

Guía de ejercicios

**Prueba de hipótesis basada en una muestra,
Prueba de hipótesis para dos poblaciones,
Prueba de Independencia – Homogeneidad de proporciones – Bondad de Ajuste,
Pruebas de normalidad, Pruebas no paramétricas**

1. Cierta fábrica de ventanas y puertas comerciales cuenta con una maquinaria que está configurada para producir una sección de acero de 0.65 cm de espesor. Se seleccionó una muestra aleatoria de 14 secciones de acero y se registró el espesor de cada una en la siguiente tabla:

0.541	0.757	0.599	0.823	0.645	0.688	0.518
0.640	0.780	0.754	0.765	0.739	0.564	0.625

Suponga que la distribución del espesor de la sección es normal.

- a) Evaluar mediante una prueba de hipótesis si el verdadero espesor medio de la sección de acero es superior a 0.65 centímetros.
- Indique la hipótesis nula y alternativa en términos de la media poblacional.
 - Identificar los dos posibles errores que se pueden cometer.
 - ¿Cuánto es el valor del p-value asociado a la prueba?
 - Usando $\alpha = 0.01$, ¿cuál es la conclusión?
- b) Si la máquina está produciendo secciones de aceros con tamaños de espesor muy variados, la producción debe detenerse y realizar una inmediata calibración a la maquinaria. Al 5% de significación, ¿hay alguna evidencia que sugiera que la desviación estándar de los espesores de las secciones de acero sea inferior a 0.20 centímetros y que deba detenerse la producción?
2. La cámara de comercio de una ciudad se encuentra interesada en estimar la cantidad promedio de dinero que gasta la gente que asiste a convenciones que se llevan a cabo en la ciudad, se seleccionaron 16 personas y se les preguntó la cantidad que gastaban por día. Se obtuvo la siguiente información en dólares:

150	175	163	148	142	189	135	174
168	152	158	184	134	146	155	163

Un reporte de una investigadora de mercados afirma que la desviación estándar real es mayor a 4.5. Pruebe a un nivel de significación del 5% si la investigadora de mercados está en lo cierto. Además, suponga que la cantidad de dinero gastado en un día es una variable aleatoria con distribución Normal.

3. Una de las últimas innovaciones en el campo de la iluminación es el proyector de cabeza móvil. Este proyector está compuesto por un foco halógeno que permite variar la intensidad luminosa. El gerente de ventas de la empresa Visual System S.A. ha decidido este año exportar este novedoso proyector si la **intensidad luminosa de estos proyectores es superior a 2600 lúmenes**. Al seleccionar 12 proyectores y medir la intensidad luminosa de cada proyector se obtuvieron los siguientes datos:

2600	2800	3000	2000	2400	2800	2700	2600	3000	3200	3100	2500
------	------	------	------	------	------	------	------	------	------	------	------

Suponiendo que la intensidad luminosa tiene distribución normal, ¿Visual System S.A. deberá decidir realizar la exportación del lote? Use $\alpha = 0.03$

4. Algunos docentes creen que el éxito escolar está directamente relacionado con la cantidad de tiempo dedicado a completar las tareas escolares. El archivo **“Tareas.xlsx”** contiene información sobre los puntajes de progreso obtenidos por una muestra de estudiantes de nivel secundaria y el tiempo empleado en el desarrollo de sus tareas secundarias (Menos de una hora, Más de dos horas).
 - a) ¿Es razonable concluir que el puntaje de progreso promedio de los estudiantes que dedican más de dos horas a sus tareas escolares sea inferior a 312? Use $\alpha = 0.05$
 - b) ¿Hay alguna evidencia de que el puntaje de progreso promedio de los estudiantes que dedican menos de una hora a sus tareas escolares sea mayor a 280? Use $\alpha = 0.05$
 - c) Al 3% de significación, ¿existe evidencia para afirmar que la proporción de estudiantes que dedican más de dos horas a sus tareas escolares es superior a 0.35?
 - d) Al 1% de significación, ¿se puede concluir que los estudiantes de nivel secundario que dedican más tiempo a sus tareas escolares obtienen mejores puntajes?
5. Una empresa dedicada a la confección de zapatos tiene tres sucursales (A, B y C) ubicadas en distritos diferentes. Debido a las restricciones presupuestales que atraviesa, desea cerrar una de sus sucursales, pero para tomar esa decisión quiere hacerlo considerando una evaluación estadística que le dé el respaldo que necesita, además de otros elementos como puede ser por ejemplo la situación actual del mercado. Dentro de la evaluación técnica la gerencia ha considerado utilizar la variable: tiempo de producción de un par de zapatos (en minutos). Se tomaron los tiempos de

