XS-2110 Métodos Estadísticos

Prácticas

PARA EXAMEN PARCIAL I

1. En un estudio sobre envejecimiento, se seleccionaron al azar 14 adultas mayores de Llorente de Tibás y se les preguntó a cuántos hijos e hijas había dado a luz durante su vida reproductiva. Las siguientes fueron las respuestas:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Identificación** | **# de hijos** |  | **Identificación** | **# de hijos** |
| **1** | 4 |  | **8** | 5 |
| **2** | 1 |  | **9** | 4 |
| **3** | 2 |  | **10** | 4 |
| **4** | 7 |  | **11** | 1 |
| **5** | 3 |  | **12** | 2 |
| **6** | 3 |  | **13** | 5 |
| **7** | 2 |  | **14** | 3 |

1. Calcule la moda de la distribución (2 puntos).
2. Al 5% de significancia, pruebe con la prueba de Shapiro Wilk si esta muestra fue seleccionada de una distribución normal. Dé repuesta completa de cualquier prueba de hipótesis. Verifique la respuesta manualmente y con R.
3. Al 5% de significancia, pruebe con la prueba de Kolmogorov Smirnov si esta muestra fue seleccionada de una distribución normal con media igual a 3.5 y desviación estándar igual a 1.3. Dé repuesta completa de cualquier prueba de hipótesis. Verifique la respuesta manualmente y con R.
4. Al 5% de significancia, pruebe la hipótesis de que el promedio de hijos es igual a 3.5. Verifique la respuesta tanto manualmente como con R.
5. Al 5% de significancia, pruebe la hipótesis de que la mediana de hijos es igual a 3.5. Verifique la respuesta tanto manualmente como con R.
6. Diseñe un cuadro de distribución de frecuencias para el número de hijos, con las categorías: 1 hijo (o menos), 2 hijos, 3 hijos, 4 hijos, 5 hijos ó más, y realice una prueba de hipótesis X2 de bondad de ajuste de la variable número de hijos a una distribución Poisson con lambda=3.5, con un α=0.05.
7. Conteste: cuál es el principal problema de usar la prueba X2 para estos datos?
8. ¿Qué problema habría de usar una prueba de Kolmogorov Smirnov a este caso para analizar la bondad de ajuste a una distribución de Poisson?
9. Una empresa de investigación de mercados seleccionó al azar a 18 personas que subían las gradas del mall San Pedro a la hora del almuerzo y observó cuántas aceptaban los volantes de ofertas de comidas y cuántos no. La empresa codificó la variable de la siguiente manera:

1= Sí tomó el volante; 0= NO tomó el volante;

La siguiente es la muestra que encontró:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Identificación** | **Tomó el volante** |  | **Identificación** | **Tomó el volante** |
| **1** | 1 |  | **10** | 0 |
| **2** | 0 |  | **11** | 1 |
| **3** | 0 |  | **12** | 0 |
| **4** | 1 |  | **13** | 0 |
| **5** | 0 |  | **14** | 0 |
| **6** | 0 |  | **15** | 0 |
| **7** | 0 |  | **16** | 0 |
| **8** | 0 |  | **17** | 1 |
| **9** | 1 |  | **18** | 0 |

Con un 10% de significancia, pruebe con una prueba exacta (binomial) que menos de un 50% de las personas aceptan uno de estos volantes de ofertas. Realice la prueba tanto manualmente como con R (5 puntos).

1. Un informático tiene un local donde alquila microcomputadoras al público para que consulten sus correos electrónicos. En promedio las personas duran 3.1 minutos y se desea incrementar dicho tiempo de duración con alguna estrategia. Luego de implementada tal estrategia, se seleccionó una muestra de 20 personas y se determinó un tiempo promedio de duración de 4.6 minutos, con una desviación estándar de 2.33 minutos. Se asume que la variable de interés se distribuye normalmente.
   1. Con un nivel de significancia del 5%, realice la prueba de hipótesis “manualmente” de que la duración promedio ha aumentado (puede utilizar cualquiera de los dos métodos: con base en cuantil, o con las unidades de medición). (5 puntos)
   2. Confirme los resultados usando R.
   3. Calcule la probabilidad de incurrir en el error tipo II si promedio verdadero fuese de 4.695 minutos (5 puntos).
2. El encargado de mercadeo de un banco quiere ver si existe preferencia entre los clientes por alguno de los 4 cajeros automáticos que tiene dentro del edificio central. Selecciona a 80 personas al azar, y con las cámaras de seguridad, las sigue para ver qué cajero automático utilizaron. Si denominamos a los 4 cajeros automáticos A, B, C y D, al 1% de significancia, pruebe la hipótesis de que la escogencia de cajeros automáticos sigue una distribución uniforme. (Escoja la prueba apropiada para las condiciones que se le dan). Los resultados de la muestra se presentan en la siguiente tabla: (5 puntos)

Muestra de clientes seleccionados para el estudio de cajeros automáticos, según cajero automático utilizado.

|  |  |
| --- | --- |
| Cajero automático | Frecuencia absoluta (fi) |
| A | 17 |
| B | 24 |
| C | 19 |
| D | 20 |

1. Según una empresa de papel higiénico, la participación de mercado (porcentaje de gente que compra un producto) de su marca estrella es de 80%. Ante la llegada de una multinacional del papel higiénico, el gerente de marca contrata a una firma de investigación de mercados para analizar si la participación de mercado ha disminuido. La firma selecciona una muestra de 400 personas, de las cuales 300 dicen que compran la marca estrella de la empresa.

* 1. ¿Con un α=0.01, hay suficiente evidencia estadística para pensar de que la participación de mercado de la marca estrella haya disminuido? (3 ptos.)
  2. Si un 75% de participación de mercado significa una disminución importante de ingresos para la empresa, ¿cuál es la potencia de la prueba con esta muestra de 400 personas para detectar una nueva participación de mercado de 75%? (5 ptos.)

1. Se decidió investigar si los niveles de HbA1C de 15 hombres. Con base en esa información, al 5% de significancia, utilice la prueba de Kolmogorov Smirnov para analizar que los datos provienen de una distribución Dagum con parámetros a=1 y b=10 y p=1 (Investiguen de dónde sacar la distribución Dagum). Los datos se encuentran abajo.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.5 | 4.0 | 4.7 | 4.9 | 5.3 | 5.5 | 5.7 | 5.8 | 6.5 | 6.8 | 7.0 | 7.2 | 7.4 | 7.6 | 8.6 |

1. Con los datos anteriores, contrate la hipótesis de que la mediana poblacional es diferente a 7 con una significancia del 10%. Verifique los resultados manualmente y con R.
2. A continuación se le presenta la distribución de frecuencias de la variable altura de la rodilla (en cm) de una muestra aleatoria de 160 cadáveres en la Morgue Judicial de San Joaquín de Flores. Realice un contraste de hipótesis de bondad de ajuste de razón de verosimilitudes para analizar si la variable altura de la rodilla proviene de una distribución uniforme continua, con un α=0.05. (5 ptos.)

