



# M1.3 Transformaciones e Inferencia Estadística

## **Integrantes**

A01068244 - Jared Andrés Silva Villa

**Fecha:** 20 de Agosto del 2024

# Índice

<b>1.- Ejercicio 1</b>	<b>3</b>
<b>a) Haga un análisis exploratorio de estos datos:</b>	<b>4</b>
a. Calcular e interpretar estadísticas descriptivas de los datos: media, mediana, moda, desviación estándar, coeficiente de variación.	4
b. ¿Cuál de las variables tiene mayor variabilidad? ¿Cuál tiene menor variabilidad? Explique, ¿cuáles estadísticas son relevantes para ello? y ¿por qué?	4
b) Utilizando la Técnica de Análisis Multifactor, obtener cuál debería ser el ranking de cada uno de los empleados para poder definir el reparto de los incentivos.	5
c) Suponga que se quiere utilizar los datos proporcionados y una regresión lineal para predecir cuáles serían las ventas generadas por 3 empleados nuevos con los siguientes valores:	6
<b>2.- Ejercicio 2</b>	<b>6</b>
a) ¿Qué tipo de variable se está midiendo? ¿Discreta o continua? Explique.	6
b) Haga un análisis exploratorio de estos datos.	7
a. Realice un histograma con al menos 2 reglas para definir el número de clases (No utilizar regla empírica).	7
b. Realice un diagrama de caja y bigotes. Analice el comportamiento de los datos. ¿Existen datos atípicos? ¿Qué se debería hacer al respecto?	8
c) Estime, con una confianza de 94%, ¿cuál sería la resistencia promedio de los envases?	8
d) Antes del estudio se suponía que la resistencia promedio era de 25kg. Dada la evidencia de los datos, ¿tal supuesto es correcto? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.	9
e) Con los datos anteriores estime, con una confianza del 98%, ¿cuál es la desviación estándar poblacional (del proceso)?	9
<b>3.-Ejercicio 3</b>	<b>9</b>
a) ¿Las muestras son dependientes o independientes? Explique.	
Las muestras son independiente, ya que cada muestra tanto hombres como mujeres se han hecho de forma separada y no tiene un emparejamiento directo entre las dos observaciones, en cuanto a su correlación tiene un valor de 0.374	10
b) ¿La temperatura promedio más confortable es igual para hombre que para mujeres? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.	11
c) ¿Los datos poseen la misma variabilidad? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.	11
<b>4.-Ejercicio 4</b>	<b>11</b>
a) ¿Las muestras son dependientes o independientes? Explique.	12
b) ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis	

correspondientes y concluya adecuadamente.	12
c) ¿Recomienda la adopción del nuevo método? Argumente su respuesta	12

## 1.- Ejercicio 1

Una pequeña empresa de manufactura estableció un sistema de incentivos para sus empleados basado en diferentes variables tanto de desempeño como de costo para la empresa. La empresa desea conocer cuál sería el ranking de los empleados tomando en cuenta todas las variables. A continuación, se presenta una tabla con los resultados obtenidos por cada empleado en cada uno de los rubros y si “más es mejor” o “menos es mejor”

a) Haga un análisis exploratorio de estos datos:

a. Calcular e interpretar estadísticas descriptivas de los datos: media, mediana, moda, desviación estándar, coeficiente de variación.

### Estadísticas

Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Varianza	Mínimo	Q1
Salario	10	0	4812.5	58.0	183.5	33656.1	4550.0	4658.3
Costo de Capacitación	10	0	401.2	17.7	56.0	3140.4	330.0	353.0
Producción Generada	10	0	9831.6	62.5	197.8	39123.6	9500.0	9672.5
Satisfacción del Cliente Intern	10	0	7.500	0.500	1.581	2.500	5.000	6.000
Ventas Generadas	10	0	75449	1178	3725	13874148	69000	72500
Ausentismo	10	0	3.600	0.452	1.430	2.044	2.000	2.000

Variable	Mediana	Q3	Máximo
Salario	4799.5	4959.0	5100.0
Costo de Capacitación	387.0	455.0	499.0
Producción Generada	9793.0	10013.3	10100.0
Satisfacción del Cliente Intern	7.500	9.000	10.000
Ventas Generadas	75750	78829	80014
Ausentismo	3.500	5.000	6.000

b. ¿Cuál de las variables tiene mayor variabilidad? ¿Cuál tiene menor variabilidad? Explique, ¿cuáles estadísticas son relevantes para ello? y ¿por qué?

