

Quizzes to accompany *Understanding Advanced Statistical Methods*, by Westfall and Henning.

Chapter 1:

1. What is true about a statistical model?  
A. The model produces data B. The data produces the model C. The model is deterministic
2. Where do you find a parameter?  
A. In the model B. In the data C. In the future
3. There is Nature, Design and Measurement, and DATA. What is “design”?  
A. A plan to collect DATA B. The type of DATA to be collected C. The process that is studied
4. Simulation is used to  
A. Produce the future B. Produce many potential futures
5. Give a definition of a statistical model.

6. You tabulated the results of a survey as follows:

Response	No	Yes	No opinion
Percent	51.2%	30.8%	18.0%

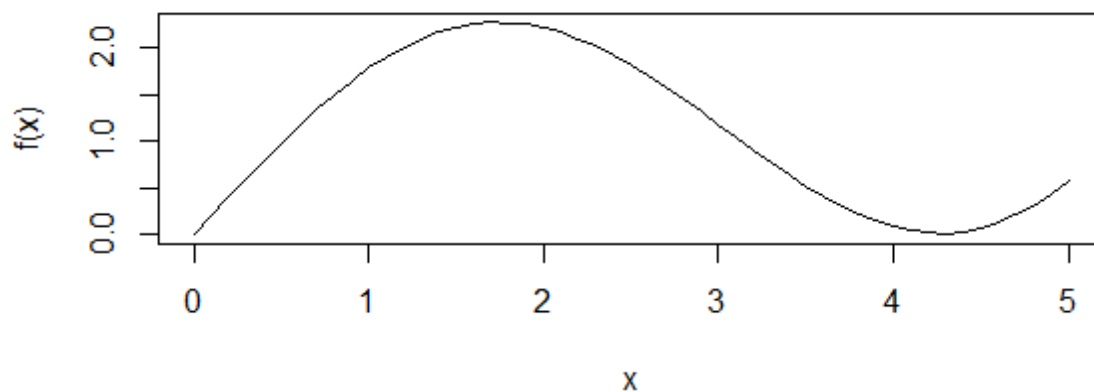
This table is called

- A. data B. DATA C. The design D. The model that produced the DATA
7. See the table in question 6. The measurement (No, Yes, No opinion) is  
A. nominal B. ordinal C. continuous
  8. When are parameters of a real data-generating process known?  
A. Before you collect the data B. After you collect the data C. Never
  9. Give an application of simulation that was discussed in Chapter 1.  
A. Minimizing queuing time B. Predicting presidential election outcomes  
C. Target marketing D. Audit sampling
  10. In the trading strategy simulation example, what was found?  
A. Stock prices are deterministic B. The best strategy is “buy and hold”  
C. The distribution of returns is Poisson
  11. In the Presidential election simulation example, what numbers were simulated using random number generators?  
A. Number of electoral votes received by a candidate  
B. Polling percentage data from a sample of voters  
C. The length of time a candidate stays in office

12. Pick the correct statement:  
A. A statistical model produces DATA    B. A statistical model is determined by data  
C. DATA produces the statistical model
13. As described in Chapter 1, what is “design”?  
A. A plan to collect DATA    B. The type and units of the DATA that will be collected  
C. The natural domain of the study    D. The observed data values
14. The model  $y = .01x$  is what kind of model?  
A. Deterministic  
B. Probabilistic  
C. A model with both deterministic and probabilistic components
15. In the model  $y = \theta x$ , which one is a parameter?  
A.  $y$     B.  $x$     C.  $\theta$
16. What model was used for stock returns in the book?  
A. A normal distribution model  
B. A capital assets pricing model  
C. A Black-Scholes martingale model
17. How were presidential elections forecast in the book?  
A. Using a simulation model  
B. Using a regression model  
C. Using a time series model
18. You plan to measure a person’s baseball knowledge using a survey, but have not done so yet, so you do not know how they will respond. The information you will collect from this person is called  
A. DATA    B. data    C. Nature    D. Design
19. What is a “good” model?  
A. One that produces realistic DATA  
B. One that is estimated using advanced statistical methods  
C. One that is statistically significant
20. What is “design”?  
A. The reality of the situation    B. A plan to collect DATA    C. A model that produces DATA
21. In the model  $y = \alpha x$ , which one is a “parameter”?  
A.  $y$     B.  $x$     C.  $\alpha$