Cuadro 1. Muestra de cadáveres humanos en la Morgue Judicial en 2009, según altura de la rodilla (en cm).

|  |  |  |
| --- | --- | --- |
| Altura de la rodilla (en cm.) | Frecuencia absoluta simple | Frecuencia relativa simple  (en %) |
| 24 a menos de 28 | 20 | 12.3 |
| 28 a menos de 32 | 29 | 17.9 |
| 32 a menos de 36 | 28 | 17.3 |
| 36 a menos de 40 | 28 | 17.3 |
| 40 a menos de 44 | 29 | 17.9 |
| 44 a menos de 48 | 28 | 17.3 |
| **Total** | **162** | **100.0** |

1. La gerente de un supermercado tradicionalmente sabe que el 20% de sus clientes compra al menos un producto de los anaqueles (o góndolas) que están justo a la par de las cajas registradoras, y se preocupa cuando algún subgrupo de clientes tiene menor probabilidad de comprar en estas góndolas. La cadena de supermercados está desarrollando una campaña de mercadeo para atraer a profesionales solteros menores de 30 años. Sin embargo, la gerente nota que muy pocos de estos profesionales solteros compran en los anaqueles a la par de las cajas. Si solo 10% de estos profesionales compran en estos anaqueles, entonces quiere cambiar el “merchandising” (estrategia de apariencia) de los mismos para incentivar la compra.
2. Si planea tomar una muestra de 30 clientes profesionales solteros y analizarlos con una significancia del 5%, ¿cuál es el tamaño de la potencia de la prueba para detectar que 10% de estos clientes compran productos en los anaqueles?
3. Si realiza la observación en estos 30 clientes, y descubre que 5 de ellos compran productos de los anaqueles en cuestión:
   * 1. Plantee la hipótesis nula y alternativa del problema
     2. Contraste la prueba de hipótesis del problema, con α=0.05
4. Un geólogo está haciendo un estudio global sobre los costos de construir torres petroleras de prospección en el Caribe. El geólogo supone que la probabilidad de encontrar petróleo en el Caribe Este es de 0.25. Sabe además que se decide construir una plataforma petrolera si después de X intentos se encuentra petróleo. Desde el punto de vista estadístico, la cantidad de intentos para encontrar petróleo se distribuye como una distribución geométrica con parámetro p. Estudia 30 proyectos petroleros escogidos al azar y analiza si la cantidad de intentos realizados en cada uno de esos proyectos se distribuye geométricamente con parámetro p=0.25.
   1. La fórmula de la probabilidad de una distribución geométrica es:

P(X=x)= (1-p)[x-1] \* p.

Llene el cuadro de abajo calculando las probabilidades [P(X=x)] con la fórmula en cuestión.

* 1. En el mismo cuadro de abajo se presenta la distribución de frecuencias de la variable “número de intentos antes de encontrar petróleo” para los 30 proyectos. Con base en las respuestas del inciso anterior y la distribución de frecuencias, contraste la hipótesis nula de que los 30 proyectos son una muestra de una población con distribución geométrica con parámetro p=0.25, usando la prueba de X2.

Cuadro 2

Muestra de 30 proyectos petroleros: Número de intentos de prospección

para encontrar petróleo.

|  |  |  |
| --- | --- | --- |
| Número de intentos para encontrar petróleo | Frecuencia absoluta simple | P(X=x) |
| 1 | 8 |  |
| 2 | 7 |  |
| 3 | 6 |  |
| 4 | 5 |  |
| 5 | 3 |  |
| 6 ó más | 1 |  |
| Total | 30 |  |

1. Pruebas efectuadas con seis modelos experimentales de motores mostraron que permanecieron operando durante 24,28,21,23,32, y 22 minutos con un galón de cierta clase de combustible. Los ingenieros diseñaron los motores para que duraran en promedio 29 minutos, pero se cree que pueden durar más. Si la probabilidad de cometer un error de Tipo I es de 0.01, ¿es esto evidencia en contra de la afirmación de que en promedio esta clase de motor operará en promedio más de durante 29 minutos por galón de esta clase de combustible? Si supone que el tiempo se distribuye normalmente, pruebe la hipótesis con una prueba t.
2. Tome los datos del problema anterior. Si no se puede suponer que el tiempo se distribuye normalmente, pruebe la hipótesis de que la mediana del tiempo es de 29 minutos, vs la hipótesis alternativa de que el tiempo mediano es mayor a 29 minutos, con un alfa de 0.01.
3. Para efectos de mejorar el servicio, un servicio de ambulancias notó que el 40% de sus llamadas son urgencias con peligro de muerte, pero se cree que este porcentaje ha disminuido en el tiempo. Se toma una muestra aleatoria de sus archivos más recientes, y se encuentra que sólo 49 de 150 llamadas son de este tipo. ¿Qué conclusión se puede sacar con esta muestra, con una significancia del 1%?
4. En una muestra de 12 automóviles que dan la vuelta a la derecha en una cierta intersección, 3 estaban alineados en el carril erróneo. Con un alfa de 0.05, pruebe la hipótesis nula de que el 30% de todos los conductores cometen esta equivocación en el cruce citado, vs. la hipótesis alternativa de que esta proporción es distinta al 0.30.

PARA EXAMEN PARCIAL II

1. Un investigador de la opinión pública realizó una encuesta probabilística hace dos años acerca de la confianza que tenían los aficionados al fútbol en Ricardo Lavolpe como seleccionador. Un 33% (66 de 200) de los encuestados decía que Ricardo Lavolpe era el mejor entrenador que había pasado por Costa Rica. Después de los resultados negativos de los últimos meses de su gestión, el investigador contactó a los mismos 200 encuestados, y un 25% (50 de 200) le contestaron que creían que Lavolpe seguía siendo el mejor entrenador para la selección. El investigador quiere probar estadísticamente que el porcentaje de personas que creen que Lavolpe es lo mejor ha cambiado? Los datos que comparan las dos muestras están en la tabla de más abajo.
   1. ¿Por qué la prueba de McNemar es apropiada para analizar estos datos?
   2. Realice la prueba de McNemar para determinar si la proporción de personas con opiniones positivas cambió con un α=0.01. Plantee las hipótesis nula y alternativa. Verifique los resultados manualmente y con R.
   3. Realice una prueba de hipótesis de signos para analizar la misma hipótesis, con un α de 0.01. (4 ptos.). Verifique los resultados manualmente y con R.

Cuadro 1. Aficionados al fútbol: Opinión sobre Lavolpe en 2009 y 2011. (Cree que Lavolpe es el mejor seleccionador para Costa Rica).

|  |  |  |  |
| --- | --- | --- | --- |
| 2006 |  | 2008 | Total |
|  | Sí | No |  |
| Sí | 48 | 18 | 66 |
| No | 2 | 132 | 134 |
| Total | 50 | 150 | 200 |

1. Clyde Bloomquist ha propuesto un cambio en la política corporativa respecto a la recolección de cuentas por cobrar. Considera que reducirá en tiempo necesario para obtener el pago de deudas pendientes de los acreedores. Los registros de la compañía demuestran que 8 de los acreedores tomaron el número de días que aparecen en la tabla, antes y después del cambio de política para remitir los fondos vencidos. ¿Está Clyde en lo cierto? ¿Debería mantenerse el cambio de política? Use un α=0.05, y analice el problema con la t-pareada, la prueba de Wilcoxon y la prueba de signos. Verifique la respuesta manualmente y con R.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Acreedor | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Antes | 18 | 27 | 32 | 23 | 31 | 36 | 18 | 35 |
| Después | 12 | 22 | 31 | 24 | 28 | 24 | 16 | 25 |

1. La Srita. Beverlee Hills debe determinar si el nivel de consumo (en dólares semanales) está asociado a cuatro grupos demográficos con los que una tienda segmenta su mercado. Los datos se dan a continuación:

|  |  |  |  |
| --- | --- | --- | --- |
| Hombres casados | Mujeres casadas | Hombres solteros | Mujeres solteras |
| 50 | 20 | 19 | 87 |
| 17 | 23 | 32 | 20 |
| 23 | 82 | 66 | 95 |
| 48 | 46 | 72 | 34 |
| 63 | 13 | 41 | 11 |

1. ¿Cuál es la medida de asociación más adecuada para analizar el problema de la Srita Hills? Justifique.
2. Calcule la medida de asociación y contraste la hipótesis nula de no asociación, con una significancia del 5%.
3. Los economistas del Mid-West Research Institute están realizando un estudio para analizar la relación entre los ingresos de las personas y sus niveles de consumo. Once consumidores reportaron las siguientes cifras en miles de dólares anuales. Con estos datos, conteste las preguntas de más abajo:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ingreso | 97 | 58 | 69 | 47 | 58 | 38 | 91 | 67 | 68 | 47 | 48 |
| Consumo | 55 | 63 | 54 | 37 | 45 | 38 | 71 | 52 | 53 | 37 | 37 |

1. Calcule e interprete el coeficiente de correlación producto-momento de Pearson, y analice la hipótesis nula de no asociación entre las variables.
2. Calcule e interprete el coeficiente de correlación de rangos de Spearman, y analice la hipótesis nula de no asociación entre las variables.
3. Calcule e interprete la tao de Kendall, y analice la hipótesis nula de no asociación entre las variables.
4. Ingrese los datos a R, analice el supuesto de normalidad y de asociación lineal de las variables, y diga qué medida de asociación recomendaría.
5. Una fábrica de herramientas está analizando la relación que existe entre la vida útil de una herramienta (en días) y la humedad a la que está expuesta. Tomó una muestra de 15 herramientas, y las expuso durante un mes a distintos niveles de humedad. Los datos se muestran abajo. Se presenta además, un gráfico de dispersión entre las dos variables.
   1. Calcule los coeficientes de correlación de Pearson, Spearman y Tau de Kendall.
   2. Realice la prueba de hipótesis de que los coeficientes de correlación de Pearson, Spearman y tau de Kendall son distintos a cero. Plantee las hipótesis nula y alternativa (α=0.05).
   3. Si suponemos efectivamente que los datos en las variables provienen de variables que se distribuyen normalmente a nivel poblacional, con base en el gráfico, cuál de los tres coeficientes de correlación escogería, y por qué?

|  |  |  |  |
| --- | --- | --- | --- |
| Unidad | vida (días) | humedad (%) |  |
| 1 | 8 | 95 |
| 2 | 7 | 90 |
| 3 | 7 | 85 |
| 4 | 8 | 80 |
| 5 | 9 | 75 |
| 6 | 10 | 70 |
| 7 | 14 | 65 |
| 8 | 12 | 60 |
| 9 | 12 | 55 |
| 10 | 13 | 50 |
| 11 | 14 | 45 |
| 12 | 14 | 40 |
| 13 | 11 | 35 |
| 14 | 16 | 30 |
| 15 | 16 | 25 |

1. Amcho Tech está considerando si comercializar un disco duro para servidores. Se realiza un experimento con 8 discos duros seleccionados aleatoriamente para determinar si hay una relación entre el número de horas en las que se prueba un disco duro antes de la venta y el número de veces que el disco falla en el proceso de terminar de correr un programa de computación intensivo en datos. Los datos se presentan a continuación:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Horas | 100 | 96 | 88 | 78 | 75 | 68 | 60 | 55 |
| Tasa de fallas | 2 | 4 | 10 | 10 | 13 | 6 | 16 | 20 |

1. Calcule e interprete el coeficiente de correlación producto-momento de Pearson, y analice la hipótesis nula de no asociación entre las variables.
2. Calcule e interprete el coeficiente de correlación de rangos de Spearman, y analice la hipótesis nula de no asociación entre las variables.
3. Calcule e interprete la tao de Kendall, y analice la hipótesis nula de no asociación entre las variables.
4. Ingrese los datos a R, analice el supuesto de normalidad y de asociación lineal de las variables, y diga qué medida de asociación recomendaría.
5. En una planta química se está analizando la pureza de oxígeno líquido licuado con la presión que tiene el oxígeno líquido envasado. Los datos se presentan abajo
   1. Calcule los coeficientes de correlación de Pearson, Spearman y Tau de Kendall.
   2. Realice la prueba de hipótesis de que los coeficientes de correlación de Pearson, Spearman y tau de Kendall son distintos a cero. Plantee las hipótesis nula y alternativa (α=0.05).
   3. Qué coeficiente escogería y por qué?

|  |  |  |
| --- | --- | --- |
| Presión | Pureza |  |
| 1.1 | 83.0 |
| 1.3 | 85.7 |
| 1.1 | 84.0 |
| 1.3 | 86.0 |
| 1.2 | 84.0 |
| 1.2 | 83.5 |
| 1.2 | 83.0 |
| 1.2 | 84.0 |
| 1.3 | 86.3 |
|  |  |
|  |  |

1. Un economista está analizando la correlación entre el salario y el nivel de ahorro, entre una muestra de 100 profesionales. Utiliza R para analizar el problema y calcula el coeficiente de correlación lineal producto-momento, el coeficiente de correlación de rangos de Spearman y el tau de Kendall. Abajo están las salidas. Llene las cajas con los espacios en blanco, e interprete tanto la magnitud como la prueba de hipótesis implícita en cada salida. ¿Hay asociación entre salario y ahorro, con un α=0.01?:

> cor.test(ahorro, salario, method="pearson")

Pearson's product-moment correlation

data: ahorro and salario

t = 10.0746, df = 9 , p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

0.6010703 0.7979086

sample estimates:

cor

0.713278

> cor.test(ahorro, salario, method="spearman")

Spearman's rank correlation rho

data: ahorro and salario

S = 49706, p-value < 2.2e-16

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.7017342

> cor.test(ahorro, salario, method="kendall")

Kendall's rank correlation tau

data: ahorro and salario

z = 7.4572, p-value = 8.837e-14

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

0.5058586

>

1. Un ingeniero estudia las características del rendimiento de combustible de cinco tipos de aditivos de gasolina. El rendimiento está dado en kilómetros por litro. Con base en las salidas que da R, calcule algún indicador que mida la asociación entre los cinco tipos de aditivos de gasolina y el rendimiento.

