

ACTIVIDAD 9

INTEGRANTES:

- Llancari Nivin Meyli
- Vera Fonseca, July
- Alejo Huaman, Melissa
- Flores Diaz, Christian
- Ccuro Minaya, Lucia

Sobre los datos para la práctica

El conjunto de datos covid19_diabetes.csv contiene información sobre pacientes con COVID-19, incluyendo variables como edad, severidad, glucosa, saturación de oxígeno, días de hospitalización, desenlace (fallecido o no), y comorbilidades como diabetes mellitus complicada o simple. En esta práctica, usaremos las variables Glucosa (mg/dL, variable dependiente) y Edad (categorizada, variable independiente) para la regresión lineal, Saturación_O2 para la prueba t, y severidad para el ANOVA.

Regresión lineal simple

Objetivo

Evaluar la asociación entre los niveles de glucosa en sangre (Glucosa) y la edad categorizada (Edad) de los pacientes.

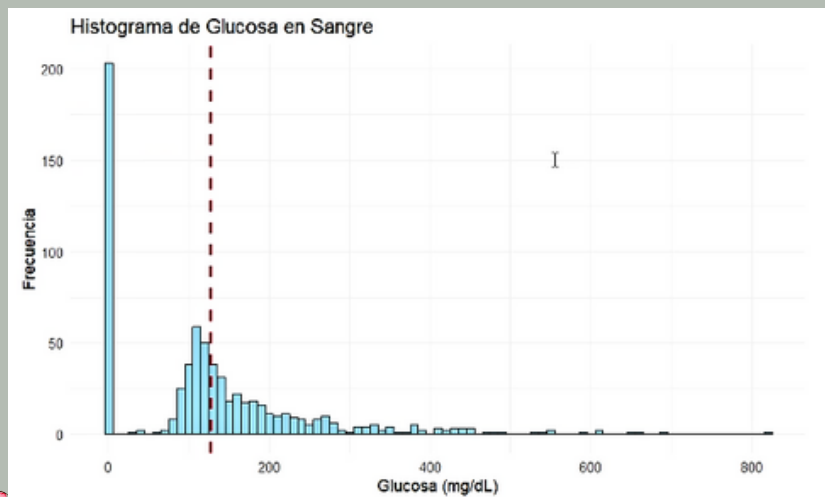
Procedimientos

Visualización de la distribución de glucosa

Creemos un histograma de la variable Glucosa con una línea vertical que indica el promedio.

```
{r}
#| label: hist-glucosa
#| fig-cap: "Distribución de los niveles de glucosa en sangre"

ggplot(data_covid, aes(x = Glucosa)) +
  geom_histogram(binwidth = 10, fill = "skyblue", color = "black") +
  geom_vline(aes(xintercept = mean(Glucosa, na.rm = TRUE)),
    color = "red", linetype = "dashed", size = 1) +
  labs(title = "Histograma de Glucosa en Sangre",
    x = "Glucosa (mg/dL)", y = "Frecuencia") +
  theme_minimal()
```



```
#
# if (!requireNamespace("car", quietly = TRUE)) { install.packages("car")}
#### Cálculo del promedio de glucosa
```

Calculamos el promedio de la variable Glucosa, ignorando valores perdidos.

```
## [r]
#| label: mean-glucosa
mean_glucosa <- mean(data_covid$Glucosa, na.rm = TRUE)
cat("El promedio de glucosa es:", round(mean_glucosa, 2), "mg/dL\n")
##
```

El promedio de glucosa es: 127.77 mg/dL

```
#### Visualización de la relación lineal
```

Creamos un diagrama de dispersión para visualizar la relación entre Glucosa y Edad (usando la variable numérica Edad para el gráfico, aunque la regresión usará la categórica).

```
## [r]
#| label: scatter-glucosa-edad
#| fig-cap: "Relación entre Glucosa y Edad"

ggplot(data_covid, aes(x = Edad, y = Glucosa)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = TRUE, color = "blue")
##
```

```
#### Visualización de la relación lineal
```

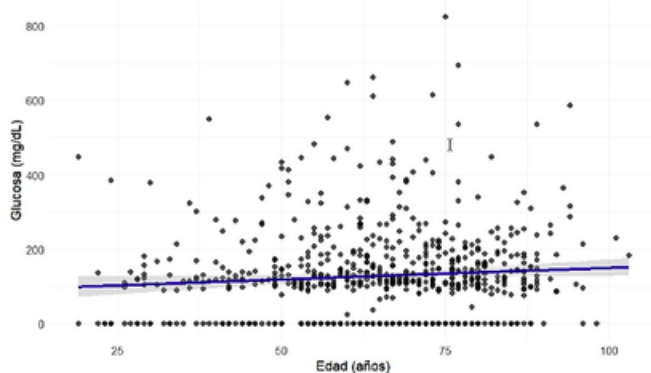
Creamos un diagrama de dispersión para visualizar la relación entre Glucosa y Edad (usando la variable numérica Edad para el gráfico, aunque la regresión usará la categórica).

```
## [r]
#| label: scatter-glucosa-edad
#| fig-cap: "Relación entre Glucosa y Edad"

ggplot(data_covid, aes(x = Edad, y = Glucosa)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(title = "Diagrama de Dispersión: Glucosa vs Edad",
       x = "Edad (años)", y = "Glucosa (mg/dL)") +
  theme_minimal()
##
```

D[38:5232m]geom_smooth() using formula = 'y ~ x'[39m

Diagrama de Dispersión: Glucosa vs Edad



```
# Ajuste del modelo de regresión lineal simple
```

Ajustamos un modelo de regresión lineal simple con Glucosa como variable dependiente y Edad (categórica) como variable independiente.

