-140
	Promedio diferencias	**-12.5**

La ventaja principal de los diseños pareados es que nos ayudan a reducir el sesgo por variables confundentes. Dependiendo del problema a investigar, podemos parear sujetos según su edad, sexo o estado de salud, entre otros. Generalmente las variables que se eligen para parear son variables que pueden influenciar la respuesta. Cuando comparamos los resultados de observaciones pareadas, los efectos de estas variables de pareamiento se cancelan.

Otro tipo de diseño pareado es cuando cada sujeto sirve como su propio control. Si se tienen dos tratamientos, cada sujeto puede recibir los dos tratamientos, en orden aleatorio. Si se tiene sólo un tratamiento, se miden las respuestas al tratamiento comparando la situación antes del tratamiento y después del tratamiento.

Diseño pareado 2: medidas en 20 individuos que fueron parte de un estudio médico para reducir la presión sanguínea.

Número	Sexo	Edad	n tabletas	Presión_antes	Presión_después
1001	M	45	2	100.2	100.1
1002	M	41	1	98.5	100.0
1003	F	51	2	100.8	101.1
.	.	.	.	.	.
.	.	.	.	.	.
1019	M	45	2	100.0	100.4
1020	M	37	3	101.5	100.8

Definición:

Tenemos muestras **pareadas** o correlacionadas cuando sabemos de antemano que una observación está relacionada con la otra. Pueden ser observaciones tomadas al mismo tiempo, diseño pareado 1, o medidas tomadas en un mismo sujeto o unidad en dos oportunidades o tiempo distintos (diseño pareado 2).

El **número de observaciones** es el número de pares.

Suponga ahora que queremos comparar los salarios de hombres y mujeres entre los profesores de la Universidad. Para comparar vamos a tener que controlar variables como jerarquía académica, años de trabajo, etc.

Diseño independiente

Marcela Paz	Pedro Letelier
Mónica Nuñez	Nemesio Lara
Inés Miranda	Antonio Jarpa
Claudia Pino	Patricio Bravo
	Carlos Pérez
	Juan González

En el esquema independiente vamos a comparar las respuestas de un grupo con las respuestas del otro grupo. El tamaño de los grupos puede ser igual o diferente y el tamaño de la muestra total es la suma de los dos grupos (número total de individuos). Calculamos medidas de resumen de uno y otro grupo y comparamos las diferencias.

Datos pareados = sueldo mensual en miles de $

Salarios mujeres	**Salarios hombres**
150	200
450	310
500	500
160	300
	720
	600
Promedio mujeres 315,0	**Promedio hombres 438,3**

Análisis de muestras pareadas

En un diseño pareado las unidades son parecidas (de hecho pueden ser las mismas), mientras que las unidades de distintos pares son diferentes. En diseños pareados analizamos las diferencias y el problema se reduce al test t para una media que vimos en el capítulo anterior.

Test t pareado

Supuestos: La muestra de las diferencias es una muestra aleatoria de una población de diferencias. El modelo de la población de las diferencias es Normal, supuesto que no es relevante si el tamaño de la muestra n es grande.

Hipótesis:
$H_0 : \mu_d = 0$ versus
$H_1 : \mu_d \neq 0$ o $H_1 : \mu_d > 0$ o $H_1 : \mu_d < 0$.

Se determina el **nivel de significancia α**

Datos: Muestra de n diferencias de la cual se calcula el promedio de las diferencias $\bar{d}$ y la desviación estándar de las diferencias s_d .

Test estadístico observado: $t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$

Este es el **test estadístico** y su distribución bajo H_0, es una distribución *t* con *n*-1 grados de libertad.

El **valor-p** del test, depende de la dirección de la hipótesis alternativa.

☑ **Extintores**

En la revista Consumo y Calidad de Vida (CCV) de 1992 que publica SERNAC, se hace un análisis comparativo entre lo rotulado y lo real de la carga de una muestra de 8 extintores. A SERNAC le interesa investigar si la diferencia entre la carga que aparece en la etiqueta (rotulado) y la carga real es distinta de cero.

Marca	**Carga rotulado** Gramos	**Carga real** Gramos	**Diferencia** gramos
Kidee	1500	1230	270
Werner	2000	1550	450
Stop-flames	2000	1650	350
The monsters	2000	1620	380
Victoria	2000	1750	250
Abc	2000	1540	460
Ing. Asociados	4000	3740	260
Asytex	2000	1980	20

Para solucionar el problema podemos usar SPSS para los cálculos. **SPSS** realiza esta prueba con los siguientes comandos: Analizar, Comparar medias, Prueba t para muestras relacionadas

Use la salida adjunta para resolver el problema de los **extintores**:

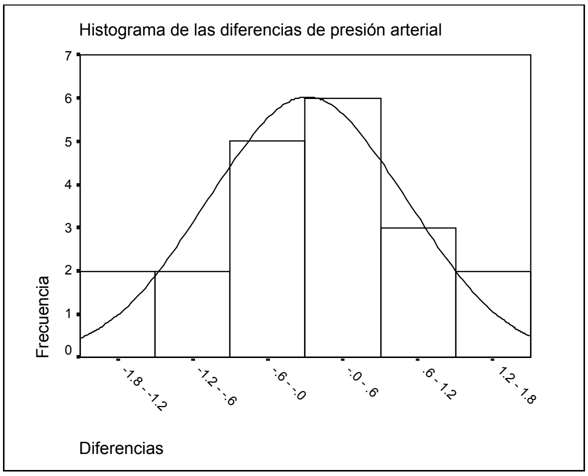
Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	Carga rotulado	2187.50	8	752.970	266.215
	Carga real	1882.50	8	779.556	275.615

Prueba de muestras relacionadas

		Diferencias relacionadas					t	gl	Sig. (bilateral)
					95% Intervalo de confianza para la diferencia				
		Media	Desviación típ.	Error típ. de la media	Inferior	Superior			
Par 1	Carga rotulado - Carga real	305.00	141.522	50.036	186.68	423.32	6.096	7	.000

Los datos del **estudio médico** sirven para investigar si disminuye significativamente la presión después del tratamiento. Use los resultados adjuntos para realizar la prueba de hipótesis de interés. Use el histograma de las diferencias para verificar el supuesto de normalidad.



Estadísticos de muestras relacionadas

		Media	N	Desviación típ.	Error típ. de la media
Par 1	Presión_antes	100.060	20	1.1614	.2597
	Presión_después	100.045	20	1.2211	.2730

Prueba de muestras relacionadas

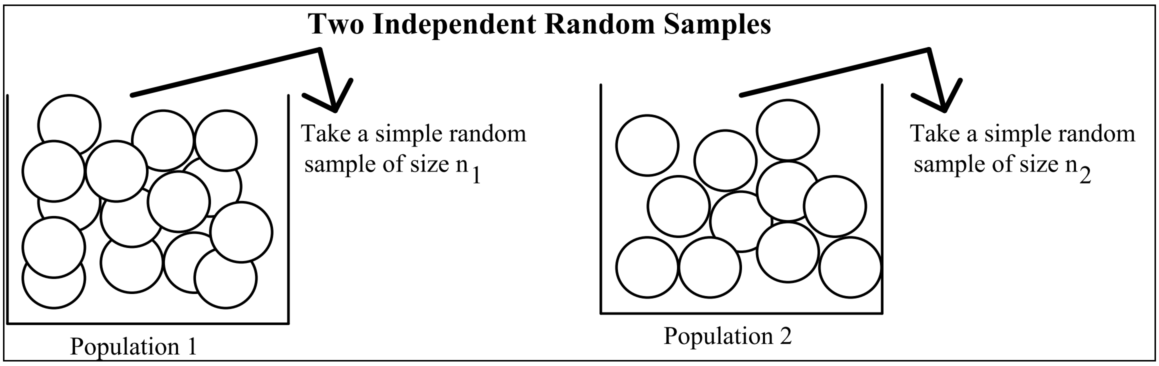
		Diferencias relacionadas							
					95% Intervalo de confianza para la diferencia				
		Media	Desviación típ.	Error típ. de la media	Inferior	Superior	t	gl	Sig. (bilateral)
Par 1	Presión_antes - Presión_después	.015	.7714	.1725	-.346	.376	.087	19	.932

Intervalo de confianza para diferencia entre medias de diseño pareados: A partir de la distribución muestral de las diferencias también podemos construir un intervalo de confianza de la forma:

$$\boxed{\bar{d} \pm t_{1-\alpha/2}(n-1)\left(s_d \sqrt{n}\right)}$$

Análisis para comparar medias de muestras independientes

Supongamos que desea comparar la media de dos poblaciones. Vamos a suponer en primer lugar que la distribución de la variable es normal en ambas poblaciones y con la misma variabilidad σ. Entonces la distribución de la variable en las dos poblaciones sólo puede diferenciarse en la media. Si las medias coinciden, tenemos una única población; si difieren, tenemos dos poblaciones que deben analizarse separadamente.



Para comparar las medias se toman muestras independientes de poblaciones con las siguientes características:

Media población 1= μ_1 Media población 2 = μ_2

Desviación estándar población 1= σ Desviación estándar población 2 = σ

Supuestos para un diseño de dos muestras independientes:

Tenemos una muestra aleatoria simple de n_1 observaciones de una población $N(\mu_1, \sigma)$.
Tenemos una muestra aleatoria simple de n_2 observaciones de una población $N(\mu_2, \sigma)$.

Las dos muestras aleatorias son **independientes**.

Estadísticas muestrales del diseño de dos muestras independientes:

n_1 = tamaño muestra 1
n_2 = tamaño muestra 2
$\bar{x}_1$ = media de la muestra 1
$\bar{x}_2$ = media de la muestra 2
s_1 = desviación estándar de la muestra 1
s_2 = desviación estándar de la muestra 2

Estamos interesados en comparar las medias poblacionales μ_1 y μ_2, el parámetro de interés es la diferencia entre las medias $\mu_1 - \mu_2$.

Para encontrar un test de hipótesis necesitamos conocer la distribución muestral de la diferencia entre las medias muestrales $\bar{X}_1 - \bar{X}_2$.

Distribución muestral de $\bar{X}_1 - \bar{X}_2$ para dos muestras independientes

1. El **promedio** de la diferencia entre medias muestrales es: $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$
2. El **error estándar** de la diferencia entre medias muestrales es: $\sigma_{\bar{X}_1 - \bar{X}_2} = \sigma\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$.
3. La **distribución** de $\bar{X}_1 - \bar{X}_2$ es normal: $\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2, \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$.

La desviación estándar de $\bar{X}_1 - \bar{X}_2$ es $\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$. De los datos tenemos dos desviaciones estándar, s_1 y s_2, las cuáles son estimadores de σ. Entonces necesitamos combinar estas dos estimaciones en un estimador común o global de la desviación estándar. Lo que se usa es un promedio ponderado denominado s_p:

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

Observemos que esta estimación es una media ponderada de las dos varianzas calculadas en cada una de las dos poblaciones. En el caso particular en que $n_1 = n_2$, la estimación s_p será la raíz de la media aritmética de las dos varianzas muestrales:

$$s_p = \sqrt{\frac{s_1^2 + s_2^2}{2}}$$

Como es de suponer al usar estimaciones de la varianza σ ya no podemos usar la distribución Normal, sino que usaremos un test t.

Test *t* para muestras independientes

Supuestos: Se tienen dos muestras independientes de poblaciones con distribución Normal, cada una con media μ_1 y μ_2 desviación estándar σ.

Hipótesis: $H_0 : \mu_1 - \mu_2 = 0$ versus
$H_1 : \mu_1 - \mu_2 \neq 0$ o $H_1 : \mu_1 - \mu_2 > 0$ o $H_1 : \mu_1 - \mu_2 < 0$.

Nivel de significación **α.**

Datos: Se tienen dos conjuntos de datos de los cuales calculamos las medias muestrales $\bar{x}_1$ y $\bar{x}_2$, y las dos desviaciones estándar s_1 y s_2.

Test estadístico observado: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ donde $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ y la distribución bajo la hipótesis nula de la variable t es una *t* de Student con *(n_1+ n_2 - 2)* grados de libertad.

Valor p: Encontramos el valor p usando la distribución *t(n_1+ n_2 - 2)*. La dirección del extremo depende de la dirección de la hipótesis alternativa.

Decisión: Si el **valor *p*** *es menor o igual a* α rechazamos H_0.

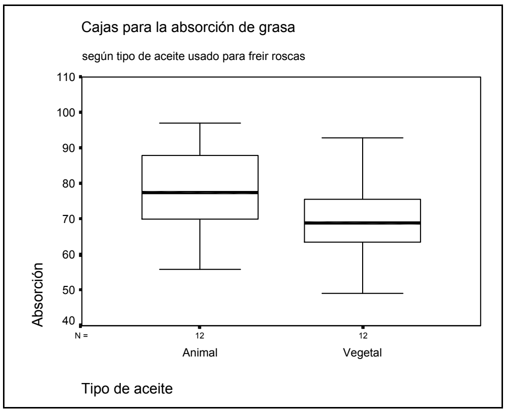
Intervalo de confianza para $\mu_1 - \mu_2$: A partir de la distribución muestral de las diferencias de medias también podemos construir un intervalo de confianza de la forma:

$$\boxed{(\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2}(n_1 + n_2 - 2)\left(s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)}$$

☑ Tipo de aceite

Se desea saber si la grasa que es absorbida por roscas fritas difiere entre 2 tipos de aceites (animal y vegetal). Para ello se frieron 12 roscas en los 2 tipos de aceites de interés. Determine si las diferencias de absorción observadas entre los 2 tipos de aceite son estadísticamente significativas.

Animal	64 72 68 77 56 95 78 91 97 82 85 77
Vegetal	66 49 64 70 68 75 93 78 71 63 76 58



Estadísticos de grupo

	Tipo de grasa	N	Media	Desviación típ.	Error típ. de la media
Absorción	Animal	12	78.50	12.428	3.588
	Vegetal	12	69.25	11.030	3.184

Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
									95% Intervalo de confianza para la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	Inferior	Superior
Absorción	Se han asumido varianzas iguales	.310	.583	1.928	22	.067	9.25	4.797	-.698	19.198
	No se han asumido varianzas iguales			1.928	21.694	.067	9.25	4.797	-.706	19.206

Supuestos de la prueba *t* para medias independientes:

1. **Independencia**

Los grupos o tratamientos son independientes entre sí. Por ejemplo en un diseño experimental, los tratamientos son asignados a grupos de personas asignados al azar. Este supuesto es parte del diseño experimental. En caso de que el estudio sea observacional se verifica en los datos.

2. **Normalidad**

La primera suposición en una prueba t para medias independientes es igual a la suposición en cualquier prueba *t*: se supone que las distribuciones poblacionales son normales. En la práctica, esto implica un problema sólo si se considera que las dos poblaciones tienen distribuciones marcadamente asimétricas y en direcciones opuestas. En general, la prueba t se aplica bastante bien en la práctica, aun cuando las formas de las distribuciones poblacionales sean moderadamente diferentes de la curva normal (se dice que la prueba t es estadísticamente robusta).

En SPSS, se verifica normalidad haciendo gráficos y pruebas de hipótesis. Para la respuesta de cada tratamiento o grupo, construya un histograma o tallo-y-hoja y verifique que no exista un sesgo pronunciado. Para tamaños de grupos n_i pequeños, estos gráficos serán de poca utilidad.

SPSS realiza dos test estadísticos para verificar normalidad, el test de Kolmogorov-Smirnov y el test de Shapiro-Wilk. El test de Kolmogorov-Smirnov es un test clásico y conocido. El test de Shapiro-Wilk es más nuevo y recomendado para tamaños muestrales mayores a 50. En todo caso, se espera que las conclusiones con cualquiera de los dos test sean las mismas.

La hipótesis será:

$$H_0 : \text{la respuesta del grupo o tratamiento } i \text{ es normal}$$
$$H_1 : \text{la respuesta del grupo o tratamiento } i \text{ NO es normal}$$

Por lo tanto si el valor p del correspondiente test es mayor que el nivel de significación, 0,05, aceptamos la hipótesis nula y concluimos que se cumple el supuesto de Normalidad. Note que en este caso especial la hipótesis de interés es la hipótesis nula.

3. **Homogeneidad de varianzas** (Homocedasticidad)

Si las desviaciones estándar de la población no son conocidas pero además no las podemos asumir iguales podemos usar el test estadístico: $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$.

La distribución muestral bajo la hipótesis nula es una distribución t pero con unos grados de libertad aproximados: $gl = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

Pero por suerte tenemos softwares que ya están programados para hacer estos cálculos.

El programa SPSS realiza un test para verificar la homogeneidad de varianzas (Test de **Levene**[1]):

Hipótesis	Test Estadístico	Distribución bajo Ho
$H_0 : \sigma_1^2 = \sigma_2^2$ $H_1 : \sigma_1^2 \neq \sigma_2^2$	F	**F de Fisher** con grados de libertad

4. **Otros tests**

Si no se cumplen los supuestos, existen tests estadísticos alternativos que no hacen supuestos acerca de la distribución de los estimadores, conocidos como métodos **no paramétricos**.

Pasos para resolver problema de comparación de medias:

1. **Descripción:** Lo primero en cualquier análisis de datos es la descripción gráfica y numérica de estos. También es importante verificar la consistencia de los datos, es decir que los datos estén bien digitados y sean valores reales.

2. **Verificación de supuestos:** En el caso de compraración de medias, asumiendo independencia entre los grupos los supuestos a verificar son:

   Normalidad: Debemos verificar este supuesto en ambos grupos.

   Homogeneidad de varianzas, es decir comparar las varianzas de ambos grupos, para lo cúal usamos el test de Levene

3. Si se acepta el supuesto de homogeneidad de varianza entonces podemos usar el test t de comparación de medias, asumiendo varianzas iguales.

4. Si NO se acepta el supuesto de homogeneidad de varianza entonces debemos usar el test t de comparación de medias, que NO asume varianzas iguales.

[1] Levene, H. (1960). In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin et al. eds., Stanford University Press, pp. 278-292.

Solución para el problema de las roscas:

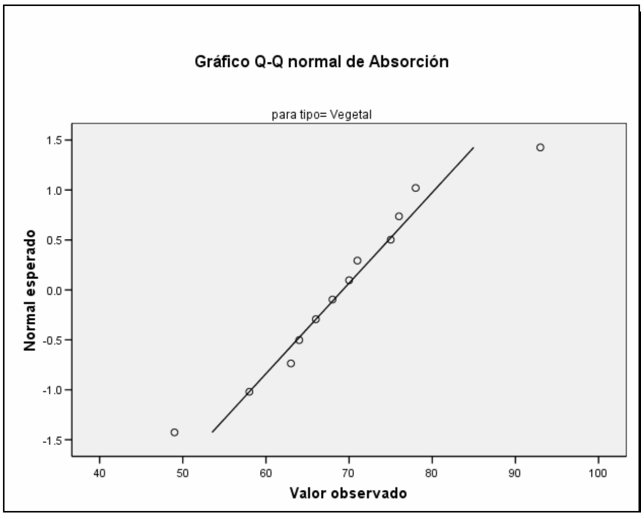
Ya tenemos la descripción. Luego verificamos los supuestos. En SPSS se puede revisar el supuesto de **normalidad** por grupos con: Analizar, Estadísticos descriptivos, Explorar, Gráficos, Gráficos con prueba de normalidad (ver ayuda SPSS, Anova)

