Universidad de los Andes
Departamento de Ingeniería Industrial
Probabilidad y Estadística I (IIND2106)
Profesor Coordinador: Mario Castillo
Profesores: Astrid Bernal, Carlos Castellanos, Fabio Lagos, María Alejandra López, Gonzalo Torres, Hernando Mutis.
Segundo semestre de 2015

Complementaria 12

Ejercicios sobre construcción, cálculo e interpretación de intervalos de confianza para la diferencia de medias, proporción y razón de varianzas. Intervalos de confianza en SPSS

Punto 1

En una fábrica de pintura especializada para techos, se diseñaron dos experimentos independientes con el propósito de comparar los tipos de pintura A y B. En el primer experimento se utilizó la pintura tipo A para pintar 18 techos y se registró el tiempo de secado en horas de cada uno. Se realizó un segundo experimento en el que se utilizó el tipo de pintura B para pintar un segundo conjunto de 15 techos. Los datos recolectados de los dos experimentos se presentan en la siguiente tabla:

	Horas de secado	
n	Tipo A	Tipo B
1	1.1	1.6
2	1.3	2.1
3	1.2	2.2
4	2.1	2.1
5	1.3	1.8
6	1.4	1.9
7	1.1	1.9
8	1.6	2.1
9	1.6	1.7
10	1.8	1.9
11	1.7	2.1
12	1.2	2
13	1.1	2.4
14	1.4	1.1
15	1.2	2
16	1.1	
17	1.2	
18	1.1	
S^2	0.0837	0.0921
$\bar{X}$	1.3611	1.9267

El tiempo de secado de cada tipo de pintura se puede representar a través de las variables aleatorias X y Y, que corresponden a los tipos de pintura A y B, respectivamente. Las variables aleatorias X y Y siguen una distribución normal de media μ_A y varianza σ_A^2 desconocidas, y de media μ_B y varianza σ_B^2 desconocidas, respectivamente.

Con el propósito de comparar los tiempos de secado de los dos tipos de pintura se quiere realizar un intervalo de confianza para la diferencia de medias poblacionales.

a. Antes de construir el IC es necesario comprobar el supuesto de varianzas iguales. Para ello, construya un intervalo de confianza del 90% para el cociente de varianzas del tiempo de secado de los dos tipos de pintura $\left(\frac{\sigma_A^2}{\sigma_B^2}\right)$.

Estadístico de la prueba:

$$X_A \rightarrow N(\mu_A, \sigma_A^2)$$

$$X_B \rightarrow N(\mu_B, \sigma_B^2)$$

$$\frac{\sigma_A^2 * S_B^2}{\sigma_B^2 * S_A^2} \sim F_{(n_A-1, n_B-1)}$$

Intervalo de confianza:

$$P\left(F_{(\alpha/2, n_B-1, n_A-1)} * \frac{S_A^2}{S_B^2} \leq \frac{\sigma_A^2}{\sigma_B^2} \leq F_{(1-\alpha/2, n_B-1, n_A-1)} * \frac{S_A^2}{S_B^2}\right) = 1 - \alpha$$

Para un 90 % de confiabilidad $F_{(\frac{0.1}{2}, 14, 17)} = 0.441,\ F_{(1-\frac{0.1}{2}, 14, 17)} = 2.328$

$$P\left(0.441 * \frac{0.0837}{0.0921} \leq \frac{\sigma_A^2}{\sigma_B^2} \leq \frac{0.0837}{0.0921} * 2.328\right) = 0.9$$

$$P\left(0.40 \leq \frac{\sigma_A^2}{\sigma_B^2} \leq 2.11\right) = 0.9$$

$$IC_{90\%} = [0.40; 2.11]$$

Con un nivel de confianza del 90% se puede observar que la relación entre el número de horas de secado entre la pintura tipo A y tipo B está entre 0.4 y 2.11. Al analizar los valores allí incluidos, el 1 está dentro de este intervalo y por esta razón se podrían considerar las varianzas poblacionales iguales.

b. Construya un intervalo de confianza del 95% para la diferencia de medias de los tiempos de secado de pinturas. Concluya.