## Chapter 2:

1. Let  $f(x) = x^2 + 1$ . Derive  $f'(x)$ . Give a brief reason for each step.
2. Let  $f(x) = x^2 + 1$ . Find  $\int f(x)dx$ . Give a brief reason for each step.
3.  $\partial(2)/\partial x =$   
A. 0      B.  $x^2/2$       C. 2      D.  $2x$
4.  $\int 2dx =$   
A. 0      B.  $x^2/2$       C. 2      D.  $2x$
5. Pick the true statement:  
A. The derivative of the cdf equals the pdf.  
B. The derivative of the pdf equals the cdf.
6. The  $U(0, 1)$  distribution is a \_\_\_\_\_ distribution.  
A. discrete      B. normal      C. continuous
7. All numbers in the following set can be produced by the Poisson distribution:  
A.  $\{0, 1, 2, \dots\}$       B.  $\{0, \pm 1, \pm 2, \dots\}$       C.  $\{y; -\infty < y < \infty\}$       D.  $\{y; 0 < y < \infty\}$
8. What is a “Rule of thumb”?  
A. A mathematical fact      B. A rough guideline      C. A requirement for publication
9. In Section 2.5, the physicist said “I am approximately \_\_\_\_\_”  
A. normally distributed      B. 1.3 years old      C. an integral      D. a light bulb
10. Consider the following graph of a function  $f(x)$ .



- Then  $f'(3) =$
- A. -1.4      B. 0      C. 1.2      D. 2.4

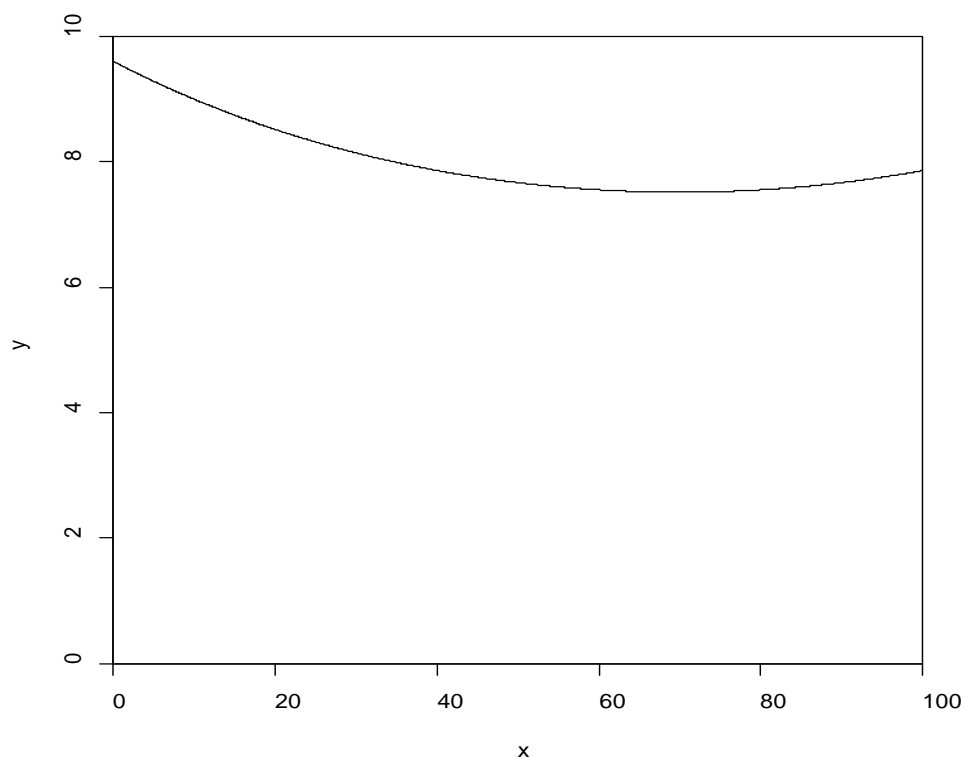
11. The expression  $P(y)$  denotes the cumulative distribution function of a random variable  $Y$ . Then  $P(y) =$   
A.  $\partial \Pr(Y \leq y) / \partial y$    B.  $\Pr(Y \leq y)$    C.  $\int \Pr(Y \leq y) dy$

12. Two points are  $(x_1, y_1) = (2, 3)$ , and  $(x_2, y_2) = (3, 5)$ . What is the slope of the line that connects those two points?  
A. 0.5   B. 1.0   C. 2.0

13. Why are the earth and the sun in “approximately” the same place?  
A. Because they are in the same solar system.  
B. Because the sun’s light heats the earth.  
C. Because it seems that way to a space alien living in a distant galaxy.

14. What is  $\Pr(Y \leq y)$  called?  
A. A probability density function.  
B. A probability mass function.  
C. A cumulative probability distribution function.

15. Guess the area under the curve shown below. Guess: \_\_\_\_\_



16. Identify the following types of DATA as either discrete or continuous (10 points each)
- A. Measured number of mobile devices owned by a student: \_\_\_\_\_
  - B. Actual number of mobile devices owned by a student: \_\_\_\_\_
  - C. Measured time of telephone conversation between two students: \_\_\_\_\_
  - D. Actual time of telephone conversation between two students: \_\_\_\_\_
17. How can you transform a continuous random variable into a discrete one?
- A. Use the logarithmic transformation      B. Use the square root transformation
  - C. Use the linear transformation              D. Round it off to the nearest digit
18. What is the sample space for the Bernoulli distribution?
- A.  $\{0, 1\}$     B. All numbers between 0 and 1.    C.  $\{\mu, \sigma\}$     D.  $\{\pi\}$
19. Suppose  $\int p(y) dy = 1$ . Then  $p(y)$  is a \_\_\_\_\_ pdf.
- A. Discrete    B. Continuous    C. Poisson    D. Bernoulli
20. The colors of cars chosen by 100 car-buyers is an example of \_\_\_\_\_ DATA.
- A. Continuous    B. Time series    C. Nominal
21. The best model for time you have to wait in line is a \_\_\_\_\_ model.
- A. Continuous    B. Time series    C. Nominal
22. If  $p(y)$  is a continuous distribution, then  $p(80) =$  \_\_\_\_\_
- A. 0    B. 1.0    C. 0.5
23. How do you calculate the area of a rectangle?
- A. Base  $\times$  Height    B.  $(1/2) \times$  Base  $\times$  Height    C.  $\int$  Base  $\times$  Height  $dx$
24. The derivative of a function is equal to
- A. the area under the curve.
  - B. the slope of the tangent line.
  - C. the intercept of the tangent line.
25. Joe says two things are “approximately” equal. Then
- A. they are equal. B. they are very different. C. they are “close enough” by anyone’s standard.
  - D. they are “close enough” by Joe’s standards.

26. The cumulative distribution function,  $P(y)$ , is equal to

- A. the slope of the probability distribution function at  $y$ .
- B. the area to the left of  $y$  under the probability distribution function.
- C. the triangular distribution evaluated at  $y$ .
- D. the number in the normal distribution table in the back of the book corresponding to  $y$ .

27. Evaluate  $\int 0.0002y \, dy$

- A.  $y$     B.  $0.0002y$     C.  $0.0002y^2$     D.  $0.0001y^2$

28. Nominal DATA is always

- A. normally distributed    B. discrete    C. continuous

29. What is a "sample space"?

- A. the set of possible DATA values
- B. the set of observations in the data set
- C. the distribution of the DATA

30. Here is a distribution, in list form.

$y$	$p(y)$
0	$1 - \pi$
1	$\pi$

Total:    1.0

What is this distribution?