> aditivo=c(rep(1,4),rep(2,4),rep(3,4),rep(4,4),rep(5,4))

> rendimiento=c(17,14,13,12,14,14,13,10,12,13,12,9,13,11,11,12,11,12,10,8)

> tapply(rendimiento,aditivo,mean)

1 2 3 4 5

14.00 12.75 11.50 11.75 10.25

> mean(rendimiento)

[1] 12.05

> sd(rendimiento)

[1] 2.012461

1. A continuación se dan los datos del índice de felicidad, según hombres o mujeres.
   1. Calcule el eta para estos datos.

|  |  |  |  |
| --- | --- | --- | --- |
| Sexo (1=H, 0=M) | Indice de felicidad | Sexo (1=H, 0=M) | Indice de felicidad |
| 1 | 9.0 | 0 | 6.6 |
| 1 | 4.5 | 0 | 2.7 |
| 1 | 6.3 | 0 | 5.2 |
| 1 | 5.8 | 0 | 8.0 |
| 1 | 8.3 | 0 | 3.6 |
| 1 | 3.1 | 0 | 5.9 |
| 1 | 2.5 | 0 | 2.4 |
| 1 | 9.1 | 0 | 6.0 |
| 1 | 7.4 | 0 | 2.0 |
| 1 | 3.3 | 0 | 3.6 |
| 1 | 4.1 | 0 | 3.0 |
| 1 | 2.1 | 0 | 5.2 |

1. Se presenta a continuación una tabla con una muestra de 300 adultos mayores, clasificados según si sufren depresión o no, y según el tipo de hogar en el que viven.
   1. Realice una prueba de hipótesis X2 para probar si la depresión está asociada con el tipo de hogar. Plantee las hipótesis nula y alternativa. (α=0.01)
   2. ¿Se podría utilizar una prueba de G2 para analizar la asociación entre las variables? ¿Por qué sí o por qué no?
   3. Calcule un coeficiente de contingencia para medir el grado de asociación.
   4. Calcule la V de Cramer para medir el grado de asociación entre las variables

Adultos mayores, por nivel de depresión, según tipo de hogar.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Depresión |  | Tipo de hogar |  |  |
|  | Nuclear | Unipersonal | Extendido | Total |
| Sí | 190 | 13 | 25 | 228 |
| No | 60 | 3 | 9 | 72 |
| Total | 250 | 16 | 34 | 300 |

1. Tome el cuadro anterior, pero excluya a los hogares nucleares.
   1. Realice una prueba de Fisher-Irwin para analizar la asociación entre depresión y tipo de hogar. Plantee las hipótesis nula y alternativa. Use un α=0.05.
   2. Calcule el odds ratio de estar deprimido entre adultos mayores de hogares extendidos y adultos mayores de hogares unipersonales. Interprete el resultado.
   3. Es significativamente distinto de uno este odds ratio a nivel poblacional? (α=0.05).
   4. Calcule el riesgo relativo de estar deprimido entre adultos mayores de hogares extendidos y adultos mayores de hogares unipersonales. Interprete el resultado.
   5. Se puede utilizar una prueba de McNemar para analizar la asociación entre estas dos variables nominales. ¿Por qué sí o por qué no?
2. Supongan que en este curso de métodos, ustedes tienen un compañero que nunca llega que se llama Pelé Maradona. Él decidió hacer el experimento que ustedes están haciendo, pero con papel higiénico. Usó dos marcas de papel higiénico (A y B) en el experimento, y a los entrevistados al final les preguntó si les gustaba. Los resultados son los siguientes.
   1. Calcule la V de Cramer
   2. Calcule el odds ratio de que al entrevistado le guste la muestra de papel higiénico de la marca A versus que le guste la muestra de la marca B.
   3. Se puede realizar una prueba de hipótesis X2 de homogeneidad para estudiar la asociación entre la marca y el que le haya gustado la muestra de papel higiénico. Por qué sí o por qué no?
   4. Realice una prueba de hipótesis de Fisher-Irwin para estudiar si hay asociación entre la marca y el que la muestra guste al entrevistado (α=0.10). Plantee las hipótesis nula y alternativa.

Personas entrevistadas, por gusto de la muestra de papel higiénico, según marca.

|  |  |  |  |
| --- | --- | --- | --- |
| Gustó la muestra |  | Marca |  |
|  | A | B | Total |
| Sí | 19 | 14 | 33 |
| No | 1 | 0 | 1 |
| Total | 20 | 14 | 34 |

1. A una muestra de hombres se les preguntó si les gustaba cocinar con microondas. Después de esta pregunta, se les dio un curso gratis de cocina rápida en microondas y se les volvió a preguntar si les gustaba cocinar. Los resultados de la muestra se ofrecen abajo.
   1. Realice una prueba de hipótesis de McNemar para analizar si la proporción que dijo que les gustaba cocinar es diferente antes y después. Plantee las hipótesis nula y alternativa (α=0.05).
   2. Con un α=0.01, realice una prueba de signos para determinar si cambió la proporción que les gusta cocinar antes y después del curso.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **id** | **antes** | **después** | | **id** | **antes** | **después** | | **id** | **antes** | **después** | | **id** | **antes** | **después** |
| 1 | Sí | Sí |  | 21 | No | No |  | 41 | Sí | No |  | 61 | No | Sí |
| 2 | Sí | Sí |  | 22 | No | No |  | 42 | Sí | Sí |  | 62 | No | No |
| 3 | Sí | Sí |  | 23 | No | No |  | 43 | Sí | Sí |  | 63 | No | No |
| 4 | Sí | Sí |  | 24 | No | No |  | 44 | No | No |  | 64 | No | Sí |
| 5 | No | Sí |  | 25 | No | No |  | 45 | No | No |  | 65 | No | Sí |
| 6 | No | Sí |  | 26 | No | No |  | 46 | No | No |  | 66 | No | No |
| 7 | No | Sí |  | 27 | No | No |  | 47 | No | No |  | 67 | No | No |
| 8 | No | No |  | 28 | No | Sí |  | 48 | No | No |  | 68 | No | Sí |
| 9 | No | No |  | 29 | No | No |  | 49 | No | Sí |  | 69 | No | No |
| 10 | No | No |  | 30 | No | No |  | 50 | No | Sí |  | 70 | No | No |
| 11 | No | No |  | 31 | No | No |  | 51 | No | Sí |  | 71 | No | No |
| 12 | No | No |  | 32 | No | Sí |  | 52 | No | No |  | 72 | No | Sí |
| 13 | No | No |  | 33 | No | No |  | 53 | No | No |  | 73 | No | No |
| 14 | No | No |  | 34 | No | No |  | 54 | No | No |  | 74 | No | No |
| 15 | No | No |  | 35 | No | No |  | 55 | No | No |  | 75 | No | No |
| 16 | No | Sí |  | 36 | No | No |  | 56 | No | No |  | 76 | No | No |
| 17 | No | Sí |  | 37 | No | No |  | 57 | No | No |  | 77 | No | No |
| 18 | Sí | Sí |  | 38 | No | Sí |  | 58 | No | No |  | 78 | No | Sí |
| 19 | Sí | No |  | 39 | No | No |  | 59 | No | No |  | 79 | No | No |
| 20 | Sí | Sí |  | 40 | No | Sí |  | 60 | No | No |  | 80 | No | Sí |

1. Un investigador de educación física está desarrollando un programa de baja intensidad para hacer que personas con sobrepeso pasen a ser personas con peso normal sin desarrollar masa muscular. Como el programa es de baja intensidad, es posible que personas también pasen de peso normal a sobrepeso. Toma una muestra de 140 personas con sobrepeso a quienes somete al programa durante 3 meses. Al final de esos tres meses vuelve a determinar si están con sobrepeso o no. La tabla más abajo presenta los resultados del programa.
   1. ¿Por qué la prueba de McNemar es apropiada para analizar estos datos en lugar de la prueba de Fisher-Irwin?
   2. Realice la prueba de McNemar para determinar si la proporción de personas con sobrepeso cambió con un α=0.10. Plantee las hipótesis nula y alternativa
   3. Realice una prueba de hipótesis de signos para analizar la misma hipótesis, con un α de 0.10.

