Trabajo Práctico Nº6:

“TEST DE HIPÓTESIS”

Ejercicio 6.1:

La producción media anual de manzanas de una zona del valle del Río Negro es una variable aleatoria que se distribuye en forma aproximadamente normal con esperanza 90,4 Tm/ha y desviación estándar 6 Tm/ha. En el contexto de la posible adquisición de un nuevo fertilizante, se registró la producción de manzanas de 16 parcelas tratadas con el fertilizante y se obtuvo una media de 94,3 Tm/ha.

a) Decidir, con un nivel de significación del 5%, si con el nuevo fertilizante se obtiene un rendimiento medio mayor que el histórico. Suponer que la desviación estándar de la población no se modificó sensiblemente.

b) Calcular el mínimo nivel de significación con el cual se rechaza la hipótesis nula.

c) Discutir la conveniencia de calcular el nivel justo de significación.

d) ¿Cuál sería la decisión de la prueba si la desviación estándar poblacional hubiera sido de 15 Tm/ha?

a) Planteamos la variable aleatoria

X: La producción media anual de manzanas de una zona del valle del Río Negro
Que sigue una distribución normal **N(μ; σ)**

Planteamos el test de hipótesis:
Teniendo en cuenta que conocemos σ, por lo tanto, se tratara de un test de Gauss y “si con el nuevo fertilizante se obtiene un rendimiento medio mayor que el histórico”:

Tests unilateral derecho

Hipótesis alternativa (H_1): $\mu > \mu_0$
Hipótesis nula (H_0): $\mu \leq \mu_0$

siendo $\mu_0 = 90,4\, Tm/ha$

$$(H_1): \mu > 90,4$$
$$(H_0): \mu \leq 90,4$$

El paso a seguir será estimar μ, a tal efecto estandarizamos $\overline{X}$ suponiendo que H_0 es cierta:

$$\overline{X} \sim N(90,4;\ \tfrac{6}{\sqrt{16}}) \Rightarrow Z\ \frac{\overline{X} - 90,4}{\frac{6}{\sqrt{16}}} \sim\ N(0;\ 1)$$

De esta manera podemos calcular el estadístico de prueba:

$$Z_m = \frac{\bar{x} - 90,4}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x} = 94,3 \Rightarrow Z_m = \frac{94,3 - 90,4}{\frac{6}{\sqrt{16}}}$ = 2,6

Sabiendo que trabajamos con un nivel de significación del 5% ($\alpha = 0,05$). Entonces nuestra zona de rechazo es:

$$Z.R. = \left\{ Z_m \in R : \left| Z_m \right| \geq Z_{\frac{0,05}{2}} = 1,645 \right\}$$

Entonces, nos quedó que el estadístico de prueba es 2,6: $\left| Z_m \right| = 2,6 > 1,645$. Con lo cual, se puede decir que se rechaza H_0 con un nivel de significación del 5% (habiendo evidencia de que con el nuevo fertilizante, se obtiene un rendimiento medio mayor que el histórico).

b) Calculamos el p-valor (con el cual podemos rechazar H_0 , en el caso de que sea menor que α)

$$p = P(Z \geq Z_m)$$

$$p = P(Z \geq 2,6) = 1 - P(Z \leq 2,6) = 1 - 0,9953 = \mathbf{0,0047}\ \%$$

Por lo tanto, el p-valor es menor que α . Entonces, decimos que rechazamos H_0 con un nivel justo de significación del 0,0047%

c) En el caso de que el desvío estándar poblacional hubiera sido de 15 Tm/ha:

$\bar{x} = 94,3 \Rightarrow Z_m = \frac{94,3 - 90,4}{\frac{15}{\sqrt{16}}}$ = 1,04

Por lo tanto, el estadístico de prueba no hubiera caído en la zona de rechazo ($\left| Z_m \right| = 1,04 < 1,645$) por ende, no se rechaza la hipótesis nula con un nivel del 5%.

Mediante el cálculo del p-valor:

$$p = P(Z \geq 1,04) = 1 - P(Z \leq 1,04) = 1 - 0,8508 = \mathbf{0,1492}\ \%$$

$p > \alpha$ no se rechaza H_0

Ejercicio 6.2:

El aumento semanal de peso de los pollos de un criadero se considera una variable aleatoria distribuida normalmente con desviación estándar 9 g. Un establecimiento que fabrica alimentos para ese criadero estudia la posibilidad de agregar un nuevo producto al alimento tradicional, que se considerará efectivo si con él se obtiene un aumento promedio de peso de por lo menos 200 g en una semana. El establecimiento no desea perder la oportunidad de fabricar el alimento si razonablemente se puede suponer que el nuevo producto es efectivo. Determinar:

a) Las hipótesis que planteará.

b) El estadístico de prueba y la zona de rechazo si trabaja con α = 0,05.

c) La decisión que tomará (¿se fabricará el alimento?) si una muestra aleatoria de 10 pollos de ese criadero alimentados con el alimento tradicional al que se agregó el nuevo producto produjo una media $\bar{x}$ = 196 g.

d) ¿Se puede informar la probabilidad de equivocarse al tomar la decisión?

e) Si ahora la premisa es: "Como el establecimiento se encuentra en etapa de control de presupuesto, sólo efectuará la inversión del cambio de producción si puede probar, con pequeño error, que el nuevo producto es realmente efectivo" ¿qué hipótesis se hubieran planteado?