Estadístico de la prueba:

$$X_A \rightarrow N(\mu_A, \sigma_A^2)$$

$$X_B \rightarrow N(\mu_B, \sigma_B^2)$$

$$\frac{\bar{X}_A - \bar{X}_B - (\mu_{A-}\mu_B)}{S_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} \sim t_{n-1}$$

$$S_p = \sqrt{\frac{S_A^2(n_A-1)+S_B^2(n_B-1)}{n_A+n_B-2}}$$

Intervalo de confianza:

$$P\left(\bar{X}_A - \bar{X}_B - t_{\left(1-\alpha/_2;(n_A+n_B-2)\right)} * S_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}} \leq \mu_{A-}\mu_B \leq \bar{X}_A - \bar{X}_B + t_{\left(1-\alpha/_2;(n_A+n_B-2)\right)} * S_p\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}\right)$$

Para un 95 % de confiabilidad $t_{\left(1-\frac{0.05}{2},31\right)} = 2.039$

Cálculo del intervalo:

$$S_p = \sqrt{\frac{0.0837(18-1)+0.0921(15-1)}{18+15-2}} = 0.296$$

$$P\left(-0.5656 - 2.039 * 0.296\sqrt{\frac{1}{18}+\frac{1}{15}} \leq \mu_{A-}\mu_B \leq -0.5656 + 2.039 * 0.296\sqrt{\frac{1}{18}+\frac{1}{15}}\right)$$

$$IC_{90\%} = [-0.77; -0.35]$$

Con un nivel de confianza del 95% se puede observar que la diferencia de medias entre el número de horas de secado entre la pintura tipo A y tipo B está entre -0.77 y -0.35. Al analizar los valores allí incluidos, el 0 no está dentro de este intervalo y por esta razón para ese nivel de confianza puede decirse que existe una diferencia entre las medias.

c. Si se conociera que $\sigma_A^2 = 0.079$ y $\sigma_B^2 = 0.086$, construya un intervalo de confianza del 95% para la diferencia de medias entre los dos tipos de pintura.

Estadístico de la prueba:

$$X_A \rightarrow N(\mu_A, 0.079)$$

$$X_B \rightarrow N(\mu_B, 0.086)$$

$$\frac{\bar{X}_A - \bar{X}_B - (\mu_{A-}\mu_B)}{\sqrt{\frac{{\sigma_A}^2}{n_A}+\frac{{\sigma_B}^2}{n_B}}} \sim N$$

Intervalo de confianza:

$$P\left(\bar{X}_A - \bar{X}_B - Z_{\left(1-\alpha/_2\right)} * \sqrt{\frac{{\sigma_A}^2}{n_A} + \frac{{\sigma_B}^2}{n_B}} \leq \mu_{A-}\mu_B \leq \bar{X}_A - \bar{X}_B + Z_{\left(1-\alpha/_2\right)} * \sqrt{\frac{{\sigma_A}^2}{n_A} + \frac{{\sigma_B}^2}{n_B}}\right)$$
$$= 1 - \alpha$$

Para un 95 % de confiabilidad $Z_{\left(1-\alpha/_2\right)} = Z_{(0,975)} = 1.96$

$$P\left(-0.5656 - 1.96 * \sqrt{\frac{0.079}{18} + \frac{0.086}{15}} \leq \mu_{A-}\mu_B \leq -0.5656 + 1.96 * \sqrt{\frac{0.079}{18} + \frac{0.086}{15}}\right)$$
$$= 0.95$$

$$IC_{95\%} = [-0.75; -0.36]$$

Punto 2

En un estudio para evaluar la resistencia de cascos de ciclismo, se tomó una muestra aleatoria de 150 cascos de ciclismo urbano y se les aplicó cierta prueba de impacto. De los 150 cascos 60 presentaron daño. Sea p la proporción de cascos de ciclismo que presentan daño al ser sometidos a la prueba de impacto, de solución a los siguientes literales:

a. Halle un intervalo de confianza del 90% para la proporción de cascos que no presentan daño. De acuerdo con el intervalo construido, ¿es correcto afirmar que la proporción de cascos que no presentan daño es superior a la proporción de cascos que sí presentan daños?