Resumen del modelo

Obtenemos los coeficientes, errores estándar, valores t, valores p y el R^2 del modelo.

```
##{r}
#| label: summary-lm
summary(modelo)
##
```

Call:
lm(formula = Glucosa ~ Edad, data = data_covid)

Residuals:

	Min	1Q	Median	3Q	Max
	-148.83	-114.76	-11.56	44.99	689.62

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	87.2871	19.4900	4.479	8.81e-06 ***
Edad	0.6279	0.2934	2.140	0.0327 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.1 on 684 degrees of freedom
Multiple R-squared: 0.006653, Adjusted R-squared: 0.0052
F-statistic: 4.581 on 1 and 684 DF, p-value: 0.03269

Interpretación de resultados

El modelo de regresión lineal simple evalúa la relación entre la glucosa en sangre y la edad categorizada. Los coeficientes indican el cambio promedio en glucosa para cada categoría de edad en comparación con la categoría de referencia (0-60 años). El valor p del predictor indica si la relación es estadísticamente significativa. El R^2 indica la proporción de variabilidad en Glucosa explicada por Edad.

Reporte de resultados

Generamos una tabla con los resultados del modelo usando `tbl_regression` y la exportamos como un archivo word.

```
##{r}
library(gtsummary)

tbl_regression(
  modelo,
  label = list(Edad ~ "Edad (categorías)"),
  show_single_row = "Edad"
) %>%
  modify_caption("***Tabla 1: Resultados de la Regresión Lineal Simple**")
##
```

i The package "broom.helpers" (>= 1.17.0) is required.
x would you like to install it?

1: Yes
2: No

> El modelo de regresión lineal simple mostró que la edad categorizada tiene una asociación significativa con los niveles de glucosa en sangre ($p < 0.05$). Por cada incremento en la categoría de edad (>60 , >70 , >80 en comparación con 0-60 años), los niveles de glucosa aumentan en promedio, con un R^2 que indica que aproximadamente $X\%$ de la variabilidad en

Prueba t de student para muestras independientes

Objetivo

Evaluar si el promedio de la saturación de oxígeno (Saturación_O2) difiere significativamente entre pacientes con diabetes mellitus simple (diabetes_mell_simple = "Si") y sin diabetes mellitus simple (diabetes_mell_simple = "No").

Procedimientos

Resumen estadístico

Calculamos el número de observaciones, la media y la desviación estándar de Saturación_O2 por grupo de diabetes mellitus simple.

```
##{r}
#| label: summary-saturacion
data_covid %>%
  group_by(diabetes_mell_simple) %>%
  summarise(
    n = n(),
    media = mean(Saturación_O2, na.rm = TRUE),
    desvio = sd(Saturación_O2, na.rm = TRUE)
  )
##
```

Visualización

Creamos un diagrama de caja para comparar la distribución de Saturación_O2 entre los grupos.

```
---[r]
#| label: boxplot-saturacion
#| fig-cap: "Distribución de Saturación de O2 por Diabetes Mellitus Simple"

ggplot(data_covid, aes(x = diabetes_mell_simple, y = Saturación_O2, fill = diabetes_mell_simple)) +
  geom_boxplot() +
  labs(title = "Saturación de O2 por Diabetes Mellitus Simple",
       x = "Diabetes Mellitus Simple", y = "Saturación de O2 (%)" ) +
  theme_minimal()
```

Pruebas preliminares

Supuesto de normalidad

Verificamos si los datos de Saturación_O2 en cada grupo siguen una distribución normal usando la prueba de Shapiro-wilk.

```
---[r]
#| label: shapiro-saturacion
shapiro.test(data_covid[Saturación_O2[data_covid$diabetes_mell_simple == "Si"]])
shapiro.test(data_covid[Saturación_O2[data_covid$diabetes_mell_simple == "No"]])
---
```

Supuesto de homogeneidad de varianzas

Evaluamos la igualdad de varianzas entre los grupos usando la prueba F.

```
---[r]
#| label: levene-severidad
leveneTest(severidad ~ pac_fue_hospital, data = data_covid)
```

Supuesto de homogeneidad de varianzas

Evaluamos la igualdad de varianzas entre los grupos usando la prueba F.

```
---[r]
#| label: levene-severidad
leveneTest(severidad ~ pac_fue_hospital, data = data_covid)
```

ANOVA de una vía

Realizamos el ANOVA para comparar las medias de severidad entre los grupos de hospitalización.

```
---[r]
#| label: anova-severidad
anova_model <- aov(severidad ~ pac_fue_hospital, data = data_covid)
```

Resumen de ANOVA

Obtenemos los resultados del ANOVA.

```
---[r]
#| label: summary-anova
summary(anova_model)
```

Realizamos el ANOVA para comparar las medias de severidad entre los grupos de hospitalización.

```
---[r]
#| label: anova-severidad
anova_model <- aov(severidad ~ pac_fue_hospital, data = data_covid)
```

Resumen de ANOVA

Obtenemos los resultados del ANOVA.

```
---[r]
#| label: summary-anova
summary(anova_model)
```

Prueba post - hoc

Si el ANOVA detecta diferencias significativas ($p < 0.05$), realizamos una prueba post-hoc de Tukey para comparaciones por pares.

```
---[r]
#| label: tukey-severidad
#| eval: false
TukeyHSD(anova_model)
```

Interpretación

Si el valor p del ANOVA es < 0.05 , concluimos que hay diferencias significativas en la severidad promedio entre los grupos de hospitalización. La prueba post-hoc identifica qué grupos difieren. Si $p > 0.05$, no hay evidencia de diferencias significativas.