

## Note: Corrections to previous version:

1) There was an important error in the solution to question 123 which has been corrected.

Sample questions for final exam

The one-hour mid-term test had 28 questions with 21 multiple choice and 7 open questions including 3 'essay' questions.

2) Questions 16, 54 and 58 cover material not on the exam.

The final exam is a two-hour exam and will have approximately 40 multiple choice questions, 5 'essay' questions, 5 'easy' open questions (e.g. name and briefly describe three type of bias in sample surveys), 2 more elaborate questions involving probabilities using tree diagrams. A number of questions will require the interpretation of output from computer analyses.

The exam will cover all the material seen in the course with weight roughly proportional to the amount of time spent on the material. The chapters covered are chapters 1 to 7, 10, 11. We excluded the part of chapter 4 devoted to simulation. Note that some material in 12, 13 and 14 was covered in earlier chapters and you are responsible for that material.

For sample questions, use

1. the sample mid-term and the mid-term itself,
2. the questions in assignment 2,
3. the questions below that focus on material covered since the mid-term.

Pay particular attention to questions in red.

Solutions:

Note that the explanations provided with multiple choice questions are not necessary for full marks. They are provided here only for pedagogical purposes.

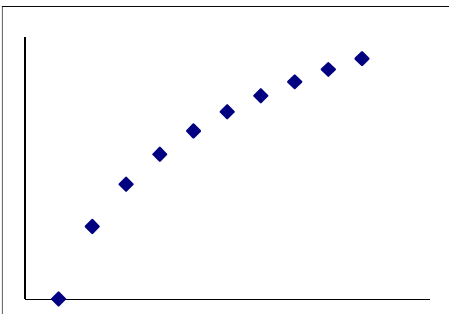
1. Which one of the following is NOT appropriate for studying the relationship between two quantitative variables?

- A. Scatterplot
- ☒ B. Bar chart
- C. Correlation
- D. Regression

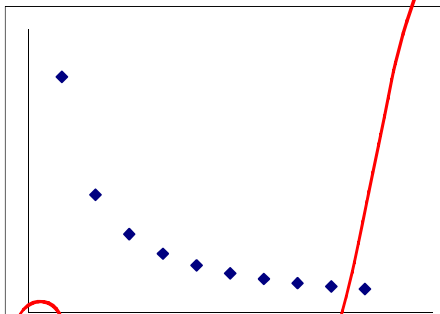
*for categorical variables*

2. Which graph shows a pattern that would be appropriately described by the equation  $\hat{y} = b_0 + b_1x$ ?

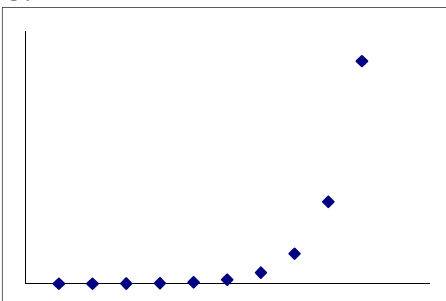
A.



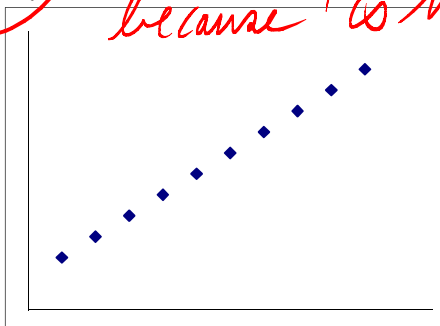
B.



C.

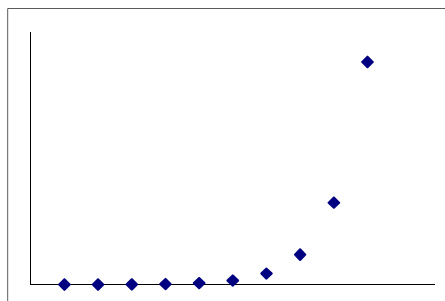


D.



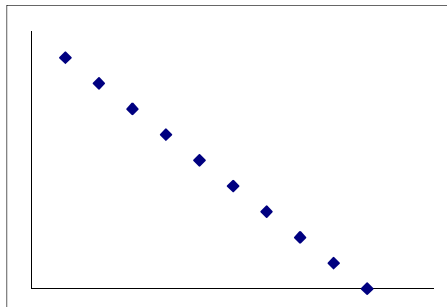
*because is the equation of a line*

3. Describe the type of association shown in the following scatterplot:



- A. Positive linear association
- B. Negative linear association
- ☒ C. Positive curvilinear association
- D. Negative curvilinear association

4. Describe the type of association shown in the following scatterplot:



- A. Positive linear association  
☒ B. Negative linear association  
C. Positive curvilinear association  
D. Negative curvilinear association
5. Which of the following can NOT be answered from a regression equation?
- A. Predict the value of y at a particular value of x.  
B. Estimate the slope between y and x.  
C. Estimate whether the linear association is positive or negative.  
☒ D. Estimate whether the association is linear or non-linear

*you need to see the data behind the equation to do this.*

**For the next 4 questions:** The simple linear regression equation can be written as  $\hat{y} = b_0 + b_1x$

6. In the simple linear regression equation, the symbol  $\hat{y}$  represents the
- ☒ A. average or predicted response (given x)  
B. estimated intercept  
C. estimated slope  
D. explanatory variable
7. In the simple linear regression equation, the term  $b_0$  represents the
- A. estimated or predicted response  
☒ B. estimated intercept  
C. estimated slope  
D. explanatory variable
8. In the simple linear regression equation, the term  $b_1$  represents the
- A. estimated or predicted response  
B. estimated intercept  
☒ C. estimated slope  
D. explanatory variable
9. In the simple linear regression equation, the symbol  $x$  represents the
- A. estimated or predicted response  
B. estimated intercept  
C. estimated slope  
☒ D. explanatory variable

**For the next 3 questions:** A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\hat{y} = 10.9 + 0.23x$$

10. One student in the sample was 73 inches tall with a foot length of 29 cm. What is the predicted foot length for this student?

A. 17.57 cm  
☒ B. 27.69 cm  
 C. 29 cm  
 D. 33 cm

$$10.9 + 0.23 \times 73 =$$

11. One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?

A. 29 cm  
☒ B. 1.31 cm  
 C. 0.00 cm  
 D. -1.31 cm

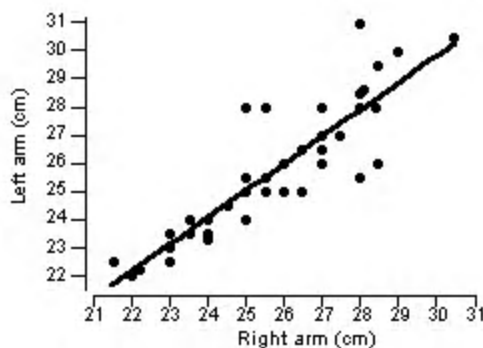
$$y - \hat{y} = 29 - 27.69 = 1.31$$

12. What is the estimated average foot length for students who are 70 inches tall?

☒ A. 27 cm  
 B. 28 cm  
 C. 29 cm  
 D. 30 cm

$$\hat{y}|_{x=70} = 10.9 + 0.23 \times 73 = 27$$

**For the next 3 questions:** The following scatterplot shows the relationship between the left and right forearm lengths (cm) for 55 college students along with the regression line, where  $y$  = left forearm length  $x$  = right forearm length.



