# **AP Statistics Review for Midterm Part 1 (Chapters 1 – 5)**

# **Chapter 1**

In the paper “Reproduction in Laboratory colonies of Bank Vole,” the authors presented the results of a study of litter size. (A vole is a small rodent with a stout body, blunt nose, and short ears.) As each new litter was born, the number of babies was recorded, and the accompanying results were obtained.

1 4 4 5 5 6 6 7 7 8

2 4 5 5 5 6 6 7 7 8

3 4 5 5 6 6 7 7 8 9

3 4 5 5 6 6 7 7 8 10

4 4 5 5 6 6 7 7 8 11

The authors also kept track of the color of the first born in each litter. (B = brown, G = gray, W = white, and T = tan)

B B T W T G G G B B

W B W B T T G B T B

G B B B B G W G T G

B B B B G G T T W G

B G T W B G T W G W

1. Which variable, litter size or color, is categorical?
2. Which variable is quantitative?
3. Make a bar chart of the colors.
4. Make a histogram of the litter sizes.
5. Make a dotplot of the litter sizes.
6. Describe the shape of the distribution of litter sizes.
7. Find the mean of the litter sizes.
8. Is the mean resistant to outliers?
9. Find the median of the litter sizes.
10. Is the median resistant to outliers?
11. Find the 5-number summary of the litter sizes.
12. What is the interquartile range?
13. Are there any outliers in the litter sizes (show work using 1.5 IQR)?
14. Make a boxplot and modified box plot (if there are outliers) of the litter sizes.
15. Use the calculator to find the variance of the litter sizes.
16. Find the standard deviation of the litter sizes, using the calculator.
17. Is standard deviation resistant to outliers?
18. Find the degrees of freedom of the litter sizes.
19. Make a **back-to-back split** stemplot of the following Reading Scores:

4th Graders 12 15 18 20 20 22 25 26 28 29

31 32 35 35 35 36 37 39 40 42

7th Graders 1 12 15 18 18 20 23 23 24 25

27 28 30 30 31 33 33 33 35 36

1. Make a comparison between 4th grade and 7th grade reading scores based on your stemplot.

**Chapter 2**

1. What is the area under any density curve?
2. The (mean or median) of a density curve is the equal-areas point, the point that divides the area under the curve in half.
3. The (mean or median) of a density curve is the balance point, at which the curve would balance if made of solid material.
4. If a density curve is skewed to the right, the (mean or median) will be further to the right than the (mean or median).
5. What is the difference between x̅ and µ ?
6. What is the difference between s and σ ?
7. Name 3 properties of normal curves are density curves:
8. How do you find the inflection points on a normal curve?
9. Sketch the graph of N(266, 16), the distribution of pregnancy length from conception to birth for humans.
10. What is the 68-95-99.7 rule?
11. Using the empirical rule (the 68-95-99.7 rule), find the length of the longest 16% of all pregnancies. Sketch and shade a normal curve for this situation.
12. Using the empirical rule, find the length of the middle 99.7% of all pregnancies. Sketch and shade.
13. Using the empirical rule, find the length of the shortest 2.5% of all pregnancies. Sketch and shade.
14. Using the empirical rule, what percentile rank is a pregnancy of 218 days?

Use the information from the distribution of pregnancy length N(266, 16), the Z formula and the Z table to answer questions 33 – 43.

1. What percentile rank is a pregnancy of 298 days?
2. What percentile is a pregnancy of 250 days?
3. What is the percentile of a pregnancy of 266 days?
4. What z-score does a pregnancy of 279 days have?
5. What percent of humans have a pregnancy lasting less than 279 days? Sketch and shade a normal curve.
6. What z-score does a pregnancy of 257 days have?
7. What percent of humans have a pregnancy lasting less than 257 days? Sketch and shade.
8. What percent of humans have a pregnancy lasting longer than 280 days? Sketch and shade.
9. What percent of humans have a pregnancy lasting between 260 and 270 days? Sketch and shade.
10. Would you say pregnancy length is a continuous or discrete variable? Justify.
11. You have normal distributions on your TI-83. Use these functions to check your answers to problems 37 to 41.
12. How long would a pregnancy have to last to be in the longest 10% of all pregnancies?
13. How short would a pregnancy be to be in the shortest 25% of all pregnancies?
14. How long would a pregnancy be to be in the middle fifth of all pregnancies?

