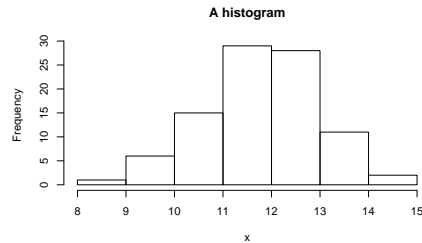


- (1) The Monte Vista Bird Refuge in the San Luis Valley is a stop in the migration path of a population of Sandhill Cranes. You would like to build a database in order to more effectively monitor this population.
- (a) Describe a data matrix that you might collect.
 - (i) What are three possible variables in the data that you might collect?
 - (ii) What are the individuals/observations?
 - (b) For each of your three variables: is it categorical or numerical?
 - (c) Give a question that you could ask about a possible correlation between two variables in these data.
 - (d) For your question directly above, what are the explanatory and response variables?
- (2) Researchers randomly assigned 72 chronic users of cocaine into three groups: desipramine (antidepressant), lithium (standard treatment for cocaine) and placebo. Results of the study are summarized below.

	no relapse relapse		total
desipramine	10	14	24
lithium	18	6	24
placebo	20	4	24
total	48	24	72

- (a) What percentage of the patients relapsed in each of the three treatment groups (desipramine, lithium, and placebo)?
 - (b) At first glance, does it appear that desipramine is more effective than a placebo for preventing relapse? Is the standard lithium treatment more effective than a placebo? Is desipramine more effective than lithium?
 - (c) Is it possible to conclude from the sample proportions alone that there is a significant difference in the efficacy of the treatments?
- (3) The values 38 53 41 55 56 61 62 48 43 47 56 65 19 61 32 105 are empathy scores recorded for 16 participants in a study relating empathy to certain types of brain activity.
- (a) Find the
 - (i) mean
 - (ii) median
 - (iii) standard deviation
 - (iv) Q1
 - (v) Q3
 - (vi) IQR
 - (b) Draw a boxplot of the data. Your plot should represent the 5 number summary, as well as any outliers falling more than $1.5 \times IQR$ away from the median.
- (4) Under what conditions is it appropriate to summarize data with the mean and standard deviation? When should you use median and quartiles?

- (5) Consider the histogram shown below. Which statistic would you use to report the center of the distribution? Which statistic would you use to report the spread? Why?



- (6) A study of characters in 100 top grossing American films from 2007–2009 examined 4,342 speaking characters for gender. The researchers found that 32.8% of the speaking characters are female. They also found that when the director of the film is female, 61.2% of the speaking characters are female. The study concludes that characters in films with a female director are more likely to be female. Identify the population of interest and the sample. Can the results of the study be generalized to the population? Can this study be used to establish a causal relationship?
- (7) In a study of prairie dog communities the number of young prairie dogs in each of 20 burrows was counted. The data are 7, 5, 3, 10, 2, 5, 13, 3, 12, 4, 4, 1, 9, 7, 5, 6, 3, 5, 12, 1
- What percent of the burrows had fewer than 8 young?
 - What percent of the burrows had 10 or more young?
 - Make a dotplot or histogram of the data.
 - Describe the distribution.
- (8) A comprehensive survey by the college reports that the true proportion of all students at the college who use drugs is 0.3. You survey 100 students in your dorm and record that the proportion of students who use drugs is 0.15. The proportion of all students at this college who use drugs is a _____ and the proportion of students who use drugs in your dorm is a _____.
- statistic; parameter
 - parameter; statistic
 - population; sample
 - measure of central tendency, measure of variability
 - none of the above

- (9) This table represents the first 8 observations from a sample of 200 individuals, who reported their age, race, income, and job satisfaction score on a scale from 0 to 100.

Age	Race	Income	Score
21	W	less than \$10,000	29
33	B	\$20,000-23,000	32
41	B	more than \$100,000	84
26	A	\$30,000-40,000	78
22	O	\$10,000-20,000	87
19	A	\$40,000-50,000	42
34	W	\$50,000-60,000	21
26	W	less than \$10,000	91
⋮	⋮	⋮	⋮

- (a) Which of the following best describes the **Income** variable?
- (i) categorical
 - (ii) geographic
 - (iii) numerical
 - (iv) logical
 - (v) observational
- (b) Which type of plot would be most useful for visualizing the relationship between **Age** and job satisfaction **Score**?
- (i) histogram
 - (ii) single box plot
 - (iii) side by side box plot
 - (iv) scatter plot
 - (v) dot plot
- (c) Below are some summary statistics from the **score** variable. Which of the following is true?
- | min | Q1 | median | Q3 | max | mean | sd | n | missing |
|-----|----|--------|----|-----|--------|----------|-----|---------|
| 30 | 57 | 69.5 | 77 | 99 | 65.075 | 16.09361 | 200 | 0 |
- (i) the standard deviation estimate is not possible because **score** is a whole number
 - (ii) there is evidence that the distribution of **score** is right-skewed
 - (iii) the minimum value of 30 would be identified as out outlier in a box plot
 - (iv) there were more survey respondents who reported job satisfaction scores less than 57 than survey respondents reported job satisfaction scores greater than 77
 - (v) none of the above are true