13. One of the four choices is the equation for the regression line in the plot. The regression equation is

☒ A.  $\hat{y} = 1.22 + 0.95x$   
 B.  $\hat{y} = 1.22 - 0.95x$   
 C.  $\hat{x} = 1.22 + 0.95x$   
 D.  $\hat{x} = 1.22 - 0.95x$

14. One of the four choices is the value of the correlation for this situation. The correlation is

A. -0.88 — not neg.  
☒ B. 0.00 — only if regression line is horizontal  
☒ C. 0.88  
 D. 1.00 — only if all points on pos. sloped line

15. The proportion of total variation explained by  $x$ ,  $r^2$ , is closest to

- A. -78.3%
- B. 0.0%
- C. 78.3%
- D. 100.0%

*impossible horizontal*

*only if all pts on line*

16. A regression equation for left palm length (y variable) and right palm length (x variable) for 55 college students gave an error sum of squares (SSE) of 10.7 and a total sum of squares (SSTO) of 85.2. The proportion of

variation explained by  $x$ ,  $r^2$ , is

- A. 11.2%
- B. 12.6%
- C. 87.4%
- D. 88.8%

$$\begin{aligned} SS_{\text{explained}} &= SSTO - SSE \\ &= 85.2 - 10.7 \end{aligned}$$

$$\begin{aligned} &= 74.5 \\ \text{Proportion explained} &= 74.5 / 85.2 = \end{aligned}$$

Next 3 questions: A regression analysis done with Rcmdr for the relationship between a prestige rating and mean education for 102 occupations yielded the following output.

```

Output Window

Call:
lm(formula = prestige ~ education, data = Prestige)

Residuals:
    Min       1Q   Median       3Q      Max
-26.0397  -6.5228   0.6611   6.7430  18.1636

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -10.732     3.677  -2.919  0.00434 **
education      5.361     0.332  16.148 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.103 on 100 degrees of freedom
Multiple R-squared:  0.7228, Adjusted R-squared:  0.72
F-statistic: 260.8 on 1 and 100 DF, p-value: < 2.2e-16
    
```

17. The regression equation to predict prestige (y) as a function of education (x) is:

- A.  $\hat{y} = -10.732 + 5.361x$
- B.  $\hat{y} = 5.361 - 10.732x$
- C.  $\hat{y} = 3.677 + 0.332x$
- D. none of the above

18. How would you describe the evidence for a relationship between prestige and education?

*The p value is < 0.00001 showing strong evidence of a relationship.*

19. The correlation between prestige and income is:

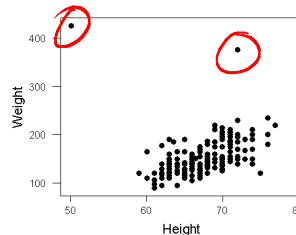
- A. -0.7228

NOT ON EXAM

$r^2 = 0.7228$  and  $r$  has same sign as slope so  
 $r = 0.8502$

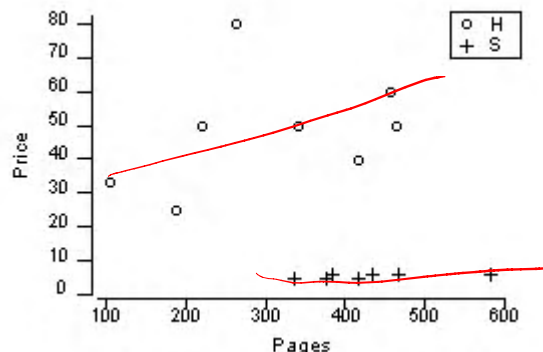
- B. 0.7228
- ☒ C. 0.8502
- D. can't tell from the output

20. A scatterplot of the self-reported weights (y variable) and self-reported heights (x variable) for 176 college students follows.



The main difficulty with using a regression line to analyze these data is

- ☒ A. Presence of one or more outliers
  - B. Inappropriately combining groups
  - C. Curvilinear data
  - D. Response variable is not quantitative
21. A scatterplot of the price of a book (y variable) versus the number of pages in the book (x variable) is shown for 15 books in a professors office. In addition, the symbol "o" shows that the book was a hardcover book, while the symbol "+" shows that the book was a softcover book.



The main difficulty with using a single regression line to analyze these data is

- A. Presence of one or more outliers
  - ☒ B. Inappropriately combining groups
  - C. Curvilinear data
  - D. Response variable is not quantitative
22. Based on 1988 census data for the 50 States in the United States, the correlation between the number of churches per State and the number of violent crimes per State was 0.85. We can conclude that
- A. There is a causal relationship between the number of churches and the number of violent crimes committed in a city.
  - ☒ B. The correlation is partly spurious because of the confounding variable of population size: both number of churches and number of violent crimes are related to the population size.
  - C. Since the data comes from a census, or nearly complete enumeration of the United States, there must be a causal relationship between the number of churches and the number of violent crimes.
  - D. The relationship is not causal because only correlations of +1 or -1 show causal relationships.

30. One interpretation of the slope is
- A. a student who scored 0 on the midterm would be predicted to score 50 on the final exam.
  - B. a student who scored 0 on the final exam would be predicted to score 50 on the midterm exam.
  - ☒ C. a student who scored 2 points higher than another student on the midterm would be predicted to score 1 point higher than the other student on the final exam.
  - D. none of the above are an interpretation of the slope.

31. Midterm exam scores could range from 0 to 100. Based on the equation, final exam scores are predicted to range from
- A. 0 to 100.
  - ☒ B. 50 to 100.
  - C. 50 to 75.
  - D. 0 to 75.

32. For which one of these relationships could we use a regression analysis? Only one choice is correct.
- ☒ A. Relationship between weight and height.
  - B. Relationship between political party membership and opinion about abortion.
  - C. Relationship between gender and whether person has a tattoo.
  - D. Relationship between eye color (blue, brown, etc.) and hair color (blond, etc.).
- both quantitative*

33. Which of the following is a possible value of  $r^2$  and indicates the strongest linear relationship between two quantitative variables?

- A. -90%
- B. 0%
- ☒ C. 80%
- D. 120%

*must be between 0 + 1 i.e., 0% and 100%*

34. The correlation between two variables is given by  $r = 0.0$ . What does this mean?

- ☒ A. The best straight line through the data is horizontal.
- B. There is a perfect positive relationship between the two variables
- C. There is a perfect negative relationship between the two variables.
- D. All of the points must fall exactly on a horizontal straight line.

35. The value of a correlation is reported by a researcher to be  $r = -0.5$ . Which of the following statements is correct?

- A. The x-variable explains 50% of the variability in the y-variable.
- B. The x-variable explains -50% of the variability in the y-variable.
- ☒ C. The x-variable explains 25% of the variability in the y-variable
- D. The x-variable explains -25% of the variability in the y-variable

36. Which one of the following statements involving correlation is possible and reasonable?

- A. The correlation between hair color and eye color is 0.80.
- ☒ B. The correlation between the height of a father and the height of his first son is 0.6
- C. The correlation between left foot length and right foot length is 2.35
- D. The correlation between hair color and age is positive.

*need 2 quant. vars and  $-1 \leq r \leq 1$*

*ok but not*

37. Correlation and regression are concerned with

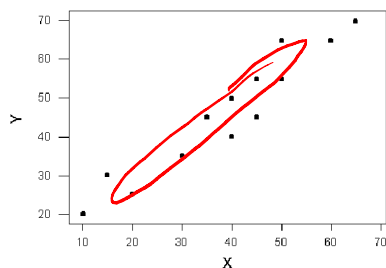
- A. the relationship between two categorical variables.
- ☒ B. the relationship between two quantitative variables.
- C. the relationship between a quantitative explanatory variable and a categorical response variable.
- D. the relationship between a categorical explanatory variable and a quantitative response variable.

38. A scatterplot is a
- A. one-dimensional graph of randomly scattered data.
  - B. two-dimensional graph of a straight line.
  - C. two-dimensional graph of a curved line.
  - ☒ D. two-dimensional graph of data values.
39. Two variables have a positive association when
- ☒ A. the values of one variable tend to increase as the values of the other variable increase.
  - B. the values of one variable tend to decrease as the values of the other variable increase.
  - C. the values of one variable tend to increase irregardless of how the values of the other variable change.
  - D. the values of both variables are always positive.
40. A researcher reports that the correlation between two quantitative variables is  $r = 0.8$ . Which of the following statements is correct?
- A. The average value of  $y$  changes by 0.8 when  $x$  is increased by 1.
  - B. The average value of  $x$  changes by 0.8 when  $y$  is increased by 1.
  - C. The explanatory variable ( $x$ ) explains 0.8 of the variation in the response variable ( $y$ )
  - ☒ D. The explanatory variable ( $x$ ) explains  $.8^2 = 0.64$  of the variation in the response variable ( $y$ )
41. What is the effect of an outlier on the value of a correlation coefficient?
- A. An outlier will always decrease a correlation coefficient.
  - B. An outlier will always increase a correlation coefficient.
  - ☒ C. An outlier might either decrease or increase a correlation coefficient, depending on where it is in relation to the other points.
  - D. An outlier will have no effect on a correlation coefficient.
42. A reviewer rated a sample of fifteen wines on a score from 1 (very poor) to 7 (excellent). A correlation of .92 was obtained between these ratings and the cost of the wines at a local store. In plain English, this means that
- A. in general, the reviewer liked the cheaper wines better.
  - B. having to pay more caused the reviewer to give a higher rating.
  - C. wines with low ratings are likely to be more expensive (probably because fewer will be sold).
  - ☒ D. in general, as the cost went up so did the rating.
43. A scatter plot and regression line can be used for all of the following EXCEPT
- A. to determine if any ( $x,y$ ) pairs are outliers.
  - B. to predict  $y$  at a specific value of  $x$ .
  - C. to estimate the average  $y$  at a specific value of  $x$ .
  - ☒ D. to determine if a change in  $x$  causes a change in  $y$ .