Estadístico de la prueba:

$$X \rightarrow Bernoulli \begin{cases} 1 & casco\ no\ presenta\ daño \\ 0 & presenta\ daño \end{cases}$$

$$\hat{\mathrm{p}} = \frac{cascos\ dañados}{total\ de\ la\ muestra} = \frac{90}{150} = 0.6$$

$$\bar{X} \rightarrow N\left(p; \frac{\hat{p}(1-\hat{p})}{n}\right)$$

$$\frac{\bar{\mathrm{X}} - \hat{\mathrm{p}}}{\sqrt{\frac{\hat{\mathrm{p}}(1-\hat{\mathrm{p}})}{\mathrm{n}}}} \sim \mathrm{N}(0,1)$$

Intervalo de confianza:

$$P\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{\bar{X} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1 - \alpha$$

$$P\left(\hat{p} - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) = 1-\alpha$$

Para un 90 % de confiabilidad, $z_{\left(1-\frac{0.1}{2}\right)} = z_{0.95} = 1.64$

Cálculo del intervalo de confianza:

$$P\left(0.6 - 1.64\sqrt{\frac{0.6(1-0.6)}{150}} \leq p \leq 0.6 + 1.64\sqrt{\frac{0.6(1-0.6)}{150}}\right) = 0.9$$

$$P(0.534 \leq p \leq 0.666) = 0.9$$

$$IC_{90\%} = [0.534; 0.666]$$

Con una confiabilidad del 90% la proporción real de cascos que no presenta daño estará entre 0.534 y 0.666. Se puede decir que la proporción de cascos que presentan daño es superior a la proporción que no presenta daño.

b. Para ampliar el estudio, se tomó otra muestra aleatoria de 120 cascos de ciclismo MTB (Mountain Bike), los cuales se sometieron a la prueba de impacto y se encontró que 42 de ellos no presentaron daño. Construya un intervalo de confianza del 95% sobre la diferencia de las medias de las proporciones de cascos de las dos muestras aleatorias que no presentan daño.

Estadístico de la prueba:

$$\hat{p}_A = \bar{X} = \frac{90}{150} = 0.6$$

$$\hat{p}_B = \bar{Y} = \frac{42}{120} = 0.35$$

$$\bar{X} \to N\left(p_1; \frac{\hat{p}_A(1-\hat{p}_A)}{n_A}\right)$$

$$\bar{Y} \to N\left(p_B; \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}\right)$$

$$\frac{\bar{X} - \bar{Y} - (p_A - p_B)}{\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}} \sim N(0,1)$$

Intervalo de confianza:

$$P\left(-Z_{\left(1-\frac{\alpha}{2}\right)} \leq \frac{\bar{X}-\bar{Y}-(p_A-p_B)}{\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A}+\frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}} \leq Z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1-\alpha$$

$$P\left(\hat{p}_A-\hat{p}_B-Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A}+\frac{\hat{p}_B(1-\hat{p}_B)}{n_B}} \leq p_A-p_B \leq \hat{p}_A-\hat{p}_B+Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A}+\frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}\right) = 1-\alpha$$

Para un 95 % de confiabilidad, $z_{\left(1-\frac{0.05}{2}\right)} = z_{0.975} = 1.96$

$$P\left(0.25-1.96\sqrt{\frac{0.6(1-0.6)}{150}+\frac{0.35(1-0.35)}{120}} \leq p_A-p_B \leq 0.25+1.96\sqrt{\frac{0.6(1-0.6)}{150}+\frac{0.35(1-0.35)}{120}}\right) = 1-\alpha$$

$$P(0.1341 \leq p_A-p_B \leq 0.3658) = 0.95$$

$$IC_{95\%} = [0.1341; 0.3658]$$

Con una confiabilidad del 95% la diferencia de proporciones entre los cascos urbanos y los cascos MTB que no presentan daños estará entre 0.1341 y 0.3658. Debido a que el intervalo para la diferencia de proporciones tiene ambos limites positivos, se puede afirmar de la muestra que el parámetro p_A es mayor que p_B, luego la probabilidad de que un casco no presente daños es mayor para los cascos tipo urbano que para los cascos tipo MTB.

Punto 3

Un viñedo está realizando un control de calidad de la producción de vinos de los últimos seis meses. Para esto, ha tomado una muestra de 180 botellas y se ha medido el porcentaje de alcohol y el porcentaje de alcalinidad de cada una. Los vinos producidos por el viñedo tienen tres cualidades que permiten describir sus propiedades finales, y cada botella producida debe ser clasificada bajo estos criterios:

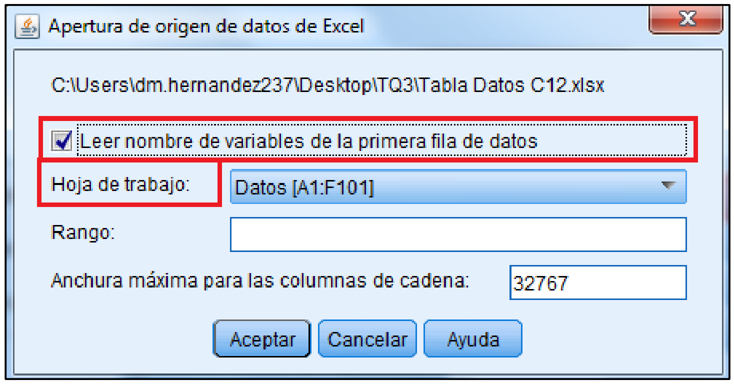
- De acuerdo al tipo de uva utilizada, los vinos pueden clasificarse en vinos tintos, rosados y blancos.
- De acuerdo con la cantidad de gas que contiene el vino se puede clasificar en vino de aguja (presión de gas entre 1 y 2 atm), vino de perla (presión de gas de 2 y 3 atm) y vinos espumosos (presión de gas mayor a 3 atm).
- De acuerdo a la cantidad de azúcar que tenga el vino se pueden clasificar como vino brut (entre 6 g y 12 g de azúcar por litro), vino seco (entre 17 g y 32 g de azúcar por litro) y vino dulce (más de 50 g de azúcar por litro).

Utilizando el software SPSS, y asumiendo que tanto la variable Porcentaje de alcohol como la variable Porcentaje de alcalinidad tienen distribución normal, responda las siguientes preguntas:

a. Construya un intervalo de confianza del 95% y otro del 99% para la media del porcentaje de alcohol.

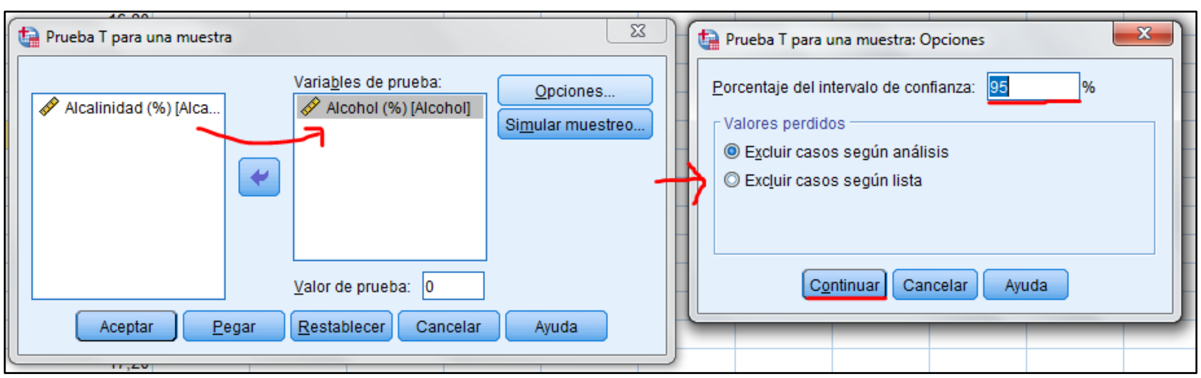
Paso 1: importar la tabla de datos apenas se abre el programa SPSS:

Paso 2: seleccionar el tipo de archivo a abrir (.xlsx)



Paso 3: ir a analizar/comparar medias/prueba t

Paso 4: ingresar los intervalos indicando el nivel de confianza requerido (95% y 99%).



Paso 5: mostrar los resultados y analizar.