## Estadísticas

Variable			Error estándar de la		Desv.Est.	Varianza	CoefVar	Mínimo
	N	N*	Media	media				
Salario	10	0	4812.5	58.0	183.5	33656.1	3.81	4550.0
Costo de Capacitación	10	0	401.2	17.7	56.0	3140.4	13.97	330.0
Producción Generada	10	0	9831.6	62.5	197.8	39123.6	2.01	9500.0
Satisfacción del Cliente Intern	10	0	7.500	0.500	1.581	2.500	21.08	5.000
Ventas Generadas	10	0	75449	1178	3725	13874148	4.94	69000
Ausentismo	10	0	3.600	0.452	1.430	2.044	39.72	2.000

Variable	Q1	Mediana	Q3	Máximo
Salario	4658.3	4799.5	4959.0	5100.0
Costo de Capacitación	353.0	387.0	455.0	499.0
Producción Generada	9672.5	9793.0	10013.3	10100.0
Satisfacción del Cliente Intern	6.000	7.500	9.000	10.000
Ventas Generadas	72500	75750	78829	80014
Ausentismo	2.000	3.500	5.000	6.000

El coeficiente de variación **CoefVar**, es la estadística que es relevante para contestar esta pregunta, esta te permite comparar la variabilidad entre diferentes variables que contengan distintas unidades o escalas.

La variable que cuenta con un coeficiente de variación más alto (por lo que es la que contiene más datos dispersos de su media) es el **Ausentismo**.

La variable que cuenta con el coeficiente de variación menor es: **Producción Generadas**.

b) Utilizando la Técnica de Análisis Multifactor, obtener cuál debería ser el ranking de cada uno de los empleados para poder definir el reparto de los incentivos.

	Menos	Menos	Más	Más	Más	Menos
	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interno	Ventas Generadas	Ausentismo
Empleado 1	0.98484848	0.93220339	0.99019802	0.7	1	0.4
Empleado 2	0.89215686	0.66132265	0.97029703	0.8	0.93733597	0.33333333
Empleado 3	1	0.73333333	0.94059406	0.6	0.86234909	0.5
Empleado 4	0.95769312	0.70212766	0.99	0.9	0.88734471	0.66666667
Empleado 5	0.93853135	0.86842105	0.96534653	0.7	0.95608269	1
Empleado 6	0.92254663	0.89189189	0.95841584	0.6	0.99750044	0.4
Empleado 7	0.90277778	1	0.96891089	0.8	0.97055515	0.5
Empleado 8	0.97409548	0.94285714	0.95544554	0.5	0.98107831	1
Empleado 9	0.96829113	0.79518072	1	0.9	0.91234034	1
Empleado 10	0.92592593	0.83756345	0.9950495	1	0.92483815	0.66666667
suma	9.46686676	8.36490129	9.73425743	7.5	9.42942485	6.46666667