a) Planteamos la variable aleatoria

X: aumento semanal en lo peso de los pollos de un criadero (gr)
Que sigue una distribución normal **N(μ; σ)**

Planteamos el test de hipótesis:
Teniendo en cuenta que conocemos σ, por lo tanto, se tratara de un test de Gauss y ***"el establecimiento no desea perder la oportunidad de fabricar el alimento si razonablemente se puede suponer que el nuevo producto es efectivo"*** con lo cual lo que queremos saber (lo que vamos a investigar) si esta suposición es falsa:

Tests unilateral izquierdo

Hipótesis nula sería que, si realmente es efectivo el producto **(H_0):** $\mu \geq \mu_0$
Hipótesis alternativa sería demostrar que el producto no es efectivo **(H_1):** $\mu < \mu_0$

siendo $\mu_0 = 200\,g$

(H_0): $\mu \geq 200$
(H_1): $\mu < 200$

b) De esta manera podemos calcular el estadístico de prueba:

$$Z_m = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Zona de rechazo: $Z_m \leq -Z_\alpha$ (osea si $Z \leq -1{,}645$)

c) $\bar{x} = 196 \Rightarrow Z_m = \frac{196 - 200}{\frac{9}{\sqrt{10}}}$ = -1,405

Calculamos el p-valor para ver si rechazamos o no la hipótesis nula (H_0):
p = P(Z ≤ -1,405) = P(Z ≥ 1,04) = 1 - 0,8508 = **0,1492** %

$p > \alpha$ no se rechaza H_0 por lo tanto, se fabricará el alimento.

pipiMorch

d) En el caso de equivocarnos se estaría cometiendo un error de tipo II (el cual se comete cuando H_0 es falsa pero no la podemos rechazar por falta de evidencia). Pero esto no ocurre en nuestro caso, ya que tenemos la evidencia para no el rechazo de H_0 proporcionada por el cálculo del p-valor.

e) ***"Como el establecimiento se encuentra en etapa de control de presupuesto, sólo efectuará la inversión del cambio de producción si puede probar, con pequeño error, que el nuevo producto es realmente efectivo"***

Hipótesis nula sería que, el producto no es efectivo por lo tanto la producción queda como estaba **(H_0):** $\mu \leq \mu_0$

Hipótesis alternativa sería demostrar que el producto es realmente efectivo con un pequeño error **(H_1):** $\mu > \mu_0$

Ejercicio 6.3:

El contenido de vitamina C de un alimento balanceado producido por un establecimiento líder A es una variable aleatoria que se distribuye normalmente con media 180 mg/kg y desviación estándar 38 mg/kg.
Otra fábrica B saca a la venta un producto equivalente. Se quiere poner a prueba la hipótesis de que el contenido medio de vitamina C del *alimento nuevo difiere de la del establecimiento líder*. Para ello, se toma una muestra de 16 paquetes de B y se obtiene una media de 202 mg/kg. Suponiendo que la varianza del contenido de vitamina C del alimento producido por el establecimiento B también es de 38 mg/kg:

a) Calcular el nivel justo de significación para decidir si los contenidos medios de vitamina C de los alimentos producidos por ambos establecimientos difieren significativamente.

b) Decidir si se rechaza la hipótesis nula, si se declara: "Se considerará diferencia significativa si la probabilidad de error es menor del 5%".

c) Igual que en b) pero ahora "Se considerará significativa toda probabilidad de error menor del 1%".

a) Planteamos la variable aleatoria

X: contenido de vitamina C de un alimento balanceado producido por un establecimiento líder A
Que sigue una distribución normal **N(μ; σ)**

Planteamos el test de hipótesis:
Teniendo en cuenta que conocemos σ, por lo tanto, se tratara de un test de Gauss y "***Se quiere poner a prueba la hipótesis de que el contenido medio de vitamina C del alimento nuevo difiere de la del establecimiento líder***":

Tests Bilateral

Hipótesis nula (H_0): $\mu = \mu_0$

Hipótesis alternativa (H_1): $\mu \neq \mu_0$

siendo $\mu_0 = 180\, mg/kg$

$\bar{x} = 202 \Rightarrow Z_m = \frac{202 - 180}{\frac{38}{\sqrt{16}}} = \frac{22}{9,5} = 2,316$

Z. R. = $\left\{Z_m \in R : |Z_m| \geq Z_{\frac{\alpha}{2}}\right\}$

Calculamos el p-valor:

$$p = P(|Z| \geq |Z_m|) = P(|Z| \geq 2,316) = 2 \cdot (1 - 0,9898) = \mathbf{0,0204}\ \%$$

$p < \alpha$ se rechaza H_0

b) Se rechaza H_0 ($p < \alpha$ se rechaza H_0)

c) No se rechaza H_0 ($p > \alpha$ no se rechaza H_0)

Ejercicio 6.4:

Se realizó una experiencia para determinar el efecto que produce una droga sobre la concentración de alcohol en la sangre de ratas en el momento de producirse cierta alteración respiratoria. Se midió esa concentración en 16 ratas tratadas con la droga y se obtuvo una media de 8,3 mg/ml y una desviación estándar de 1,2 mg/ml. Se quiere verificar si la concentración media de alcohol en la sangre de las ratas tratadas con la droga es inferior a 8,9 mg/ml.

a) Indicar las variables aleatorias y las suposiciones acerca de ellas.

b) Plantear la hipótesis nula y la alternativa.

c) Determinar el estadístico de prueba y la zona de rechazo para un nivel de significación del 5%.

d) Tomar la decisión utilizando los valores de la muestra y expresar esa decisión en términos del problema.

LA VARIANZA ES DESCONOCIDA

a) Planteamos las variables aleatorias:

X: concentración de alcohol en la sangre de ratas tratadas con la droga

Sigue una distribución t de Student (*n-1*) cuando H_0 es cierta

b) *Test unilateral izquierdo*

Hipótesis nula (H_0): $\mu \geq \mu_0$

Hipótesis alternativa (H_1): $\mu < \mu_0$

siendo $\mu_0 = 8,9$

c)

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

tenemos un nivel de significación del 5% $(\alpha = 0,05)$

$$T = \frac{8,3 - 8,9}{\frac{1,2}{\sqrt{16}}} = \frac{-0,6}{0,3} = -2$$

La zona de rechazo correspondiente, ya que el nivel de significación es del 5%:

$$Z.R. = \{T_m \in R : T_m \leq -t_{n-1,\alpha}\}$$

es de cola izq.

$$Z.R. = \{T_m \in R : T_m \leq -t_{15,\,0.05} = -1,753\}$$

	α(una cola)											
n	0,2	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
11	0,876	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437	4,863	5,453	11
12	0,873	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318	4,716	5,263	12
13	0,870	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221	4,597	5,111	13
14	0,868	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140	4,499	4,985	14
15	0,866	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073	4,417	4,880	15

d) Evidentemente se rechaza H_0 porque -2 > -1,753 (entonces hay evidencia significativa de que la concentración de alcohol en la sangre de ratas tratadas con la droga es inferior a la media planteada).

Ejercicio 6.5:

El contenido medio de los frascos de cierto antibiótico debe ser 1,5 g. Un laboratorio midió el contenido de antibiótico para una muestra de 9 frascos envasados automáticamente y obtuvo una media de 1,35 g y un desvío estándar de 0,159 g.

a) Obtener un intervalo de confianza del 95% para el contenido medio.

b) Utilizando a) decidir, con un nivel de significación del 5%, si el contenido de antibiótico de los frascos envasados automáticamente cumple la especificación pedida.

Planteamos la variables aleatoria:

X: contenido medio de frascos con cierto antibiótico

Se supone que sigue una distribución normal **N(μ; σ)**

a)

NO CONOCEMOS LA VARIANZA

$$\overline{X}_9 \sim N(1,35;\ \frac{0,159}{\sqrt{9}})$$

$$C(\overline{x} - t_{\frac{\alpha}{2};\, n-1}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\frac{\alpha}{2};\, n-1}\frac{s}{\sqrt{n}}) = 0,95$$

$$C(1,35 - t\frac{0,05}{2};\ 8\frac{0,159}{\sqrt{9}} < \mu < 1,35 + t\frac{0,05}{2};\ 8\frac{0,159}{\sqrt{9}}) = 0,95$$

$$C(1,35 - 0,122218 < \mu < 1,35 + 0,122218) = 0,95$$

$$C(1,23 < \mu < 1,47) = 0,95$$

	or.(dos colas)												
n	0,2	0,1	0,05	0,025	0,02	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
6	1,440	1,943	2,447	2,969	3,143	3,707	4,317	4,981	5,959	6,788	7,708	9,082	6
7	1,415	1,895	2,365	2,841	2,998	3,499	4,029	4,595	5,408	6,082	6,814	7,885	7
8	1,397	1,860	2,306	2,752	2,896	3,355	3,833	4,334	5,041	5,617	6,234	7,120	8
9	1,383	1,833	2,262	2,685	2,821	3,250	3,690	4,146	4,781	5,291	5,830	6,594	9
10	1,372	1,812	2,228	2,634	2,764	3,169	3,581	4,005	4,587	5,049	5,533	6,211	10

b)

Tests de Student unilateral izquierdo

Hipótesis alternativa será que no se cumple que el contenido medio del frasco tenga 1,5 g

$$(H_1): \mu < \mu_0$$

Hipótesis nula será que el contenido de antibiótico de los frascos envasados cumple con los 1,5 g

$$(H_0): \mu \geq \mu_0$$

siendo $\mu_0 = 1,5\ gramos$

$$(H_1): \mu < 1,5$$
$$(H_0): \mu \geq 1,5$$

Planteamos el estadístico de prueba:

$$T = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

tenemos un nivel de significación del 5% $(\alpha = 0,05)$

$$T = \frac{1,35 - 1,5}{\frac{0,159}{\sqrt{9}}} = \frac{-0,15}{0,053} = -2,830$$

La zona de rechazo correspondiente, ya que el nivel de significación es del 5%:

$$Z.R. = \{T_m \in R : T_m \leq -t_{n-1,\alpha}\}$$

es de cola izq.

$$Z.R. = \{T_m \in R : T_m \leq -t_{8,\,0.05} = -1,860\}$$

	α(una cola)											
n	0,2	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
6	0,906	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959	6,788	8,025	6
7	0,896	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408	6,082	7,063	7
8	0,889	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041	5,617	6,442	8
9	0,883	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781	5,291	6,010	9

Evidentemente se rechaza H_0 porque -1,860 > -1,753 (entonces hay evidencia significativa de que el contenido de antibiótico de los frascos envasados automáticamente no cumple la especificación pedida).

