

ACTIVIDAD 7- DESARROLLO:

COVID 19_ DIABETES

- Llancari Nivin, Meyli
- Vera Fonseca, Juana July
- Alejo Huamán, Melissa
- Flores Díaz, Christian
- Ccuro Minaya, Lucía

TEMA: BONDAD DE AJUSTE

GRUPO: 7

INTEGRANTES:

- Llancari Nivin, Meyli
- Vera Fonseca, Juana July
- Alejo Huamán, Melissa
- Flores Díaz, Christian
- Ccuro Minaya, Lucía

Glosario introductorio

1. Modelo estadístico
Es una representación matemática que describe relaciones entre variables. Se utiliza para hacer inferencias, predicciones o entender el comportamiento de los datos.

Glosario introductorio

1. Modelo estadístico
Es una representación matemática que describe relaciones entre variables. Se utiliza para hacer inferencias, predicciones o entender el comportamiento de los datos.

2. Variable continua
Tipo de variable numérica que puede tomar un número infinito de valores dentro de un rango (por ejemplo: saturación de oxígeno o nitrógeno ureico en sangre).

3. Distribución de probabilidad
Describe cómo se distribuyen los valores posibles de una variable aleatoria. Algunas distribuciones comunes incluyen la normal, binomial y chi-cuadrado.

4. Hipótesis nula (H_0)
Proposición que se plantea al inicio de una prueba estadística y que indica la ausencia de efecto o diferencia. Se rechaza o no según el valor p obtenido.

5. Valor p
Probabilidad de obtener un resultado igual o más extremo que el observado, si la hipótesis nula fuera cierta. Un valor p menor a 0.05 suele considerarse estadísticamente significativo.

¿Qué son las pruebas de bondad de ajuste?

Las pruebas de bondad de ajuste evalúan qué tan bien los datos observados se ajustan a los valores esperados según un modelo estadístico.

La bondad de ajuste puede evaluarse en al menos dos escenarios principales:

1. En modelos de regresión

Por ejemplo, un estudiante podría aplicar un modelo de regresión lineal para evaluar la relación entre el número de los pacientes con COVID-19 y su nivel de saturación de oxígeno. Para determinar si el modelo es adecuado para explicar esta relación, se

caso, el nivel de saturación de oxígeno) que es explicado por el modelo de regresión. Cuanto mayor sea el valor de R^2 , mejor será el ajuste del modelo a los datos observados.

2. En distribuciones de probabilidad

En algunos casos, el modelo estadístico que se desea aplicar requiere que los datos sigan una distribución de probabilidad específica, como la distribución normal.

Por otro lado, muchas pruebas de hipótesis utilizan **estadísticos de prueba** (no necesariamente modelos completos). Por ejemplo:

- Las **pruebas t** (t de Student) usan el estadístico t .
- El **ANOVA** usa el estadístico F .
- Las **pruebas de chi-cuadrado** usan el estadístico χ^2 .

Estas pruebas se basan en las distribuciones teóricas de estos estadísticos para calcular los valores p, los cuales permiten decidir si aceptar o rechazar la hipótesis nula.

Esta sesión práctica se enfocará en el segundo escenario.

Cargamos los paquetes necesarios

```
{r}  
library(rio)  
library(here)
```

Cargar los datos

```
{r}  
data_covid_19_diabetes <- import(here("data", "covid_19_diabetes.csv")  
)
```

1. Para datos continuos

La prueba t de Student y el ANOVA son dos pruebas estadísticas ampliamente utilizadas que permiten evaluar si el valor promedio de una variable numérica difiere

Quarto

data_glucosa_circun 200 obs. of 11 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

C:\Users\USER\Desktop\ESTADISTICA_UPSJB\ESTADISTICA_U

Name

- ☐ Trabajo grupal Semana 7 - GRUPO 7- Estadística.qmd
- ☐ practica_semana_04.qmd
- ☐ nombre_de_objeto.png
- ☐ GRUPO 7 COVID Y DIABETES.qmd
- ☐ 07_prueb_de_ajuste.qmd
- ☐ 06_muestra_potencia.qmd
- ☐ 05_est_descript_visualización.qmd
- ☐ 04_est_descript_resumen.qmd

1. Para datos continuos

La prueba t de Student y el ANOVA son dos pruebas estadísticas ampliamente utilizadas que permiten evaluar si el valor promedio de una variable numérica difiere entre dos o más grupos o categorías.

Ambas pruebas asumen que la variable continua sigue una distribución normal. Pero, ¿cómo podemos comprobar si esta condición se cumple? Mediante una prueba de bondad de ajuste.