	Menos	Menos	Más	Más	Más	Menos
	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interno	Ventas Generadas	Ausentismo
Empleado 1	0.10403109	0.11144225	0.10172302	0.09333333	0.10605101	0.06185567
Empleado 2	0.09423993	0.07905923	0.09967859	0.10666667	0.09940542	0.05154639
Empleado 3	0.10563157	0.0876679	0.0966272	0.08	0.09145299	0.07731959
Empleado 4	0.10116263	0.08393735	0.10170267	0.12	0.0941038	0.10309278
Empleado 5	0.09913854	0.10381725	0.09917002	0.09333333	0.10139353	0.15463918
Empleado 6	0.09745005	0.10662312	0.09845803	0.08	0.10578593	0.06185567
Empleado 7	0.09536183	0.11954714	0.09953619	0.10666667	0.10292835	0.07731959
Empleado 8	0.10289524	0.11271587	0.09815289	0.06666667	0.10404434	0.15463918
Empleado 9	0.10228211	0.09506158	0.10272997	0.12	0.09675461	0.15463918
Empleado 10	0.09780701	0.10012831	0.10222141	0.13333333	0.09808002	0.10309278
suma	1	1	1	1	1	1

	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interno	Ventas Generadas	Ausentismo
Importancia	0.06	0.03	0.16	0.25	0.40	0.10

	Menos	Menos	Más	Más	Más	Menos	
	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interno	Ventas Generadas	Ausentismo	Promedio
Empleado 1	0.00624187	0.00334327	0.01627568	0.02333333	0.0424204	0.00618557	0.01630002
Empleado 2	0.0056544	0.00237178	0.01594857	0.02666667	0.03976217	0.00515464	0.01592637
Empleado 3	0.00633789	0.00263004	0.01546035	0.02	0.0365812	0.00773196	0.01479024
Empleado 4	0.00606976	0.00251812	0.01627243	0.03	0.03764152	0.01030928	0.01713518
Empleado 5	0.00594831	0.00311452	0.0158672	0.02333333	0.04055741	0.01546392	0.01738078
Empleado 6	0.005847	0.00319869	0.01575329	0.02	0.04231437	0.00618557	0.01554982
Empleado 7	0.00572171	0.00358641	0.01592579	0.02666667	0.04117134	0.00773196	0.01680065
Empleado 8	0.00617371	0.00338148	0.01570446	0.01666667	0.04161774	0.01546392	0.01650133
Empleado 9	0.00613693	0.00285185	0.0164368	0.03	0.03870184	0.01546392	0.01826522
Empleado 10	0.00586842	0.00300385	0.01635543	0.03333333	0.03923201	0.01030928	0.01801705

	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interna	Ventas Generadas	Ausentismo	Promedio
Empleado 9	0.0061369	0.0028518	0.0164368	0.03	0.0387018	0.0154639	0.0182652
Empleado 10	0.0058684	0.0030038	0.0163554	0.0333333	0.039232	0.0103093	0.0180171
Empleado 5	0.0059483	0.0031145	0.0158672	0.0233333	0.0405574	0.0154639	0.0173808
Empleado 4	0.0060698	0.0025181	0.0162724	0.03	0.0376415	0.0103093	0.0171352
Empleado 7	0.0057217	0.0035864	0.0159258	0.0266667	0.0411713	0.007732	0.0168006
Empleado 8	0.0061737	0.0033815	0.0157045	0.0166667	0.0416177	0.0154639	0.0165013
Empleado 1	0.0062419	0.0033433	0.0162757	0.0233333	0.0424204	0.0061856	0.0163
Empleado 2	0.0056544	0.0023718	0.0159486	0.0266667	0.0397622	0.0051546	0.0159264
Empleado 6	0.005847	0.0031987	0.0157533	0.02	0.0423144	0.0061856	0.0155498
Empleado 3	0.0063379	0.00263	0.0154604	0.02	0.0365812	0.007732	0.0147902

c) Suponga que se quiere utilizar los datos proporcionados y una regresión lineal para predecir cuáles serían las ventas generadas por 3 empleados nuevos con los siguientes valores:

Empleados Nuevos	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interna	Ventas Generadas	Ausentismo
Empleado 11	4700	420	9800	8	?	3
Empleado 12	4900	450	9600	7	?	5
Empleado 13	4850	380	10000	8	?	4



