

Diego Olalla Carrión

Estadística

### Ejercicio 1

**a) Obtener, usando algún programa estadístico, las medidas de centralización y dispersión para cada uno de los dos grupos de control para el nivel de glucosa basal, especificando para cada uno de los casos si la media es o no representativa.**

Se calcularon los siguientes coeficientes de variación (CV):

Grupo 1:  $CV = 0.104$

Grupo 2:  $CV = 0.082$

Ambos valores son menores a 0.3, lo que indica que la dispersión relativa de los datos es baja. Sin embargo, la media sigue siendo representativa en ambos casos.

El coeficiente de variación es una medida de la dispersión de los datos respecto a la media. Un  $CV < 0.3$  significa que la desviación estándar es pequeña en relación con la media, lo que sugiere que los datos están concentrados alrededor de la media.

La dispersión dentro de ambos grupos es pequeña, por lo que los datos de estos grupos están cerca de la media. Por lo tanto, la media puede considerarse una medida aceptable de tendencia central y representativa en ambos casos.

**b) Estudiar la simetría y la curtosis del nivel de glucosa basal en los adultos (grupo de control 2)**

El coeficiente de asimetría para el nivel de glucosa basal en el Grupo 2 (Adultos) es -0.01.

La distribución presenta una asimetría negativa (sesgo a la izquierda).

El coeficiente de curtosis del nivel de glucosa basal en el Grupo 2 (Adultos) es -0.61.

La distribución es platicúrtica, es decir, menos concentrada que una distribución normal.

**c) Indicar para cada una de las variables de estudio (nivel glucosa basal y nivel glucosa pasados 60 min) y en el grupo de control 1 el valor de los cuartiles y su significado y obtener el box- plot (diagrama de cajas) correspondiente. Estudiar la presencia de valores atípicos**

Comparando ambas variables, el nivel de glucosa a los 60 minutos tiene una dispersión más baja que el nivel de glucosa basal, lo que sugiere que los valores son más consistentes a los 60 minutos. Aunque ambos conjuntos tienen medias distintas, los rangos intercuartílicos reflejan

que los datos están razonablemente agrupados y permiten identificar valores extremos con claridad.

La distribución del Nivel de Glucosa Basal está concentrada en el rango esperado y es consistente para el grupo de control.

En cuanto al Nivel de Glucosa a los 60 Minutos, se identificaron dos valores atípicos: 172 (alto) y 131 (bajo). Esto sugiere que algunas mediciones fuera del rango normal probablemente reflejan una variación amplia en las respuestas de ciertos individuos a los 60 minutos. Estos valores atípicos pueden influir en las estadísticas descriptivas de esta variable y deben considerarse en el análisis general para no distorsionar las conclusiones sobre el comportamiento típico del grupo.

#### **d) Estudiar la normalidad de los datos de cada uno de los grupos de control estudiados para el nivel de glucosa pasados 60 minutos.**

Para determinar si las muestras del Nivel de Glucosa a los 60 Minutos en los dos grupos de control tienen una distribución normal, comenzamos visualizando los datos mediante histogramas. En ambos casos, los histogramas mostraron distribuciones que parecían aproximadamente normales, aunque esta herramienta no es suficiente para confirmar la normalidad de forma concluyente. Posteriormente, generamos Q-Q plots como otra forma de evaluar la normalidad. En estos gráficos, las distribuciones de los datos en ambos grupos se alinearon razonablemente bien con la línea de referencia, lo que sugiere una buena aproximación a una distribución normal. Para confirmar esto, realizamos la prueba estadística de Shapiro-Wilk, que evalúa si una muestra proviene de una población distribuida normalmente. En esta prueba, la hipótesis nula asume normalidad, mientras que la hipótesis alternativa sugiere que los datos no siguen una distribución normal. Para el Grupo 1, obtuvimos un  $p$ -valor 0.2734  $p$ -valor=0.2734, y para el Grupo 2, un  $p$ -valor 0.7592  $p$ -valor=0.7592. Ambos valores son mayores que el nivel de significación de 0.05 0.05, lo que indica que no podemos rechazar la hipótesis nula. Por lo tanto, las muestras de ambos grupos parecen provenir de una distribución normal. Adicionalmente, empleamos la prueba de Kolmogorov-Smirnov, que verifica si los datos de una muestra coinciden con una distribución conocida, en este caso, la normal. Para el Grupo 1, obtuvimos un  $p$ -valor 0.3393  $p$ -valor=0.3393, y para el Grupo 2, un  $p$ -valor 0.8340  $p$ -valor=0.8340. Estos resultados también están por encima del nivel de significación de 0.05 0.05, lo que respalda que las muestras son consistentes con una distribución normal. Tanto los análisis visuales (histogramas y Q-Q plots) como las pruebas estadísticas de Shapiro-Wilk y Kolmogorov-Smirnov confirman que los datos del Nivel de Glucosa a los 60 Minutos en ambos grupos de control siguen una distribución normal.

