

Prueba evaluación curso 2024-2025

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.

Medidas de tendencia central (Nivel glucosa basal)			
Grupo de Control 1		Grupo de Control 2	
Media	84,69	Media	89,40
Mediana	82,00	Mediana	90,00
Moda	90,00	Moda	88,00

El Grupo 1 muestra una media (84,69) inferior a la moda (90), sugiriendo una ligera asimetría a la derecha, mientras que en el Grupo 2 la media (89,40) y la mediana (90) están más próximas, indicando una distribución más simétrica.

Medidas de dispersión (Nivel glucosa basal)			
Grupo de Control 1		Grupo de Control 2	
Mínimo	65	Mínimo	77
Máximo	103	Máximo	104
Rango	38	Rango	27
Percentil 25	78,5	Percentil 25	84,5
Percentil 75	90,5	Percentil 75	94
R. Intercuartílico	12	R. Intercuartílico	9,5
Varianza	78,22	Varianza	53,97
Desv. Típica	8,84	Desv. Típica	7,35
CV	0,10	CV	0,08

El Grupo de Control 1 muestra mayor dispersión en los niveles de glucosa basal que el Grupo 2, con un rango, varianza, desviación estándar y coeficiente de variación superiores.

Aun así, ambos presentan un bajo coeficiente de variación ($CV1 = 0,10$ y $CV2 = 0,08$). Se suele considerar que la media es representativa si el coeficiente de variación es menor que 0,3. Por lo tanto, **podemos afirmar que la media es representativa en ambos grupos** de control para los niveles de glucosa basales.

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

Grupo control 2 (Nivel glucosa basal)	
Asimetría	-0,0075
Curtosis	-0,6127

Con base en los valores proporcionados:

1. **Asimetría (-0.0075): Es prácticamente cero**, indicando que la distribución del nivel de glucosa basal en el Grupo 2 **es simétrica**.
2. **Curtosis (-0.6127): Es negativa**, lo que sugiere que **la distribución tiene colas más planas y un pico ligeramente menos pronunciado** que una distribución normal (platicúrtica).

Conclusión: La distribución del nivel de glucosa basal en el Grupo 2 es simétrica y ligeramente más plana 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.

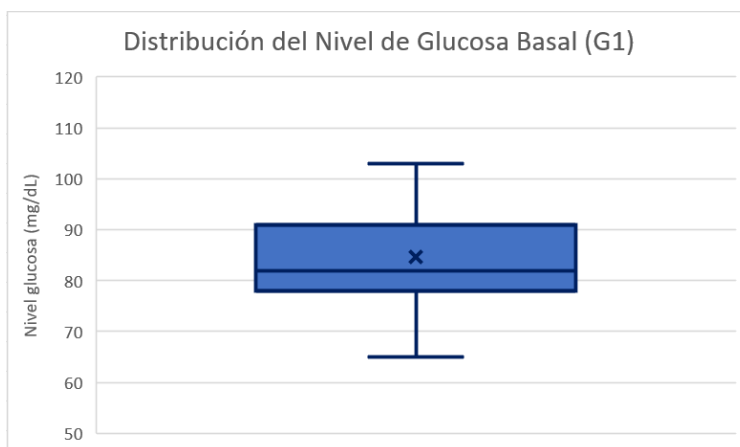
Grupo de Control 1			
Nivel glucosa basal		Nivel glucosa 60 min	
Mínimo	65,0	Mínimo	131,0
Primer Cuartil	78,5	Primer Cuartil	146,5
Mediana	83,0	Mediana	150,0
Tercer Cuartil	93,0	Tercer Cuartil	156,5
Máximo	103,0	Máximo	190,0

Interpretación de los resultados:

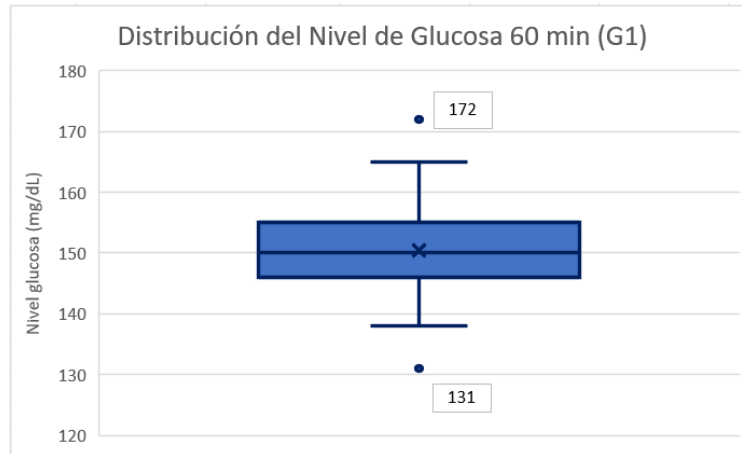
- **Nivel basal (antes de la administración de glucosa):**
 - El nivel de glucosa basal va de **65.0** a **103.0** (mínimo a máximo), lo que indica los niveles de glucosa en ayunas, antes de la ingesta de los 100 gramos de glucosa.
- **Nivel a los 60 minutos (después de la carga de glucosa):**
 - A los 60 minutos, el nivel de glucosa sube significativamente, con un rango de **131.0** a **190.0**.
 - La **mediana** en este momento es de **150.0**, lo que muestra un aumento notable respecto al nivel basal (**83.0 en la mediana basal**).
 - Este aumento es esperado, ya que la glucosa ingerida se absorbe en el torrente sanguíneo y genera un incremento en los niveles de glucosa.

Conclusión sobre los cuartiles:

- **Nivel basal:** Los datos están más dispersos en el nivel basal, con una mayor variabilidad en los valores (rango intercuartílico de 14.5).
- **Nivel a los 60 minutos:** Los cuartiles muestran una distribución más concentrada, con menos variabilidad en los valores (rango intercuartílico de 10.0), lo que puede indicar que, a pesar de los aumentos, los valores se agrupan más cerca de la mediana en este punto.



El box-plot mostrado, perteneciente al nivel de glucosa basal (G1), **no muestra valores atípicos** (outliers).



En cambio, el box-plot perteneciente al nivel de glucosa pasados 60 min, nos indica **dos valores atípicos**. Por el lado inferior el **131** y por el lado superior en **172**.

Esto sugiere que algunos individuos tienen respuestas de glucosa que se **desvían significativamente** del resto 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.

```
> shapiro_test_g1 <- shapiro.test(nivel_glucosa_60_g1)
> print(shapiro_test_g1)
```

Shapiro-wilk normality test

```
data: nivel_glucosa_60_g1
W = 0.96261, p-value = 0.2734
```

```
> shapiro_test_g2 <- shapiro.test(nivel_glucosa_60_g2)
> print(shapiro_test_g2)
```

Shapiro-wilk normality test

```
data: nivel_glucosa_60_g2
W = 0.97761, p-value = 0.7592
```

Interpretación de los resultados:

- **Grupo 1 (nivel_glucosa_60_g1):**
 - El valor p es **0.2734**, que es **mayor que 0.05**. Esto significa que **no rechazamos la hipótesis nula**, lo que indica que **los datos del grupo 1 siguen una distribución normal**.
- **Grupo 2 (nivel_glucosa_60_g2):**
 - El valor p es **0.7592**, que también es **mayor que 0.05**. Esto significa que **no rechazamos la hipótesis nula** y, por lo tanto, **los datos del grupo 2 también siguen una distribución normal**.

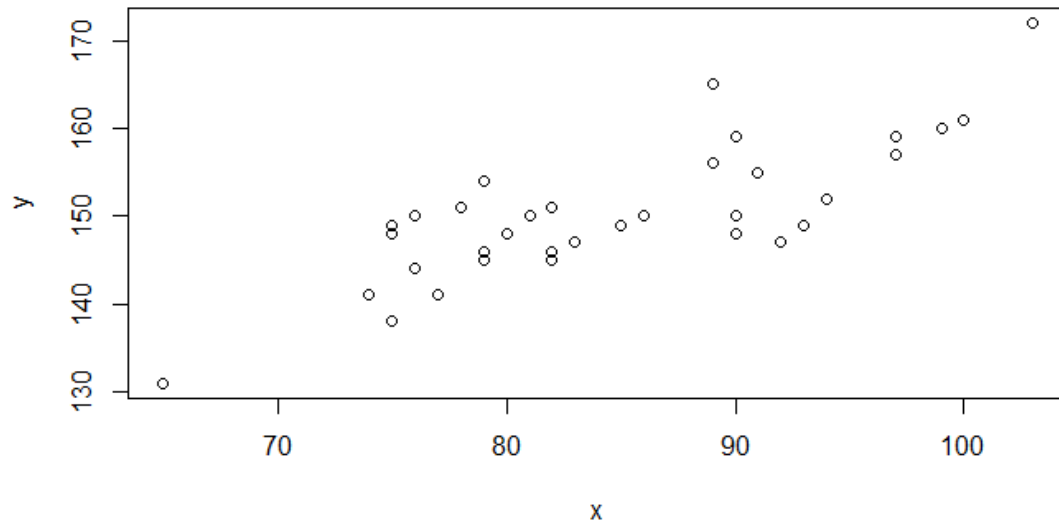
Ejercicio 2

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.

Tras filtrar los datos para seleccionar aquellos valores correspondientes al grupo de control 1 y seleccionar '**x**' (**variable independiente**) como el **nivel de glucosa basal** e '**y**' (**variable dependiente**) como el **nivel de glucosa pasados 60 min**, procedemos a ejecutar el siguiente análisis utilizando el programa RStudio.

Grafico de Dispersión



Fijándonos en el gráfico de dispersión (obtenido a través de R), que tiene por valores independientes (x) el nivel basal del grupo (G1) y por valores dependientes (y) el nivel de glucosa a los 60 min (G1), observamos **que es probable que haya algún tipo de correlación positiva entre las variables**. Ya que, los puntos parecen formar una línea ascendente en diagonal y seguir algún tipo de patrón.

Hallaremos el coeficiente de correlación para tener un dato más claro sobre esto.

```
> correlacion <- cor(x, y)
> print(correlacion)
[1] 0.7964053
```

Una correlación de **0.7964** nos indica una relación lineal **fuerte y positiva** entre los valores de glucosa a nivel basal y los valores de glucosa 1 hora después. Concordando este dato, con el gráfico inicial de dispersión que ya nos hacía intuir estos resultados.

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

```
> summary(recta)
```

```
Call:
```

```
lm(formula = y ~ x)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-8.4972 -3.0284 -0.3754  2.5325 11.5934
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  91.38365    7.84251   11.652 3.08e-13 ***
x              0.69689    0.09212    7.565 1.06e-08 ***
---

```

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

```
Residual standard error: 4.751 on 33 degrees of freedom
```

```
Multiple R-squared:  0.6343,    Adjusted R-squared:  0.6232
```

```
F-statistic: 57.23 on 1 and 33 DF,  p-value: 1.056e-08
```

Basándonos en los datos obtenidos a través de R, obtenemos que nuestro modelo lineal sigue la siguiente formula:

$$y = 91,38 + 0,70x$$

Coeficientes:

1. Intercepto (91,38):

- Este es el valor predicho de y (glucosa 1 hora después) cuando x=0.

2. Pendiente (0,70):

- Indica que por cada unidad de aumento en x (glucosa basal), se espera que y (glucosa 1 hora después) aumente, en promedio, en 0,70 unidades.
- La relación es positiva.

Para estudiar la estimación para un paciente cuyo nivel basal es de 83 mg/Dl simplemente tendríamos que sustituir el valor de la x por 83 y realizar el cálculo.

$$y = 91,38 + 0,70 \cdot 83 = 149,48$$

Es decir, el modelo predice que nuestro paciente con 83 mg/Dl de nivel de glucosa basal **tendría un nivel, pasados los 60 min, de aproximadamente 149,48 mg/Dl.**

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

Basándonos en los datos obtenidos con R en el apartado b), obtenemos que:

Calidad del ajuste:

1. R^2 (0,6343):

- Indica que el modelo explica el 63.43% de la variabilidad de los valores de glucosa 1 hora después.
- Es un ajuste moderado-alto.

Si lo que necesitamos es la parte del modelo que no queda explicado, **tendremos que restarle a 1 (el total) el R^2 (la parte explicada por el modelo).**

$$1 - R^2 = 1 - 0,6343 = 0,3657$$

Esto significa que **el 36.57% de la variabilidad** en los valores de glucosa 1 hora después no está explicado por el modelo. Esta parte puede deberse a:

- Factores no incluidos en el modelo (por ejemplo, otras variables relevantes como peso, dieta, etc.).
- Variabilidad aleatoria o ruido en los datos.

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?

Para responder a esta pregunta, usamos la **pendiente** (β_1) del modelo de regresión lineal:

$$y = \beta_0 + \beta_1 \cdot x$$

La **pendiente** del modelo ($\beta_1=0.70$) indica cuánto cambia 'y' (glucosa 1 hora después) por cada unidad de cambio en x (nivel basal).

$$\Delta y = 0.7 \cdot 5 = 3,5$$

Si el nivel basal de un paciente aumenta en 5 mg/Dl, su nivel de glucosa 1 hora después **aumentaría en promedio en 3,5 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.

```
> resultado_95 <- t.test(nivel_basal_g1, mu = 88, conf.level = 0.95)
> print(resultado_95)
```

One Sample t-test

```
data: nivel_basal_g1
t = -2.217, df = 34, p-value = 0.03342
alternative hypothesis: true mean is not equal to 88
95 percent confidence interval:
 81.64759 87.72384
sample estimates:
mean of x
 84.68571
```

```
> resultado_99 <- t.test(nivel_basal_g1, mu = 88, conf.level = 0.99)
> print(resultado_99)
```

One Sample t-test

```
data: nivel_basal_g1
t = -2.217, df = 34, p-value = 0.03342
alternative hypothesis: true mean is not equal to 88
99 percent confidence interval:
 80.60687 88.76456
sample estimates:
mean of x
 84.68571
```

Resultados del contraste al 95%

- **p-valor:** 0.03342
- **Intervalo de confianza:** [81,65 a 87,72]
- **Media muestral:** 84.69

Interpretación al 95%:

1. Decisión sobre H_0 :

- Dado que el **p-valor (0.03342)** es menor que **$\alpha = 0.05$** , se **rechaza la hipótesis nula** al nivel de confianza del 95%.
- **No se podría admitir que el nivel medio de glucosa en sangre sea de 88 mg/Dl** con este nivel de confianza (95%).

2. Intervalo de confianza:

- El valor **88 no está dentro del intervalo [81,65 a 87,72]**, lo que también refuerza la decisión de rechazar H_0 .

Resultados del contraste al 99%

- **p-valor:** 0.03342
- **Intervalo de confianza:** [80,61 a 88,76]
- **Media muestral:** 84.69

Interpretación al 99%:

1. Decisión sobre H_0 :

- En este caso, el **p-valor (0.03342)** es mayor que **$\alpha = 0.01$** , por lo que **no se rechaza la hipótesis nula** al nivel de confianza del 99%.
- **Se podría admitir que el nivel medio de glucosa en sangre sea de 88 mg/Dl** con este nivel de confianza (99%).

2. Intervalo de confianza:

- El **88** está **dentro del intervalo [80,61 a 88,76]**, lo que respalda la decisión de no rechazar H_0 .

Conclusión General

- Al **95% de confianza**, rechazamos H_0 y concluimos que la media no es igual a 88.
- Al **99% de confianza**, no rechazamos H_0 , indicando que no hay evidencia suficiente para concluir que la media es distinta de 88.

Esto refleja cómo el nivel de confianza afecta nuestras decisiones: un nivel de confianza más alto requiere un p-valor más pequeño para rechazar H_0 .

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.

```
> resultado_diferencia <- t.test(
+   nivel_basal_g1,
+   nivel_basal_g2,
+   var.equal = TRUE,
+   conf.level = 0.95
+ )
> resultado_diferencia
```

Two Sample t-test

```
data: nivel_basal_g1 and nivel_basal_g2
t = -2.3138, df = 63, p-value = 0.02395
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.7858509 -0.6427205
sample estimates:
mean of x mean of y
 84.68571  89.40000
```

Intervalo de confianza al 95% para la diferencia de medias:

- El intervalo es $[-8,79 \text{ a } -0,64]$. **Este intervalo no incluye el 0**, lo que nos indica que **hay una diferencia significativa** entre las medias de los dos grupos.
- El hecho de que el intervalo esté completamente negativo indica que, en promedio, los niveles de glucosa en los jóvenes son más bajos que en los adultos.

p-valor (0.02395): Dado que el **p-valor es menor que 0.05** (nivel de significación del 5%), **rechazamos la hipótesis nula de que las medias de los dos grupos son iguales**. Esto nos refuerza la anterior conclusión de que existe una diferencia significativa entre los niveles basales de glucosa de los jóvenes y los adultos.

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%. 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%.

```
> nivel_glucosa_basal <- EVALMASTER$`Nivel glucosa basal`
> proporcion <- sum(nivel_glucosa_basal > 95) / length(nivel_glucosa_basal)
> proporcion
[1] 0.1692308
```

El cálculo de la proporción de la muestra con glucosa superior a 95 mg/Dl (prediabetes) es de 16,92%. Es decir, **un 16,92% de los pacientes tiene prediabetes.**

```
> ic_proporcion <- prop.test(sum(nivel_glucosa_basal > 95), length(nivel_glucosa_basal), conf.level = 0.98)
> print(ic_proporcion)
```

1-sample proportions test with continuity correction

```
data: sum(nivel_glucosa_basal > 95) out of length(nivel_glucosa_basal), null probability 0.5
X-squared = 27.138, df = 1, p-value = 1.894e-07
alternative hypothesis: true p is not equal to 0.5
98 percent confidence interval:
 0.08225785 0.31042485
sample estimates:
              p
0.1692308
```

Intervalo de confianza (al 98%):

- El intervalo de confianza para la proporción con glucosa mayor a 95 mg/Dl, con un nivel de confianza del 98%, es: **[0,0823 a 0,3104]**

Esto significa que, con un **98% de confianza**, la muestra calculada anteriormente de **0,1692 estaría dentro del intervalo**, indicando que es un valor correcto **según el intervalo de confianza**.

p-valor ($1,89 \cdot 10^{-7}$):

- El p-valor nos muestra un resultado muy pequeño y, por supuesto, inferior al valor de significación $\alpha = 0.05$.

Por lo tanto, según el p-valor, **se rechazaría la hipótesis nula**, lo que nos indicaría que **0,15 no es valor correcto**.

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 cabo de 60 minutos en el grupo de los adultos 4 (contraste de muestras emparejadas). Plantea el correspondiente contraste de hipótesis considerando un nivel de significación del 5%.

```
> diferencias_adultos <- nivel_glucosa_60_g2 - nivel_basal_g2  
> resultado_contraste <- t.test(diferencias_adultos, mu = 0, alternative = "two.sided")  
> print(resultado_contraste)
```

One Sample t-test

```
data: diferencias_adultos  
t = 34.374, df = 29, p-value < 2.2e-16  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
 77.52862 87.33805  
sample estimates:  
mean of x  
 82.43333
```

Resultados del contraste de hipótesis (Prueba t para muestras emparejadas):

- **P-valor** = $2,2e-16$ (muy pequeño, menor que 0.05)
- **Intervalo de confianza al 95%** = (77,53 a 87,34)
- **Estimación de la media de las diferencias** = 82,43

Conclusión:

Dado que el valor p es muy pequeño, rechazamos la hipótesis nula de que no hay diferencia en los niveles de glucosa y concluimos que **hay una variación significativa en el nivel de glucosa en sangre en los adultos después de 60 minutos** de la ingestión de glucosa. El aumento promedio es de aproximadamente **82.43 mg/DL**.