Empleados Nuevos	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interna	Ventas Generadas	Ausentismo
Empleado 11	4700	420	9800	8		3
Empleado 12	4900	450	9600	7		5
Empleado 13	4850	380	10000	8		4
	Menos	Menos	Más	Más	Más	Menos
Empleados Nuevos	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interna	Ventas Generadas	Ausentismo
Empleado 11	1	1	0.98	1		1
Empleado 12	0.959183673	0.933333333	0.96	0.875		0.6
Empleado 13	0.969072165	1.105263158	1	1		0.75
suma	2.92825584	3.03859649	2.94	2.875		2.35
	Menos	Menos	Más	Más	Más	Menos
Empleados N	Salario	Costo de Cap	Producción G	Satisfacción	Ventas Gener	Ausentismo
Empleado 11	0.34150022	0.32909931	0.33333333	0.34782609	0.46017004	0.42553191
Empleado 12	0.32756143	0.30715935	0.32653061	0.30434783	0.46662243	0.25531915
Empleado 13	0.33093835	0.36374134	0.34013605	0.34782609	0.49316664	0.31914894
suma	1	1	1	1	1.41995911	1

Lo intente en Excel profe, tengo unas dudas jaja

	Salario	Costo de Capacitación	Producción Generada	\
0	4700	420	9800	
1	4900	450	9600	
2	4850	380	10000	
	Satisfacción del Cliente Interna	Ausentismo	Ventas Generadas	
0	8	3	71178.649796	
1	7	5	72703.543875	
2	8	4	78412.096115	

Python

## 2.- Ejercicio 2

En la elaboración de envases de plástico es necesario garantizar que cierto tipo de botella en posición vertical tenga una resistencia mínima de 20kg de fuerza. Para garantizar esto,

se aplica fuerza a la botella hasta que ésta cede, y el equipo registra la resistencia que alcanzó la botella. Se obtuvieron los siguientes datos de la resistencia máxima alcanzada de cada botella mediante pruebas destructivas

28.3	26.8	26.6	26.5	28.1	24.8	27.4	26.2	29.4	28.6	24.9	25.2	30.4	27.7	27.0	26.1	28.1
26.9	28.0	27.6	25.6	29.5	27.6	27.3	26.2	27.7	27.2	25.9	26.5	28.3	26.5	29.1	23.7	29.7
26.8	29.5	28.4	26.3	28.1	28.7	27.0	25.5	26.9	27.2	27.6	25.5	28.3	27.4	28.8	25.0	25.3
27.7	25.2	28.6	27.9	28.7												

a) ¿Qué tipo de variable se está midiendo? ¿Discreta o continua? Explique.

Son variables **continuas**, ya que puede tomar cualquier valor dentro de un intervalo específico, entre cada par de valores, siempre puede existir otro valor intermedio.

b) Haga un análisis exploratorio de estos datos.

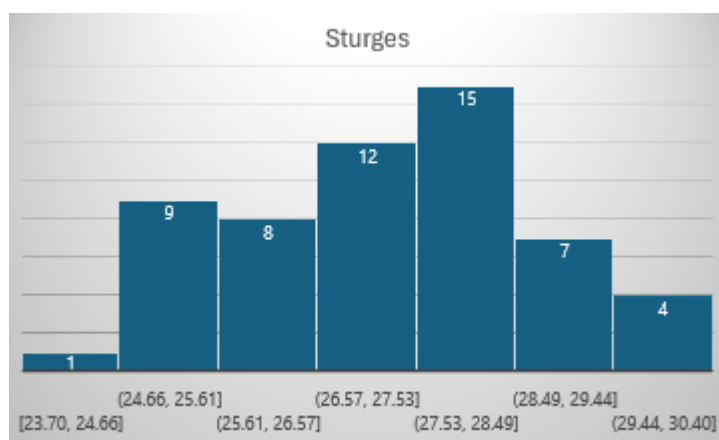
## Estadísticas

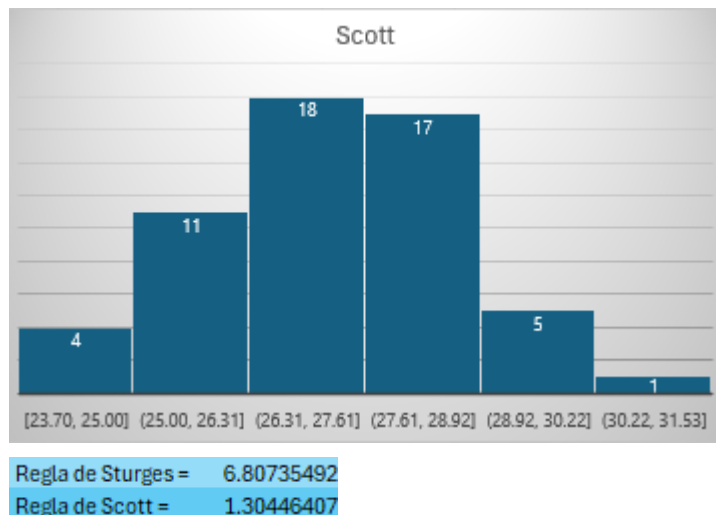
Variable	N	N*	Media	Error estándar de la media	Desv.Est.	Varianza	CoefVar	Mínimo	Q1
Resistencia Máxima	56	0	27.246	0.191	1.430	2.046	5.25	23.700	26.225

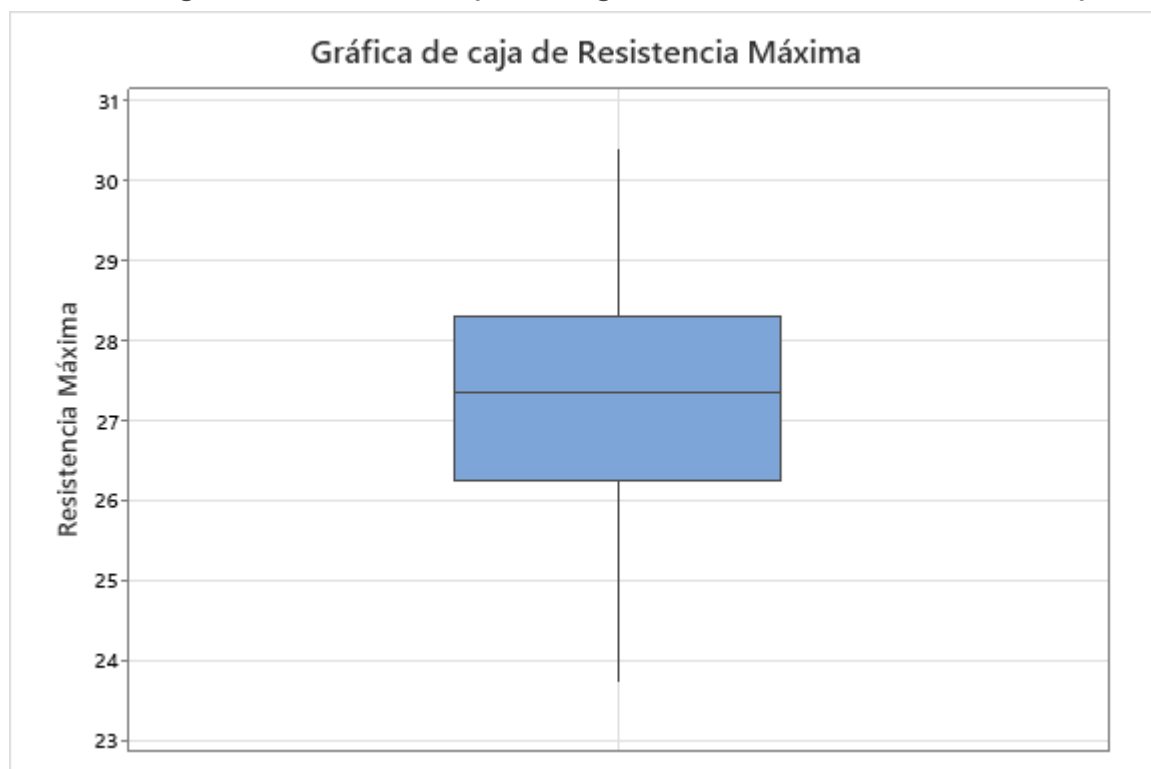
Variable	Mediana	Q3	Máximo	Asimetría
Resistencia Máxima	27.350	28.300	30.400	-0.15