## **Ejercicio 2**

**Con los datos del fichero anterior, se quiere estudiar la relación existente entre el nivel basal y el nivel de glucosa que tienen los pacientes sanos jóvenes (grupo 1) una hora después de tomar el preparado de glucosa. Se pide:**

**a) Estudiar la relación lineal existente entre estas dos variables de estudio gráficamente y mediante algún valor estadístico de forma razonada.**

El gráfico evidencia una relación lineal entre el nivel de glucosa basal y el nivel de glucosa a los 60 minutos en los pacientes del Grupo 1. En general, al aumentar el nivel de glucosa basal, también se observa un incremento en el nivel de glucosa medido a los 60 minutos.

El coeficiente de correlación calculado es 0.796, lo que sugiere una conexión fuerte y positiva entre estas dos variables. Esto indica que el nivel de glucosa basal influye significativamente en el nivel de glucosa tras una hora, reforzando la relación entre estas mediciones en este grupo.

**b) Obtener un modelo lineal que explica el nivel de glucosa en sangre a los 60 minutos en función del nivel basal del paciente y realizar la estimación para un paciente cuyo nivel basal es 83 mg/Dl**

$$Y = 91.38 + 0.70 * \text{Nivel Glucosa Basal}$$

En esta ecuación, Y representa el Nivel de Glucosa a los 60 Minutos.

Para un Nivel de Glucosa Basal de 83 mg/dL, la estimación del Nivel de Glucosa a los 60 Minutos es: 149.23 mg/dL

**c) ¿Qué tanto por ciento del nivel de glucosa en sangre pasados 60 minutos queda no queda explicado por el anterior modelo?**

Coefficiente de Determinación ( $R^2$ ): 0.6343

El 36.57% del nivel de glucosa en sangre a los 60 minutos no es explicado por este modelo lineal.

**d) Si aumentásemos el nivel basal de un paciente en 5 mg/Dl ¿Qué variación experimentaría su nivel de glucosa al cabo de 60 minutos?**

Si el nivel basal de un paciente aumenta en 5 mg/dL, el nivel de glucosa a los 60 minutos se incrementa en aproximadamente 3.48 mg/dL.

### **Ejercicio 3**

**a) Se quiere estudiar si se puede admitir que el nivel medio de glucosa en sangre en el momento de la ingestión en los jóvenes es 88 mg/Dl. Obtener el intervalo de confianza al 95% y al 99% para el nivel medio de glucosa en sangre de los jóvenes**

**y posteriormente contesta a la cuestión planteada con los resultados obtenidos o con un contraste de hipótesis.**

interpretación del intervalo de confianza al 95%

Para los jóvenes (Grupo 1), el intervalo de confianza al 95% para el nivel medio de glucosa basal se estima entre 81.648 y 87.724. Esto significa que, con un 95% de confianza, podemos afirmar que el verdadero promedio del nivel medio de glucosa basal en esta población se encuentra dentro de este rango. En términos técnicos, si tomáramos múltiples muestras aleatorias y calculáramos el intervalo de confianza al 95% para cada una, aproximadamente el 95% de estos intervalos incluirían el verdadero promedio poblacional. Este intervalo refleja la incertidumbre inherente al uso de datos muestrales para inferir parámetros poblacionales, considerando tanto el tamaño de la muestra como su variabilidad.

