
Taller

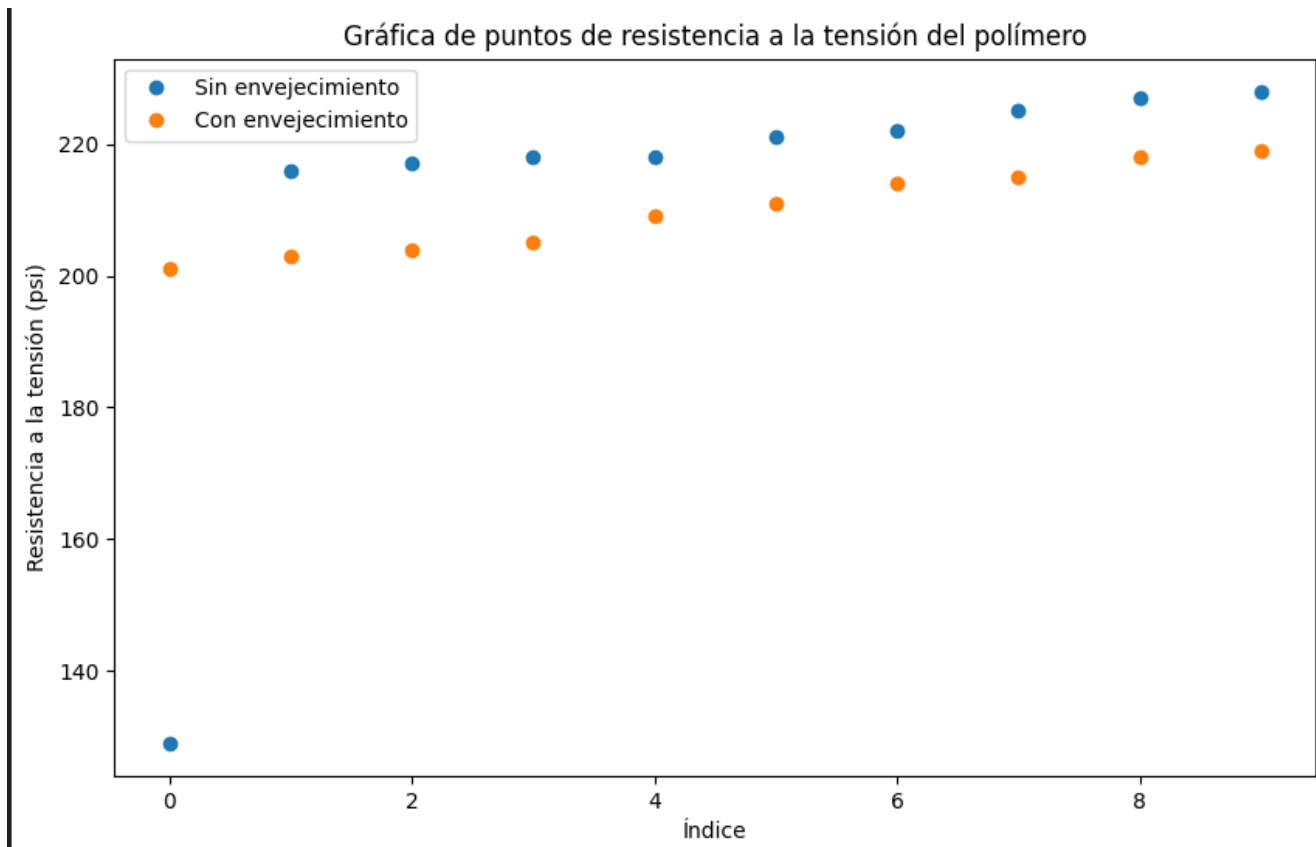
- El trabajo se recibirá en un documento en formato *pdf*.
- Se debe evidenciar el código que se utilizó, en caso de usar algún software, y los procedimientos deben quedar descritos.
- El documento debe estar ordenado y debe ser claro
- Se debe trabajar en grupos de mínimo 2 personas y máximo 3

El siguiente taller se trabajó en un documento Jupyter de visual studio code, el cual será entregado junto con este taller.

Se utilizó la librería numpy de Python tanto para estructurar los datos como para realizar gráficas y cálculos estadísticos.

Se utilizó la librería matplotlib para generar gráficos de los datos de los ejercicios.

1. Se utiliza cierto polímero para los sistemas de evacuación de los aviones. Es importante que el polímero sea resistente al proceso de envejecimiento. Se utilizaron 20 especímenes del polímero en un experimento. 10 Se asignaron aleatoriamente para exponerse a un proceso de envejecimiento acelerado del lote, el cual implica la exposición a altas temperaturas durante 10 días. se hicieron las mediciones de resistencia a la tensión de los especímenes y se registraron los siguientes datos sobre resistencia a la tensión en *psi*.
 - Sin envejecimiento: 227, 222, 218, 217, 225, 218, 216, 129, 228, 221
 - Con envejecimiento: 219, 214, 215, 211, 209, 218, 203, 204, 201, 205
- a) Elabore la gráfica de puntos de los datos



b) ¿En la gráfica que obtuvo parece que el proceso de envejecimiento, tuvo un efecto en la resistencia a la tensión de este polímero? Explique su respuesta

Si, en base a la gráfica se puede afirmar que, para después de 200 psi, los polímeros sin envejecimiento fueron más resistentes a la tensión.

c) Calcule la resistencia a la tensión de la media de la muestra en las dos muestras.

```
## c) Calcule la resistencia a la tensión de la media de la muestra en las dos muestras.

media_sin_envejecimiento = np.mean(polimero_sin_envejecimiento)

media_con_envejecimiento = np.mean(polimero_con_envejecimiento)

print("Resistencia a la tensión de la media del polímero sin envejecimiento:", media_sin_envejecimiento)
print("Resistencia a la tensión de la media del polímero con envejecimiento:", media_con_envejecimiento)
```

✓ 0.0s

Resistencia a la tensión de la media del polímero sin envejecimiento: 212.1
Resistencia a la tensión de la media del polímero con envejecimiento: 209.9

Se obtuvo que la resistencia a la tensión media de los polímeros sin envejecimiento fue de 212.1 psi.

Mientras que la de los polímeros con envejecimiento fue de 209.9 psi

Lo cual concuerda con la gráfica.

- d) Calcule la mediana de ambas. Analice la similitud o falta de similitud entre la media y la mediana de cada grupo.

```
## d) Calcule la mediana de ambas. Analice la similitud o falta de similitud entre la media y la mediana de cada grupo.

mediana_sin_envejecimiento = np.median(polimero_sin_envejecimiento)
mediana_con_envejecimiento = np.median(polimero_con_envejecimiento)

print("Mediana del polímero sin envejecimiento:", mediana_sin_envejecimiento)
print("Mediana del polímero con envejecimiento:", mediana_con_envejecimiento)
```

✓ 0.0s

Mediana del polímero sin envejecimiento: 219.5
Mediana del polímero con envejecimiento: 210.0

Se obtuvo que la mediana de los polímeros sin envejecimiento fue de 219.5 psi

Mientras que la mediana de los polímeros con envejecimiento fue de 210.0 psi

Ambos datos concuerdan con la gráfica.

- e) Calcule la varianza de la muestra y su desviación estándar en cuanto a la resistencia a la tensión en ambas muestras

```
import math
## e) Calcule la varianza de la muestra y su desviación estándar en cuanto a la resistencia a la tensión en ambas muestras

varianza_sin_envejecimiento = np.var(polimero_sin_envejecimiento, ddof=1) #se utiliza el parametro ddof = 1 para manejar los grados de libertad
varianza_con_envejecimiento = np.var(polimero_con_envejecimiento, ddof=1)

desviacion_estandar_sin_envejecimiento = math.sqrt(varianza_sin_envejecimiento)
desviacion_estandar_con_envejecimiento = math.sqrt(varianza_con_envejecimiento)

print("varianza de los polimeros sin envejecimiento:", varianza_sin_envejecimiento)
print("varianza de los polimeros con envejecimiento:", varianza_con_envejecimiento)

print("desviación estandar de los polimeros sin envejecimiento", desviacion_estandar_sin_envejecimiento)
print("desviación estandar de los polimeros con envejecimiento", desviacion_estandar_con_envejecimiento)
```

✓ 0.0s

varianza de los polimeros sin envejecimiento: 870.3222222222222
varianza de los polimeros con envejecimiento: 42.1
desviación estandar de los polimeros sin envejecimiento 29.501224080065256
desviación estandar de los polimeros con envejecimiento 6.488451279003334

Para la varianza:

Los polímeros sin envejecimiento tienen una varianza de 870.32222...

Los polímeros con envejecimiento tienen una varianza de 42.1

Para la desviación estándar:

Los polímeros sin envejecimiento tienen una desviación estándar de 29.501224080065256 psi

Los polímeros con envejecimiento tienen una desviación estándar 6.488451279003334 psi

Esto concluye que los datos de los polímeros sin envejecimiento están más dispersos que los datos de los polímeros con envejecimiento.

- f) ¿Parece haber alguna evidencia de que el envejecimiento afecta a la variabilidad en la resistencia a la tensión.

Basados en la gráfica y en los resultados de la varianza y la desviación estándar parece que el envejecimiento ayuda a que los datos no estén tan dispersos.

