Poll 2 - Data Visualization, Histograms, Measures of Center, CLT

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on October 11, 2018

This live poll was conducted on September 24, 2018. This is meant to review some of the main concepts we have learned up untill now. Correct answers are indicated by check marks. Number of votes and number of participants are indicated in the figure legend. For some questions, several selections were allowed.

Descriptive stats | Confidence Interval | p-value | Gaussian distribution | CLT | Simple linear regression

1. Which of the following aesthetics can be used to represent continuous data (select all that apply)?

1. position **(Correct)**
2. shape
3. size **(Correct)**
4. color **(Correct)**
5. line width **(Correct)**
6. line type

2. Which of the following statement characterizes a scale (select all that apply)?

1. Defines a unique mapping between data & aesthetics **(Correct)**
2. Must be 1-to-many between data & aesthetics
3. Must be 1-to-1 between data & aesthetics **(Correct)**

3. Which color palette would you use to distinguish groups of data from each other (select all that apply)?

1. Sequential (Brewer palette)
2. Qualitative (Brewer palette) **(Correct)**
3. Diverging (Brewer palette)
4. Viridis

4. Which color palette would you use to represent continuous data values, such as income, temperature, or speed (select all that apply)?

1. Sequential (Brewer palette) **(Correct)**
2. Qualitative (Brewer palette)
3. Diverging (Brewer palette) **(Correct)**
4. Viridis **(Correct)**

5. The area under a frequency histogram must equal to 1

1. TRUE
2. FALSE **(Correct)**

6. A boxplot can show whether a data set is

1. symmetric
2. skewed
3. symmetric and skewed **(Correct)**

7. If one side of the box is longer than the other, it means that side contains more data.

1. TRUE
2. FALSE **(Correct)**

8. To construct a boxplot, we need (select all that apply)

1. Interquartile range
2. Minimum **(Correct)**
3. Maximum **(Correct)**
4. Standard deviation
5. Mean
6. Median **(Correct)**
7. Mode
8. Skewness
9. 1st quartile (Q1) **(Correct)**
10. 3rd quartile (Q3) **(Correct)**

9. In a distribution with a long left tail, the mean is

1. greater than the median
2. less than the median **(Correct)**
3. equal to the median

10. Which of the following are true concerning a parameter (select all that apply)?

1. A numerical constant **(Correct)**
2. Pertains to a population **(Correct)**
3. Is unknown **(Correct)**
4. is known
5. Is a statistic
6. $\bar{y}$ and p (the sample proportion) are parameters

11. The standard error of the mean (SEM) describes (select all that apply)

1. How far $\bar{y}$ could typically deviate from μ **(Correct)**
2. How far an individual y typically deviates from μ or from $\bar{y}$

12. The Central Limit Theorem states that (select all that apply)

1. $\bar{y} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$
2. $\bar{y} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$ for large enough n
3. $\bar{y} \sim \mathcal{N}(\mu, \sigma/\sqrt{n})$ for large enough n and finite variance **(Correct)**
4. The sampling distribution of $\bar{y}$ is, for a large enough n and finite variance, close to Gaussian in shape no matter what the shape of the distribution of individual Y values **(Correct)**