↑  
association is  
not necessarily causal



44. Consider the following graph.



For these data, there is

- ☒ A. positive correlation.
- ☐ B. negative correlation.
- ☐ C. biased correlation.
- ☐ D. zero correlation.

45. Which of the following is a deterministic relationship?

- ☐ A. The relationship between hair color and eye color.
- ☐ B. The relationship between father's height and son's height.
- ☒ C. The relationship between height in inches and height in centimeters.
- ☐ D. The relationship between height as determined with a ruler and height as determined by a tape measure.

46. A scatter plot of number of teachers and number of people with college degrees for cities in Pennsylvania reveals a positive association. The most likely explanation for this positive association is:

- ☐ A. Teachers encourage people to get college degrees, so an increase in the number of teachers is causing an increase in the number of people with college degrees.
- ☐ B. Teaching is a common profession for people with college degrees, so an increase in the number of people with college degrees causes an increase in the number of teachers.
- ☐ C. Cities with higher incomes tend to have more teachers and more people going to college, so income is a confounding variable, making causation between number of teachers and number of people with college degrees difficult to prove.
- ☒ D. Larger cities tend to have both more teachers and more people with college degrees, so the association is explained by a third variable, the size of the city.

50. A professor found a negative correlation between the number of hours students came to her office hours and their score on her final exam. What possible confounding variable could explain the observed negative correlation?

*Difficulty with the course material could cause both ↑ in office hours and a ↓ in score on final.*

51. Suppose a study of employees at a large company found a negative correlation between weight and distance walked on an average day. Explain why a conclusion that walking reduces weight cannot be made from this study (i.e. What confounding variable could also explain the observed negative correlation?)

*Fitness could cause both*

**Next 4 questions:** A regression was done for 20 cities with latitude as the explanatory variable (x) and average January temperature as the response variable (y). The latitude is measured in degrees and average January temperature in degrees Fahrenheit. The latitudes ranged from 26 (Miami) to 47 (Duluth) The regression equation is

$$\hat{y} = 49.4 - 0.313x$$

52. The city of Pittsburgh, Pennsylvania has latitude 40 degrees with average January temperature of 25 degrees Fahrenheit. What is the estimated average January temperature for Pittsburgh, based on the regression equation? What is the residual?

$$\hat{y} = 49.4 - 0.313 \times 40 = 36.88 \quad \text{residual} = 25 - 36.88 = -11.88$$

53. The city of Miami, Florida has latitude 26 degrees with average January temperature of 67 degrees Fahrenheit. What is the estimated average January temperature for Miami, based on the regression equation? What is the residual?

$$\hat{y} = 49.4 - 0.313 \times 26 = 41.262 \quad \text{residual}$$

54. The total sum of squares (SSTO) = 4436.6, and the error sum of squares (SSE) = 1185.8. What is the value of  $r^2$ , the proportion of variation explained by x?

$$r^2 = \frac{SSTO - SSE}{SSTO} = \frac{4436.6 - 1185.8}{4436.6} = 0.7327$$

55. Mexico City has latitude 20 degrees. What is the problem with using the regression equation to estimate the average January temperature for Mexico City?

The observed x-value range from 26 to 47. Mexico City's latitude of 20 is too far outside the range.

**Next 3 questions:** The next four questions are based on a regression equation for 55 college students with x = left forearm length (cm) and y = height. The forearm lengths ranged from 22 cm to 31 cm. The regression equation is

$$\hat{y} = 30.3 + 1.49x$$

56. One student's left forearm length was 27 cm, and his height was 75 inches. What is the estimated height for this student, based on the regression equation? What is the residual?

$$\hat{y} = 30.3 + 1.49 \times 27 = 70.53 \quad \text{resid} = y - \hat{y} = 75 - 70.53 = 4.47$$

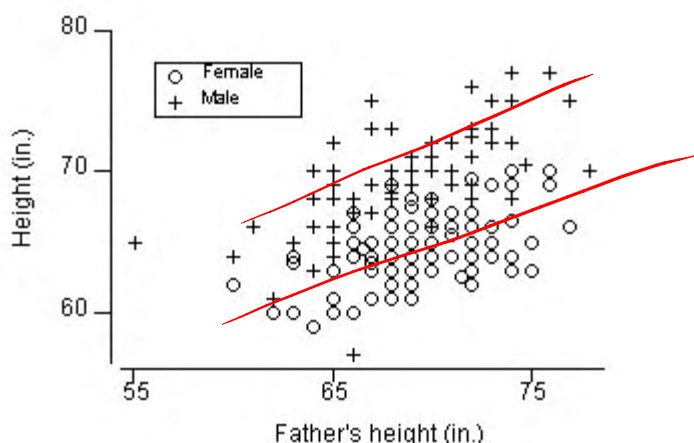
57. One student's left forearm length was 22 cm, and her height was 63 inches. What is the estimated height for this student, based on the regression equation? What is the residual?

$$\hat{y} = 30.3 + 1.49 \times 22 = 63.08 \quad \text{resid} = 63 - 63.08 = -0.08$$

58. The total sum of squares (SSTO) = 1054.8, and the error sum of squares (SSE) = 464.5. What is the value of  $r^2$ , the proportion of variation explained by x?

$$r^2 = (1054.8 - 464.5) / 1054.8 = 55.96\%$$

**Next 3 questions** The scatterplot below shows student heights (y axis) versus father's heights (x axis) for a sample of 173 college students. The symbol "+" represents a male student and the symbol "o" represents a female student.



59. Based on the scatterplot, what is the problem with using a regression equation for all 173 students?

There are really two subgroups. Each would need a separate regression or gender should be used in a multiple regression.

60. The correlation between father's heights and student's heights for the 79 male students is  $r = 0.669$ . What is the proportion of variation explained by father's heights?

$$r^2 = 0.669^2 = 0.448$$

61. The regression equation between father's heights and student's heights for the 79 male students is  $\hat{y} = 30 + .58x$ . One student was 72 inches tall; his father's height was 65 inches. What is the estimated height for this student, and what is the residual?

$$\hat{y} = 30 + .58 \times 65 = 67.7$$

$$\text{resid} = 72 - 67.7 = 4.3$$

## Chapter 6

**Next 4 questions:** The table below shows the counts by gender and highest degree attained for 498 respondents in the General Social Survey.

	No High School Degree	High School Degree Juni	or College	Bachelor Degree	Graduate Degree Total	
Male	49 95		11	39 23		217
Female	52 166	14		37	12	281
Total 10	1	261	25	76	35	498

62. What percent of the sample were males?

- A. 43.6%  
B. 48.5%  
C. 56.4%  
D. 77.2%

$$\frac{217}{498}$$

63. What percent of the sample were males with no high school degree?

- A. 9.8%  
B. 20.3%  
C. 22.6%  
D. 48.5%

$$\frac{49}{498}$$

64. What percent of the sample did not graduate from high school?

- A. 18.5%  
B. 20.3%  
C. 22.6%  
D. 52.4%

$$\frac{101}{498}$$

65. What percent of females had a graduate degree?

- A. 2.4%  
B. 4.3%  
C. 7.0%  
D. 56.4%

$$\frac{12}{281}$$

**Questions 5 to 8:** The table below shows the number of Olympic medals won by the three countries with the most medals during the 2000 Olympics in Sydney, Australia. There were a total of 244 medals won by the three countries.

Country	Gold	Silver	Bronze	Total
USA	39	25	33	97
Russia	32	28	28	88
China	28	16	15	59
Total	99	69	76	244

66. What percent of the medals won among the three countries were won by the USA?

- A. 39.4%
- ☒ B. 39.8%
- C. 40.2%
- D. 40.6%

$$\frac{97}{244}$$

67. What percent of the medals won among the three countries were gold?

- A. 39.4%
- B. 39.8%
- ☒ C. 40.2%
- D. 40.6%

$$\frac{99}{244}$$

68. What percent of the medals won by the USA were gold?

- A. 39.4%
- B. 39.8%
- ☒ C. 40.2%
- D. 40.6%

$$\frac{39}{97}$$

69. What percent of the medals won by China were silver?

- A. 6.6%
- B. 24.2%
- ☒ C. 27.1%
- D. 28.3%

$$\frac{16}{59}$$

**Next 5 questions:** The table and output below shows the responses from a sample of 680 people in the General Social Survey to the question, “Do you sometimes drink more than you think you should?”

Gender Y	es	No	Total
Male 15	1	177	328
Female 92		260	352
Total 24	3	437	680

Output Window
Submit

```

> .Table # Counts
      Yes  No
Male   151 177
Female  92 260

> .Test <- chisq.test(.Table, correct=FALSE)

> .Test

      Pearson's Chi-squared test

data:  .Table
X-squared = 29.2788, df = 1, p-value = 6.268e-08

```

70. What is the risk (or percentage) of men thinking they drank more than they should?

- A. 22.2%
- ☒ B. 35.7%
- C. 46.0%
- D. 62.1%

$$\frac{151}{328}$$

71. What is the risk (or percentage) of women thinking they drank more than they should?

- A. 13.5%
- ☒ B. 26.1%
- C. 35.7%
- D. 37.9%

$$\frac{92}{352}$$

72. What is the relative risk for women thinking they drank more than they should compared to men?

- A. 0.41
- ☒ B. 0.57
- C. 1.76
- D. 2.41

$$\frac{26.1\%}{46.0\%}$$

73. What is the odds ratio for women thinking they drank more than they should compared to men?

- ☒ A. 0.41
- B. 0.57
- C. 1.76
- D. 2.41

$$\frac{\text{odds for W}}{\text{odds for M}} = \frac{92/260}{151/177} =$$

74. Report whether there is evidence of a difference between sexes in the proportion who report that they drank more than they should. Report the p-value and interpret it.

p value is 0.0000000628  
providing very strong evidence of a  
difference in sexes

**Next 4 questions:** The table below summarizes, by gender of respondent, the responses from 1,033 people to the question, "Do you smoke?"

Gender Yes		No	Total
Male 142		302	444
Female 171		418	589
Total 313		720	1033

75. What are the odds of smoking (to not smoking) for a man?

- A. 0.14
- B. 0.32
- C. 0.45
- D. 0.47

$$\frac{142}{302}$$

76. What are the odds of smoking (to not smoking) for a woman?

- A. 0.17
- B. 0.39
- C. 0.41
- D. 0.55

$$\frac{171}{418}$$

77. What is the odds ratio for women smoking (to not smoking) compared to men?

- A. 0.87
- B. 0.91
- C. 1.10
- D. 1.15

$$\frac{171/418}{142/302}$$

78. What is the odds ratio for men smoking (to not smoking) compared to women?

- A. 0.87
- B. 0.91
- C. 1.10
- D. 1.15

$$\frac{142/302}{171/418}$$

79. What is the relative risk for women smoking compared to men smoking?

- A. 0.87
- B. 0.91
- C. 1.10
- D. 1.15

$$\frac{171/589}{142/444}$$

80. Based on the National Household Survey on Drug Abuse, the percentage of 17-year olds who ever tried cigarette smoking is 56.2%. The relative risk of ever smoking for a 17-year old versus a 12-year old is 3.6. What is the risk of smoking for a 12-year-old (i.e. what was the percentage of 12-year olds who ever tried smoking)?

- A. 14.1%
- B. 15.6%
- C. 50.0%
- D. 56.2%

Relative Risk =  $\frac{\text{risk for 17yo}}{\text{risk for 12yo}}$

So  $3.6 = \frac{56.2\%}{?}$  . Therefore  $? = \frac{56.2\%}{3.6}$

**Next 4 questions:** A study done by the Center for Academic Integrity at Rutgers University surveyed 2116 students at 21 colleges and universities. Some of the schools had an "honor code" and others did not. Of the students at schools with an honor code, 7% reported having plagiarized a paper via the Internet, while at schools with no honor code, 13% did so. (*Sacramento Bee*, Feb 29, 2000, D1.)

81. For this study, the relative risk of a student having plagiarized a paper via the Internet at a school with no honor code, compared to a school with an honor code is:

- A.  $13/7 = 1.857$
- B.  $7/13 = 0.538$
- C.  $13/87 = 0.149$
- D.  $87/13 = 6.692$

$$\frac{13-7}{7} = \frac{6}{7} = 85.7\%$$

82. Which of the following statements about *percent increase in risk* is correct for this study?

- A. There is a 185.7% increase in the risk of plagiarism at a school with no honor code, compared to a school with an honor code.
- B. There is an 85.7% increase in the risk of plagiarism at a school with no honor code, compared to a school with an honor code.
- C. There is a 53.8% increase in the risk of plagiarism at a school with no honor code, compared to a school with an honor code.
- D. There is a 6% increase in the risk of plagiarism at a school with no honor code, compared to a school with an honor code.

83. Although the data provided are not sufficient to carry out a chi-square test of the relationship between whether or not a school has an honor code and whether or not a student would plagiarize a paper via the Internet, suppose such a test were to show a statistically significant relationship on the basis of this study. The correct conclusion would be:

- A. Because this is an observational study, it can be concluded that implementing an honor code at a college or university will reduce the risk of plagiarism.
- B. Because this is a randomized experiment, it can be concluded that implementing an honor code at a college or university will reduce the risk of plagiarism.
- C. Because this is an observational study and confounding variables are likely, it cannot be concluded that implementing an honor code at a college or university will reduce the risk of plagiarism.
- D. Because this is a randomized experiment and confounding variables are likely, it cannot be concluded that implementing an honor code at a college or university will reduce the risk of plagiarism.

84. Simpson's Paradox occurs when

- A. No baseline risk is given, so it is not known whether or not a high relative risk has practical importance.
- B. A confounding variable rather than the explanatory variable is responsible for a change in the response variable.
- C. The direction of the relationship between two variables changes when the categories of a confounding variable are taken into account.
- D. The results of a test are statistically significant but are really due to chance.

85. Pick the choice that best completes the following sentence. If a relationship between two variables is called statistically significant, it means the investigators think the variables are

- A. related in the population represented by the sample.
- B. not related in the population represented by the sample.
- C. related in the sample due to chance alone.
- D. very important.

86. The statement that changes in the explanatory variable will have no effect on the response variable is an example of a(n) \_\_\_\_\_ hypothesis.

- A. negative
- B. dependent
- C. null
- D. alternative

87. A study on the use of seat belts versus belted booster seats for children ages 4 and 5 reported that "Using seat belts instead of booster seats was associated with increased risk for serious injury in an accident; the relative risk was 2.4." Based on this, it can be concluded that for this study:
- A. Children ages 4 and 5 in a booster seat were 2.4 times more likely to have serious injuries in an accident than were children wearing seatbelts.
  - ☒ B. Children ages 4 and 5 wearing seatbelts were 2.4 times more likely to have serious injuries in an accident than were children in a booster seat.
  - C. The percent of children ages 4 and 5 in a booster seat was 2.4 times higher than the percent of children wearing seatbelts.
  - D. The percent of children ages 4 and 5 wearing seatbelts was 2.4 times higher than the percent of children in a booster seat.
88. In a newspaper article about whether the regular use of Vitamin C reduces the risk of getting a cold, a researcher is quoted as saying that Vitamin C performed better than placebo in an experiment, but the difference was not larger than what could be explained by chance. In statistical terms, the researcher is saying the results are \_\_\_\_.
- A. due to non-sampling errors.
  - B. definitely due to chance.
  - ☒ C. statistically significant.
  - D. not statistically significant.
89. The relative risk of allergies for children of parents who smoke compared to children of parents who don't smoke is 3.0. Suppose that the risk of allergies for the children of non-smokers is 0.15 (15%). What is the risk of allergies for the children of smokers?
- $3.0 \times 15\%$
- A. 3%
  - B. 5%
  - C. 30%
  - ☒ D. 45%
90. In a sample of 1000 students majoring in the humanities, 660 were female. The 66% (660/1000) chance of a humanities major being female is a
- ☒ A. Subjective probability
  - B. Relative frequency probability
91. A quarter is flipped 2000 times and results in 500 heads. The 25% (500/2000) chance of heads is a
- ☒ A. Subjective probability
  - B. Relative frequency probability
92. Among 5000 new tires sold by a tire company, 20% (1000/5000) lasted more than 100,000 miles. The 20% chance that a new tire will last more than 100,000 miles is a
- A. Subjective probability
  - ☒ B. Relative frequency probability
93. A football fan believes that the Oakland Raiders have a 50% chance of winning the next Super bowl. The 50% is a
- ☒ A. Subjective probability
  - B. Relative frequency probability
94. A college basketball player has made 53% of his shots from 3-point range. The probability that he will make a 3-point shot, 53%, is a
- ☒ A. Subjective probability
  - B. Relative frequency probability



**Next 4 questions:** A statistics class has 4 teaching assistants (TAs): three female assistants (Lauren, Rona, and Leila) and one male assistant (Josh). Each TA teaches one tutorial section.