Ejercicio 6.6:

En el marco de un estudio sobre hidroarsenicismo crónico regional endémico (HACRE), se tomaron muestras de agua en dos partidos de la provincia de Córdoba y se midió la concentración de Arsénico (As) en g/litro. En 18 muestras del partido de Unión se obtuvo una media de 120 g/litro y en 20 muestras del partido de San Justo se obtuvo una media de 150 g/litro.
Se supone que el contenido de As sigue una distribución aproximadamente normal con varianzas de 1600 (g/litro)^2 para Unión y 2025 (g/litro)^2 para San Justo.
¿Se puede afirmar que existe diferencia significativa entre los contenidos medios de As de los dos partidos?
Calcular el valor P (mínimo nivel de significación con el cual se rechaza H_0) y tomar la decisión sobre la base de un nivel de significación del 5%.

Planteamos las variables aleatorias:

X_1: concentración de Arsénico (As) en g/litro en muestra de agua del partido de Unión

X_2: concentración de Arsénico (As) en g/litro en muestra de agua del partido de San Justo

Siguen una distribución normal **N(μ; σ)**

Planteamos el test de Gauss para la diferencia de medias *(TEST Bilateral)*

VARIANZAS CONOCIDAS

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0;\ 1)\ si\ vale\ H_0$$

La hipótesis nula será que no hay diferencias significativas entre los contenidos medios de As en las muestras y la hipótesis alternativa será que los contenidos medios de As difieren.

Nos queda:

$$H_1: \mu_1 \neq \mu_2$$
$$H_0: \mu_1 = \mu_2$$

Los datos que nos aporta el enunciado son:

$\overline{x}_1 = 120\ g/L$

$\overline{x}_2 = 150\ g/L$

$\sigma_1^2 = 1600\ (g/L)^2$

$\sigma_2^2 = 2025\ (g/L)^2$

$n = 18$

$m = 20$

$$Z_m = \frac{120 - 150}{\sqrt{\frac{1600}{18} + \frac{2025}{20}}} = \frac{-30}{13{,}789} = \mathbf{-2{,}17}$$

Calculamos el p-valor:

$$p = P(|Z| \geq |Z_m|)$$

$$p = P(|Z| \geq -2,17) = 2 \cdot (1 - 0,9850) = \mathbf{0{,}03}$$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857

$p < \alpha$ por lo cual se rechaza H_0, podemos decir que entonces, existe una diferencia en los contenidos medios de las muestras de As.

Ejercicio 6.7:

Sea X una variable aleatoria que sigue una <u>distribución F de Fisher</u>. Mediante el uso de tablas hallar:
a) a para que $P(X > a) = 0,10$; si $n_1 = 30$, $n_2 = 40$.

P(X > a) = 0,10 → a = **1,54**

Grados de libertad en el numerador (n_1)

Grados de libertad en el denominador (n_2)	α	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30
29	0,10	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62
29	0,05	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85
29	0,01	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41
30	0,10	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61
30	0,05	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84
30	0,01	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39
40	0,10	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54
40	0,05	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74

Ejercicio 6.8:

Un laboratorio produce una nueva droga que, se supone, disminuye el nivel de colesterol en sangre. Como esa droga puede tener efectos adversos de cierta gravedad, sólo se continuará investigando sobre ella si se puede admitir su efectividad con probabilidad de error pequeño. Para decidir sobre la cuestión, efectuaron un estudio preclínico en conejos de la siguiente manera: 28 conejos se mantuvieron durante un tiempo prudencial con una dieta hipercolesterolémica. Luego, a 13 de ellos tomados como control se les midió el nivel de colesterol en sangre y se obtuvo una media de 241,5 mg/dl y una varianza de 1264 $(mg/dl)^2$. Al grupo restante de 15 conejos se los trató con la droga y se obtuvo una media de 220,9 mg/dl y una varianza de 875 $(mg/dl)^2$.

a) Identifique las variables aleatorias y las suposiciones necesarias para poder efectuar el análisis. Plantear las hipótesis: nula y alternativa, atento a la condición impuesta, definir el estadístico de prueba y la zona de rechazo si se fija el nivel de significación en 1%.

b) Efectuar el test correspondiente para verificar si se cumple la homogeneidad de varianzas.

c) Si se verifica la homogeneidad de varianzas, efectuar el test de Student para muestras independientes para decidir si el laboratorio continuará investigando sobre la droga.

d) ¿Cuál es la decisión en cuanto a administrar la droga si el nivel de significación se fija en 1%?

e) ¿Le parece adecuado el diseño de la prueba? ¿Qué propondría?

Planteamos la variable aleatoria:

X_1: disminución del nivel de colesterol en sangre (mg/dl) sin aplicar la droga

X_2: disminución del nivel de colesterol en sangre (mg/dl) aplicando la droga

X_1 y X_2 son independientes → $X_1 \sim N(\mu_1, \sigma_1)$, $X_2 \sim N(\mu_2, \sigma_2)$

Suponemos que siguen una distribución t de Student $(n + m - 2)$ grados de libertad

Planteamos el test de Student para diferencias de medias *(Test Unilateral derecho)*

LOS DESVÍOS SON DESCONOCIDOS

La hipótesis alternativa será comprobar la efectividad de la droga, es decir que el nivel de colesterol con posibles efectos adversos de μ_2 debe ser menor al de μ_1 y la **hipótesis nula** será que la droga no tiene efecto; por ende no hay disminución.

Nos queda:

$$H_1: \underbrace{\boxed{\mu_1}}_{\text{sin aplicar la droga}} > \underbrace{\boxed{\mu_2}}_{\text{aplicando la droga}}$$

$$H_0: \mu_1 \leq \mu_2$$

Datos:

$\bar{x}_1 = 241,5\ mg/dl$

$\bar{x}_2 = 220,9\ mg/dl$

$s_1^2 = 1264\ (mg/dl)^2$

$s_2^2 = 875\ (mg/dl)^2$

$n = 13$

$m = 15$

Chequeamos la Homocedasticidad (si las varianzas son o no homogéneas):

Para verificar este supuesto se pone a prueba:

$$H_0: \sigma_1 = \sigma_2 \qquad H_1: \sigma_1 \neq \sigma_2$$

Calculamos el estadístico de prueba:

$$F_m = \frac{S^2_{Mayor}}{S^2_{Menor}} \sim F_{n\,Mayor - 1\,;\ m\,Menor - 1} \quad si\ vale\ H_0$$

$$F_m = \frac{1264}{875} = 1,44$$

Si consideramos un $\alpha = 0,20$

$$F_{14,\,12;\,\frac{0,20}{2}} = 2,05$$

Grados de libertad en el numerador (n_1)

Grados de libertad en el denominador (n_2)	α	1	2	3	4	5	6	7	8	9	10	12
9	0,10	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38
9	0,05	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07
9	0,01	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11
10	0,10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28
10	0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91
10	0,01	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71
11	0,10	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21
11	0,05	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79
11	0,01	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40
12	0,10	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15
12	0,05	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69
12	0,01	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16
13	0,10	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10
13	0,05	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60
13	0,01	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96
14	**0,10**	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	$\boxed{2,05}$
14	0,05	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53
14	0,01	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80

Se rechaza la homogeneidad de varianzas con un nivel $\alpha = 0.20$ si:

$$Z.R. = \left\{F_m \in R: F_m \geq F_{n\,Mayor - 1,\, n\,menor - 1;\, \alpha/2}\right\}$$

Entonces, siendo 1,44 < 2,05 no se rechaza H_0

Podemos suponer que ambas variables ***son homogéneas*** $\sigma_1 \approx \sigma_2$

Una vez comprobada la homogeneidad , podemos seguir con el respectivo test

$$T = \frac{\overline{X}_1 - \overline{X}_2}{Sp\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \quad si\ vale\ H_0$$

$$Sp^2 = \frac{(n-1)\cdot S^2_{x1} + (m-1)\cdot S^2_{x2}}{n+m-2} = \frac{(13-1)\cdot 1264 + (15-1)\cdot 875}{13+15-2} = \frac{27418}{26} = 1054,5$$ ***aplicamos √Sp²***

Planteamos el estadístico de prueba:

$$T_m = \frac{241,5 - 220,9}{32,47\sqrt{\frac{1}{13} + \frac{1}{15}}} = \frac{20,6}{12,30} = \mathbf{1,674}$$

Definimos la zona de rechazo como: $T_m \geq t_{s, \alpha}$

Entonces, cómo se utiliza un $\alpha = 0,01$ se observa que:

$$1,674 \in \left\{T_m \in R : T_m \geq t_{26,\, 0.01} = 2,479\right\}$$

n	α(una cola) 0,2	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
26	0,856	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707	3,974	4,324	26
27	0,855	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,690	3,954	4,299	27
28	0,855	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674	3,935	4,275	28
29	0,854	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,659	3,918	4,254	29

No se rechaza la hipótesis nula (H_0)

También podemos calcular el p-valor:

$p = P(t_s \geq T_m) = P(t \geq 1,674) = 0,05$ *p-valor* > α

Conclusión: No se rechaza la hipótesis nula ya que la evidencia es demasiado fuerte para hacerlo. Por ende, la droga no disminuye el nivel de colesterol en sangre, por lo que no se continuará investigando sobre ella.

Ejercicio 6.9:

La empresa Eco andino se lanza al mercado a comercializar agua mineral sin agregado de sodio. Afirma que el contenido de sodio de su agua es significativamente menor que el de la empresa líder IceG. Para demostrarlo, toma 20 botellas de agua de cada empresa y mide los contenidos de sodio de ambas muestras. Para el agua IceG obtiene una media de 11,06 mg/dL y una desviación estándar de 2,26 mg/dL; para el agua Eco andino, obtiene una media de 9,25 mg/dL con una desviación estándar de 3,85 mg/dL.
Testear la homogeneidad de varianzas.