Interpretación del intervalo de confianza al 99%

Al aumentar el nivel de confianza al 99%, el intervalo se amplía a (80.607, 88.765). Este intervalo más amplio incluye un rango mayor de valores posibles para el promedio poblacional. Este aumento se debe a que, para alcanzar un mayor nivel de confianza, es necesario incluir una mayor proporción de valores posibles, lo que reduce la precisión. Sin embargo, asegura que en aproximadamente el 99% de las muestras aleatorias tomadas de esta población, el verdadero promedio estará contenido en este intervalo. En este caso, el valor hipotético de 88 mg/dL está dentro del intervalo al 99%, lo que sugiere que es plausible, aunque no está cerca del centro del rango.

Evaluación del valor hipotético y su relación con los intervalos

El valor hipotético de 88 mg/dL no se encuentra dentro del intervalo al 95%, lo que indica que, con este nivel de confianza, no es razonable asumir que el nivel medio de glucosa basal en los jóvenes sea 88 mg/dL. Esto implica que el valor probablemente está fuera del rango esperado para el promedio poblacional, basado en los datos observados. Sin embargo, al considerar el intervalo al 99%, el valor de 88 mg/dL apenas está incluido, lo que indica que podría ser plausible, aunque es probable que el promedio real sea menor.

Contraste de hipótesis

Dado que el valor hipotético no está contenido en el intervalo al 95%, se rechaza la hipótesis nula con este nivel de significación. Esto indica que los datos no son consistentes con la afirmación de que el promedio de glucosa basal es 88 mg/dL. Aunque el intervalo al 99% incluye este valor, el análisis refuerza que el promedio real es más probable que esté por debajo de 88 mg/dL.

**b) Obtener los intervalos de confianza al 95% para la diferencia de medias en el nivel basal de glucosa entre adultos y jóvenes e interpreta los resultados. ¿Se puede concluir que el nivel basal de glucosa de los jóvenes y los adultos es el mismo con nivel de significación del 5%. Suponiendo que se cumplen las condiciones iniciales teóricas para obtener los intervalos de confianza**

Intervalo de Confianza al 95%: (0.701, 8.728)

Como 0 no está dentro del intervalo de confianza al 95%,

se rechaza la hipótesis nula. Existe evidencia de una diferencia significativa.

El intervalo de confianza al 95% para la diferencia de medias es (0.701, 8.728). Esto indica que, con un nivel de confianza del 95%, podemos afirmar que la verdadera diferencia en el nivel basal de glucosa entre jóvenes y adultos está dentro de este rango.

Como el valor 0 no está incluido dentro del intervalo de confianza, se puede rechazar la hipótesis nula con un nivel de significación del 5%. Esto demuestra que existe una diferencia estadísticamente significativa entre los niveles basales de glucosa de jóvenes y adultos.

Con un nivel de significación del 5%, no se puede asumir que el nivel basal de glucosa de jóvenes y adultos sea igual. Los datos sugieren que los adultos tienen un nivel basal de glucosa significativamente mayor que los jóvenes, como lo demuestra el intervalo de confianza, que excluye el valor 0.

**c) Se quiere estudiar la proporción de la población con un nivel basal de glucosa superior a 95 mg/dl (prediabetes). A partir de la muestra del fichero (tomando todos los datos) obtener un intervalo de confianza al 98% y contrastar la hipótesis que la proporción de la población con glucosa superior a 95 mg/dl es 0,15 con nivel de significación del 5%.**

El intervalo de confianza al 98% para la proporción de la población con un nivel basal de glucosa superior a 95 mg/dL se calculó en el rango de (0.061, 0.277). Esto significa que, con un nivel de confianza del 98%, se puede afirmar que la verdadera proporción de la población con glucosa superior a 95 mg/dL se encuentra dentro de este intervalo. En términos generales, si se tomaran muchas muestras aleatorias y se calcularan intervalos de confianza al 98% para cada una, aproximadamente el 98% de ellos contendrían la verdadera proporción poblacional.