Sin embargo, no considero esta información como concluyente, creería que habría que analizar muestras más grandes de la población, analizar que pasa en psi bajos y en psi altos por ejemplo.

1

2. Con base en los datos recolectados en la tabla que se realizó en Excel realice los siguientes puntos.

- a) Identifique los tipos de variables

ID: Numérico discreto

País Favorito: Categórico

Deporte Favorito: Categórico

Altura: Numérico continuo

Edad: Numérico discreto

Tiempo de trayecto a la U: Numérico continuo

Licencia de conducir: binaria

- b) Realice un diagrama de caja con los datos numéricos.

Diagrama de Caja: Alturas

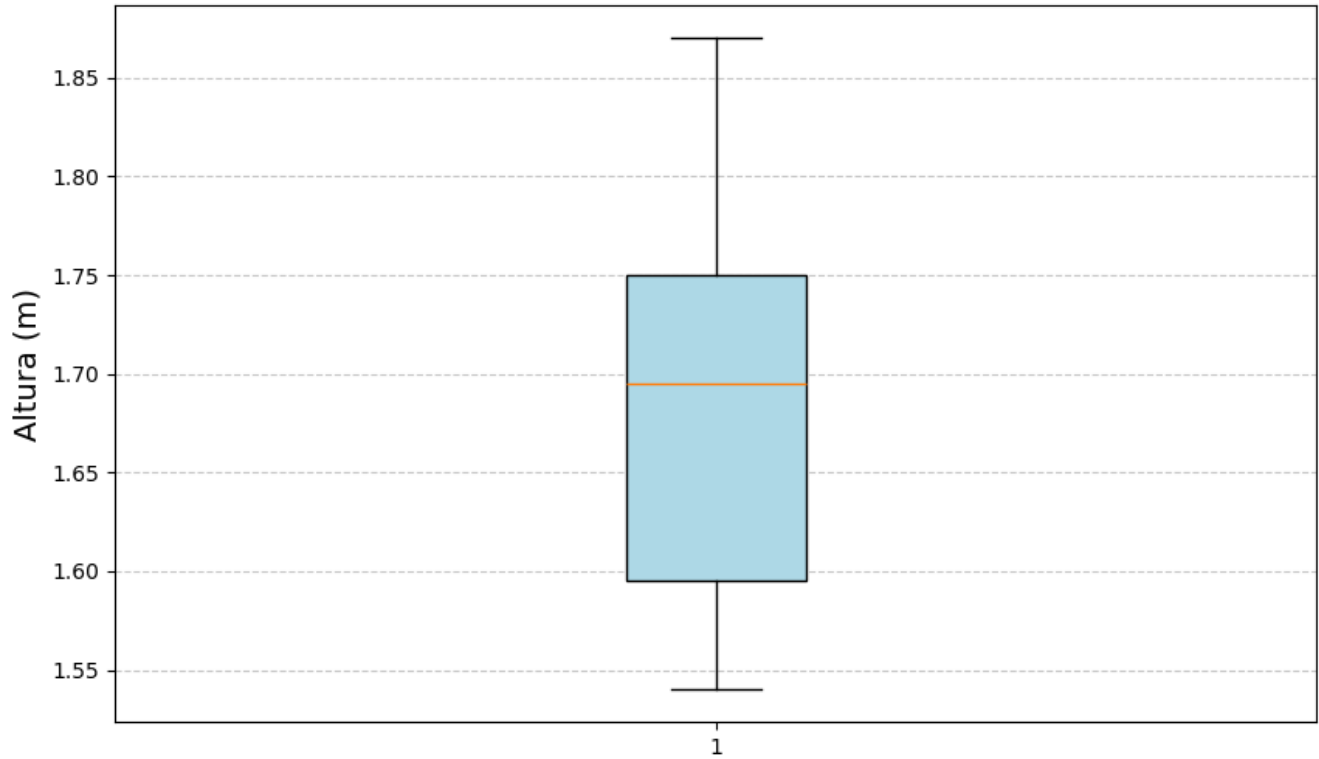


Diagrama de Caja: Edades

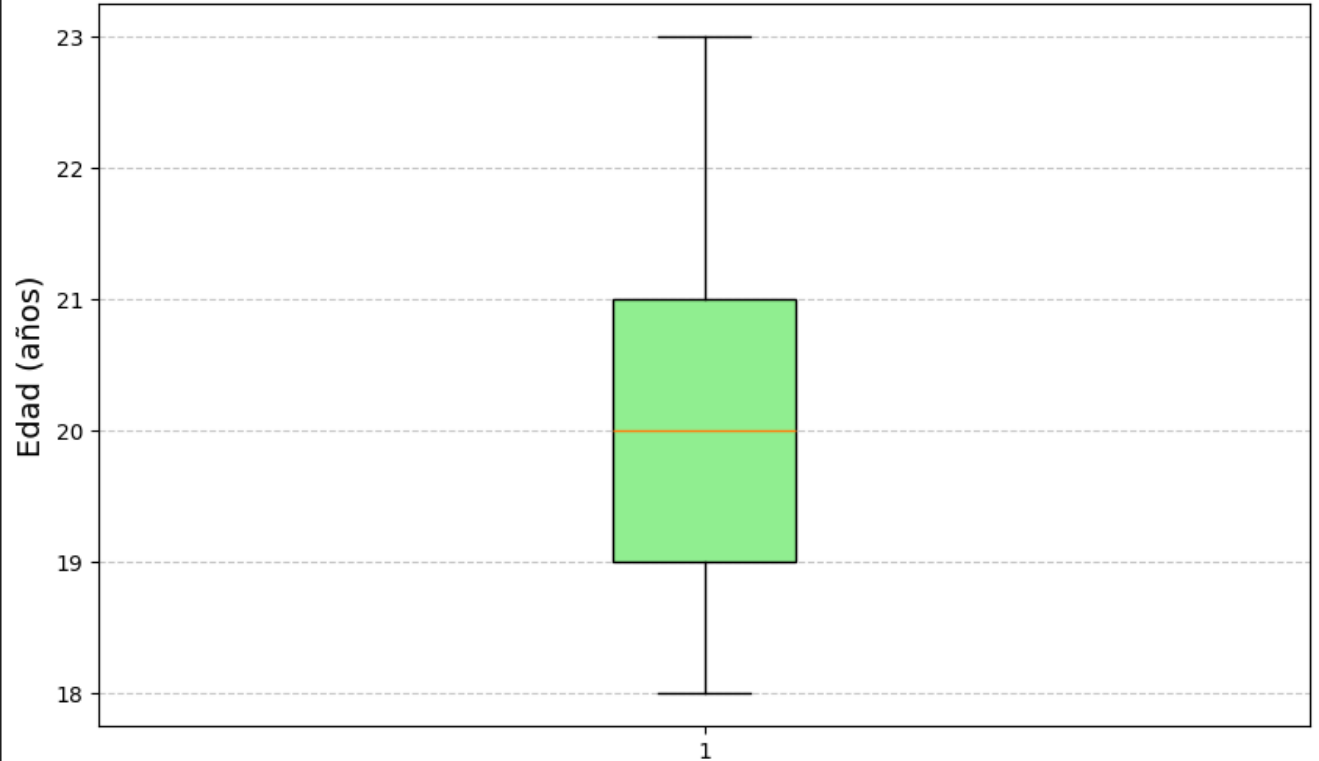
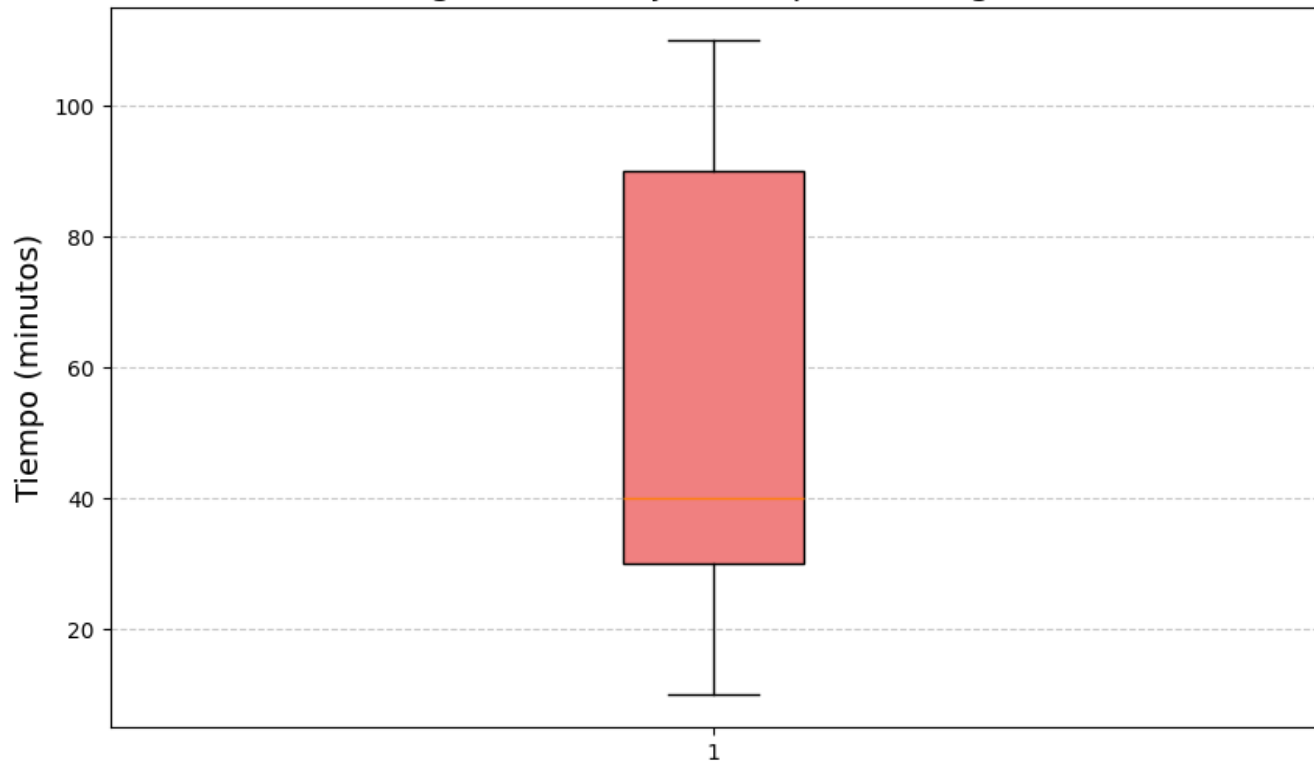
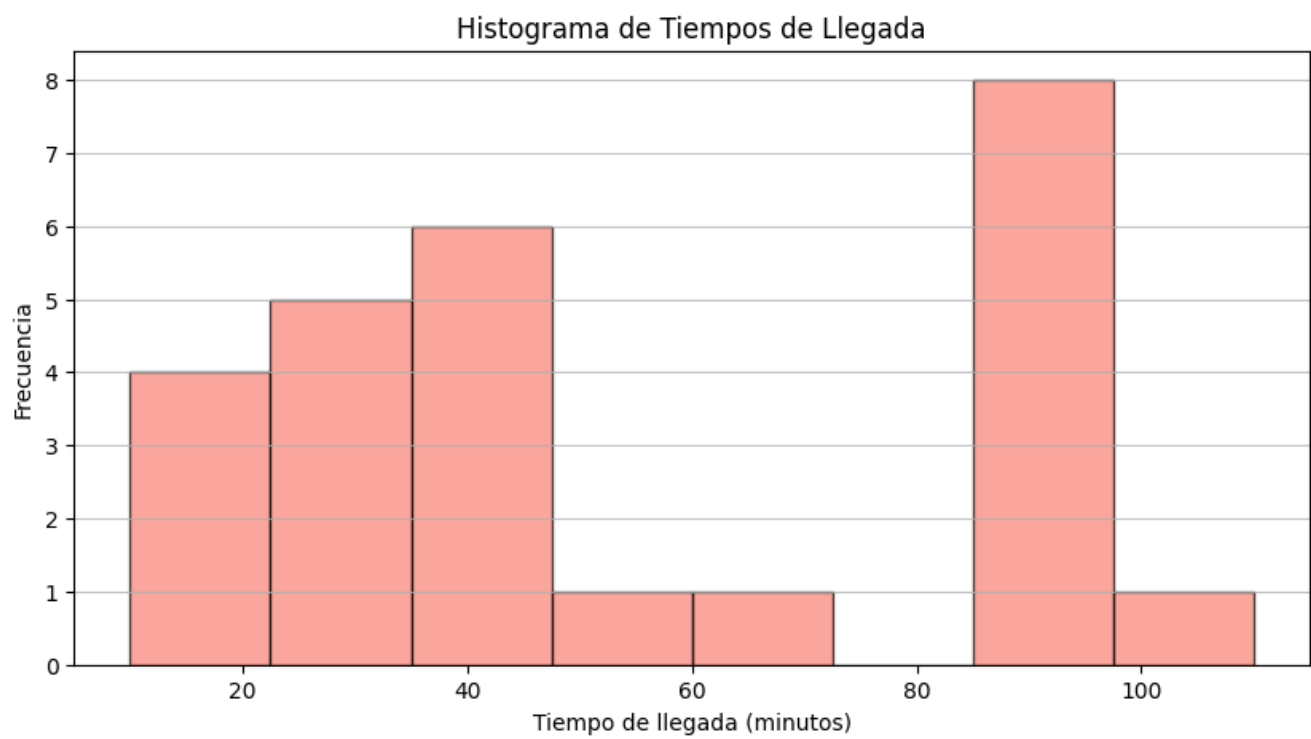
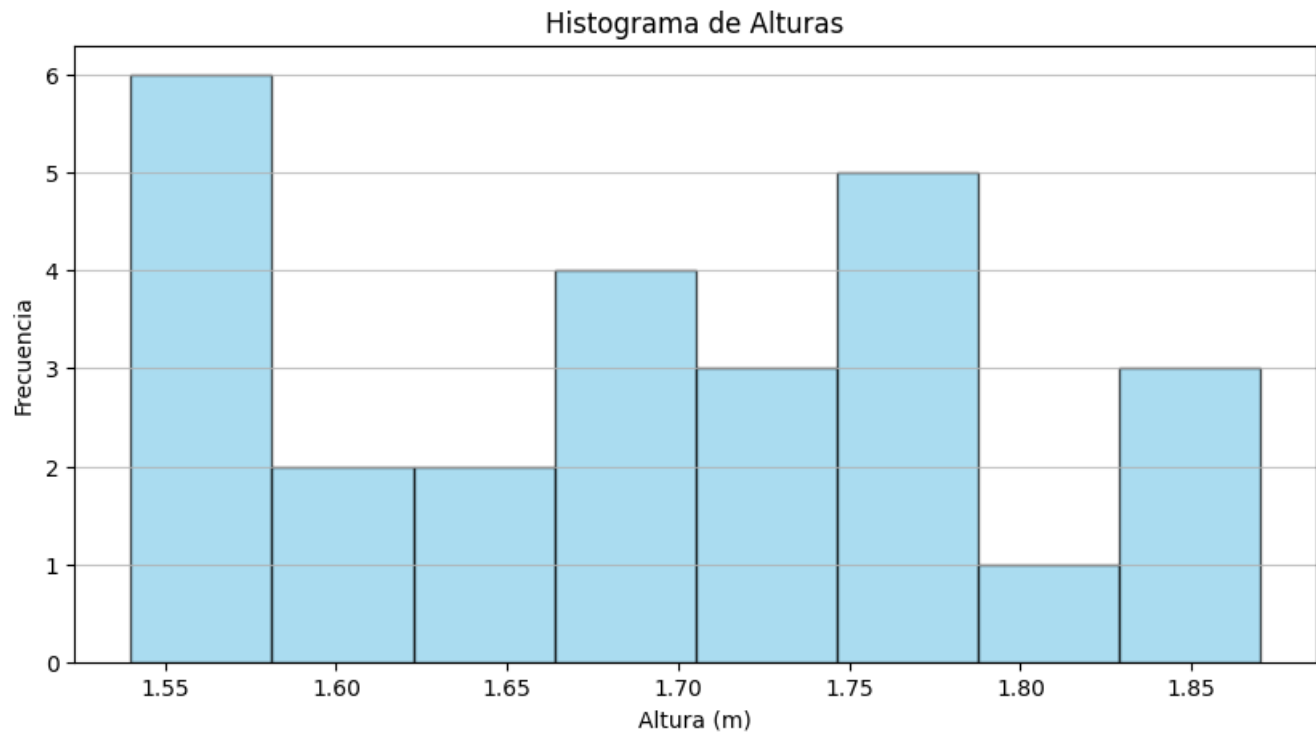


Diagrama de Caja: Tiempos de Llegada



c) Realice un histograma con los datos numéricos continuos.



d) ¿Que información puede observar de en los gráficos?

La información de la altura está menos dispersa que la de los tiempos de llegada

La altura más común está entre 1.52 aprox y 1.575 aprox

A la mayoría le toma una hora o más llegar a la universidad.

e) Halle la media y mediana para los datos numéricos.

```
## e) Halle la media y mediana para los datos numéricos.
#medias
media_altura = np.mean(alturas)
media_edades = np.mean(edades)
media_tiempos = np.mean(tiempos_de_llegada)

#medianas

mediana_alturas = np.median(alturas)
mediana_edades = np.median(edades)
mediana_tiempos = np.median(tiempos_de_llegada)

print("Media de alturas:", media_altura)
print("Mediana de alturas:", mediana_alturas)

print("Media de edades:", media_edades)
print("Mediana de edades:", mediana_edades)

print("Media de tiempos de llegada:", media_tiempos)
print("Mediana de tiempos de llegada:", mediana_tiempos)
```

✓ 0.0s

```
Media de alturas: 1.6888461538461539
Mediana de alturas: 1.6949999999999998
Media de edades: 19.92
Mediana de edades: 20.0
Media de tiempos de llegada: 54.0
Mediana de tiempos de llegada: 40.0
```

Para la altura la media fue de 1.68 m y la mediana de 1.69499... m

Para las edades la media fue de 19.92 años mientras que la mediana fue de 20 años