Está demás

Planteamos la variable aleatoria:

X_1: comercialización de agua mineral sin agregado de sodio empresa líder IceG
X_2: comercialización de agua mineral sin agregado de sodio empresa Ecp andino

X_1 y X_2 son independientes → $X_1 \sim N(\mu_1, \sigma_1)$, $X_2 \sim N(\mu_2, \sigma_2)$

Suponemos que siguen una distribución t de Student $(n + m - 2)$ grados de libertad

Planteamos el test de Student para diferencias de medias *(Test Unilateral derecho)*

LAS VARIANZAS SON DESCONOCIDAS

La hipótesis alternativa es que la nueva empresa Eco lanza un agua cuyo contenido de sodio es significativamente menor al de la empresa líder.

Nos queda:

$$H_1: \underbrace{\boxed{\mu_1}}_{\text{empresa líder}} > \underbrace{\boxed{\mu_2}}_{\text{empresa Eco}}$$

$$H_0: \mu_1 \leq \mu_2$$

Testeamos la homogeneidad de varianzas:

$$H_0: \sigma_1 = \sigma_2$$

$$H_1: \sigma_1 \neq \sigma_2$$

$$\sigma_2 = 3,85\, mg/dl \qquad \sigma_1 = 2,26\, mg/dl$$

$$F_m = \frac{\sigma^2_{Mayor}}{\sigma^2_{menor}} \sim F_{n\,Mayor-1\,;\,m\,menor-1} \; si\ vale\ H_0$$

$$F_m = \frac{14,8225}{5,1076} = 2,90$$

Si consideramos un α = 0.20 entonces:

$$F_{19,\,19;\,\frac{0.20}{2}} = 1,81$$

Siendo la zona de rechazo:

$$Z.R. = \left\{F_m \in R: F_m \geq F_{n\,Mayor-1\,;\,m\,menor-1;\,\alpha/2}\right\}$$

Se rechaza H_0, por lo tanto se puede suponer que la dispersión de ambas variables no son homogéneas.

Ejercicio 6.10:

En un estudio sobre 15 mujeres posmenopáusicas sin ningún tipo de tratamiento, 9 de ellas se diagnosticaron con osteoporosis y las restantes 6, sin osteoporosis, se consideraron controles. Se midió la producción de Factor de Necrosis Tumoral (TNF) liberado (ng/ml) en todas las mujeres. En el grupo con osteoporosis resultó una media de 1,33 y un desvío estándar de 0,54; en el grupo control, una media de 2,68 y un desvío estándar de 0,69. ¿Se puede afirmar que hay diferencia significativa entre las concentraciones medias de TNF liberado del grupo control y del grupo con osteoporosis? Basar la decisión en el cálculo del nivel justo de significación, si se declara: "Se considerará significativa toda probabilidad de error menor del 5%".

Se realizó un estudio sobre 15 mujeres posmenopáusicas sin ningún tipo de tratamiento, de las cuales: 9 tienen osteoporosis y 6 no (consideradas control).

Plantamos a la variable aleatoria como:

X_1: concentración media de TNF liberado (ng/ml) en mujeres con osteoporosis
X_2: concentración media de TNF liberado (ng/ml) en mujeres sin osteoporosis

Sabiendo que ambas muestras (ambos grupos de mujeres con y sin osteoporosis) fueron tomadas de un mismo grupo de mujeres, por lo tanto, tiene un "lazo" en común que no las hace independientes:

No importa la distribución de X_1 y X_2 sino la distribución de la diferencia: $D \sim N(\mu_D, \sigma_D)$

Es un test de Student porque <u>no conocemos la varianza</u>

Planteamos el test de hipótesis, sabiendo que lo que se busca es si hay o no una diferencia significativa entre las concentraciones medias:

Test bilateral (Contraste para la diferencia de muestras apareadas)

$$H_0: \mu_D = \mu_0$$

$$H_1: \mu_D \neq \mu_0$$

Datos:

$\bar{x}_1 = 1,33\, ng/ml$

$\bar{x}_2 = 2,68\, ng/ml$

$s_1 = 0,54\, ng/ml$

$s_2 = 0,69\, ng/ml$

$n = 9$

$m = 6$

Calculamos el estadístico de prueba:

$$T = \frac{\bar{D} - \mu_0}{\sqrt{\frac{S_D^2}{n}}}$$

$$S_D^2 = \frac{(n-1)\cdot S_{x1}^2 + (m-1)\cdot S_{x2}^2}{n+m-2} = \frac{(9-1)\cdot 0{,}2916 + (6-1)\cdot 0{,}4761}{9+6-2} = 0,36$$ ***aplicamos √S$_D^2$***

$$T_m = \frac{(2{,}68 - 1{,}33) - 0}{0{,}6\left(\sqrt{\frac{1}{9}+\frac{1}{6}}\right)} = \frac{1{,}35}{0{,}316} = 4{,}272$$

Realizamos el cálculo del p-valor: $p = P(|t_{n-1}| \geq |T_m|)$

$p = P(t_{14} \geq 4,27) = 0,0005 < p - valor < 0,001$

n	0,2	0,1	0,05	0,025	0,02	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
	or.(dos colas)												
11	1,363	1,796	2,201	2,593	2,718	3,106	3,497	3,895	4,437	4,863	5,306	5,921	11
12	1,356	1,782	2,179	2,560	2,681	3,055	3,428	3,807	4,318	4,716	5,128	5,694	12
13	1,350	1,771	2,160	2,533	2,650	3,012	3,372	3,735	4,221	4,597	4,984	5,513	13
14	~~1,345~~	~~1,761~~	~~2,145~~	~~2,510~~	~~2,624~~	~~2,977~~	~~3,326~~	~~3,675~~	~~4,14~~ 4,272	,499	4,865	5,363	14
15	1,341	1,753	2,131	2,490	2,602	2,947	3,286	3,624	4,073	4,417	4,766	5,239	15