- A. Poisson    B. Normal    C. Bernoulli    D. Uniform    E. generic

31. Suppose  $x/2 > 20$ . Then

- A.  $x < 10$     B.  $x > 10$     C.  $x < 40$     D.  $x > 40$

32. Find  $\partial(2 + 3x)/\partial x$ .

- A. 2    B. 3    C.  $x$     D. 5

33. Find  $\int_0^1 y \, dy$ .

- A. 0.5    B. 1.0    C. 2.0    D.  $\infty$

### Chapter 3:

1. Explain the advantage of integration over simulation for calculating probabilities.
2. The “inverse cdf method” is used to generate random numbers from any continuous distribution. The method first requires
  - A. a sample from the normal distribution with mean 0 and variance 1.
  - B. a sample from the exponential distribution with mean  $\lambda=1$ .
  - C. a sample from the Bernoulli distribution with  $p=0.5$ .
  - D. a sample from the uniform distribution over the (0,1) range.
3. What is the “inverse cdf method” used for?
4. What does the expression  $\Pr(Y \in A)$  refer to?
  - A. The probability of the sample space
  - B. The probability of an event
  - C. The quantile function
  - D. The Poisson distribution
5. What does  $\frac{\#Y_i^* \in A}{NSIM}$  refer to?
  - A. The integral of  $p(y)$  over the range from 0 to  $NSIM$ .
  - B. An estimate of a probability
  - C. A deterministic model’s output
  - D. The fact that a statistical model has unknown parameters
6. What is the inverse cdf method used for?
  - A. To find the parameters of the model
  - B. To generate random data
  - C. To relate a dependent variable  $Y$  to an independent variable  $X$
  - D. To find the mean of the pdf
7. What does the uniform distribution look like, when drawn in a graph?
  - A. A triangle
  - B. A bell curve
  - C. A flat line
  - D. An upside-down “U” shape
8. What kinds of numbers do you plug into the inverse cdf function?
  - A. mean values
  - B. uniform random numbers
  - C. normal distribution function values
  - D. integrals of pdf functions
9. You have seen  $p(y)$  and  $P(y)$ . What is  $P(y)$ ?
  - A. A probability distribution function
  - B. A cumulative distribution function
  - B. A quantile function
  - D. A random number function
10. What does the symbol “ $\Pr(Y \in A)$ ” refer to?
  - A.  $Y$  is predicted to be in the set  $A$ .
  - B. The probability of  $Y$  is equal to  $A$ .
  - C. The probability that  $Y$  is in the set  $A$ .
  - D. The probability that  $Y$  is less than  $A$ .

## Chapter 4:

1. The “stoplight example” shows a case where data values come from which distribution?  
A. Normal B. Bernoulli C. Poisson D. Uniform
2. The “stoplight example” creates a probability distribution for which measurement?
3. Your data set consists of either “Success” or “Failure” of business ventures. Select the appropriate distribution that might have produced these data.  
A. Normal B. Poisson C. Bernoulli D. Uniform E. Exponential
4. Which distribution was suggested as a possible distribution for the number of insects found in traps?  
A. Normal B. Poisson C. Bernoulli D. Uniform E. Exponential
5. The quantile of a distribution is the  
A. Mean B. Standard Deviation C. Inverse pdf D. Inverse cdf E. Derivative of the pdf
6. What does a uniform distribution look like, when drawn in a graph?  
A. A triangle B. A bell curve C. A flat line D. You can’t tell from the information given
7. What does a generic distribution look like, when drawn in a graph?  
A. A triangle B. A bell curve C. A flat line D. You can’t tell from the information given
8. Explain what the 0.30 quantile of a distribution (also called  $y_{0.30}$ ) is in a real example of your own choosing. That means reality, not simulation or calculus. Two sentences maximum.
9. Pick the best model for  $Y = \text{grade on this quiz}$ .  
A. Bernoulli B. Normal C. Poisson D. Uniform E. Generic
10. What happens to histograms with a larger sample size,  $n$ ?  
A. They become closer to a normal distribution.  
B. They become closer to an exponential distribution.  
C. They become more accurate estimates of the pdf that produced the data.
11. What information is not shown by a histogram?  
A. The center of the data  
B. The range of the data  
C. The strength of relationship between  $X$  and  $Y$
12. According to an ugly rule of thumb, how many intervals should you pick for the histogram?  
A.  $n^{1/2}$  B. 30 C. no more than 5
13. What does the “ $q$ ” refer to in a “ $q$ - $q$  plot”?  
A. query B. question C. quantile
14. The data produced by  $p(y)$  are 3.2, 3.5, 4.2. Then 4.2 is an estimate of the \_\_\_\_ percentile of  $p(y)$ .  
A. 100<sup>th</sup> B. 50<sup>th</sup> C. 83.3<sup>rd</sup>
15. What the ideal appearance of the  $q$ - $q$  plot?  
A. A flat line B. An “S”-shaped curve C. A 45° line