Y para los tiempos de llegada, el promedio llega en 54 minutos y la mediana fue de 40 minutos

f) Halle las medidas de variabilidad de los datos numéricos.

```
## f) Halle las medidas de variabilidad de los datos numéricos.

#varianzas

varianza_alturas = np.var(alturas, ddof=1 )
varianza_edades = np.var(edades, ddof=1)
varianza_tiempos = np.var(tiempos_de_llegada, ddof=1)

#desviación estandar

desviacion_alturas = math.sqrt(varianza_alturas)
desviacion_edades = math.sqrt(varianza_edades)
desviacion_tiempos = math.sqrt(varianza_tiempos)

print("Varianza de alturas:", varianza_alturas)
print("Desviación estándar de alturas:", desviacion_alturas)

print("Varianza de edades:", varianza_edades)
print("Desviación estándar de edades:", desviacion_edades)

print("Varianza de tiempos de llegada:", varianza_tiempos)
print("Desviación estándar de tiempos de llegada:", desviacion_tiempos)
```

✓ 0.0s

```
Varianza de alturas: 0.009690615384615386
Desviación estándar de alturas: 0.0984409233226476
Varianza de edades: 1.8266666666666669
Desviación estándar de edades: 1.351542328847553
Varianza de tiempos de llegada: 943.2
Desviación estándar de tiempos de llegada: 30.711561340967346
```

Según la información, la muestra de la altura es la menos dispersa de todas, seguida de las edades que tampoco está tan dispersa, sin embargo, los tiempos de llegada tienen una desviación estándar alta, lo que indica una alta dispersión de datos.

g) ¿Cuáles son los cuartiles de los datos?

```
## g) ¿Cuáles son los cuartiles de los datos?

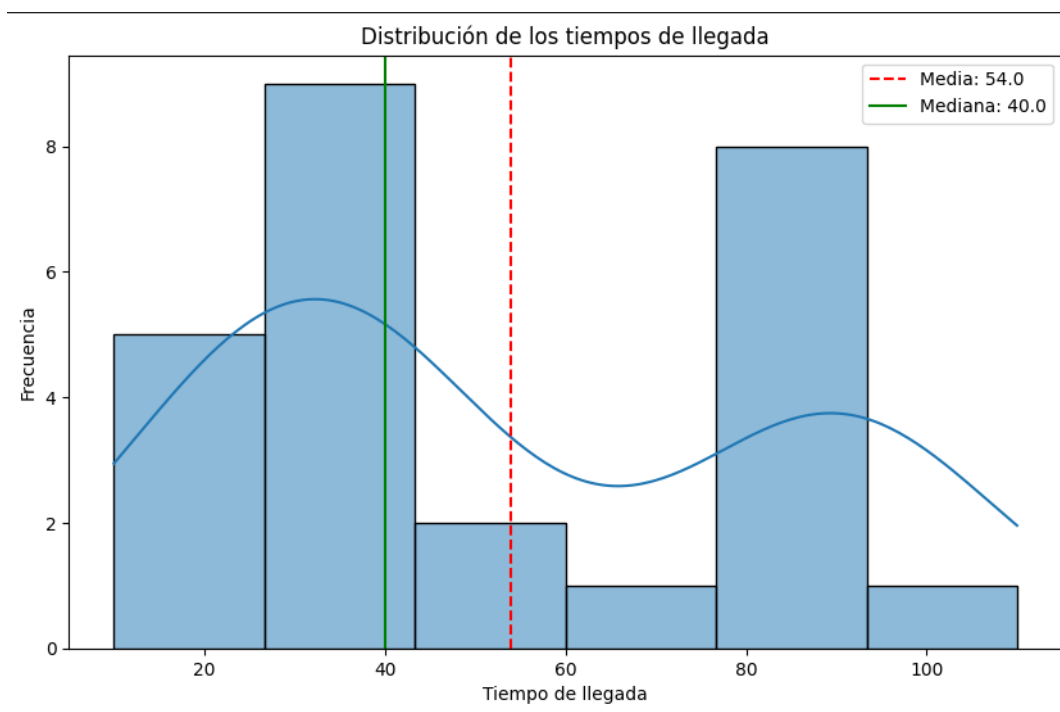
#cuartiles
cuartiles_alturas = np.percentile(alturas, [25, 50, 75])
cuartiles_edades = np.percentile(edades, [25, 50, 75])
cuartiles_tiempos = np.percentile(tiempos_de_llegada, [25, 50, 75])

print("Cuartiles de alturas:", cuartiles_alturas)
print("Cuartiles de edades:", cuartiles_edades)
print("Cuartiles de tiempos de llegada:", cuartiles_tiempos)
```

✓ 0.0s

Cuartiles de alturas: [1.595 1.695 1.75]
Cuartiles de edades: [19. 20. 21.]
Cuartiles de tiempos de llegada: [30. 40. 90.]

h) ¿Son los datos de tiempo de trayecto a la U simétricos?

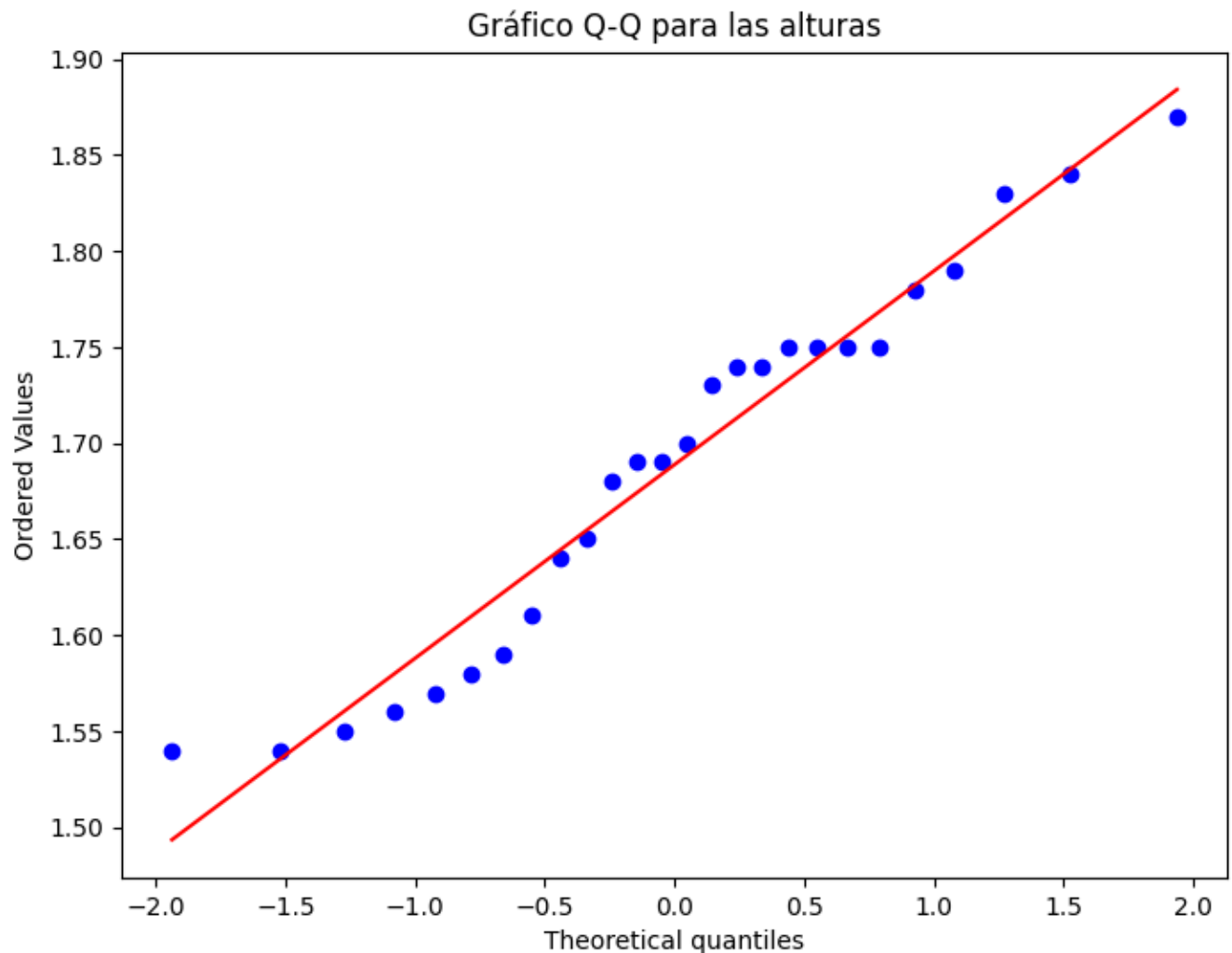


Según la línea de densidad se puede observar como la distribución de los tiempos de llegada está sesgada a la derecha, ya que la mayoría de los datos se encuentran a la izquierda

3. Mencione un tipo de grafica que no se haya hablado en clase con un ejemplo en donde se halla usado

Los gráficos Q-Q son un tipo de grafico echo para comparar los cuantiles de una muestra contra los cuantiles teóricos, si los datos siguen la distribución teórica, los puntos en el grafico se alinearán aproximando a una recta.

Para el caso de la altura:



Se puede observar como los datos siguen una distribución teórica.

No están muy dispersos.

Se puede afirmar que los datos son válidos al seguir esta distribución.