$p=0.15$  está contenido dentro del intervalo de confianza al 98% (0.061, 0.277). Esto significa que, con un nivel de significación del 2% (1–nivel de confianza), no se puede rechazar la hipótesis nula. Los datos respaldan la posibilidad de que la proporción real de la población con un nivel basal de glucosa superior a 95 mg/dL sea 0.15

Aunque no se puede afirmar que la proporción exacta sea 0.15, los datos observados, representados en el intervalo de confianza, indican que este valor es plausible. Por lo tanto, con un nivel de significación del 5%, no se rechaza la hipótesis nula, concluyendo que no hay evidencia suficiente para afirmar que la proporción real de la población con glucosa superior a 95 mg/dL sea diferente de 0.15.

Se realizó un contraste de hipótesis para determinar si la proporción de la población con un nivel basal de glucosa superior a 95 mg/dL (prediabetes) es 0.15, utilizando un nivel de

significación del 5%. La hipótesis nula establece que la proporción verdadera es 0.15, mientras que la hipótesis alternativa plantea que la proporción no es igual a 0.15.

Dado que el p-valor es mayor que el nivel de significación ( $p > 0.05$ ), no se puede rechazar la hipótesis nula. Esto indica que los datos observados no proporcionan evidencia suficiente para concluir que la proporción real de la población con un nivel basal de glucosa superior a 95 mg/dL sea distinta de 0.15.

**d) (VOLUNTARIO) ¿Se detecta una variación significativa del nivel de glucosa en sangre en el grupo de los adultos después de la toma?**

**Nota: Se pretende comprobar si, como muestran los datos, los niveles d de glucosa en sangre son distintos para los adultos en el momento de la ingestión y a los 60 minutos, por ello para contestar a la pregunta has de considerar las series de datos obtenidas a partir de las diferencias entre el nivel basal y el nivel de glucosa al cabode 60 minutos en el grupo de los adultos (contraste de muestras emparejadas).**

**Plantea el correspondiente contraste de hipótesis considerando un nivel de significación del 5%.**

Se buscó determinar si los niveles de glucosa en sangre en adultos son significativamente diferentes antes y después de la ingestión de glucosa. Para ello, se planteó un contraste de hipótesis considerando las diferencias individuales en los niveles de glucosa en dos momentos específicos, utilizando un enfoque de datos emparejados. Este método es adecuado porque compara mediciones de la misma población en dos condiciones distintas: antes y 60 minutos después de la ingestión. El nivel de significación utilizado fue del 5%.

La hipótesis nula establece que no hay diferencia en las medias de los niveles de glucosa antes y después de la ingestión. La hipótesis alternativa plantea que existe una diferencia significativa entre estos niveles. Para evaluar estas hipótesis, se calculó un intervalo de confianza al 95% para las diferencias en los niveles de glucosa, obteniéndose el rango (77.5286, 87.338). Dado que el valor 0 no se encuentra dentro de este intervalo de confianza, se rechazó la hipótesis nula al nivel de significación del 5%. Esto implica que existe suficiente evidencia estadística para concluir que hay una diferencia significativa en los niveles de glucosa antes y después de la ingestión. Es decir, los datos indican que el proceso de ingestión provoca un cambio real en los niveles de glucosa.

El intervalo de confianza también proporciona información sobre la magnitud de esta diferencia. Con un 95% de confianza, se estima que el aumento promedio en los niveles de glucosa tras la ingestión está entre 77.5286 mg/dL y 87.338 mg/dL. Este resultado muestra un efecto consistente y relevante de la ingestión en los niveles de glucosa, siendo estadísticamente significativo y relevante en términos prácticos.

Estos resultados confirman que la ingestión de glucosa tiene un impacto significativo en los niveles de glucosa en sangre en adultos. La metodología utilizada, basada en el contraste de datos emparejados y el intervalo de confianza, permite concluir con un alto grado de certeza que

los niveles de glucosa en sangre no son iguales antes y después de la ingestión. Esto valida que la intervención tiene un efecto medible y relevante en esta población específica.