- (10) A political scientist is interested in the effect of government type on economic development. She wants to use a sample of 30 countries evenly represented among the Americas, Europe, Asia, and Africa to conduct her analysis. What type of study should she use to ensure that countries are selected from each region of the world?
- (a) Observational - simple random sample
 - (b) Observational - cluster
 - (c) Observational - stratified
 - (d) Experimental

- (11) A researcher is interested in seeing if there is an association between whether or not an individual uses a smart phone right before bed and how well an individual sleeps. Participants in the study report whether or not they used a smart phone before bed and then rate their quality of sleep as either “very poor,” “poor,” “average,” “good,” or “very good.” The researcher concludes at the end of the study that there is an association between the two variables. Which of the following statements are true?
- (a) The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can say the response and explanatory variables are independent.
 - (b) The explanatory variable is sleep quality and the response variable is whether or not a participant uses a smart phone. An association was present, so the researcher can say the response and explanatory variables are not independent.
 - (c) The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can say the response and explanatory variables are not independent.
 - (d) The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can determine that smart phone use right before bed causes change in quality of sleep.
 - (e) The explanatory variable is sleep quality and the response variable is whether or not a participant uses a smart phone. An association was present, so the researcher can determine that smart phone use right before bed causes change in quality of sleep.
- (12) The table below describes residents of an Atlanta neighborhood based on their car ownership and public transportation usage.

	Owns car	Does not own car	Total
Uses public transport	34	94	128
Does not use public transport	126	17	143
Total	160	111	271

- (a) What is the probability that a randomly selected resident of this neighborhood both owns a car and uses public transport?
 - (i) 0.125
 - (ii) 0.153
 - (iii) 0.213
 - (iv) 0.266
- (b) Which proportions should we examine if we want to compare the proportion of individuals who use public transport among those who do and do not own a car?
 - (i) $34/128$ vs $126/143$
 - (ii) $160/271$ vs $128/271$
 - (iii) $34/271$ vs $126/271$
 - (iv) $34/160$ vs $94/111$
 - (v) none of the above
- (c) Is owning a car independent of using public transportation?
 - (i) Yes, because $P(\text{uses public transit} \mid \text{owns car}) = P(\text{uses public transit})$
 - (ii) Yes, because $P(\text{owns car AND uses public transit}) = P(\text{uses public transit}) \times P(\text{owns car})$
 - (iii) No, because $P(\text{uses public transit} \mid \text{owns car}) = P(\text{uses public transit})$
 - (iv) No, because $P(\text{owns car AND uses public transit}) \neq P(\text{uses public transit}) \times P(\text{owns car})$
 - (v) Both (iii) and (iv)

- (13) A researcher would like to study the effect of eating breakfast on a cognitive function. Volunteers are recruited through the study by posting flyers on campus. He randomly assigns subjects to two groups, one told to eat before participating in the study and one asked to eat breakfast following the study, however, he suspects whether or not the person typically eats breakfast affects this relationship (their typical breakfast habits). In order to address this, what should he do prior to assigning subjects to experimental groups?
- Cluster on typical breakfast habits.
 - Randomly assign subjects to typical breakfast habits and do a multifactor experiment.
 - Sample from each strata, typical breakfast eater and not.
 - Block on typical breakfast habits.
- (14) The plot below displays the distribution of the percent of days spent sleeping by male fruit flies. Which of the following are valid estimates of the mean and median of this distribution?



- mean = 24, median = 18
 - mean = 18, median = 24
 - mean = 18, median = 18
 - mean = 20, median = 40
 - mean = 25, median = 35
- (15) A disease has a prevalence of 1%. A test for the disease gives a positive result 94% of the time when administered to an infected patient. The test gives a negative result 98% of the time when administered to a healthy patient. Draw a tree diagram representing these given probabilities, and fill in the four joint probabilities on the leaves. What is the probability that a patient with a positive test result is infected with the disease?
- (16) A gambler plays a game where she must pay \$10 to play. She has a 50% chance of not winning anything, a 25% chance of winning her \$10 back, a 20% chance of winning \$20, and a 5% chance of winning \$100. What is the expected value of playing the game? What should she expect to win if she plays 100 games?