a. Realice un histograma con al menos 2 reglas para definir el número de clases (No utilizar regla empírica).





b. Realice un diagrama de caja y bigotes. Analice el comportamiento de los datos. ¿Existen datos atípicos? ¿Qué se debería hacer al respecto?



**No se presentan datos atípicos**, sin embargo si se presentaran dependería del contexto y se decidiría si se eliminan o se conservan para el análisis.

c) Estime, con una confianza de 94%, ¿cuál sería la resistencia promedio de los envases?

### Estadísticas descriptivas

			Error estándar de la	
N	Media	Desv.Est.	media	IC de 94% para $\mu$
56	27.246	1.430	0.191	(26.879, 27.614)

$\mu$ : media de población de Resistencia Máxima

La resistencia promedio de los envases sería de 27.246

d) Antes del estudio se suponía que la resistencia promedio era de 25kg. Dada la evidencia de los datos, ¿tal supuesto es correcto? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.

Dado que el valor P es  $< 0.05$  esto significa que tenemos la suficiente evidencia para rechazar la hipótesis nula, por lo que se rechaza la suposición de que la resistencia promedio es 25 kg.

### Prueba

Hipótesis nula  $H_0: \mu = 25$   
Hipótesis alterna  $H_1: \mu \neq 25$

Valor T	Valor p
11.75	0.000

e) Con los datos anteriores estime, con una confianza del 98%, ¿cuál es la desviación estándar poblacional (del proceso)?

## Estadísticas descriptivas

		Error estándar de la media IC de 98% para $\mu$		
N	Media	Desv.Est.		
56	27.246	1.430	0.191	(26.788, 27.704)

$\mu$ : media de población de Resistencia Máxima

Con un 98% de confianza, podemos ver que la variable media se encuentra entre 26.788 y 27.704.

## 3.-Ejercicio 3

En un laboratorio bajo condiciones controladas, se evaluó, para 10 hombres y 10 mujeres, la temperatura que cada persona encontró más confortable. Los resultados en grados Fahrenheit fueron los siguientes:

Mujer	75	77	78	79	77	73	78	79	78	80
Hombre	74	72	77	76	76	73	75	73	74	75

## Estadísticas

		Error estándar de la media								
Variable	N	N*	Media	Desv.Est.	Varianza	CoefVar	Mínimo	Q1	Mediana	
Mujer	10	0	77.400	0.653	2.066	4.267	2.67	73.000	76.500	78.000
Hombre	10	0	74.500	0.500	1.581	2.500	2.12	72.000	73.000	74.500
Variable	Q3		Máximo	Asimetría						
Mujer	79.000		80.000	-1.15						
Hombre	76.000		77.000	-0.00						

a) ¿Las muestras son dependientes o independientes? Explique.

Las muestras son independiente, ya que cada muestra tanto hombres como mujeres se han hecho de forma separada y no tiene un emparejamiento directo entre las dos observaciones, en cuanto a su correlación tiene un valor de 0.374

### Correlaciones

	Mujer
Hombre	0.374

Esto nos quiere decir que existe una correlación débil, ya que es un número muy bajo.

b) ¿La temperatura promedio más confortable es igual para hombre que para mujeres? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.

Para este tipo de problemas, podemos utilizar la prueba t de dos muestras, con el objetivo de hacer la hipótesis nula de si las diferencias de medias son iguales a 0.

### Estadísticas descriptivas

				Error estándar de la media
Muestra	N	Media	Desv.Est.	
Mujer	10	77.40	2.07	0.65
Hombre	10	74.50	1.58	0.50

### Prueba

Hipótesis nula  $H_0: \mu_1 - \mu_2 = 0$   
Hipótesis alterna  $H_1: \mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
3.53	16	0.003

El valor es  $p < 0.05$  por lo que se rechaza la hipótesis nula y nos sugiere que si hay una diferencia significativa.

c) ¿Los datos poseen la misma variabilidad? ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.

Para este ejercicio se debe hacer pruebas de igualdad de varianza, la hipótesis nula si las dos variables son iguales.

### Prueba

Hipótesis nula  $H_0: \sigma_1^2 / \sigma_2^2 = 1$

Hipótesis alterna  $H_1: \sigma_1^2 / \sigma_2^2 \neq 1$

Nivel de significancia  $\alpha = 0.05$

Estadística				
Método de prueba	GL1	GL2	Valor p	
Bonett	0.39	1	0.530	
Levene	0.03	1 18	0.860	

Dado que las dos pruebas tanto Bonett y Levene tienen un valor  $P > 0.05$  se puede concluir que no hay suficiente evidencia para decir que las varianzas son diferentes.

## 4.-Ejercicio 4

La prueba actual de un solo disco se tarda 2 minutos. Se supone un nuevo método de prueba que consiste en medir solamente los radios 24 y 57, donde casi es seguro que estará el valor mínimo buscado. Si el método nuevo resulta igual de efectivo que el método actual se podrá reducir en 60% el tiempo de prueba. Se plantea un experimento donde se mide la densidad mínima de metal en 18 discos usando tanto el método actual como el método nuevo. Los resultados están ordenados horizontalmente por disco. Así, 1.88 y 1.87 es el resultado para el primer disco con ambos métodos.

Método Actual	1.88	1.84	1.83	1.90	2.19	1.89	2.27	2.03	1.96	1.98	2.00	1.92	1.83	1.94	1.94	1.95	1.93	2.01
Método Nuevo	1.87	1.90	1.85	1.88	2.18	1.87	2.23	1.97	2.00	1.98	1.99	1.89	1.78	1.92	2.02	2.00	1.95	2.05

a) ¿Las muestras son dependientes o independientes? Explique.

Los métodos no se relacionan directamente entre sí, lo que quiere decir que uno no depende del otro para funcionar o producir sus resultados, en cuanto a las **mediciones** cada par son **dependientes** del disco que se mida.

b) ¿Qué tipo de prueba estadística se debe realizar? Plantee las hipótesis correspondientes y concluya adecuadamente.

Para este problema, podemos realizar una **prueba t** de dos muestras, esta prueba compara las **medias** de ambos métodos para determinar si hay alguna diferencia significativa entre ellas.

### Estadísticas descriptivas

Muestra	N	Media	Desv.Est.	Error estándar de la media
Método Actual	18	1.961	0.115	0.027
Método Nuevo	18	1.963	0.112	0.026

### Prueba

Hipótesis nula  $H_0: \mu_1 - \mu_2 = 0$

Hipótesis alterna  $H_1: \mu_1 - \mu_2 \neq 0$

Valor T	GL	Valor p
-0.06	33	0.954

Aquí podemos observar un valor **p > 0.05**, por lo que no se rechaza la hipótesis nula, y podemos concluir que las medias no difieren significativamente.

c) ¿Recomienda la adopción del nuevo método? Argumente su respuesta

Si se recomienda cambiar ya que el texto nos dice “Si el método nuevo resulta igual de efectivo que el método actual se podrá reducir en 60% el tiempo de prueba”, ya que los dos métodos son igual de efectivos si se recomienda cambiarlo, ya que aumentaría el 60% el tiempo de prueba.