Una de las pruebas más comunes para evaluar la normalidad de una variable numérica es la prueba de Shapiro-Wilk. Esta prueba permite determinar si los datos provienen de una distribución normal, lo cual es un requisito clave antes de aplicar pruebas como la t de Student o el ANOVA.

Para la variable Saturación_O2

La saturación de oxígeno describe el contenido de oxígeno en la sangre. Indica qué

En R, usamos la función nativa `shapiro.test()` para realizar la prueba de Shapiro-Wilk.

```
{r}  
shapiro.test(data_covid_19_diabetes$saturación_o2)
```

shapiro-wilk normality test

data: data_covid_19_diabetes\$saturación_o2
W = 0.50088, p-value < 2.2e-16

W = 0.50088, p-value < 0.00000000000000022

p-value (< 0.00000000000000022) < $\alpha = 0.05$

Se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alternativa (H_1), por lo tanto, los datos no siguen una distribución normal.

Para la variable Nitrógeno_ureico_sangre

Para la variable Nitrógeno_ureico_sangre

La variable nitrógeno ureico en sangre (BUN) es un indicador clínico que refleja la cantidad de urea nitrogenada en la sangre y se utiliza comúnmente para evaluar la función renal. Los valores de BUN pueden variar ligeramente entre laboratorios, pero típicamente se consideran normales entre: 7 a 20 mg/dL en adultos.

```
{r}  
shapiro.test(data_covid_19_diabetes$Nitrógeno_ureico_sangre)
```

shapiro-wilk normality test

data: data_covid_19_diabetes\$Nitrógeno_ureico_sangre
W = 0.70391, p-value < 2.2e-16

W = 0.70391, p-value < 0.00000000000000022

p-value (< 0.00000000000000022) < $\alpha = 0.05$

Se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alternativa (H_1), por lo tanto, los datos no siguen una distribución normal.

Respecto a la interpretación de los dos resultados

Las hipótesis de la prueba de Shapiro-Wilk

- La hipótesis nula (H_0) establece que la muestra proviene de una distribución normal.
- La hipótesis alternativa (H_1) plantea que la muestra no proviene de una distribución normal.

El valor de p encontrado para la variable **Saturación de oxígeno** ($p < 0.00000000000000022$) fue menor a 0.05, entonces el resultado de la evaluación de normalidad para esta variable indica que NO tiene una distribución normal, aceptándose la hipótesis alternativa (H_1).

De igual manera, el resultado para la variable **Nitrogeno ureico** en sangre ($p < 0.00000000000000022$) indica que la muestra NO proviene de una distribución normal, aceptándose la hipótesis alternativa (H_1).

2. Para datos categóricos

El dataset para esta sesión contiene información sobre el estado de desenlace fallecidos. En esta muestra, el número de fallecidos es 168 de un total de 686.

```
{r}  
table(data_covid_19_diabetes$desenla_fallecido)  
  
No  Sí  
518 168
```

Un estudio previo realizado en Perú en el año 2020, titulado "**Manejo del paciente hospitalizado con diabetes mellitus y COVID-19**", reportó una prevalencia de fallecidos del 9,1% en pacientes con diabetes y COVID-19. (DOI: <https://doi.org/10.15381/anales.v81i2.17781>)

En este caso, la prevalencia del estudio previo representa el valor esperado, mientras que la prevalencia observada en nuestro conjunto de datos representa el valor observado.

Uno de los objetivos de nuestro análisis es evaluar si la proporción observada de fallecidos difiere significativamente de la proporción esperada. Para ello, utilizamos la prueba de bondad de ajuste de Chi-cuadrado.

Las hipótesis de esta prueba son las siguientes:

- **Hipótesis nula (H_0):** No existe una diferencia significativa entre la proporción observada y la esperada.
- **Hipótesis alternativa (H_1):** Existe una diferencia significativa entre la proporción observada y la esperada.

En R, esta prueba se realiza mediante la función `chisq.test()`, a la cual se deben proporcionar los valores observados y las proporciones esperadas para llevar a cabo la comparación.

```
chisq.test(x = c(168, 518), p = c(0.091, 0.909))
```

Chi-squared test for given probabilities

```
data: c(168, 518)  
X-squared = 196.42, df = 1, p-value < 2.2e-16
```

X-squared = 196.42, df = 1, p-value < 0.00000000000000022

p-value (< 0.00000000000000022) < $\alpha = 0.05$

Se rechaza la hipótesis nula (H_0) y se acepta la hipótesis alternativa (H_1).

Interpretación

Dado que el valor de p es menor a 0.05, podemos concluir que las proporciones observadas son significativamente diferentes de las proporciones esperadas, aceptándose la hipótesis alternativa (H_1).