

Always, Sometimes, or Never (15 points)

(15 points) The following statements are always, sometimes, or never true. Mark each statement with Always, Sometimes, or Never, depending on the truth of each statement.

1. A categorical variable is a nominal variable. [Always](#)
2. A dataset without any prominent peaks is known as multimodal. [Sometimes](#)
3. The residual for a point above the line of best fit is positive. [Always](#)
4. Statistics are characteristics of a population. [Never](#)
5. Two variables X and Y have a correlation near 0. X and Y have no relationship. [Sometimes](#)
6. The intercept describes the expected value of the response variable when the explanatory variable is equal to 0. [Always](#)
7. Two variables X and Y are linearly related with negative correlation. Large values of X are associated with smaller values of Y . [Always](#)
8. The value of $P(3)$ in the following table is unknown. The table contains a valid probability distribution. [Always](#)

X	0	1	2	3
$P(X)$	0.5	0.2	0.2	$P(3)$

9. For two events A and B , $P(A \text{ and } B) = P(A) \times P(B)$. [Sometimes](#)
10. The mean is resistant to outliers or extreme values. [Never](#)
11. For two events A and B with $P(A) = 0.8$ and $P(B) = 0.1$, $P(A \text{ and } B) = 0$. [Never](#)
12. A discrete variable is a numerical variable. [Always](#)
13. The value of $P(3)$ in the following table is unknown. The table contains a valid probability distribution. [Never](#)

X	0	1	2	3
$P(X)$	0.2	0.6	0.3	$P(3)$

14. A categorical variable is a continuous variable. [Never](#)
15. In a right-skewed dataset, the median will be greater than the mean. [Never](#)

Matching

16. Label each dataset with the most appropriate type of graph. The types may be used once, more than once, or not at all. (6 points)

- The weight of a hardcover book compared to the total word count inside for Wikipedia's list of the 100 longest novels ever published. [Scatterplot](#)
- The times 50 readers take to complete an unabridged copy of Les Miserables by Victor Hugo (1,456 pages). [Box Plots](#)
- Preferred genre of book (historical fiction, memoir, science fiction, thriller) for members of a local book club. [Bar Chart](#)
- The original publication languages of the books on Wikipedia's list of the 100 longest novels ever published (Tamil, Urdu, English, French...) [Bar Chart](#)
- Pages per book for a sample of historical fiction books, memoirs, science fiction books, and thrillers. [Bar chart](#)
- Amounts of money spent by customers at a local independent bookshop. [Histogram](#)

a. Histogram b. Box Plots c. Scatterplot d. Bar Chart

17. A librarian is interested in estimating the amount of books read per year by library card holders. Since card holders might buy or be gifted books in addition to checking them out from the library, they decide to take a sample of card holders and ask them directly rather than relying on their records. Label each scenario with the most appropriate type of sampling. The types may be used once, more than once, or not at all. (4 points)

- The librarian randomly selects library card numbers to form their sample, and reaches out to everyone included, promising a coupon to a local coffee shop upon survey completion. [Simple Random Sampling](#)
- The librarian suspects that card holders with children might read more books per year (since children's books tend to be shorter). They sort card holders into groups based on the number of children's books checked out in the last year (0, 1-10, 11-20, 21-30, 30+), and randomly sample within each group so that the sample is representative of the population of library card holders. All sample members are promised a coupon to the local coffee shop if they complete the survey. [Cluster Sampling](#)
- The librarian randomly selects five days over the next month. During each of those days, all card holders entering the library are asked how many books they read per year. When they complete the survey, the card holders are given coupons to the local coffee shop. [Multistage Sampling](#)
- The librarian posts a sign near the check out desk, promising a coupon to a local coffee shop if card holders fill out the survey. [Convenience Sampling](#)

a. Simple Random Sampling b. Stratified Sampling c. Cluster Sampling
d. Multistage Sampling e. Convenience Sampling

Test 1

Stat 140
Fall 2021

18. Match the notation given below to the correct name. The names may be used once, more than once, or not at all. (5 points)

- | | | |
|---|---|--------------------------------------|
| • s Sample Standard deviation | • $\hat{p}$ Sample Proportion | • b_1 Sample Slope |
| • β_0 Sample Intercept | • μ Population Mean | |
-
- | | | |
|----------------------------------|-------------------------|------------------------------|
| a. Sample Mean | b. Population Mean | c. Sample Standard Deviation |
| d. Population Standard Deviation | e. Sample Proportion | f. Population Proportion |
| g. Sample Intercept | h. Population Intercept | i. Sample Slope |
| j. Population Slope | k. Sample Correlation | l. Population Correlation |

19. A team of researchers is collecting data to see if readers have higher emotional empathy levels than non-readers.

Scenario A: The researchers enroll a sample of 150 grade school children. Each child is asked to keep a reading log for 8 weeks. At the end of the study, the children take a test to measure empathy levels.

Scenario B: The researchers enroll a sample of 150 grade school children. 50 randomly selected children are told to behave as normal, 50 randomly selected children are told to read an additional 10 minutes per night, and 50 randomly selected children are told to read an additional 30 minutes per night. Children take a test after 8 weeks to measure empathy levels.

Scenario C: One of the researchers noticed that one of their children, who reads regularly, tends to be more sensitive, and another child, who prefers physical play, to be less empathetic.

Label each description with the *most appropriate* vocabulary word, which may be used once, more than once, or not at all. (10 points)

- The amount of additional reading per night in Scenario B. [Levels](#)
 - Scenario B. [Open Label](#)
 - The attitude towards reading of the parents of the study subjects in Scenario A. [Lurking Variable](#)
 - Scenario A. [Observational study](#)
 - The type of reading a study subject is doing (for example, non-fiction vs. fiction) in Scenario A. [Treatment Groups](#)
 - Normal behavior, 10 minutes per night, 30 minutes per night in Scenario B. [Levels](#)
 - The groups assigned to behave as normal in Scenario B. [Control Group](#)
 - The groups assigned to read an additional 10 or 30 minutes per night in Scenario B. [Treatment Groups](#)
 - Empathy scores for each study subject, measured using a commonly used test. [Response variable](#)
 - Scenario C. [Anecdotal Evidence](#)
-
- | | | | |
|------------------------|-----------------|-------------------------|----------------------|
| a. Observational Study | b. Experiment | c. Anecdotal Evidence | d. Control Group |
| e. Treatment Group(s) | f. Placebo | g. Explanatory Variable | h. Response Variable |
| i. Lurking Variable | j. Double-Blind | k. Open Label | l. Levels |