producción de 31 pares de zapatos elegidos al azar de cada uno de las sucursales y se obtiene el siguiente reporte:

Variables	N	Media	Desviación	Calidad muy buena
Tiempo "A"	31	14.184	3.132	22
Tiempo "B"	31	15.020	3.984	16
Tiempo "C"	31	14.882	4.156	18

- ¿Se puede afirmar que el tiempo promedio de fabricación de un par de zapatos en la sucursal C es mayor de 14 minutos? Use un nivel de significancia del 0.05.
 - ¿Se puede afirmar que la proporción de calzados de Calidad Muy Buena de la sucursal C es mayor de 0.55? Use $\alpha = 0.05$
6. Se desea probar si el salario medio mensual (en dólares) de los empleados oficinistas de 2 empresas del ramo de servicios turísticos son iguales o no. Para ello, se toman muestras de ambas empresas y los datos correspondientes se muestran a continuación:

Empresa 1	945	882	741	946	817	971	744	816	900	755	846	823	997	916	801
Empresa 2	923	760	983	949	916	982	500	500	899	706	824	798	958	934	712

- Pruebe si las varianzas son iguales. Use $\alpha = 0.05$.
 - Pruebe si existe diferencia entre los salarios de los empleados de ambas empresas. Use $\alpha = 0.05$.
7. En una clínica de reducción de peso se afirma que su programa permite reducir el peso. En la tabla siguiente se muestra el resultado que obtuvieron 10 personas. Compruebe si la afirmación de la clínica es correcta, con un nivel de significación de 5 por ciento.

Cliente	Peso antes	Peso después
1	85.9	77.1
2	91.7	86.5
3	100.2	96.7
4	94.1	87.4
5	88.2	81.7
6	80.3	73.3
7	87.7	79.1
8	91.9	85.1
9	94.6	84.6
10	105.9	92.6

8. En un proceso de producción se encontraron 35 artículos defectuosos dentro de una muestra aleatoria de 500, y se identificaron 20 defectuosos en otra muestra de 400 artículos provenientes de un proceso similar que se lleva a cabo en otra fábrica. Compruebe la hipótesis que afirma que los dos procesos producen la misma proporción de artículos defectuosos, con un nivel de significación de 1%.
9. Cada vez más personas utilizan las redes sociales para establecer contactos y el uso de motores de búsqueda de empleo en la web para encontrar una próxima oferta laboral. Cierta consultora recientemente ha realizado una encuesta sobre el uso de LinkedIn para búsqueda de trabajo y cuenta también con datos de una encuesta anterior realizada hace cinco años, los cuales se resumen a continuación:

Encuesta	Tamaño de muestra	Número de personas que usan LinkedIn para buscar trabajo
Reciente	1320	620
Hace cinco años	1050	530

- a) ¿Cuál es la proporción estimada de personas que usan LinkedIn para buscar trabajo para cada encuesta realizada?
- b) ¿Hay alguna evidencia que sugiera que la verdadera proporción de personas que usan LinkedIn para buscar trabajo hoy en día se ha incrementado en comparación con la de hace cinco años? Use $\alpha = 0.03$.
10. Un estudio evaluó a novecientos cincuenta escolares los cuales se clasificaron de acuerdo a sus hábitos alimenticios y a su coeficiente intelectual:

País	Nivel de satisfacción				Total
	< 80	80 - 90	91 - 99	> 99	
Nutrición buena	245	228	177	219	869
Nutrición pobre	31	27	13	10	81
Total	276	255	190	229	950

A un nivel de significación del 5%, pruebe si hay relación entre el nivel de coeficiente intelectual y el estado nutricional de los niños.

11. Una empresa de investigación de mercados está interesada en analizar la relación entre la edad de una persona y la fuente preferida de noticias de la persona. Los resultados de una muestra aleatoria de 1000 personas se encuentran en el archivo “**Noticias.csv**”, que contiene las siguientes variables:

- **Edad:** Categoría de edad de la persona (Joven, Adulto, Adulto mayor).
- **Fuente:** Fuente preferida de noticias de la persona (Periódico, Radio/TV, Internet).

- a) Mostrar la tabla de frecuencias observada y esperadas.
- b) A un nivel de significancia del 5%, ¿existe evidencia de asociación entre las dos variables mencionadas?
- c) Usando $\alpha = 0.05$, evaluar si la fuente preferida de noticias de la persona se distribuye de la misma manera.
- d) ¿Hay alguna evidencia de que la proporción de personas jóvenes que prefieren internet como fuente de noticias sea superior a la proporción de personas adultas que prefieren internet como fuente de noticia? Use $\alpha = 0.05$.

12. Una empresa multinacional desea conocer si existen diferencias significativas entre sus trabajadores en distintos países en el grado de satisfacción en el trabajo. Para ello se toman muestras aleatorias simples de trabajadores, obteniendo los siguientes resultados:

País	Nivel de satisfacción			
	Muy satisfecho	Satisfecho	Insatisfecho	Muy insatisfecho
España	200	300	300	100
Francia	300	400	350	150
Italia	350	300	250	150

¿Puede admitirse con un nivel de significación del 5% que la satisfacción en el trabajo es similar en los tres países?

13. El encargado del área de marketing de un nuevo cine realizó un estudio para determinar la preferencia de jóvenes en una opción sobre 3 tipos de películas. Para ello se tomó una muestra aleatoria de 200 jóvenes y se obtuvieron los siguientes resultados:

Tipo de película	Terror	Comedia	Acción
Frecuencia	80	50	70

- a) Usando un nivel de significación de 0.05, probar si la preferencia de los jóvenes se distribuye de la misma manera. Concluya de acuerdo a los resultados.
 - b) Usando un nivel de significación de 0.05, probar la preferencia de los jóvenes se distribuye en la proporción de 2/6, 1/6, 3/6. Concluya de acuerdo a los resultados.
14. Durante las primeras 13 semanas de la temporada de televisión, se registraron las audiencias de sábado por la noche, de 8:00 p.m. a 9:00 pm. Como sigue: ABC 29%, CBS 28%, NBC 25% y otros 18%. Dos semanas después, una muestra de 300 hogares seleccionados aleatoriamente arrojó los siguientes resultados de audiencia: ABC 95 hogares, CBS 70 hogares, NBC 89 hogares y otros 46 hogares. Pruebe, con nivel de significación 0.05, si han cambiado las proporciones de telespectadores.

15. Se realizó una encuesta para saber si existe una relación entre el género y la confianza que la gente tiene en los alimentos transgénicos. Los resultados de una muestra aleatoria se presentan en la siguiente tabla:

	Confianza en los alimentos transgénicos		
Género	Mucha	Regular	Muy Poca
Hombre	115	56	29
Mujer	175	91	31

Use un nivel de significación del 0.05 para probar si existe una relación del género y la confianza en los alimentos transgénicos.

16. En un estudio para determinar la opinión de los ingenieros sobre un nuevo tipo de proceso en la verificación de maquinarias se tomó una muestra aleatoria de 400 ingenieros en una región, obteniéndose los siguientes resultados:

Opinión	Muy bueno	Bueno	Regular	Malo	Muy malo
Frecuencia	25	60	175	120	20

Probar si la opinión de los ingenieros respecto al nuevo tipo de proceso de verificación no se distribuye en la proporción del 10%, 20%, 30%, 25% y 15%, respectivamente. Use nivel de significación del 0.01.

17. En el año 2000 se fundó la compañía “AGROSERV.S.A.”, encargada de la fabricación y distribución de implementos de placas de reemplazo de cierto tipo de artefacto, la cual tiene sucursales en los distritos de Lince, San Borja, San Miguel y San Martín. La gerencia realizó un estudio para poder evaluar el nivel de las ventas, la satisfacción del cliente y el proceso de ventas. Los clientes en ciertos casos solicitan descuentos por la compra de los implementos y esto puede cambiar según la ubicación de la sucursal. La forma de pago puede ser al crédito o al contado. Se realizó una encuesta piloto a los clientes de las cuatro sucursales, para evaluar la calidad del servicio brindado (Regular, Buena, Muy buena). Se tomó una muestra distinta de clientes atendidos en cada una de las sucursales (80 del Lince, 110 de San Borja, 105 de San Miguel y 75 de San Martín).

	Distritos			
Calidad del servicio	Lince	San Borja	San Miguel	San Martín
Regular	15	20	30	20
Buena	25	40	35	25
Muy buena	40	50	40	30

Realice la prueba más adecuada para este caso. Plantee las hipótesis respectivas y concluya. Use un nivel de significación del 0.05.

18. Un actuario que trabaja para una compañía de seguros desea examinar los registros de reclamos por robo de dinero hechos por personas aseguradas bajo una póliza de bienes familiares. En el pasado, la mediana de la cantidad de dinero producida en los reclamos fue de 570 soles. Se tomó una muestra aleatoria de 11 reclamos y los resultados obtenidos son los siguientes:

810	600	350	1100	540	570	540	1600	590	520	720
-----	-----	-----	------	-----	-----	-----	------	-----	-----	-----

- Pruebe si los datos se distribuyen normalmente.
 - Con los resultados de la parte a), utilice la prueba más adecuada para probar si existe evidencia de que la mediana de los reclamos ha aumentado significativamente. Use $\alpha = 0.05$.
19. Según las estadísticas encontradas en estudios anteriores, generalmente una persona desempleada encuentra un nuevo trabajo después de buscar aproximadamente 36 semanas. Sin embargo, recientemente el INEI sugiere que este tiempo de búsqueda se ha incrementado. Para evaluar esto, se toma una muestra aleatoria de 15 adultos desempleados y se registra el número de semanas que están desempleados.

35	40	38	43	38	45	35	47	41	37	39	43	49	39	34
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

- Evaluar visualmente y mediante la prueba de hipótesis respectiva si los datos se distribuyen normalmente.
 - ¿Hay alguna evidencia que sugiera que la mediana de duración del desempleo es mayor a 36 semanas? Use $\alpha = 0.10$.
20. Los siguientes datos muestran horas anuales perdidas por enfermedad de dos muestras entre hombres y mujeres de la empresa Textil Cotton. Al nivel de significancia de 0.05, ¿existe alguna diferencia atribuible al sexo? Asuma que los datos no se distribuyen normalmente.

Hombres	31	44	25	30	70	63	54	42	
Mujeres	38	34	33	47	58	83	18	36	41

21. La gerente de Recursos Humanos de una empresa desea comparar el número de tardanzas acumuladas por el personal en dos de sus sucursales. Pruebe a un nivel de significancia de 0.05 que la mediana del número de tardanzas de la sucursal A es superior a la sucursal B.

Sucursal A	4	3	3	5	5	3	3	5	5	3	4	3	3	5	5
Sucursal B	3	2	3	3	4	4	4	2	4	3	3	2	3	3	4

22. Los siguientes datos muestran los índices de trabajo defectuoso de los empleados antes y después de un cambio en el plan de incentivos al salario. Compare los dos conjuntos de datos siguientes para ver si el cambio disminuyó las unidades defectuosas producidas. Utilice el nivel de significancia de 0.10.

