**Ayudantía 1**

**El Viejo Confiable**

**03 de enero 2023**

**Nombre:**

**Parte 1**

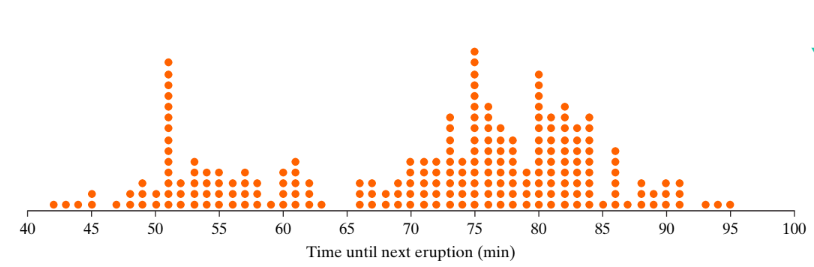
Millones de personas de todo el mundo acuden al Parque de Yellowstone para contemplar las erupciones del géiser “El Viejo Confiable” (Old Faithful). Pero, ¿hasta qué punto es fiel este géiser? ¿Es previsible? ¿Cuánto suele esperar una persona entre dos erupciones?

Supongamos que el encargado le da una predicción para el tiempo de la próxima erupción, y luego esa erupción ocurre cinco minutos después de la hora prevista. ¿Concluiría que las predicciones del encargado no son muy precisas?

Esperamos que no, porque eso sería utilizar pruebas anecdóticas. Para investigar estas cuestiones sobre la fiabilidad del Viejo Confiable es mucho mejor recopilar datos.

Los investigadores recopilaron datos de 222 erupciones del Viejo Confiable en agosto de 1978 y agosto de 1979. La figura 1 contiene un gráfico (llamado gráfico de puntos o dotplot) que muestra ***el tiempo transcurrido hasta la siguiente erupción (en minutos)*** de estas 222 erupciones. Cada punto del gráfico representa una única erupción.

Figura 1: Dotplot del tiempo transcurrido hasta la siguiente erupción.



1. ¿Cuáles son las unidades observacionales y la variable de este estudio?
2. ¿La variable es cuantitativa o categórica?
3. ¿Cuál es el menor tiempo que se registró entre dos erupciones?
4. ¿Cuál es el mayor tiempo que se registró entre dos erupciones?

1. ¿La distribución es simétrica? ¿Tiene forma de montículo?
2. ¿Hay grupos de observaciones?

1. ¿Cómo son esos grupos de observaciones? ¿En torno a qué valores se mueven?
2. ¿Hay observaciones inusuales o atípicas?

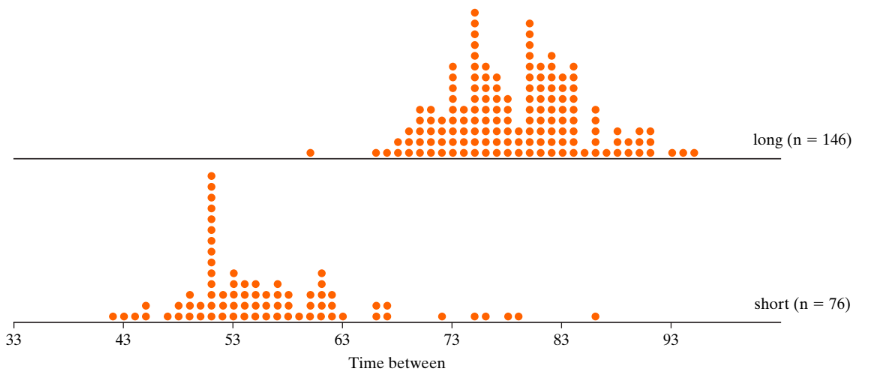
**Parte 2**

¿Cuáles son las posibles explicaciones de la variabilidad de los tiempos? Una idea es que

parte de la variabilidad en los tiempos hasta la siguiente erupción podría explicarse considerando la duración de la erupción anterior. Parece tener sentido que después de una erupción particularmente larga el Viejo Confiable podría necesitar más tiempo para acumular suficiente presión para producir otra erupción. Del mismo modo, después de una erupción más corta, el Viejo Confiable podría estar listo para entrar en erupción de nuevo antes, sin tener que esperar mucho tiempo. Afortunadamente, los investigadores registraron una segunda variable sobre cada erupción: ***la duración de la erupción***, que es otra variable cuantitativa. Para simplificar, podemos clasificar la duración de cada erupción como corta (menos de 3,5 minutos) o larga (3,5 minutos o más), una variable categórica.

En la figura 2 se muestran los gráficos de puntos de las ***distribuciones de tiempo hasta la siguiente erupción*** para las erupciones cortas y largas por separado.

Figura 2: Dotplots del tiempo transcurrido hasta la siguiente erupción para erupciones largas (long) y cortas (short)



1. ¿Cuántas erupciones fueron clasificadas como largas y cuántas cómo cortas?

1. Aproximadamente, ¿cuánto tiempo demora la siguiente erupción cuando las erupciones son largas?

1. Aproximadamente, ¿cuánto tiempo demora la siguiente erupción cuando las erupciones son cortas?

1. ¿Cómo se comparan las formas de las distribuciones de datos respecto a?
   1. Forma
   2. Centros de la distribución
   3. Valores inusuales
   4. Variabilidad

**Parte 3**

Una forma de medir el centro de una distribución es con la ***media***, también llamada ***promedio***. Una forma de medir la variabilidad es con la ***desviación estándar***, que puede interpretarse aproximadamente como la distancia típica (en valor absoluto) entre los valores de los datos y la media de la distribución. Por lo tanto, cuanto más variado o disperso sea un conjunto de datos, mayor será su desviación típica. Estos valores para el tiempo hasta la siguiente erupción, tanto para la distribución global como para las erupciones cortas y largas por separado, se dan en la Tabla 1.

Tabla 1: Medias y desviaciones estándar de la variable ***tiempo transcurrido hasta la siguiente erupción***

|  | Media | Desviación Estándar |
| --- | --- | --- |
| Todas las erupciones | 71.0 min | 12.8 min |
| Erupciones cortas | 56.3 min | 8.5 min |
| Erupciones largas | 78.7 min | 6.3 min |

1. Considerando la Figura 2 y la Tabla 1 ¿le parecen coherentes los resultados? ¿por qué?
2. ¿Qué aprendimos sobre el Géiser “El Viejo Confiable”?

**Terminología Básica**

De este ejemplo deberías haber aprendido que un gráfico como el diagrama de puntos puede mostrar la distribución de una variable cuantitativa. Algunos aspectos a buscar en esa distribución son:

* ***Forma***: ¿La distribución es simétrica? ¿Tiene forma de montículo? ¿Hay varios picos o grupos?
* ***En el centro***: ¿Dónde se centra la distribución? ¿Cuál es un valor típico?
* ***Variabilidad***: ¿Cuál es la dispersión de los datos? ¿Están la mayoría dentro de un determinado rango de valores?
* **Observaciones inusuales:** ¿Hay ***valores atípicos*** que se desvían notablemente del patrón general de los demás valores de datos? Si es así, identifícalos para ver si puedes explicar por qué esas observaciones son tan diferentes. ¿Hay otras características inusuales en la distribución?

Nota:

Hay una webcam en directo de El Viejo Confiable y los géiseres circundantes en:

