

AP STATISTICS: SEMESTER 1 REVIEW

Chapter 1

1. What type of graphs would be appropriate to display categorical data?
2. What 3 things must always be mentioned when describing a quantitative distribution?
3. Which is affected by extreme values: mean or median?
4. What is the rule for justifying outliers?
5. What is the difference between a frequency distribution and a relative frequency distribution?

Chapter 2

6. What is the area under a density curve?
7. The (mean or median) of a density curve is the equal-areas point, the point that divides the area under the curve in half.
8. The (mean or median) of a density curve is the balance point, at which the curve would balance if made of solid material.
9. If a density curve is skewed to the right, the (mean or median) will be further to the right than the (mean or median).
10. Identify what each variable represents: \bar{x} , μ , s , and σ ?
11. The distribution of pregnancy length from conception to birth for humans follows a Normal $N(266, 16)$.
 - a. Sketch the graph.
 - b. What is the 68-95-99.7 rule?
 - c. What z-score does a pregnancy of 257 days have?
 - d. What percent of humans have a pregnancy lasting less than 257 days?
 - e. What percent of humans have a pregnancy lasting longer than 280 days?
 - f. What percent of humans have a pregnancy lasting between 260 and 270 days?
 - g. How long would a pregnancy have to last to be in the longest 10% of all pregnancies?

Chapter 3

12. The scatterplot shows calories vs sodium (mg) for hotdogs.

Sodium (mg)



calories

- a. What is the response variable?
 - b. What is the explanatory variable?
 - c. Describe the strength and direction?
 - d. Are there outliers? If so, are they influential?
13. A linear regression was performed: $\text{sodium} = -85.41 + 3.11(\text{calories})$; $r = .9195$, $R^2 = .845$
- a. What is the slope of this line, and what does it tell you in this context?
 - b. Predict the amount of sodium in a hot dog with 155 calories.
 - c. When is a residual negative?
 - d. Interpret r and r^2 .
 - e. If I change the units on sodium to grams instead of milligrams, what happens to the correlation?
 - f. If I switch the explanatory and response variables, what happens to the correlation?
14. What two things does correlation tell us about a scatter plot?
15. What is the highest correlation possible?

16. What is the lowest correlation possible?
17. Is correlation resistant to outliers?
18. Does a high correlation indicate a strong cause-effect relationship?
19. How can you use a residual plot to tell if a line is a good model for data?
21. What is a lurking variable?

Chapter 5

22. Why is a two-way table called a two-way table?
23. The table shows smoking status and the highest education level completed by a random sample of adults.

Education	Smoking Status		
	Never smoked	Smoked, but quit	Smokes
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

- a. Fill in the marginal distributions for this table.
- b. What percent of these people smoke?
- c. What % of never-smokers completed high school?
- d. What % of those with 4 or more years of college have quit smoking?
- e. What % of smokers did not finish high school?
- f. What conclusion can be drawn about smoking and education from this table?

Chapter 4

24. What is the difference between an observational study and an experiment?
25. What is a voluntary response sample and why is it biased?
26. How are a population and a sample related but different?
27. Name and define a SRS.
28. What is a stratified random sample?
29. What is a cluster sample?
30. What is undercoverage?
31. What is nonresponse?
32. What is response bias?
33. Biased is reduced by _____, and Variability is reduced by _____.
34. Explanatory variables in experiments are often called _____.
35. If I test a drug at 100 mg, 200 mg, and 300 mg, I am testing one variable at three _____.
36. What is the placebo effect?
37. What is the purpose of a control group?
38. What are the 3 principles of experimental design?
39. What does double-blind mean?
40. What is block design and why is it important?

Chapter 5

41. What makes events independent?
42. You are going to flip a coin 3 times and note how many heads you get. What is the sample space?
43. You are going to flip a coin 3 times and note what you get on each flip. What is the sample space?
44. How many different four-digit numbers can you make?
45. How many different four-digit numbers can you make without repeating digits?
46. Any probability is a number between (and including) _____ and _____.
47. If S is the sample space, $P(S) = \underline{\hspace{1cm}}$.
48. What are complementary events?
49. What are disjoint events?
50. Beth can beat Erica in tennis 9% of the time. Erica can swim faster than Beth 8% of the time.
 - a. What is the probability that Beth would beat Erica in a tennis match and in a swimming race?
 - b. What assumption are you making in (a)? Do you think this assumption is valid?
52. What is the union of two events?
53. What is an intersection of two events?
54. Perform an independence test on the smoking/education chart from problem #25 to show that smoking status and education are not independent.
55. Make a Venn diagram for the following situation: Safety engineers must determine whether industrial workers can operate a machine's emergency shutoff device. Among a group of test subjects, 66% were successful with their left hands, 82% with their right hands, and 51% with both hands.
 - a. What percent of these workers could not operate the switch with either hand?
 - b. If the worker can operate the switch with their right hand, what is the probability that they can operate the switch with their left hand?
56. Consider the process of a drawing a card from a standard deck and replacing it. Let A be drawing a heart, B be drawing a king, and C be drawing a spade.
 - a. Are the events A and B disjoint? Explain.
 - b. Are the events A and B independent? Explain.
 - c. Are the events A and C disjoint? Explain.
 - d. Are the events A and C independent? Explain.
57. What does the symbol \cup mean?
58. What does the symbol \cap mean?

Chapter 6

59. Given the probability distribution:
 - a. what is $P(X > 2)$?
 - b. what is $P(X \geq 2)$?
 - c. Find the mean and SD.

Grade	0	1	2	3	4
Probability	0.05	0.28	0.19	0.32	0.16

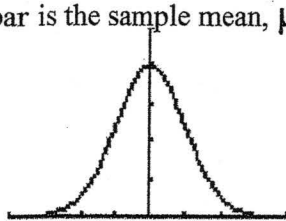
60. Normal distributions are (continuous or discrete).
61. Expected value is another name for _____.

62. What is the law of large numbers?
63. Before you can use the rules for variances you must make sure the variables are _____.
64. Consider the scores for 2 different tests in a statistics class. For Test 1, the class average was 80 with a standard deviation of 10. For Test 2, the class average was 70 with a standard deviation of 12.
- What is the mean combined score for the tests?
 - What is the sd for the combined scores?
 - What is the difference in the mean of the tests?
 - What is the sd for the difference of the tests?
 - If I cut the test scores on Test 2 in half and add 50, what is the new mean and SD?
65. What are the four conditions of a binomial distribution?
66. How is a geometric distribution different from a binomial distribution?
67. The probability that a child born to a certain set of parents will have blood type AB is 25%. The parents have four children. X is the number of those children with blood type AB.
- Is this binomial or geometric?
 - Find the mean and SD.
 - find $P(X = 2)$
 - find $P(X \geq 1)$
68. Parents continue having children until they have a child with type AB blood. X is the number of children they have to give birth to in order to have one child with type AB blood.
- Is this binomial or geometric?
 - Find the mean.
 - find $P(X = 1)$
 - find $P(X > 5)$

AP Statistics: Exam Review Answer Key

1. pie graph, bar graph
2. center, shape, spread
3. mean
4. lower fence: $Q1 - 1.5(IQR)$ upper fence: $Q3 + 1.5(IQR)$ Any value beyond fences is an outlier
5. a frequency distribution shows counts and relative frequency shows proportions.
6. area = 1
7. median
8. mean
9. mean, median
10. \bar{x} is the sample mean, μ = population mean, s = sample SD, σ = population SD

11. a.



- b. About 68% of the observations will fall within one sd of the mean.
About 95% of the observations will fall within two sd of the mean.
About 99.7% of the observations will fall within three sd of the mean

218 234 250 266 282 298 314

- c. -0.5625 d. 28.69% e. 19.08% f. 24.49% g. 286.5 days
12. a. sodium b. calories c. positive and strong linear d. yes, the point in the lower left, if it was removed the slope of the LSRL would decrease and correlation would increase.
 13. a. slope = 3.11: for each increase in 1 calorie, the sodium is predicted to increase by 3.11mg
b. 395.64 mg c. a residual is negative when the predicted y is larger than the actual y
d. $r = .9195$ which indicates a strong, positive linear relationship, $R^2 = .845$ 84.5% of the variation in sodium can be explained by the change in calories. e. correlation stays the same
f. correlation stays the same = 0.9195
 14. the strength and direction of a linear relationship
 15. 1 or -1
 16. 0
 17. no, not resistant to outliers
 18. no, correlation does not necessarily imply causation
 19. The residuals should be randomly scattered and relatively close to zero.
 21. A lurking var. is a variable that may influence values of the variables in a study, but it is not part of the study.
 22. There are two variables.
 - 23 a. Marginal distributions are the totals for each row and column.
b. $312/825 = 37.82\%$ c. $97/357 = 27.17\%$ d. $63/186 = 33.87\%$ e. $113/312 = 36.22\%$
f. The more education a person has completed, the less likely they are to smoke; 53% of those who did not complete high school smoke, 45% of those who completed high school smoke, 30% of those with 1-3 years of college smoke, and 20% of those with 4 or more years of college smoke.
 24. In an experiment, a treatment is imposed.
 25. The subjects select themselves to be in the sample. Usually people with strong feelings (negative) will want to be in the sample.
 26. A sample is a subset of the population.
 27. Simple Random Sample: each member of the population is equally likely to be selected and each possible sample is equally likely to be selected.
 28. The population is divided in to groups, then a random sample is chosen from each group.
 29. The population is divided into groups. Certain groups are chosen randomly and all members of the chosen groups are sampled.
 30. One or more groups with similar characteristics do not have a chance to be chosen for the sample.
 31. A subject chosen to be in the sample does not respond or refused to participate.
 32. The subject is influenced to respond a certain way (by the wording of the question, interviewer etc)

33. Bias is reduced by random sampling, variability is reduced by large samples
34. Factors
35. Levels
36. Some individuals respond to any form of treatment, regardless of whether it is "real" or "fake".
37. To reduce or eliminate the effects of lurking variables.
38. Control the effects of lurking variables by comparing several treatments.
Randomly assign subjects to treatment groups.
Replicate the experiment on many subjects to reduce chance variation.
39. In a double-blind experiment, neither the subjects nor the persons recording the observations knows which subject received which treatment. This reduces bias.
40. A block design divides the sample into groups of similar characteristics to reduce the effects of lurking variables. Within each group, units/subjects are randomly assigned to each of the treatment groups.
41. Two events are independent if knowing that one occurs does not change the probability that the other occurs. $P(A | B) = P(A)$ and $P(B | A) = P(B)$
42. $S = \{0,1,2,3\}$
43. $S = \{ HHH HHT HTH HTT THH THT TTH TTT \}$
44. 10^4
45. $(10)(9)(8)(7) = 5050$
46. 0, 1
47. 1
48. If A is the event that something occurs, then A complement is the event that it does not occur.
49. Disjoint events have no outcomes in common. The events cannot happen at the same time.
50. a. $(0.09)(0.92) = 0.0828$ b. independence is probably not valid since the strength of one player can impact the other player
52. The event that either one or both occur.
53. The event that both occur.
54. Let A = smokes and B = 4 + years of college; $P(A | B) = 0.1989$ $P(A) = 0.3782$
55. a) 3% b) 0.622
56. a) No, the king of hearts is a member of A and B.
- b. Yes, $P(H) = P(H|K)$, both probabilities = .25. (other ways to show independence)
- c. Yes, a card cannot be both a heart and a spade.
- d. No, knowing the card is a heart changes the probability that it will be a spade. (disjoint events cannot be independent)
57. Union of events ("or")
58. Intersection of events ("and")
59. a) 0.48 b) 0.67 c. 2.26, 1.17
60. continuous
61. mean
62. As the number of observations increases, the sample mean approaches the population mean and the expected value approaches the population mean.
63. independent
64. a. 150 b. 15.6 c. 10 d. 15.6 e. 85, 6
65. 2 outcomes: success or failure, there is a fixed number of observations (n), observations are independent, the probability of success (p) remains constant
66. not a fixed number of trials, the variable of interest is the number of trials required to obtain the first success
67. a. binomial b. 1, .866 c. $\text{binompdf}(4, 0.25, 2) = 0.2109$
- d) $\text{binompdf}(4, 0.25, 1) + \text{binompdf}(4, 0.25, 2) + \text{binompdf}(4, 0.25, 3) + \text{binompdf}(4, 0.25, 4) = 0.6835$
- OR $1 - \text{binompdf}(4, 0.25, 0) = 0.6835$
68. a. geometric; b. 4 c. $\text{geometpdf}(0.25, 1) = 0.25$ d) $1 - \text{geometcdf}(0.25, 5) = 0.2373$