Participantes en programa de educación física, por sobrepeso antes y después del progama.

|  |  |  |  |
| --- | --- | --- | --- |
| Sobrepeso antes |  | Sobrepeso después | Total |
|  | Sí | No |  |
| Sí | 60 | 20 | 80 |
| No | 4 | 56 | 60 |
| Total | 64 | 76 | 140 |

1. Se presume que el nivel de escolaridad de los empleados de una empresa tiene asociación con su puesto, para ello se tomó la siguiente información de una muestra de 11 funcionarios:

nivel de escolaridad: 1 4 3 2 2 3 4 1 3 2 2

puesto que ocupa: 1 5 4 2 3 3 5 1 4 2 2

donde el nivel de escolaridad 1 = primaria 2 = secundaria

3 = universidad incompleta 4 = universidad completa

donde el puesto que ocupa 1 = limpieza-mensajería 2 = apoyo administrativo 3 = técnicos

4 = jefes medios 5 = gerentes

Con base en los resultados anteriores calcule el coeficiente de correlación de Spearman y pruebe la hipótesis para comprobar si existe asociación entre las variables estudiadas. Use α = 5%

1. Del Montgomery (p.57), doce inspectores midieron el diámetro de un cojinete de bolas, utilizando cada uno dos tipos diferentes de calibradores. O sea, se usó el calibrador 1 y el calibrador 2 en los mismos cojinetes.
   1. Si supone que los diámetros de cojinetes tienen en la población una distribución normal, se tiene que usar una muestra t-pareada, por qué?
   2. Con un α=0.01, realice una prueba de hipótesis para ver si hay diferencias en el diámetro de los cojinetes según calibrador. Plantee las hipótesis nula y alternativa
   3. Con el mismo α y las mismas hipótesis, realice una prueba de signos.
   4. Con el mismo α y las mismas hipótesis, realice una prueba de rangos de Wilcoxon.
   5. Encontró diferencias en los resultados. ¿Por qué sí o por qué no?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **id** | **Calib 1** | **Calib 2** |  | **id** | **Calib 1** | **Calib 2** |
| 1 | 0.265 | 0.264 |  | 7 | 0.267 | 0.264 |
| 2 | 0.265 | 0.265 |  | 8 | 0.267 | 0.265 |
| 3 | 0.266 | 0.264 |  | 9 | 0.265 | 0.265 |
| 4 | 0.267 | 0.266 |  | 10 | 0.268 | 0.267 |
| 5 | 0.267 | 0.267 |  | 11 | 0.268 | 0.268 |
| 6 | 0.265 | 0.268 |  | 12 | 0.265 | 0.269 |

1. (De Práctica de Estadística General II: Variación): Los corredores de bienes raíces a menudo quieren saber como el avalúo de una casa cambia según el tamaño de la misma. Un corredor de bienes raíces cree que la correlación entre estas dos variables es de 0.95. Se tomó una muestra de 200 propiedades, se midió su superficie (en metros cuadrados) y su precio de avalúo. Se calculó el coeficiente de correlación producto-momento de Pearson y se obtuvo un valor de 0.90. Con un nivel de significancia de 0.05, se puede decir que la hipótesis del corredor de bienes raíces es aceptable.
2. Los índices económicos proporcionan medidas de fluctuaciones en el mercado. La revista de junio de 1977 del U.S. News and World Report publicó una muestra aleatoria de índices económicos para mayo de 1977 y para mayo de 1976, para proporcionar los datos que permitieran medir las fluctuaciones económicas durante el año:
   1. Escriba las hipótesis nula y alternativa para analizar si los indicadores económicos promedio han cambiado entre 1976 y 1977. (2 ptos.)
   2. Realice la prueba t-pareada a fin de probar la hipótesis anterior, con un alfa de 0.05. (5 ptos.)
   3. Utilice la prueba de rangos de Wilcoxon para probar la misma hipótesis, con el mismo α. (5 ptos.)
   4. ¿Qué supuesto tiene que hacer para aceptar los resultados de la prueba t-pareada qué no tiene que hacer para aceptar los resultados de la prueba de Wilcoxon? (2 ptos.)

|  |  |  |
| --- | --- | --- |
| Áreas: Producción de... | 1977 | 1976 |
| Acero | 99.3 | 106.4 |
| Automóviles | 123.9 | 118.7 |
| Petróleo crudo | 91.0 | 93.2 |
| Madera | 117.1 | 112.6 |
| Transportes | 84.8 | 82.2 |
| Energía eléctrica | 169.5 | 167.0 |

PARA EXAMEN PARCIAL III

1. Una nutricionista está investigando cuánto afectan el consumo de alimentos y el peso de los bebés en el tiempo que duermen durante la siesta vespertina. Seleccionan al azar a 300 bebés de 12 a 18 meses, les miden las kilocalorías que consumieron al almuerzo (kilocalorías) y su peso en kilogramos (peso), y por último les toman en minutos el tiempo que duermen durante la siguiente siesta tomada después del almuerzo. Analizaron los datos con un modelo de regresión múltiple con el programa R y los resultados están abajo.
   1. Interprete la pendiente para la variable peso. (3 ptos.)

Por cada cambio de una caloría consumida aumenta 2,4324 minutos la siesta manteniendo constantes las otras variables independientes.

* 1. ¿Qué proporción de la variabilidad del tiempo de la siesta es explicado por el modelo? (2 ptos.)

El modelo de regresión explica el 13,67% de la variabilidad de la siesta de los bebés.

> summary(regsiesta)

Call:

lm(formula = siesta ~ kilocalorias + peso)

Residuals:

Min 1Q Median 3Q Max

-143.240 -37.009 1.078 37.245 136.206

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -25.1728 13.7251 -1.834 0.0676 .

kilocalorias 0.4563 0.2825 1.615 0.1074

peso 2.4324 0.3671 6.626 1.62e-10 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 51.14 on 297 degrees of freedom

Multiple R-squared: 0.1367, Adjusted R-squared: 0.1309

F-statistic: 23.52 on 2 and 297 DF, p-value: 3.296e-10

1. Un investigador realiza un experimento para ver si a los hombres se les sube más la presión arterial jugando un partido de fútbol-5 o jugando un partido de basketball. Se tomó a un grupo de 200 hombres, a 100 aleatoriamente se les puso a jugar fútbol 5 y a otros 100 aleatoriamente basketball (obviamente en equipos de 5). A los 25 minutos de empezados los juegos, se les tomó la presión sistólica. Se utilizó un Análisis de Variancia para ver si había diferencia en la presión arterial entre los jugadores de fut-5 y los de basket. Los datos se analizaron con R.
   1. Diga 2 de los supuestos (aparte de asignación aleatoria) que se deben de cumplir para poder analizar los datos con un ANDEVA. (4 ptos.)

Los residuos de las variables deben distribuirse de manera normal.

Las observaciones que se toman en cuenta para el ANOVA deben de ser independientes.

* 1. Plantee las hipótesis nula y alternativa respectivas, dado que se va a utilizar un ANDEVA. (2 ptos.)

H0: μ1 = μ2

H1: μ1 = μ2 y al menos un μi es diferente <> 0

* 1. Dadas las salidas de computadora que tiene abajo, qué problemas habría para interpretar el ANDEVA? (3 ptos.)

Como no hay homogeneidad de varianzas…

> andeva1=lm(presion~deporte)

>

> anova(andeva1)

Analysis of Variance Table

Response: presion

Df Sum Sq Mean Sq F value Pr(>F)

deporte 1 7170 7170 2.8862 0.09091 .

Residuals 198 491865 2484

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> levene.test(presion,deporte)

Levene's Test for Homogeneity of Variance

Df F value Pr(>F)

group 1 43.871 3.233e-10 \*\*\*

198

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Warning message:

In levene.test(presion, deporte) : deporte coerced to factor.

>

1. En una planta química se está analizando si la presión (psi) determina la pureza de oxígeno líquido licuado (puntos porcentuales). Los datos se presentan abajo
   1. Estime a mano un modelo de regresión lineal para estos datos.
   2. Interprete los coeficientes de regresión.
   3. Estime un modelo lineal gaussiano con computadora para describir estos datos y verifique que obtiene los mismos resultados que el punto a.
   4. Plantee la hipótesis nula y la alternativa de los coeficientes de regresión y la prueba global del ANDEVA.
   5. Usando los p-values, haga conclusiones de las tres pruebas de hipótesis planteadas en el punto anterior.
   6. Con base en el error estándar de estimación que aparece en las salidas de regresión, recalcule los estadísticos t calculados que aparecen en la salida de computadora.

|  |  |  |
| --- | --- | --- |
| Presión | Pureza |  |
| 1.1 | 83.0 |
| 1.3 | 85.7 |
| 1.1 | 84.0 |
| 1.3 | 86.0 |
| 1.2 | 84.0 |
| 1.2 | 83.5 |
| 1.2 | 83.0 |
| 1.2 | 84.0 |
| 1.3 | 86.3 |
|  |  |
|  |  |

1. (Del Montgomery 3.10). Se determinó el tiempo de respuesta en milisegundos para tres diferentes tipos de circuitos que podrían usarse en un mecanismo de desconexión automática. Abajo se presenta el gráfico de cajas para el tiempo de resupuesta y el gráfico de normalidad para los residuos de un ANDEVA.

Gráfico de cajas de tiempo de respuesta por tipo de circuito (1, 2, y 3)

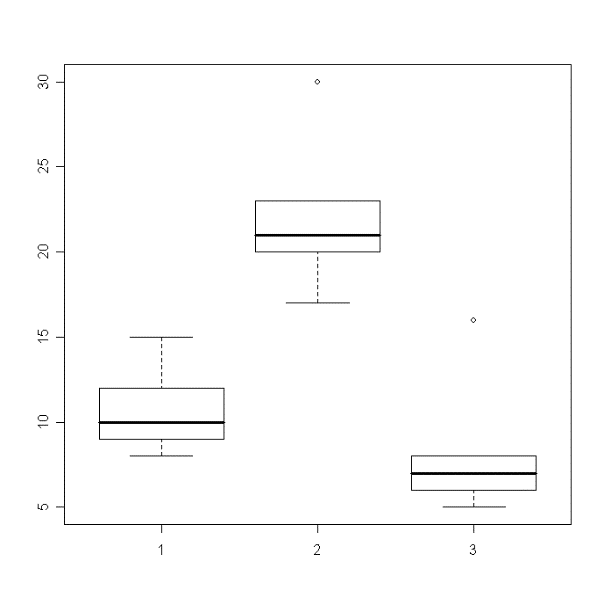
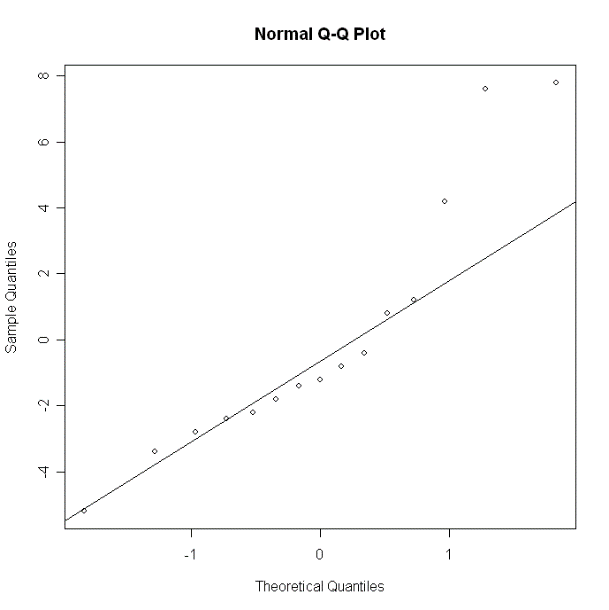


Gráfico de normalidad (qqnorm) de los residuos de un ANDEVA paramétrico



1. Qué concluye del gráfico de cajas en términos de las diferencias entre circuitos y la variabilidad.

Se podría concluir que entre las 3 variables hay homocedasticidad porque el tamaño de las cajas y las patas incluyendo los valores extremos son muy semejantes. Y sus medias son muy diferentes, arrojando una media mayor el circuito 3

1. Según el gráfico de normalidad, qué problemas podría haber para realizar un ANDEVA paramétrico para determinar las diferencias en los tiempos de respuestas entre circuitos?
2. Se presentan las pruebas de levene y bartlett para inspeccionar la homoscedasticidad en los tiempos de respuesta por circuitos. Plantee las hipótesis nula y alternativa. ¿Qué sugieren estas pruebas?

Que no se rechazn

> bartlett.test(densidad,temp)

Bartlett test of homogeneity of variances

data: densidad and temp

Bartlett's K-squared = 1.3366, df = 3, p-value = 0.7205

> levene.test(densidad,temp)

Levene's Test for Homogeneity of Variance

Df F value Pr(>F)

group 3 0.2955 0.828

14

1. Se presentan los resultados en R del ANDEVA y de la prueba de Kruskal y Wallis. Cuál escogería? Hace alguna diferencia?

> anova(ej.circuito)

Analysis of Variance Table

Response: tiempo

Df Sum Sq Mean Sq F value Pr(>F)

circuito 2 543.6 271.8 16.083 0.0004023 \*\*\*

Residuals 12 202.8 16.9

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> ej.circuito.2=kruskal.test(tiempo,circuito)

> ej.circuito.2

Kruskal-Wallis rank sum test

data: tiempo and circuito

Kruskal-Wallis chi-squared = 10.3735, df = 2, p-value = 0.00559

1. Se presentan los resultados en R de las pruebas de comparaciones múltiples de Tukey. Plantee las hipótesis nulas y alternativas. Qué grupo tiene un tiempo de respuesta diferente al de los demás? Pudo haber adivinado algo de alguno de los gráficos?

> TukeyHSD(circuito.anova)

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = tiempo ~ circuito)

$circuito

diff lwr upr p adj

2-1 11.4 4.463555 18.336445 0.0023656

3-1 -2.4 -9.336445 4.536445 0.6367043

3-2 -13.8 -20.736445 -6.863555 0.0005042

1. Una fábrica de herramientas está analizando la relación que existe entre la vida útil de una herramienta (en días) y la humedad a la que está expuesta. Tomó una muestra de 15 herramientas, y las expuso durante un mes a distintos niveles de humedad. Los datos se muestran abajo. Se presenta además, un gráfico de dispersión entre las dos variables.
   1. Estime a mano un modelo de regresión en el que la humedad prediga la vida útil de la herramienta.
   2. Calcule el error estándar de estimación.
   3. Calcule e interprete el coeficiente de determinación.
   4. Calcule e interprete los dos coeficientes de regresión.
   5. Haga la prueba de hipótesis de que β1 es igual a 0, con un α=0.
   6. Con R, grafique el gráfico qqplot para los residuos del modelo. ¿Qué puede decir acerca de la normalidad de los errores?

|  |  |  |  |
| --- | --- | --- | --- |
| Unidad | vida (días) | humedad (%) |  |
| 1 | 8 | 95 |
| 2 | 7 | 90 |
| 3 | 7 | 85 |
| 4 | 8 | 80 |
| 5 | 9 | 75 |
| 6 | 10 | 70 |
| 7 | 14 | 65 |
| 8 | 12 | 60 |
| 9 | 12 | 55 |
| 10 | 13 | 50 |
| 11 | 14 | 45 |
| 12 | 14 | 40 |
| 13 | 11 | 35 |
| 14 | 16 | 30 |
| 15 | 16 | 25 |

1. Suponga que se está analizando los gastos de consumo de electricidad de una muestra de 50 hogares (use el archivo electricidad.Rdata). Se desea analizar si el número de miembros del hogar, si el hogar es propio (alquilado=0) o alquilado (alquilado=1) y los metros cuadrados de construcción de la vivienda determinan el gasto en electricidad (en Kw).
   1. Haga gráficos de dispersión entre cada variable independiente y el consumo de electricidad.
   2. Estime un modelo de regresión con la variable dependiente y las independientes antes señaladas. Escriba la ecuación.
   3. Sin mirar los p-values de los coeficientes, interprete los 4 coeficientes de regresión que tiene en la ecuación.
   4. Plantee las hipótesis nulas y alternativas tanto de los coeficientes como la hipótesis nula global (la prueba F) implícitas en la salida de regresión.
   5. Interprete los resultados de cada una de las pruebas de hipótesis usando como indicador los p-values.
   6. Interprete el coeficiente de determinación.
   7. Interprete el gráfico de residuos contra los predichos (puede suponer homoscedasticidad).
   8. Realice la prueba de Shapiro para los residuos. Se rechaza o no se rechaza H0, con un α=0.05.
2. Un ingeniero estudia las características del rendimiento de combustible de cinco tipos de aditivos de gasolina. Selecciona varios vehículos y les asigna un aditivo al azar.
   1. Con un ANDEVA, y al 5% de significancia, pruebe si hay diferencias en las medias de rendimiento entre los cinco tipos de aditivos de gasolina y el rendimiento.
   2. Evalúe la normalidad condicional de los residuos. ¿Qué significa esto para el contraste de hipótesis que está realizando?
   3. Evalúe el supuesto de homoscedasticidad.

> aditivo=c(rep(1,4),rep(2,4),rep(3,4),rep(4,4),rep(5,4))

> rendimiento=c(17,14,13,12,14,14,13,10,12,13,12,9,13,11,11,12,11,12,10,8)

> tapply(rendimiento,aditivo,mean)

1 2 3 4 5

14.00 12.75 11.50 11.75 10.25

> mean(rendimiento)

[1] 12.05

> sd(rendimiento)

[1] 2.012461

1. (Del manual de prácticas de Estadística General II): Una fábrica de refrigeradoras tiene tres plantas de producción y desea probar si existe diferencia en por lo menos dos promedios de producción de dichas plantas. Para ello, se recolectaron las producciones (cantidad de refrigeradoras que se fabrican diariamente) y se presentaron los siguientes resultados en la producción por planta:

|  |  |  |
| --- | --- | --- |
| Alajuela | Heredia | San José |
| 10 | 15 | 12 |
| 12 | 17 | 17 |
| 15 | 18 | 15 |
| 18 | 12 | 15 |
| 9 | 13 | 18 |
| 17 | 11 | 12 |
| 15 | 12 | 13 |
| 12 | 11 | 14 |
| 18 | 12 | 14 |

1. Realice la prueba paramétrica respectiva utilizando un nivel de significancia del 5%
2. Evalúe con una prueba el supuesto de normalidad condicional.
3. Evalúe la homoscedasticidad.
4. A continuación se dan los datos de un índice de felicidad, según hombres o mujeres.
   1. Suponiendo que la muestra es al azar, con un modelo lineal pruebe si hay diferencias en la felicidad entre hombres y mujeres con un (α=0.05).
   2. Evalúe el supuesto de homoscedasticidad.
   3. Evalúe el supuesto de normalidad gráficamente.
   4. Evalúe el supuesto de normalidad condicional con la prueba de Shapiro.

|  |  |  |  |
| --- | --- | --- | --- |
| Sexo (1=H, 0=M) | Indice de felicidad | Sexo (1=H, 0=M) | Indice de felicidad |
| 1 | 9.0 | 0 | 6.6 |
| 1 | 4.5 | 0 | 2.7 |
| 1 | 6.3 | 0 | 5.2 |
| 1 | 5.8 | 0 | 8.0 |
| 1 | 8.3 | 0 | 3.6 |
| 1 | 3.1 | 0 | 5.9 |
| 1 | 2.5 | 0 | 2.4 |
| 1 | 9.1 | 0 | 6.0 |
| 1 | 7.4 | 0 | 2.0 |
| 1 | 3.3 | 0 | 3.6 |
| 1 | 4.1 | 0 | 3.0 |
| 1 | 2.1 | 0 | 5.2 |

1. Los siguientes datos corresponden a las mediciones de la capacidad de producción de dos máquinas que fabrican chips de computadoras (en miles de unidades):

Máquina 1: 8.26 8.13 8.385 8.07 8.34

Máquina 2: 7.95 7.89 7.9 8.14 7.92 7.84

1. Con un modelo lineal y “a mano”, estime un modelo de regresión que determine las diferencias entre Máquina 1 y Máquina 2.
2. Con computadora y con un modelo lineal, estime si hay diferencias en la capacidad de producción entre las dos máquinas.
3. Gráficamente, evalúe el supuesto de normalidad condicional.
4. Con una prueba de Shapiro, evalúe el supuesto de normalidad condicional.
5. Evalúe gráficamente el supuesto de homoscedasticidad.
6. Supongan que en este curso de métodos, ustedes tienen un compañero que nunca llega que se llama Pelé Maradona. Él decidió hacer el experimento que ustedes están haciendo, pero con papel higiénico. Usó dos marcas de papel higiénico (A y B) en el experimento, y a los entrevistados al final les preguntó si les gustaba. Los resultados son los siguientes.
   1. Realice una prueba de hipótesis de que la proporción que les gustó la muestra de papel higiénico es mayor entre los que probaron la marca A y la marca B

H0: P1=P2

H1: P1<>P2

Zc= [Ptecho1-Ptecho2-(P1-P2)]/ raíz(pbarra\*qbarra\*(1/n1+1/n2))

Personas entrevistadas, por gusto de la muestra de papel higiénico, según marca.

|  |  |  |  |
| --- | --- | --- | --- |
| Gustó la muestra |  | Marca |  |
|  | A | B | Total |
| Sí | 19 | 14 | 33 |
| No | 1 | 0 | 1 |
| Total | 20 | 14 | 34 |

1. Se decidió investigar si los niveles de HbA1C (hemoglobina glicosilada) son menores entre las personas con dietas vegetarianas que entre las personas con dietas no vegetarianas. Se realizó este estudio en una muestra de 30 personas que llegaron a la consulta a un nutricionista. Los datos se encuentran abajo.
   1. Suponiendo que la distribución poblacional de HbA1C es normal, realice una prueba de hipótesis t con muestras independientes para determinar si los niveles de HbA1C son menores entre las personas con dieta vegetariana y las personas con otra dieta (α=0.05). Plantee las hipótesis nula y alternativa.
   2. Con las mismas hipótesis nula y alternativa y el mismo nivel de significancia, realice con los mismos datos una prueba de U de Mann-Whitney.
   3. Hay alguna diferencia entre los resultados de la prueba t para muestras independientes y la prueba U de Mann Whitney.

|  |  |  |  |
| --- | --- | --- | --- |
| grupo | h | grupo | h |
| 1 | 3 | 0 | 8 |
| 1 | 3.5 | 0 | 4 |
| 1 | 3 | 0 | 5 |
| 1 | 3.5 | 0 | 6.5 |
| 1 | 4.5 | 0 | 3.5 |
| 1 | 5.5 | 0 | 7 |
| 1 | 2 | 0 | 5 |
| 1 | 4.5 | 0 | 7.5 |
| 1 | 5.5 | 0 | 5 |
| 1 | 6.5 | 0 | 6 |
| 1 | 3.5 | 0 | 6.5 |
| 1 | 4 | 0 | 3.5 |
| 1 | 4 | 0 | 4.5 |
| 1 | 3 | 0 | 4 |
| 1 | 2.5 | 0 | 7 |
| 1 | 6 | 0 | 5.5 |

1. Los siguientes datos corresponden a las mediciones de la capacidad de producción de dos máquinas que fabrican chips de computadoras (en miles de unidades):

Máquina 1: 8.26 8.13 8.385 8.07 8.34

Máquina 2: 7.95 7.89 7.9 8.14 7.92 7.84

1. Realice una prueba de U de Mann-Whitney para ver si existen diferencias en la producción de chips entre la máquina 1 y la máquina 2.
2. Realice una prueba de hipótesis para determinar si existen diferencias en la variabilidad de la capacidad de producción de la máquina 1 y de la máquina 2.
3. Tomando en cuenta los resultados de la prueba anterior, y asumiendo que la capacidad de producción de chips es una variable aleatoria con distribución normal, realice una prueba t-de Student para diferencia de medias. Calcule la t tabular como si estuviera suponiendo homogeneidad de variancias.
4. En los últimos 10 años un banco ha determinado que la proporción de morosos en las sucursales San Pancracio y Santa Enriqueta eran parecidos. Sin embargo, el gerente cree que la proporción de morosos últimamente es mayor en la sucursal San Pancracio. Selecciona una muestra de 100 clientes de San Pancracio y 150 clientes de Santa Enriqueta, y encuentra que 15 eran morosos en la primera muestra y 15 eran morosos en la segunda sucursal. Con un α=0.10, realice la prueba de hipótesis correspondiente con una prueba paramétrica.
5. (Del Manual de Prácticas de XS-0215). En una investigación sobre salud dental en niños menores de seis años se tomaron muestras aleatorias de tres comunidades, obteniéndose la siguiente información:

Comunidad Niños sin Tamaño de

caries muestra

Valle Hondo 48 125

Monte Verde 36 125

Guacalito 44 125

¿Son similares las condiciones en la salud dental de las tres comunidades? Trabaje con una significancia del 5%.

H0: P1=P2=P3

H1: Al menos un P\_i <> P\_j para todo i<>j

X2

1. (Del manual de Prácticas de XS-0215: Variación): En 200 pacientes diagnosticados con envenenamiento por metilmercurio, en el Centro de Control de Envenenamientos de un hospital metropolitano, se registró los valores promedio de la velocidad de conducción de un nervio motor. Se hizo también, mediciones similares en 250 personas aparentemente sanas. Los promedios y las desviaciones estándar obtenidos fueron:

Promedio Desviación estándar Tamaño de la muestra

Envenenados 55 16 200

Sanos 63 15 250

* 1. Pruebe la diferencia en los promedios de las dos poblaciones, use un nivel de significancia de 5%. Interprete el resultado. Especifique la hipótesis del investigador, la hipótesis nula, la hipótesis alternativa, la estadística de prueba y su valor tabular.

1. (Del manual de prácticas de XS-0215). Supóngase que se desea averiguar si la aspirina y un producto de comparación son igualmente eficaces para el alivio de los síntomas de la gripe. Para ello, se registran los tiempos (en minutos) desde la toma de la medicina hasta el momento en que el paciente dice sentirse mejor. Los resultados son:

|  |  |  |
| --- | --- | --- |
|  | Aspirina | Producto de comparación |
|  | 15,2 | 13,4 |
| s | 8,7 | 2,9 |
| n | 20 | 28 |

Son igualmente rápidos en el alivio ambos productos? Para comprobarlo utilice un 1% de significancia.

1. (Del manual de prácticas de Estadística General II). Un ingeniero de control de calidad, que trabaja en la División de Limpiaparabrisas de una empresa está examinando dos nuevos productos de goma sintética; se le ha encargado investigar si tiene diferente durabilidad. Para poder efectuar las pruebas correspondientes se seleccionaron aleatoriamente 22 limpiaparabrisas, 12 con la sustancia A y el resto con la sustancia B. Al final se presentaron los siguientes resultados:

|  |  |  |
| --- | --- | --- |
| DURACIÓN EN DIAS | | |
|
| Promedio | | Desviación estándar |
| 221  254 | | 36  27 |
| ¿Existe diferencia significativa entre las duraciones de los limpiaparabrisas fabricados con las dos sustancias? | | |
| Asuma α = 0,05 y concluya en términos del problema | | |
| ¿Cuál error se puede cometer al tomar la decisión de a) y por qué? | | |

1. A una muestra probabilística de pasajeros de los buses de Empresarios Caribeños, se les preguntó cuántos habían comprado “paty” en su viaje de San José a Limón. Se les clasificó según si eran residentes de Limón o de San José. El siguiente cuadro tiene las respuestas:

Muestra de pasajeros de la ruta San José-Limón, por consumo de paty, según lugar de residencia:

|  |  |  |  |
| --- | --- | --- | --- |
| Consumo de paty | Limón | San José | Total |
| Sí | 350 | 578 | 928 |
| No | 43 | 113 | 156 |
| Total | 393 | 691 | 1084 |

Se analizaron los datos con R, y estos fueron los resultados:

> fisher.test(tabla)

Fisher's Exact Test for Count Data

data: tabla

p-value = 0.01507

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

1.080570 2.374975

sample estimates:

odds ratio

1.590642

> prop.test(tabla)

2-sample test for equality of proportions with continuity correction

data: tabla

X-squared = 5.5239, df = 1, p-value = 0.01876

alternative hypothesis: two.sided

95 percent confidence interval:

0.01073302 0.09749970

sample estimates:

prop 1 prop 2

|  |  |
| --- | --- |
|  |  |

* 1. Plantee las hipótesis nula y alternativa de la prueba exacta de Fisher. (2 ptos.)
  2. ¿Por qué se le llama “prueba exacta” a la prueba exacta de Fisher? (2 ptos.)
  3. Al 1% de significancia, qué se concluye de acuerdo a las salidas de la prueba exacta de Fisher? (2 ptos.)
  4. Rellene los campos de la salida para la prueba de igualdad de proporciones (los campos que están en dos cuadrados) (2 ptos.)
  5. Interprete la estimación puntual del OR en la prueba de Fisher (4 ptos.)