3F-1M

95. A student picks a discussion section. The two events  $A = [\text{the TA is a woman}]$  and  $B = [\text{the TA is Josh}]$  are

- A. Independent events
- ☒ B. Disjoint (mutually exclusive) events
- C. Each simple events
- D. None of the above

96. A student picks a discussion section. The two events  $A = [\text{the TA is a woman}]$  and  $B = [\text{the TA is a man}]$  are

- A. Independent events
- ☒ B. Disjoint (mutually exclusive) events
- C. Each simple events
- D. None of the above

97. A student picks a discussion section. The events  $A = [\text{the TA is Lauren}]$  and  $B = [\text{the TA is a woman}]$  are

- A. Independent events
- B. Disjoint (mutually exclusive) events
- ☒ C. Each simple events
- ☒ D. None of the above

98. Two students, Kimiko and Tom, who don't know each other, each pick a discussion section. The two events  $A = [\text{Kimiko's TA is Lauren}]$  and  $B = [\text{Tom's TA is a woman}]$  are

- ☒ A. Independent events
- B. Disjoint (mutually exclusive) events
- C. Each simple events
- D. None of the above

99. Two students, Michelle and Charles, who don't know each other, each choose a channel. The two events,  $N = [\text{Charles watches news}]$  and  $F = [\text{Michelle watches football}]$  are

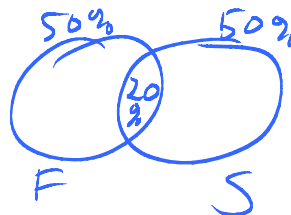
- ☒ A. Independent events
- B. Disjoint (mutually exclusive) events
- C. Each simple events
- D. None of the above

100. Lauren wants to wear something warm when she leaves for class. She reaches into her coat closet without looking and grabs a hanger. Based on what she has in her coat closet, she has a 30% chance of picking a sweater, a 50% chance of picking a coat, and a 20% chance of picking a jacket. What is the probability that she will pick a sweater or a coat?

- A. 15%
- B. 30%
- ☒ C. 50%
- ☒ D. 80%

101. Michael wants to take French or Spanish, or both. But classes are closed, and he must apply and get accepted to be allowed to enroll in a language class. He has a 50% chance of being admitted to French, a 50% chance of being admitted to Spanish, and a 20% chance of being admitted to both French and Spanish. If he applies to both French and Spanish, the probability that he will be enrolled in French or Spanish (or possibly both) is

- A. 70%
- ☒ B. 80%
- C. 90%
- D. 100%



$$.5 + .5 - .2 = .8$$

102. The California Super Lotto Plus game is played by picking 5 numbers between 1 and 47 and one MEGA number between 1 and 27. Assuming that the winning numbers are picked at random, how many different sets of winning lottery numbers are possible?

- A. 74
- B. 262
- C. 6345
- ☒ D.  $6.2 \times 10^9$

$$47^5 \times 27$$

103. A test to detect prostate cancer in men had a **sensitivity** of 95%. This means that

- A. 95% of the men who test positive will actually have prostate cancer.
- ☒ B. 95% of the men with prostate cancer will test positive.
- C. 95% of the men who do not have prostate cancer will test negative.
- D. 95% of the men who test negative will actually not have prostate cancer.

104. A test to detect prostate cancer in men had a **specificity** of 80%. This means that

- A. 80% of the men who test positive will actually have prostate cancer.
- B. 80% of the men with prostate cancer will test positive.
- ☒ C. 80% of the men who do not have prostate cancer will test negative.
- D. 80% of the men who test negative will actually not have prostate cancer.

105. Which of the following statements is TRUE for 6 tosses of a fair coin, where H = Heads and T = Tails?

- A. Sequences with all heads, like HHHHHH, are less likely than specific sequences with 3 heads and 3 tails, like HTHTHT
- B. Sequences with all heads, like HHHHHH, are more likely than specific sequences with 3 heads and 3 tails, like HTHTHT
- ☒ C. All specific sequences six letters long (made up of H's and T's), regardless of the number of H's and T's, are equally likely.
- D. None of the above

106. Which of the following statements is TRUE for tossing a fair coin (i.e. Probability of Heads =  $\frac{1}{2}$ ) if the first 100 tosses of the coin result in 100 heads?

- A. The chance that the next toss will be heads is almost certain.
- B. The chance that the next toss will be heads is nearly 0.
- ☒ C. The chance that the next toss will be heads is  $\frac{1}{2}$ .
- D. None of the above

107. A birth is selected at random. Define events A=baby is a boy and B=the mother had the flu during her pregnancy. The events A and B are

- A. Disjoint but not independent.
- ☒ B. Independent but not disjoint.
- C. Disjoint and independent.
- D. Neither disjoint nor independent.

*assuming having the flu does not affect the prob. of a miscarriage differently for male or female fetuses.*

108. A soft drink company holds a contest in which a prize may be revealed on the inside of the bottle cap. The probability that each bottle cap reveals a prize is 0.2 and winning is independent from one bottle to the next. What is the probability that a customer must open three or more bottles before winning a prize?

- A.  $(.2)(.2)(.8) = .032$
- B.  $(.8)(.8)(.2) = .128$
- ☒ C.  $(.8)(.8) = .64$
- D.  $1 - (.2)(.2)(.8) = .968$

*"Open 1st with no prize" and "Open 2nd with no prize" then ... it doesn't matter"*

109. A six-sided die is made that has four Green sides and two Red sides, all equally likely to land face up when the die is tossed. The die is tossed three times. Which of these sequences (in the order shown) has the highest probability?

- ☒ A. Green, Green, Green
- B. Green, Green, Red

$$\frac{4}{6} \times \frac{4}{6} \times \frac{4}{6}$$

$$\frac{2}{6} \times \frac{4}{6} \times \frac{2}{6}$$

- C. Green, Red, Red
- D. They are all equally likely

$$= \frac{1}{100} \times \frac{1}{100}$$

110. Suppose two different states each pick a two-digit lottery number between 00 and 99 (there are 100 possible numbers). What is the probability that both states pick the same number?

- A. 2/100
- ☒ B. 1/100
- C. 1/200
- D. 1/10,000

$$P(\text{both pick } 00) + P(\text{both pick } 01) + \dots + P(\dots 99) \\ = \frac{1}{10,000} + \frac{1}{10,000} + \dots + \frac{1}{10,000} \\ = 100 \times \frac{1}{10,000} = \frac{1}{100}$$

111. Suppose two different states each pick a two-digit lottery number between 00 and 99 (there are 100 possible numbers). What is the probability that both states pick the number 13?

- A. 2/100
- B. 1/100
- C. 1/200
- ☒ D. 1/10,000

$$P(\text{Both pick } 13) = \frac{1}{100} \times \frac{1}{100}$$

112. A survey was given to sophomores at a university and one of the questions asked was "What is the probability that you will leave school before you graduate?" The answer to this question for an individual student is an example of:

- A. A relative frequency probability based on long-run observation
- B. A relative frequency probability based on physical assumptions
- ☒ C. A personal probability
- D. A probability based on measuring a representative sample and observing relative frequencies that fall into various categories.

113. A thumbtack is tossed repeatedly and observed to see if the point lands resting on the floor or sticking up in the air. The goal is to estimate the probability that a thumbtack would land with the point up. That probability is an example of:

- A. A relative frequency probability based on long-run observation
- B. A relative frequency probability based on physical assumptions
- C. A personal probability
- ☒ D. A probability based on measuring a representative sample and observing relative frequencies that fall into various categories.

114. A short quiz has two true-false questions and one multiple-choice question with four choices. A student guesses at each question. Assuming the choices are all equally likely, what is the probability that the student gets all three correct?

- A. 1/32
- B. 1/3
- C. 1/8
- ☒ D. 1/16

$$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{4}$$

115. A medical treatment has a success rate of .8. Two patients will be treated with this treatment. Assuming the results are independent for the two patients, what is the probability that neither one of them will be successfully cured?

- A. .5
- B. .36
- C. .2
- ☒ D. .04

$$.2 \times .2$$

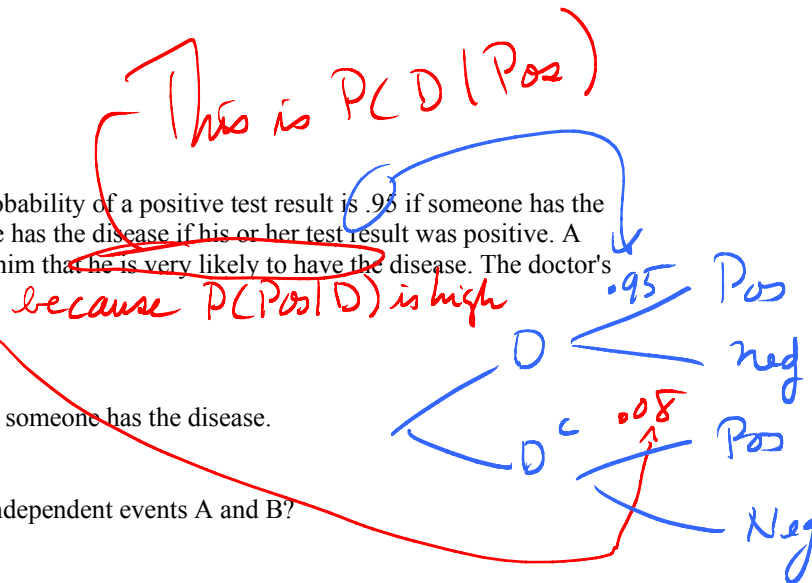
116. Elizabeth has just put 4 new spark plugs in her car. For each spark plug, the probability that it will fail in the

next 50,000 miles is  $1/100$  (which is .01), and is independent from one spark plug to the next. What is the probability that none of the spark plugs will fail in the next 50,000 miles?

- A.  $(.01)(.01)(.01)(.01)$
- B.  $1 - (.01)(.01)(.01)(.01)$
- ☒ C.  $(.99)(.99)(.99)(.99)$
- D.  $1 - (.99)(.99)(.99)(.99)$

117. Imagine a test for a certain disease. Suppose the probability of a positive test result is .95 if someone has the disease, but the probability is only .08 that someone has the disease if his or her test result was positive. A patient receives a positive test, and the doctor tells him that he is very likely to have the disease. The doctor's response is:

- ☒ A. An example of "Confusion of the inverse."
- B. An example of the "Law of small numbers."
- C. An example of "The gambler's fallacy."
- D. Correct, because the test is 95% accurate when someone has the disease.



118. Which of the following is not necessarily true for independent events A and B?

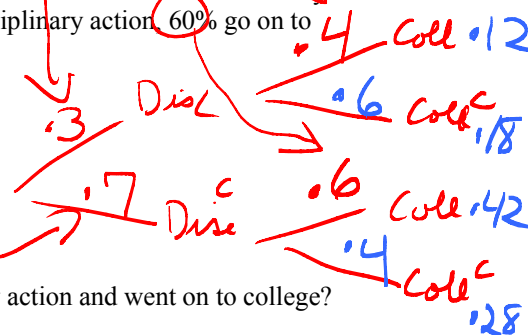
- A.  $P(A \text{ and } B) = P(A)P(B)$
- B.  $P(A|B) = P(A)$
- C.  $P(B|A) = P(B)$
- ☒ D.  $P(A \text{ or } B) = P(A) + P(B)$

**Next 3 questions:** Thirty percent of the students in a high school face a disciplinary action of some kind before they graduate. Of those students, 40% go on to college. Of the 70% who do not face a disciplinary action, 60% go on to college.

119. What percent of the students from the high school go on to college?

- A. 12%
- B. 42%
- C. 50%
- ☒ D. 54%

$$.12 + .42$$



120. What is the probability that a randomly selected student both faced a disciplinary action and went on to college?

- ☒ A. .12
- B. .40
- C. .42
- D. .84

$$.12$$

121. Given that a randomly selected student goes on to college, what is the probability that he or she faced a disciplinary action?

- A. .12
- ☒ B. .22
- C. .30
- D. .78

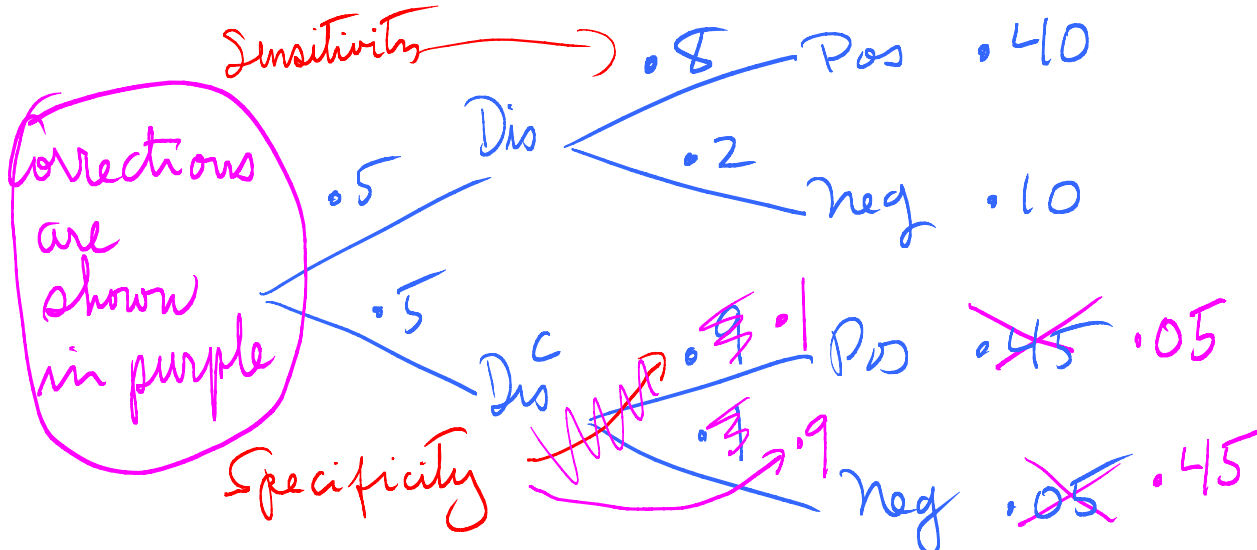
$$\frac{.12}{.12 + .42} = \frac{.12}{.54}$$

122. Students at a university who apply for campus housing can be assigned to live in a dormitory room, a suite in a dormitory or an apartment. If Alice has a .30 chance of being assigned to a dormitory room and a .50 chance of being assigned to a suite in a dormitory, what is the probability that she will be assigned to an apartment?

$$1 - (.3 + .5) = .2$$

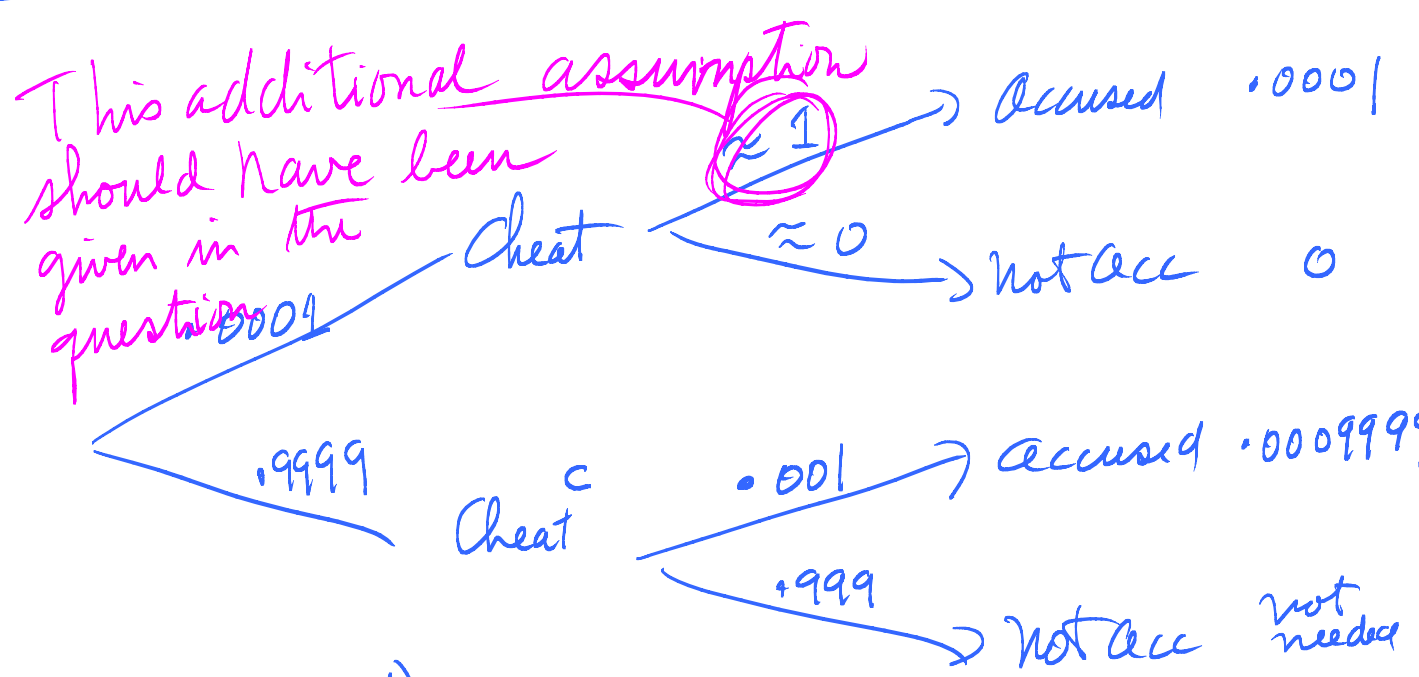
123. A test for Lyme's Disease has specificity 0.9 and sensitivity 0.8. John has gone hiking in Maine where the tick that causes Lyme's Disease is wide spread and has returned with a rash on his leg. John's physician feels that these symptom's give John a 50% chance of having Lyme's disease. The physician administers the test which comes back negative. What are John's chances of having Lyme's disease given the negative outcome for the test.

*NOTE: Error in original solution was pointed out by a student at the Tutorial*



$$\begin{aligned}
 P(\text{Dis} | \text{Neg}) &= \frac{P(\text{Dis and Neg})}{P(\text{Neg})} \\
 &= \frac{.10}{.10 + .05} = \frac{2}{3} = \frac{2}{11} \text{ answer}
 \end{aligned}$$

Solution for #124 below



$$\begin{aligned}
 P(\text{Cheat} | \text{Accused}) &= \frac{P(\text{Cheat \& Accused})}{P(\text{Accused})} = \frac{.0001}{.0001 + .0009999} = .09 \text{ answer}
 \end{aligned}$$

124. A computer program analyzes answers on a multiple choice test to attempt to detect cheaters. In a class of 100 students, the computer program computes the probability of similar wrong answers for every pair of students assuming that the students are answering independently. Mark and Phillip were the pair of students that achieved the lowest probability of 0.0005. A professor follows the policy of accusing students of cheating if the probability is lower than 0.001.

If the probability that a particular pair of students is cheating in such a class is  $1/10,000$ . (there are 4,950 pairs of students in a class of 100). what is the probability that Mark and Philip were cheating given that the professor has decided to accuse them of cheating?

*See previous page*

**Next 3 questions:** Suppose that a 95% confidence interval for the proportion of first-year students at a school who played in intramural sports is 35% plus or minus 5%.

125. The **confidence level** for the confidence interval is

A. 5%  
B. 35%  
☒ C. 95%  
D. None of the above

126. The **margin of error** for the confidence interval is

☒ A. 5%  
B. 35%  
C. 95%  
D. None of the above

127. The **95% confidence interval** for the proportion of students playing intramural sports is

A. 25% to 45%  
B. 30% to 35%  
☒ C. 35% to 40%  
D. 30% to 40%

128. Which statement is NOT true about confidence intervals?

A. A confidence interval is an interval of values computed from sample data that is likely to include the true population value.  
B. An approximate formula for a 95% confidence interval is *sample estimate  $\pm$  margin of error*.  
☒ C. A confidence interval between 20% and 40% means that the population proportion lies between 20% and 40%.  
D. A 99% confidence interval procedure has a higher probability of producing intervals that will include the population parameter than a 95% confidence interval procedure.

129. You plan to determine an approximate 95% confidence interval for a population proportion; you plan to use a conservative margin of error of 2%. How large a sample size do you need?

A. 100  
B. 400  
☒ C. 2500  
D. None of the above

$$\frac{1}{\sqrt{n}} = 0.02 \quad \text{so } n = \left(\frac{1}{0.02}\right)^2$$

130. You plan to determine an approximate 95% confidence interval for a population proportion; you plan to use a conservative margin of error of 10%. How large a sample size do you need?

☒ A. 100  
B. 400  
C. 2500  
D. None of the above

131. If the confidence level is increased, which of the following must also be increased?

- ☐ A. sample estimate
- ☒ B. multiplier
- ☐ C. standard error
- ☐ D. none of the elements would be increased

132. If the sample size is increased but the sample proportion remains the same, which of the following must also be increased?

- ☐ A. sample estimate
- ☐ B. multiplier
- ☐ C. standard error
- ☒ D. none of the elements would be increased

133. The *multiplier* for a confidence interval for a population proportion is determined by:

- ☒ A. The desired level of confidence and the sample size
- ☐ B. The desired level of confidence but not the sample size
- ☐ C. The sample size but not the desired level of confidence
- ☐ D. Neither the sample size nor the level of confidence

**Next 4 questions:** A 95% confidence interval for the proportion of young adults who skip breakfast is .20 to .27.

134. Which of the following is a correct interpretation of the 95% *confidence level*?

- ☒ A. In about 95% of all studies for which this procedure is used, the confidence interval will cover the true population proportion, but there is no way to know if this interval covers the true proportion or not.
- ☐ B. There is a 95% probability that the proportion of young adults who skip breakfast is between .20 and .27.
- ☐ C. If this study were to be repeated with a sample of the same size, there is a 95% probability that the sample proportion would be between .20 and .27.
- ☐ D. The proportion of young adults who skip breakfast 95% of the time is between .20 and .27.

135. Which of the following is the correct interpretation of the 95% confidence interval?

- ☐ A. There is a 95% probability that the proportion of young adults who skip breakfast is between .20 and .27.
- ☐ B. If this study were to be repeated with a sample of the same size, there is a 95% probability that the sample proportion would be between .20 and .27.
- ☐ C. We can be 95% confident that the proportion of young adults in the sample who skip breakfast is between .20 and .27.
- ☒ D. We can be 95% confident that the proportion of young adults in the population who skip breakfast is between .20 and .27.

136. From the information provided, we can determine that:

- ☒ A.  $\hat{p} = .235$  and margin of error = .035
- ☐ B.  $\hat{p} = .235$  and margin of error = .07
- ☐ C.  $p = .235$  and margin of error = .035
- ☐ D.  $p = .235$  and margin of error = .07

137. A poll is done to estimate the proportion of adult Americans who like their jobs. The poll is based on a random sample of 400 individuals. What is the “conservative” margin of error of this poll?

- ☒ A. 0.10
- ☐ B. 0.05
- ☐ C. 0.04
- ☐ D. 0.025

$$= \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{400}} = \frac{1}{20} = 0.05$$

138. In a randomly selected sample of 500 students at a University, 80% say they plan to graduate within 4 years. Calculate a 95% confidence interval for the proportion of all students who plan to graduate within 4 years.

(using conservative margin of error)

$$.8 \pm \frac{1}{\sqrt{n}} \quad \text{i.e.} \quad .8 \pm .045$$

139. In a past General Social Survey, responders to the question "Have you ever been beaten or punched by another person" gave the following 95% confidence intervals for the population proportions of women and men who would say yes as .195 to .265 and .493 to .587, respectively. The sample sizes for men and women were large (444 men and 591 women). Is it reasonable to conclude that there is a difference in the population between men and women in the proportion that have ever been beaten or punched by another person?

**Questions 1 to 4:** For each statement, determine if the statement is a typical null hypothesis ( $H_0$ ) or alternative hypothesis ( $H_a$ ).

140. There is no difference between the proportion of overweight men and overweight women in America.

- ☒ A. Null Hypothesis
- ☐ B. Alternative Hypothesis

141. The proportion of overweight men is greater than the proportion of overweight women in America.

- ☐ A. Null Hypothesis
- ☒ B. Alternative Hypothesis

142. The average price of a particular statistics textbook over the internet is the same as the average price of the textbook sold at all bookstores in a college town.

- ☒ A. Null Hypothesis
- ☐ B. Alternative Hypothesis

143. The average time to graduate for an undergraduate English major is less than the average time to graduate for a history major.

- ☐ A. Null Hypothesis
- ☒ B. Alternative Hypothesis

144. If the results of a hypothesis test give a p-value of 0.03, then with a significance level ( $\alpha$ ) = .05 the results are said to be

- ☐ A. Not statistically significant because  $p \leq \alpha$
- ☒ B. Statistically significant because  $p \leq \alpha$
- ☐ C. Practically significant because  $p \leq \alpha$
- ☐ D. Not practically significant because  $p \leq \alpha$

145. The likelihood that a statistic would be as extreme or more extreme than what was observed is called a

- ☐ A. statistically significant result
- ☐ B. test statistic
- ☐ C. significance level
- ☒ D. p-value

146. The designated level (typically set at .05) for which the p-value is compared to decide whether the alternative hypothesis is accepted (or not) is called a

- ☐ A. statistically significant result
- ☐ B. test statistic
- ☒ C. significance level
- ☐ D. none of the above

147. When the p-value is less than or equal to the designated level of 0.05, the result is called a

- ☒ A. statistically significant result
- ☐ B. test statistic
- ☐ C. significance level



D. none of the above

148. Which of the following conclusions is NOT equivalent to rejecting the null hypothesis?

- A. The results are statistically significant.
- ☒ B. The results are not statistically significant.
- C. The alternative hypothesis is accepted.
- D. The  $p$ -value  $\leq$  alpha (the significance level)

149. If the result of a hypothesis test for a proportion is statistically significant, then

- ☒ A. The null hypothesis is rejected.
- B. The alternative hypothesis is rejected.
- C. The population proportion must equal the null value.
- D. None of the above

150. The smaller the  $p$ -value, the

- A. stronger the evidence against the alternative hypothesis
- B. stronger the evidence for the null hypothesis
- ☒ C. stronger the evidence against the null hypothesis
- D. none of the above

**Next 4 questions:** The significance level for a hypothesis test is  $\alpha = 0.05$ .

151. If the  $p$ -value is 0.001, the conclusion is to

- ☒ A. Reject the null hypothesis.
- B. Accept the null hypothesis
- C. Not reject the null hypothesis
- D. None of the above

152. If the  $p$ -value is 0.049, the conclusion is to

- ☒ A. Reject the null hypothesis.
- B. Accept the null hypothesis
- C. Not reject the null hypothesis
- D. None of the above

153. If the  $p$ -value is 0.500, the conclusion is to

- A. Reject the null hypothesis.
- B. Accept the null hypothesis
- ☒ C. Not reject the null hypothesis
- D. None of the above

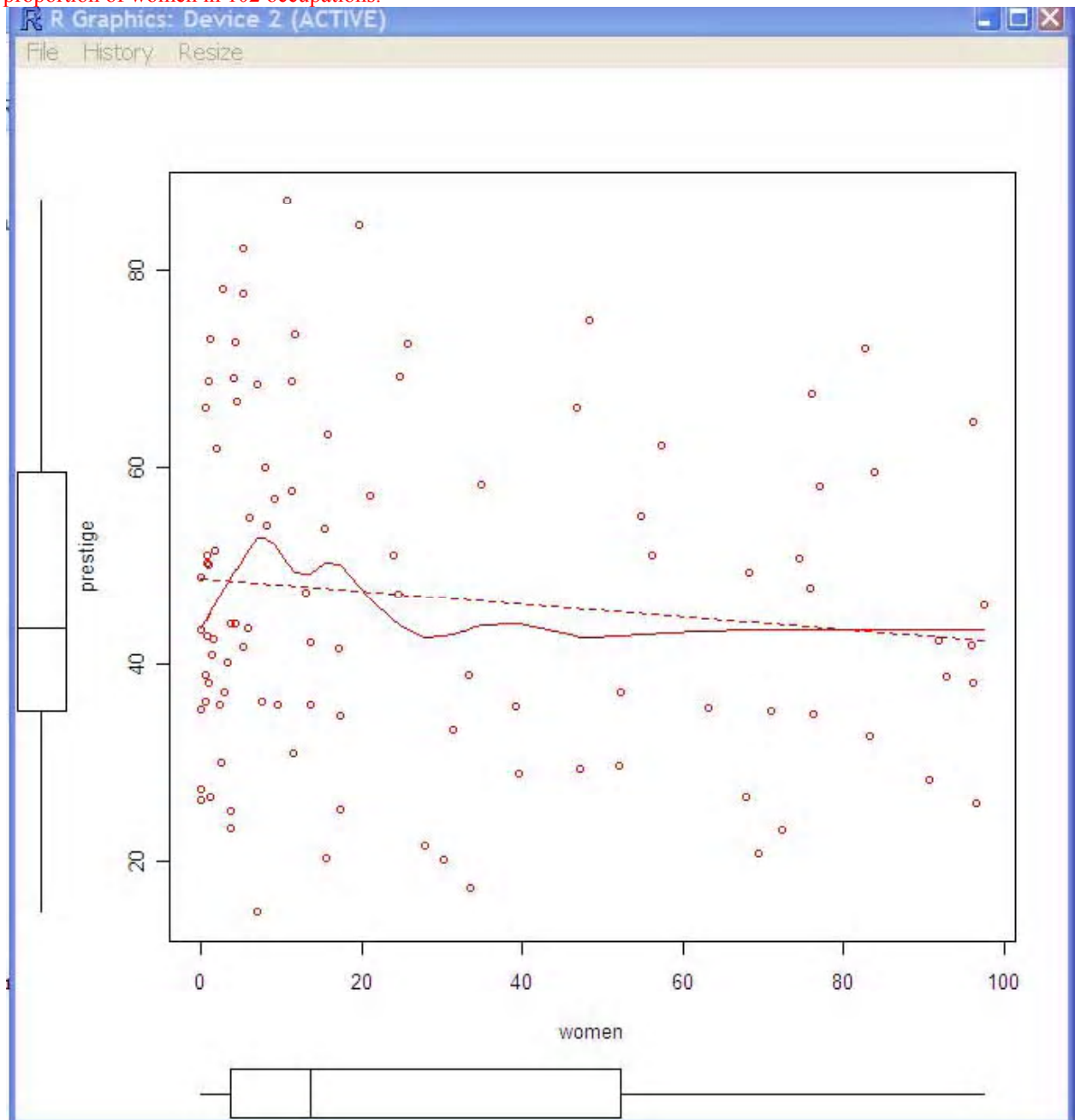
154. In hypothesis testing, a Type 1 error occurs when

- A. The null hypothesis is not rejected when the null hypothesis is true.
- ☒ B. The null hypothesis is rejected when the null hypothesis is true.
- C. The null hypothesis is not rejected when the alternative hypothesis is true.
- D. The null hypothesis is rejected when the alternative hypothesis is true.

155. In hypothesis testing, a Type 2 error occurs when

- A. The null hypothesis is not rejected when the null hypothesis is true.
- B. The null hypothesis is rejected when the null hypothesis is true.
- ☒ C. The null hypothesis is not rejected when the alternative hypothesis is true.
- D. The null hypothesis is rejected when the alternative hypothesis is true.

**Next 4 questions.** You perform a regression analysis to study the relationship between the prestige and the proportion of women in 102 occupations.



## Output Window

Submit

```

Call:
lm(formula = prestige ~ women, data = Prestige)

Residuals:
    Min       1Q   Median       3Q      Max
-33.444 -12.391  -4.126  13.034  39.185

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.69300    2.30760   21.101  <2e-16 ***
women       -0.06417    0.05385   -1.192    0.236
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17.17 on 100 degrees of freedom
Multiple R-squared:  0.014, Adjusted R-squared:  0.004143
F-statistic:  1.42 on 1 and 100 DF,  p-value: 0.2362

```

- same sign

156. What is the correlation between prestige and the proportion of women?

$$r = -\sqrt{0.015} = -0.118$$

157. Is there evidence of a relationship between prestige and the proportion of women in occupations?

p-value = 0.236 so we would say that

158. Is it reasonable to use a simple linear regression to investigate this relationship?

there is no such evidence.  
Looking at the scatterplot, points are roughly evenly distributed above and below the fitted regression line and no other pattern is apparent in the scatterplot. Therefore, it is reasonable to use a simple linear regression.