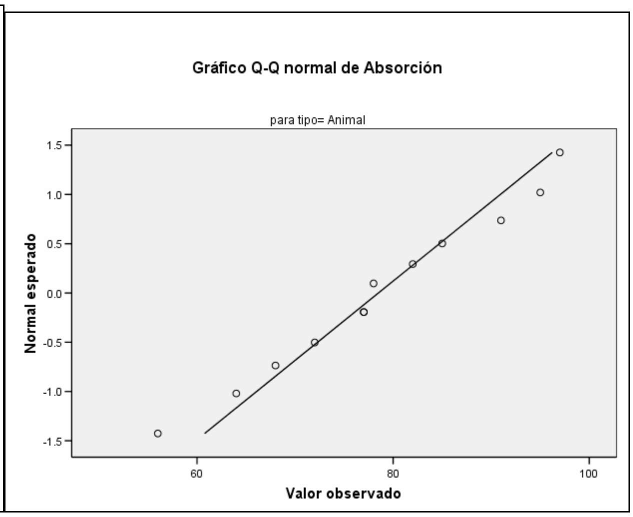
Pruebas de normalidad

	Tipo de aceite	Kolmogorov-Smirnov[a] Estadístico	gl	Sig.	Shapiro-Wilk Estadístico	gl	Sig.
Absorción	Animal	.119	12	.200*	.975	12	.958
	Vegetal	.130	12	.200*	.972	12	.926

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

a. Corrección de la significación de Lilliefors





Prueba de muestras independientes

		Prueba de Levene para la igualdad de varianzas F	Sig.	Prueba T para la igualdad de medias t	gl	Sig. (bilateral)	Diferencia de medias	Error típ. de la diferencia	95% Intervalo de confianza para la diferencia Inferior	Superior
Absorción	Se han asumido varianzas iguales	.310	.583	1.928	22	.067	9.25	4.797	-.698	19.198
	No se han asumido varianzas iguales			1.928	21.694	.067	9.25	4.797	-.706	19.206

Test homogeniedad de varianzas:

$$H_0 : \sigma_1^2 - \sigma_2^2 = 0$$
$$H_0 : \sigma_1^2 - \sigma_2^2 \neq 0$$

De la salidad de SPSS, obtenemos: Test estadístico observado: F=0,31, valor-p=0,583 > 0,05

Por lo tanto concluímos que se cumple el supuesto de homocedasticidad.

Finalmente, procedemos con la **hipótesis de interés**:

$$\begin{array}{|l|}H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \\ \hline\end{array}$$

Estadístico t=1,928, valor-p=0,067, por lo tanto no tenemos suficiente evidencia para concluir que sea distinta la absorción de grasa.

Tabla resumen:

Situación	**Parámetro(s)**	**Test Estadístico**	**Intervalo de Confianza**
Comparar **medias** de muestras **pareadas**	Media de las diferencias μ_d	$t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$	$\bar{d} \pm t_{1-\alpha/2}(\text{n - 1})\left(s_d / \sqrt{\text{n}}\right)$
Comparar **varianzas** de muestras independientes	Diferencia entre varianzas $\sigma_1^2 - \sigma_2^2$	Test F de Levene	
Comparar **medias** de muestras independientes (**homocedasticidad**)	Diferencia entre medias $\mu_1 - \mu_2$	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2}(n_1 + n_2 - 2)\left(s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\right)$
Comparar **medias** de muestras independientes (**heterocedasticidad**)	Diferencia entre medias $\mu_1 - \mu_2$	$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$(\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2}(gl)\left(\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}\right)$

Donde: $s_p = \sqrt{\dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$ y $gl = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$