#### Chapter 3

1. Graph the following hot dog data:

Calories Sodium (milligrams)

108 149

130 350

132 345

1. 360
2. 360

140 375

1. 380
2. 390
3. 400

163 415

1. 400
2. 420
3. 450
4. 500
5. 505

195 500

200 515

1. What is the explanatory variable?
2. What is the response variable?
3. Describe the association of the scatterplot (strength, direction, and form)
4. Are there outliers? (Outliers in a scatterplot have large residuals.)
5. If there are outliers, are they influential?
6. Use your calculator to compute the correlation.
7. What percent of the variation of sodium is explained by the LSRL on calories?
8. What two things does correlation tell us about a scatterplot?
9. If I change the units on sodium to grams instead of milligrams, what happens to the correlation?
10. What is the highest correlation possible?
11. What is the lowest correlation possible?
12. Correlation only applies to what type(s) of relationship(s)?
13. Is correlation resistant to outliers?
14. Does a high correlation indicate a strong cause-effect relationship?
15. Sketch a scatterplot with a correlation of about 0.8.
16. Sketch a scatterplot with a correlation of about –0.5.
17. Find the least-squares regression line (LSRL) for the hot dog calories-sodium data.
18. Draw the LSRL on your scatterplot.
19. What is the slope of this line, and what does it tell you in this context?
20. What is the y-intercept of this line, and what does it tell you in this context?
21. Predict the amount of sodium in a hot dog with 155 calories.
22. Predict the amount of sodium in a hot dog with 345 calories.
23. Find the error in prediction (residual) for a hot dog with 180 calories.
24. Find the residual for 195 calories.
25. Sketch the residual plot of calories vs. residuals and use it to determine if the LSRL is a good model for the data:
26. Why is the prediction for 155 calories acceptable but the prediction for 345 calories not?
27. What point is always on the LSRL? Find this point, and label it on your scatterplot.
28. Find the standard deviation of the calories.
29. Find the standard deviation of the sodium.
30. Using the equations for slope and y-intercept, verify the slope and intercept of the LSRL.

#### Chapter 4

1. If you know a scatterplot has a curved shape, how can you decide whether to use a power model or an exponential model to fit data?
2. Sketch a graph the following data:

Time (days) Mice

1. 19
2. 60
3. 195
4. 597
5. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get a transformed linear equation (show a sketch of each scatterplot):
6. Make a residual plot to support your choice.
7. Transform your best LSRL equation back to exponential or power:
8. Check your equation (from above problem) by using your calculator’s power and exponential regression functions
9. Graph the following data:

Diameter (inches) Cost (dollars)

1. 3.50
2. 8.00
3. 14.50
4. 22.50
5. 39.50
6. Perform the appropriate logarithmic transformation (power or exponential) on the above data to get a transformed linear equation (show a sketch of each scatterplot):
7. Make a residual plot to support your choice.
8. Transform your best LSRL equation back to exponential or power:
9. Check your equation (from above problem) by using your calculator’s power and exponential regression functions
10. What is extrapolation, and why shouldn’t we trust predictions using extrapolation?
11. What is interpolation, are these predictions acceptable?
12. What is a lurking variable?
13. Why should we avoid using averaged data for regression and correlation?
14. What is causation? Give an example.
15. What is common response? Give an example.
16. What is confounding? Give an example.
17. What type of data do we put in a two-way table?

Use this table for the next few questions:

 Smoking Status

### Education Never smoked Smoked, but quit Smokes Total

Did not complete high school 82 19 113

Completed high school 97 25 103

1 to 3 years of college 92 49 59

4 or more years of college 86 63 37

Total:

1. Fill in the marginal totals in the table above.
2. Fill in the marginal distributions in the table above
3. What percent of these people smoke?
4. What percent of never-smokers completed high school?
5. What percent of those with 4 or more years of college have quit smoking?
6. What percent of those with some college smoke?
7. What percent of smokers did not finish high school?
8. Create segmented bar graphs for each smoking status
9. Are smoking status and education independent?
10. What is Simpson’s Paradox?
11. What do you look for in a data table that might “suggest” a Simpson’s Paradox is present?

**Chapter 5**

1. What is the difference between an observational study and an experiment?
2. What is a voluntary response sample?
3. How are a population and a sample related but different?
4. Why is convenience sampling biased?
5. SRS stands for what kind of sample? Name and define.
6. Discuss how to choose a SRS of 4 towns from this list:

Allendale Bangor Chelsea Detour Edmonton Fennville

Gratiot Hillsdale Ionia Joliet Kentwood Ludington

1. What is a stratified random sample?
2. What is a multistage sample?
3. What is undercoverage?
4. What is nonresponse?
5. What is response bias?
6. Why is the wording of questions important? Give an example.
7. How are experimental units and subjects similar but different?
8. What is the placebo effect?
9. What is the purpose of a control group?
10. Give an example of when we may not want to use a placebo/control group.
11. What are matched pairs experiments? Give an example.
12. What are the components of a well-designed experiment?
13. What does double-blind mean, and why would we want an experiment to be double-blind?
14. What is block design and how is it different than stratified random samples?
15. I want to test the effects of aerobic exercise on resting heart rate. I want to test two different levels of exercise, 30 minutes 3 times per week and 30 minutes 5 times per week. I have a group of 20 people to test, 10 men and 10 women. I will take heart rates before and after the experiment. Draw a diagram for this experimental design.
16. Why is simulation useful?
17. What are the five steps of a simulation?
18. Design and perform a simulation of how many children a couple must have to get two sons. (A simulation involves many trials. For this simulation, perform 10 trials.). Show your results below:
19. If John makes 70% of his free throw shots, use your calculator to simulate how many shots he will make in 10 trials. Do this simulation 5 times and show your results below: