

Transformaciones e Inferencia Estadística

Resuelva los siguientes problemas de forma individual. Para cada problema, incluir la formulación de las hipótesis; gráficas y tablas necesarias; y la interpretación de los resultados. Puede utilizar Excel, Minitab o cualquier otro software o lenguaje (Python o R) como apoyo para su solución.

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”:

	Menos Salario	Menos Costo de Capacitación	Más Producción Generada	Más Satisfacción del Cliente Interna	Más Ventas Generadas	Menos Ausentismo
Empleado 1	4620	354	10001	7	80014	5
Empleado 2	5100	499	9800	8	75000	6
Empleado 3	4550	450	9500	6	69000	4
Empleado 4	4751	470	9999	9	71000	3
Empleado 5	4848	380	9750	7	76500	2
Empleado 6	4932	370	9680	6	79814	5
Empleado 7	5040	330	9786	8	77658	4
Empleado 8	4671	350	9650	5	78500	2
Empleado 9	4699	415	10100	9	73000	2
Empleado 10	4914	394	10050	10	74000	3

Previamente, y con apoyo de la junta directiva, se aplicó la metodología AHP para definir los pesos de cada una de las variables y se obtuvieron los siguientes porcentajes:

	Salario	Costo de Capacitación	Producción Generada	Satisfacción del Cliente Interna	Ventas Generadas	Ausentismo
Importancia	6%	3%	16%	25%	40%	10%

a) Haga un análisis exploratorio de estos datos:

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

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Variance	CoefVar	Minimum	Q1
Salario	10	0	4812.5	58.0	183.5	33656.1	3.81	4550.0	4658.3
Costo de Capacitación	10	0	401.2	17.7	56.0	3140.4	13.97	330.0	353.0
Producción Generada	10	0	9831.6	62.5	197.8	39123.6	2.01	9500.0	9672.5
Satisfacción del Cliente Intern	10	0	7.500	0.500	1.581	2.500	21.08	5.000	6.000
Ventas Generadas	10	0	75449	1178	3725	13874148	4.94	69000	72500
Ausentismo	10	0	3.600	0.452	1.430	2.044	39.72	2.000	2.000

Variable	Median	Q3	Maximum	Mode	N for Mode
Salario	4799.5	4959.0	5100.0	*	0
Costo de Capacitación	387.0	455.0	499.0	*	0
Producción Generada	9793.0	10013.3	10100.0	*	0
Satisfacción del Cliente Intern	7.500	9.000	10.000	6, 7, 8, 9	2
Ventas Generadas	75750	78829	80014	*	0
Ausentismo	3.500	5.000	6.000	2	3

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

- La variable con mayor coeficiente de variación y por lo tanto mayor variabilidad es **Ausentismo**.
- La variable con menor variabilidad es **Producción Generada**

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.

	Salario_3	Costo de Capacitación_3	Producción Generada_3	Satisfacción del Cliente Interna_3	Ventas Generadas_3	Ausentismo_3	Promedio
Empleado 1	0.006241866	0.003343267	0.016275682	0.023333333	0.042420403	0.006185567	0.01630002
Empleado 2	0.005654396	0.002371777	0.015948574	0.026666667	0.039762169	0.005154639	0.01592637
Empleado 3	0.006337894	0.002630037	0.015460352	0.02	0.036581196	0.007731959	0.01479024
Empleado 4	0.006069758	0.002518121	0.016272428	0.03	0.03764152	0.010309278	0.017135184
Empleado 5	0.005948312	0.003114518	0.015867204	0.023333333	0.040557413	0.015463918	0.017380783
Empleado 6	0.005847003	0.003198694	0.015753285	0.02	0.04231437	0.006185567	0.01554982
Empleado 7	0.00572171	0.003586414	0.01592579	0.026666667	0.04117134	0.007731959	0.016800647
Empleado 8	0.006173714	0.003381476	0.015704463	0.016666667	0.041617737	0.015463918	0.016501329
Empleado 9	0.006136927	0.002851847	0.016436796	0.03	0.038701845	0.015463918	0.018265222
Empleado 10	0.005868421	0.003003849	0.016355425	0.033333333	0.039232007	0.010309278	0.018017052

Ranking	
0.018265222	Empleado 9
0.018017052	Empleado 10
0.017380783	Empleado 5
0.017135184	Empleado 4
0.016800647	Empleado 7
0.016501329	Empleado 8
0.01630002	Empleado 1
0.01592637	Empleado 2
0.01554982	Empleado 6
0.01479024	Empleado 3

Utilizando la Técnica de Análisis Multifactor en Excel pude obtener el ranking de todos los empleados del mayor al menor. Y como se puede ver en la imagen, el empleado número 9 es el mejor de los 10.

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
-------------	------	-----	-------	---	---	---

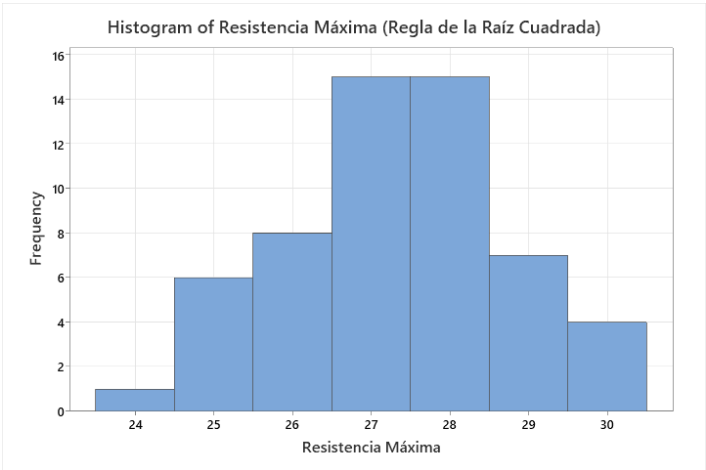
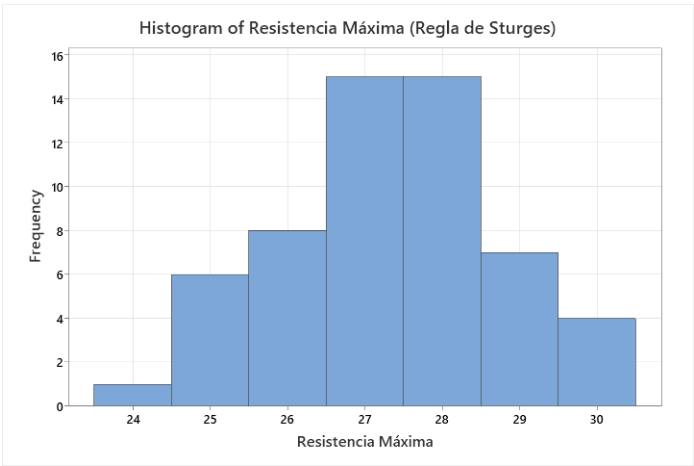
Tip 1: Utilizar la transformación MinMax Scaler para las variables predictoras antes de realizar la regresión. **Tip 2:** Transformar los datos de los nuevos empleados con los mismos parámetros de las variables originales para después meterlos en la ecuación de regresión.

C15-T	C16	C17	C18	C19	C20	C21
Empleados Nuevos	Salario_1	Costo de Capacitación_1	Producción Generada_1	Satisfacción del Cliente Inte_1	Ventas Generadas_1	Ausentismo_1
Empleado 11	0.272727	0.532544	0.500000	0.6	71178.6	0.25
Empleado 12	0.636364	0.710059	0.166667	0.4	72703.5	0.75
Empleado 13	0.545455	0.295858	0.833333	0.6	78412.0	0.50

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:

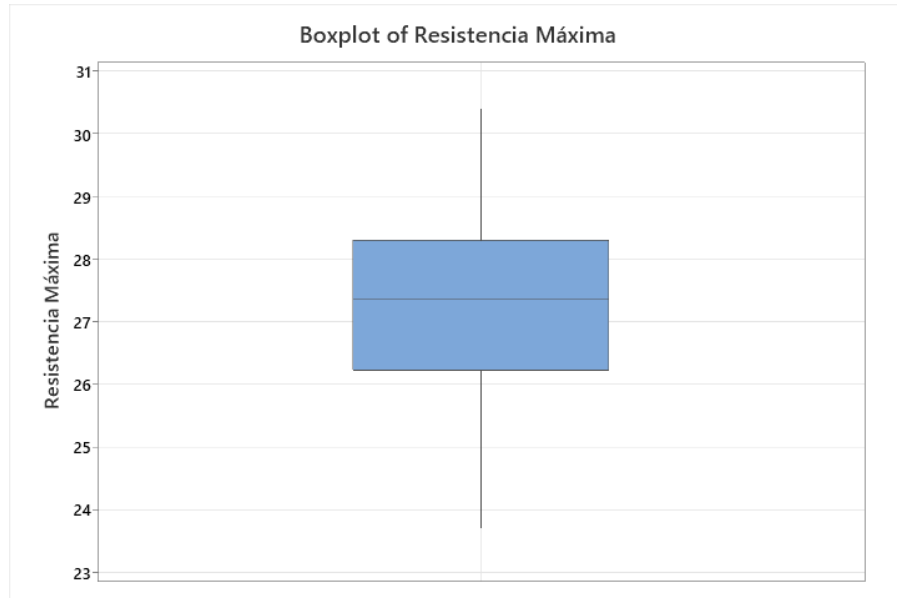
28.3	26.8	26.6	26.5	28.1	24.8	27.4	26.2	29.4	28.6	24.9	25.2	30.4	27.7	27.0	26.1	28.1
26.9	28.0	27.6	25.6	29.5	27.6	27.3	26.2	27.7	27.2	25.9	26.5	28.3	26.5	29.1	23.7	29.7
26.8	29.5	28.4	26.3	28.1	28.7	27.0	25.5	26.9	27.2	27.6	25.5	28.3	27.4	28.8	25.0	25.3
27.7	25.2	28.6	27.9	28.7												

- a) ¿Qué tipo de variable se está midiendo? ¿Discreta o continua? Explique.
Variable continua: La resistencia medida es una variable continua porque puede tomar cualquier valor en un rango, no se limita a valores enteros específicos. En este caso, se mide en kilogramos de fuerza, y puede tomar valores como 28.3, 26.8, etc.
- b) Haga un análisis exploratorio de estos datos.
 - a. Realice un histograma con al menos 2 reglas para definir el número de clases (No utilizar regla empírica). Describa la forma y analice el comportamiento de los datos.



- Podemos observar una forma de campana para los 2 histogramas de las diferentes reglas. Lo que nos inclina a creer que los datos tienen una distribución normal.
- Las 2 reglas me dieron como resultado que tenía que poner 7 intervalos al histograma, es por eso por lo que se ven iguales.

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



- Con este boxplot podemos observar que este data set no cuenta con ningún dato outlier.
- Mediana: La línea dentro de la caja representa la mediana de la resistencia máxima. Esto indica el valor central de los datos y como podemos apreciar se encuentra aproximadamente en el 27.5.

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

Descriptive Statistics

N	Mean	StDev	SE Mean	94% CI for μ
56	27.246	1.430	0.191	(26.879, 27.614)

μ : population mean of Resistencia Máxima

La resistencia promedio sería 27.246 por lo que podemos estar razonablemente seguros de que el proceso de fabricación de las botellas cumple con el estándar requerido.

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.

Test

Null hypothesis $H_0: \mu = 25$

Alternative hypothesis $H_1: \mu \neq 25$

T-Value	P-Value
11.75	0.000

Realizando una prueba de hipótesis se obtuvo un p-value de 0 por lo cual se rechaza la hipótesis nula, con esta evidencia se puede concluir que la resistencia promedio de los envases no es 25 kg, es decir, la resistencia promedio es diferente de 25 kg.

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

Dado que los intervalos de confianza para la desviación estándar poblacional se superponen bastante entre los dos métodos podemos decir que la desviación estándar poblacional del proceso está estimada dentro del rango de aproximadamente 1.17 a 1.83.

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
--------	----	----	----	----	----	----	----	----	----	----

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

En este caso, se tienen dos grupos distintos: hombres y mujeres. No hay ninguna indicación de que las mediciones en los hombres estén emparejadas con las mediciones en las mujeres (por ejemplo, no se especifica que se hayan tomado parejas de datos de hermanos, parejas, etc.). Por lo tanto, las muestras se consideran independientes.

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.

Test

Null hypothesis $H_0: \mu_1 - \mu_2 = 0$

Alternative hypothesis $H_1: \mu_1 - \mu_2 \neq 0$

T-Value	DF	P-Value
3.53	16	0.003

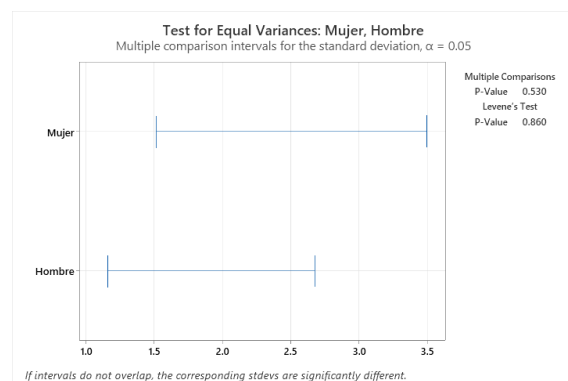
- Dado que estamos comparando las medias de dos muestras independientes, la prueba estadística apropiada es la prueba t para dos muestras independientes.
- Rechazamos la hipótesis nula (H_0), ya que el valor p es menor que 0.05.
- Hay una diferencia significativa en la temperatura promedio más confortable entre hombres y mujeres. En promedio, las mujeres encuentran una temperatura más confortable en comparación con los hombres.

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

d)

Tests

Method	Test Statistic	P-Value
Multiple comparisons	0.39	0.530
Levene	0.03	0.860



- Para comparar la variabilidad entre los dos grupos, la prueba estadística adecuada es la prueba de Levene o la prueba F para dos varianzas.
- Dado que el valor p es mayor que 0.05, no se rechaza la hipótesis nula. Esto indica que no hay evidencia suficiente para afirmar que las varianzas de las temperaturas más confortables son diferentes entre hombres y mujeres. Las varianzas parecen ser similares.

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.

Las muestras son dependientes porque las mediciones del método actual y del método nuevo se realizan en los mismos discos. Cada par de valores corresponde al mismo disco, lo que significa que las mediciones están emparejadas.

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

Test

Null hypothesis $H_0: \mu_{\text{difference}} = 0$

Alternative hypothesis $H_1: \mu_{\text{difference}} \neq 0$

T-Value P-Value

-0.24 0.814

- Dado que las muestras son dependientes (emparejadas), se debe realizar una Prueba T de Diferencias de Medias para Muestras Relacionadas.
- El valor de p obtenido es 0.814. Dado que este valor es mucho mayor que el nivel de significancia comúnmente utilizado ($\alpha = 0.05$), no se rechaza la hipótesis nula.
- Esto sugiere que no hay evidencia estadística suficiente para afirmar que hay una diferencia significativa en la densidad mínima medida entre el Método Actual y el Método Nuevo.

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

- Dado que la prueba estadística no encontró una diferencia significativa entre los resultados obtenidos por el Método Actual y el Método Nuevo (el valor de p es 0.814, lo que indica que las diferencias observadas son probablemente debidas al azar), se puede recomendar la adopción del nuevo método.
- La razón principal para recomendarlo es que el nuevo método reduce el tiempo de prueba en un 60%, lo que implica una mejora significativa en la eficiencia operativa sin comprometer la precisión de las mediciones.