Prueba de muestra única

	Valor de prueba = 0

	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Alcohol (%)	212,903	178	,000	12,99268	12,8723	13,1131

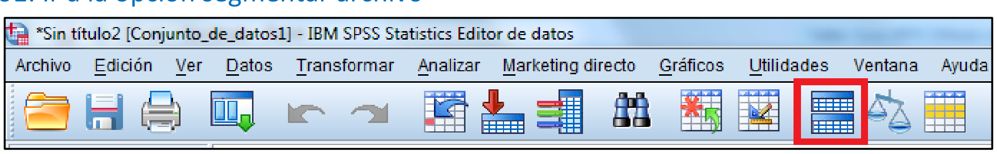
Prueba de muestra única

	Valor de prueba = 0					
	t	gl	Sig. (bilateral)	Diferencia de medias	99% de intervalo de confianza de la diferencia	
					Inferior	Superior
Alcohol (%)	212,903	178	,000	12,99268	12,8338	13,1516

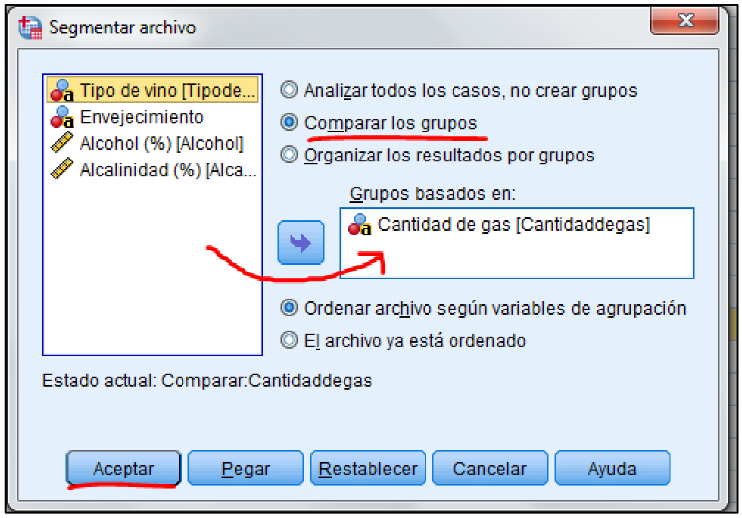
b. Construya el intervalo de confianza del 95% para la media del porcentaje de alcohol de los vinos aguja y los vinos espumosos.

Dado que se busca realizar un intervalo segmentando los datos según tipo de vino por cantidad de gas, se debe utilizar la opción Segmentar Archivo de SPSS. Esta opción permitirá que, al momento de correr el análisis de las medias, los resultados se obtengan por grupos.

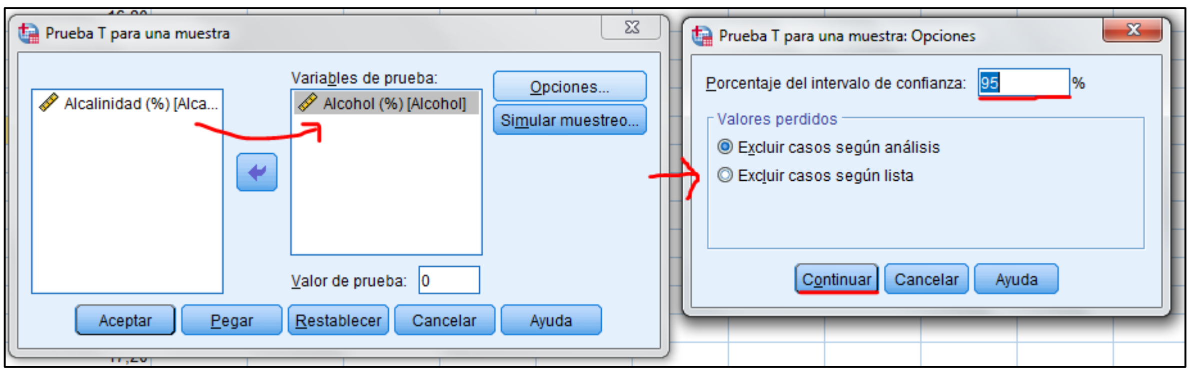
Paso1: Ir a la opción segmentar archivo



Paso2: Seleccionar la variable de interés, en este caso porcentaje de alcohol basado en la cantidad de gas.



Paso3: Ir a analizar/ diferencia de medias/prueba t…



Paso 3: mostrar los resultados y concluir.

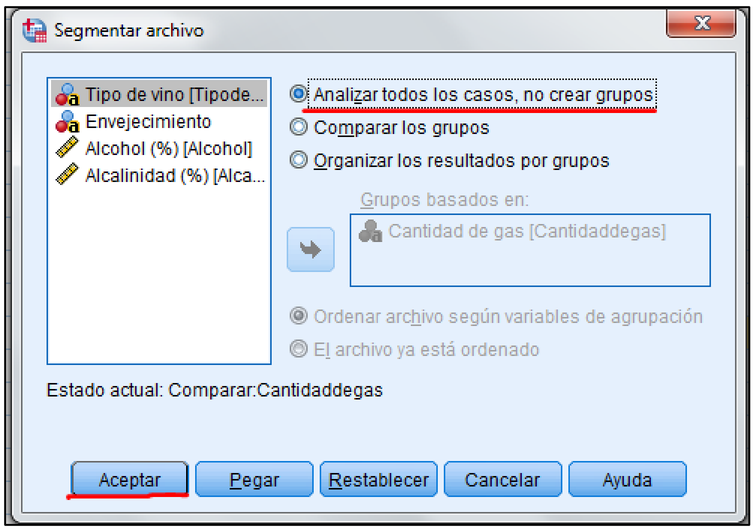
Prueba de muestra única

Cantidad de gas		Valor de prueba = 0					
		t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia: Inferior	Superior
Aguja	Alcohol (%)	126,412	59	,000	13,05383	12,8472	13,2605
Espumoso	Alcohol (%)	124,339	58	,000	13,04237	12,8324	13,2523
Perla	Alcohol (%)	118,041	59	,000	12,88267	12,6643	13,1010

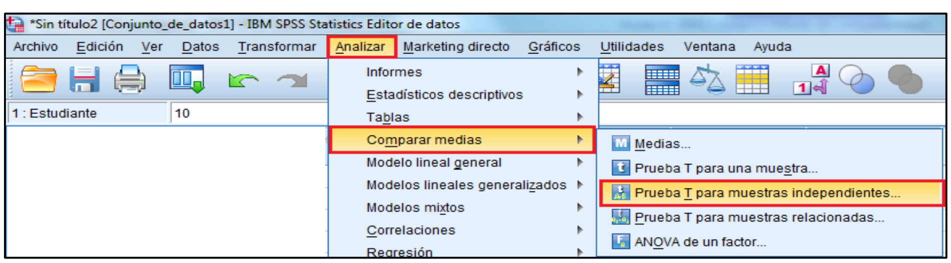
Con un nivel de confianza del 95%, la media del porcentaje de alcohol de los vinos aguja estará entre 12.8472% y 13.2605%, y la media del porcentaje de alcohol de los vinos espumosos entre un 12.8324% y 13.2523%

c. Durante una cata realizada en el viñedo, un grupo de enólogos afirmó que el porcentaje de alcalinidad en los vinos jóvenes es mayor al porcentaje de alcalinidad de los añejos, lo que implicaría un problema de calidad en la fabricación de los mismos. Construya un intervalo de confianza del 95% para la diferencia de medias del porcentaje de alcalinidad y concluya si dicha hipótesis es verdadera.

Paso 1: dejar de segmentar los datos.



Paso 2: construir el intervalo de diferencia de medias. Ir a: analizar/comparar medias/ prueba T...



Paso 3: seleccionar como variable para contrastar Porcentaje de Alcalinidad y como Variable de Agrupación, Envejecimiento. En la opción Definir grupos debemos escribir los nombres de las categorías de la variable de agrupación, en este caso Joven y Reserva.

Paso 4: mostrar los resultados y concluir.

Prueba de muestras independientes

		Prueba de Levene de calidad 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
Alcalinidad (%)	Se asumen varianzas iguales	,306	,581	-1,238	177	,217	-,61675	,49830	-1,60012	,36661
	No se asumen varianzas iguales			-1,237	173,099	,218	-,61675	,49869	-1,60104	,36754

El intervalo de confianza obtenido contiene el cero, por lo cual con una certeza del 95%, no existe diferencia significativa entre la diferencia de medias del porcentaje de alcalinidad entre los vinos jóvenes y los vinos de reserva.