Antes	8	7	6	9	7	10	8	6	5	8	10	8
Después	6	5	8	6	9	8	10	7	5	6	9	8

23. En una clase de una escuela de enseñanza superior hay 48 alumnos varones, de los cuáles 12 viven en el campo y 36 en la ciudad. Se desarrolló una prueba para determinar la condición física de los alumnos. Se aplicó esta prueba a los 48 alumnos y se asignó a cada uno una puntuación. Una puntuación baja indica mala condición física. Los resultados obtenidos son los siguientes:

Campo		Ciudad					
14.8	10.6	2.2	16.9	7.6	2.4	6.2	1.2
7.3	12.5	14.2	7.9	2	6.4	6.1	10.6
5.6	12.9	12.6	16	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5
9	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1	3.2	5.9	4

- Evaluar gráficamente y por prueba de hipótesis si las puntuaciones de los alumnos que viven en el campo y en la ciudad provienen de una distribución normal. Considere un nivel de significación del 5%.
 - Pruebe con un nivel de significación del 10% si los alumnos que viven en el campo tienen mejor condición física que los que viven en la ciudad. Se conoce que la condición física no sigue una distribución Normal.
24. La administración de una empresa desea modificar el procedimiento de producción que utiliza actualmente (A) por otro procedimiento que se ha propuesto (B). Con el propósito de evaluar si el nuevo procedimiento (B) mejora los rendimientos se eligen a al azar a 8 de trabajadores, con los cuales se hacen las mediciones de la cantidad producida (en kilogramos) en un periodo de dos horas. Las cantidades producidas con cada procedimiento se proporciona en el siguiente cuadro.

Procedimiento	Trabajadores							
	1	2	3	4	5	6	7	8
Actual (A)	3	4	3.5	4	3.5	4	5	6
Nuevo (B)	4	5.5	4.5	5	4.5	5	6.5	6

¿Recomendaría el cambio de procedimiento de producción? Explique y sustente su respuesta con nivel de significación de 0.05.

25. El archivo “**MarketingDirecto.csv**” contiene datos de 1000 clientes de un vendedor de marketing directo el cuál vende sus productos sólo a través de correos electrónicos personalizados. El vendedor envía catálogos a los clientes con las características de los productos, y estos ordenan directamente de los catálogos. El responsable de marketing ha desarrollado registros de clientes para aprender qué hace que algunos clientes gasten más que otros, considerando las siguientes variables:

- **Edad:** Grupo etario del cliente (Adulta/Media/Joven).
- **Genero:** Género del cliente (Masculino/Femenino).
- **Vivienda:** Si el cliente es dueño de su casa (Propia/Alquilada)
- **Ecivil:** Estado civil (Soltero/Casado).
- **Ubicación:** Ubicación de un negocio que vende productos similares en términos de distancia (Lejos/Cerca).
- **Salario:** Sueldo anual de los clientes (en dólares).
- **Hijos:** Número de hijos (0-3).
- **Historial:** Historial del volumen de compra anterior (Bajo/Medio/Alto/NA).
NA significa que este cliente aún no ha adquirido ningún producto.
- **Catalogos:** Número de catálogos enviados.
- **Monto:** Gasto en dólares.

Responda a las siguientes preguntas para un nivel de significancia de 0.05:

- Evalúe si existe asociación entre el grupo etario del cliente y el historial del volumen de compra anterior.
- Evalúe visualmente y mediante la prueba de hipótesis respectiva si el sueldo anual de los clientes tiene distribución normal.
- ¿Es posible afirmar que los clientes tienen un sueldo medio mayor a 53 mil dólares?
- ¿Es posible concluir que la proporción de clientes jóvenes es menor a 0.30?
- Pruebe si los hombres tienen un mayor gasto que las mujeres.
- ¿Es posible afirmar que la proporción de clientes que gastan más de 2 mil dólares es menor en los clientes solteros que en los casados?