Entonces, decimos que p < 0,001 sabiendo que se considerará significativa toda probabilidad de error menor del 5%:

p-valor < α se rechaza H_0

Hay una diferencia significativa entre la concentración de medias

Ejercicio 6.11:

Los datos de esta base se encuentran en el archivo Ej.6.11.
Un investigador midió la disminución de ácidos grasos libres (Eq/l) después de 30 minutos de aplicada una inyección de glucosa, en perros normales y en perros hipertiroideos. Los datos obtenidos fueron:

Normales	10	30	48	60	16	58
Hipertiroideos	79	142	126	50	68	75

El investigador sospecha que la disminución media es mayor en los hipertiroideos que en los normales, y desea confirmar o no su sospecha.

a) Efectuar el test correspondiente para verificar la suposición acerca de las varianzas.

b) Según el resultado de a), realizar el test adecuado indicando el nivel justo de significación y expresar la conclusión en términos del problema.

a)

X_1: disminución de ácidos grasos libres (Eq/l) en perros normales
X_2: disminución de ácidos grasos libres (Eq/l) en perros hipertiroideos

suponemos que X_1 y X_2 son independientes → $X_1 \sim N(\mu_1, \sigma_1)$, $X_2 \sim N(\mu_2, \sigma_2)$

Suponemos que siguen una distribución t de Student $(n + m - 2)$ grados de libertad

Planteamos el test de Homocedasticidad para ver si las varianzas son iguales o no (homogeneidad de varianzas):

$$H_0: \sigma_1 = \sigma_2 \qquad H_1: \sigma_1 \neq \sigma_2$$

Calculamos el estadístico de prueba:

$$F_m = \frac{S^2_{Mayor}}{S^2_{Menor}} \sim F_{n\,Mayor - 1\,;\ m\,Menor - 1} \quad si\ vale\ H_0$$

$$F_m = \frac{1286}{462} = 2,78$$

Si consideramos un $\alpha = 0,20$

$$F_{5,5;\frac{0,20}{2}} = 3,45$$

Grados de libertad en el numerador (n_1)

Grados de libertad en el denominador (n_2)

	α	1	2	3	4	5	6	7	8	9	10	12
1	0,10	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71
1	0,05	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	243,91
1	0,01	4052,18	4999,50	5403,35	5624,58	5763,65	5858,99	5928,36	5981,07	6022,47	6055,85	6106,32
2	0,10	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41
2	0,05	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41
2	0,01	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42
3	0,10	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22
3	0,05	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74
3	0,01	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05
4	0,10	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90
4	0,05	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91
4	0,01	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37
5	0,10	4,06	3,78	3,62	3,52	$\boxed{3,45}$	3,40	3,37	3,34	3,32	3,30	3,27
5	0,05	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68
5	0,01	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89

Siendo la zona de rechazo:

$$Z.R. = \left\{ F_m \in R : F_m \geq F_{n\,Mayor-1\,;\,m\,menor-1;\,\alpha/2} \right\}$$

$F_m = 2,78 \geq 3,45$ No se rechaza H_0 por ende, las varianzas ***son homogéneas***

Test Unilateral derecho (Test de Student para diferencia de medias)

$$H_0: \boxed{\mu_1} > \boxed{\mu_2}$$

perros normales — perros hipertiroideos

$$H_1: \mu_1 \leq \mu_2$$

Datos:

$\bar{x}_1 = 37\ Eq/l$

$\bar{x}_2 = 90\ Eq/l$

$\sigma_1^2 = 462\ Eq/l$

$\sigma_2^2 = 1286\ Eq/l$

$n = 6$

$m = 6$

Calculamos el estadístico de prueba:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{Sp\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2} \ si\ vale\ H_0$$

$$Sp^2 = \frac{(n-1)\cdot S_{x1}^2 + (m-1)\cdot S_{x2}^2}{n+m-2} = \frac{(6-1)\cdot 462 + (6-1)\cdot 1286}{6+6-2} = \frac{8740}{10} = 874$$ ***aplicamos √Sp²***

$$T_m = \frac{37-90}{29,56\sqrt{\frac{1}{6}+\frac{1}{6}}} = \frac{-53}{17,066} = \mathbf{-3,105}$$

Calculamos el p-valor: $p = P(t_s \geq T_m)$ ***siendo s = n + m - 2***

n	α(una cola) 0,2	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
6	0,906	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959	6,788	8,025	6
7	0,896	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408	6,082	7,063	7
8	0,889	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041	5,617	6,442	8
9	0,883	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781	5,291	6,010	9
10	0,879	1,372	1,812	2,228	2,76 3,105	,169	3,581	4,144	4,587	5,049	5,694	10

$$0,0005 \ < \ p - valor \ < \ 0,001$$

p-valor < α se rechaza H_0
La disminución media es mayor en los hipertiroideos que en los normales

Ejercicio 6.13:

Los datos se encuentran en el archivo Ej.6.13.
En un estudio se incorporaron 21 pacientes que consultaron por infertilidad y con diagnóstico de varicocele clínico, y 8 varones con fertilidad reconocida. Se les efectuó un estímulo de GnRH y se midió LH a los 30 y 60 minutos. En todos hubo hiper-respuesta de LH. Los resultados están expresados como incremento máximo respecto del basal. Decidir si el incremento máximo de LH respecto del basal medio en los pacientes con varicocele es mayor que en los controles.

a) Efectuar el test correspondiente para verificar la suposición acerca de las varianzas.

b) Realizar el test adecuado indicando el nivel justo de significación.

Variable	n	Media	D.E.	Var(n-1)
INFÉRTIL	21	616,29	215,09	46262,71
FÉRTIL	8	383,50	106,53	11349,43

Planteamos las variables aleatorias:

X_1: estímulo de GnRH en los pacientes fertilidad reconocida (control)
X_2: estímulo de GnRH en los pacientes con infertilidad y varicocele

Suponemos que siguen una distribución t de Student $(n + m - 2)$ grados de libertad

Planteamos el test de Homocedasticidad para ver si las varianzas son iguales o no (homogeneidad de varianzas):

$$H_0: \sigma_1 = \sigma_2 \qquad H_1: \sigma_1 \neq \sigma_2$$

Calculamos el estadístico de prueba:

$$F_m = \frac{S^2_{Mayor}}{S^2_{Menor}} \sim F_{n\,Mayor - 1\ ;\ \ m\,Menor - 1} \quad si\ vale\ H_0$$

$$F_m = \frac{46262{,}71}{11349{,}43} = 4,076$$

Si consideramos un $\alpha = 0,20$

$$F_{7,20;\frac{0,20}{2}} = 2,59$$

Siendo la zona de rechazo:

$$Z.R. = \left\{F_m \in R: F_m \geq F_{n\,Mayor-1\,;\,m\,menor-1;\,\alpha/2}\right\}$$

$F_m = 4,076 \geq 2,59$ Se rechaza H_0 por ende, ***no se puede suponer que las varianzas son homogéneas***.

Se debe aplicar la corrección de Satterthwaite

Ejercicio 6.16:

Los datos siguientes corresponden a la prueba de tolerancia a la glucosa realizada en 7 pacientes varones acromegálicos:

Paciente	1	2	3	4	5	6	7
Glucemia (mg/dl) en ayunas	81	77	85	82	97	86	78
Glucemia (mg/dl) a los 120 minutos	76	84	83	95	92	97	84

Diferencia ($\bar{d}$)	5	-7	2	-13	5	-11	-6

Se quiere analizar si la glucemia varía al final de la prueba. Hallar el nivel justo de significación e indicar la conclusión en términos del problema.

Tenemos a las variables aleatorias:

X_1: pacientes con glucemia (mg/dl) en ayunas
X_2: pacientes con glucemia (mg/dl) a los 120 minutos

No importa la distribución de X_1 y X_2 sino la distribución de la diferencia: $D \sim N(\mu_D, \sigma_D)$
Es un test de Student

Test bilateral (Contraste para la diferencia de muestras apareadas)

$$H_0: \mu_D = \mu_0$$

$$H_1: \mu_D \neq \mu_0$$

RECORDAR: *Cómo ingresar los datos en calculadora f_x-82MS*

Media, Varianza y Desviación Estándar usando la calculadora CASIO fx-82MS para dato...

colocar para números negativos los paréntesis: (- nº)

Datos:

$\bar{d} = -3,57$

$S_D = 7,52$

$n = 7$

Calculamos el estadístico de prueba:

$$T = \frac{\overline{D} - \mu_0}{\sqrt{\frac{S_D^2}{n}}}$$

$$T_m = \frac{-3,57 - 0}{7,52\left(\sqrt{\frac{1}{7}}\right)} = \frac{-3,57}{2,84} = \text{-1,26}$$

Realizamos el cálculo del p-valor: $p = P(|t_{n-1}| \geq |T_m|)$

$p = P(t_6 \geq 1,26)$ = $p - valor > 0,2$

las áreas disminuyen →

	α (dos colas)													
n	> 0,2	0,2	0,1	0,05	0,025	0,02	0,01	0,005	0,0025	0,001	0,0005	0,00025	0,0001	n
6	1,26	1,440	1,943	2,447	2,969	3,143	3,707	4,317	4,981	5,959	6,788	7,708	9,082	6
7		1,415	1,895	2,365	2,841	2,998	3,499	4,029	4,595	5,408	6,082	6,814	7,885	7
8		1,397	1,860	2,306	2,752	2,896	3,355	3,833	4,334	5,041	5,617	6,234	7,120	8
9		1,383	1,833	2,262	2,685	2,821	3,250	3,690	4,146	4,781	5,291	5,830	6,594	9
10		1,372	1,812	2,228	2,634	2,764	3,169	3,581	4,005	4,587	5,049	5,533	6,211	10

Si consideramos un α = 0.05 entonces: *p-valor > α*

Conclusión: No se rechaza la hipótesis nula ya que la evidencia es demasiado fuerte para hacerlo: *no se puede afirmar que la glucemia varía al final de la prueba*.