16. Pick the most appropriate model for the outcome ( $Y = 0$  for Tails,  $Y = 1$  for Heads) of a toss of a bent coin.

- A.  $\pi$     B.  $p(y)$     C. 0.5    D. Bernoulli( $\pi$ )    E. Bernoulli(0.5)

17. Which distribution is called “generic”?

- A.  $p(y)$     B.  $y^2/2$     C.  $N(\mu, \sigma^2)$     D. Exponential( $\lambda$ )    E. Uniform

18. Which distribution models time until the light turns red in the stoplight example?

- A. Poisson(0.5)    B. Student's  $T$     C.  $N(\mu, \sigma^2)$     D. Exponential( $\lambda$ )    E.  $U(0, \theta)$

19. A histogram is an estimate of

- A. a mean    B. a pdf    C. a cdf    D. an inverse cdf

20. When there is “^” on top of a symbol, then the resulting symbol is called

- A. a true value  
B. an estimate of a true value  
C. a mean value  
D. a quantile value

21. Why is there a “-0.5” in the symbol  $\hat{y}_{(i-.5)/n}$ ?

- A. To divide the result by 2  
B. To correspond with the “-.5” in the exponent of the normal distribution  
C. Because your potential DATA can be larger than the maximum of your existing data

22. Even when the data are produced by a normal distribution, the histogram does not look like a perfect bell curve. Why not?

- A. Because the true distribution  $p(y)$  is skewed.  
B. Because the true distribution  $p(y)$  is discrete.  
C. Because of randomness.

23. Even when the data are produced by a normal distribution, the normal q-q plot does not look like a perfect straight line. Why not?
- A. Because the true distribution  $p(y)$  is skewed.
  - B. Because the true distribution  $p(y)$  is discrete.
  - C. Because of randomness.
24. Identify the most appropriate model for producing data that looks like the results of flipping a bent coin.
- A.  $N(\mu, \sigma^2)$     B.  $N(0, 1)$     C. Bernoulli( $\pi$ )    D. Bernoulli(0.5)
25. Identify the most appropriate distribution for producing data that looks like the number of insects caught in a trap.
- A. Normal    B. Bernoulli    C. Poisson    D. Uniform    E. Generic
26. Identify the most appropriate distribution for producing data that looks like the time until the light turns red in the stoplight example.
- A. Normal    B. Bernoulli    C. Poisson    D. Uniform    E. Generic
27. The histogram is an estimate of what?
- A. A continuous probability distribution function
  - B. A discrete probability distribution function
  - C. A continuous cumulative distribution function
  - D. A discrete cumulative distribution function
28. Which equation defines the  $p$  quantile,  $y_p$ ?
- A.  $\Pr(Y \leq y_p) = p$     B.  $\Pr(Y = y_p) = p$     C.  $\Pr(Y \geq y_p) = p$     D.  $\int_{\text{all } y} p(y) dy = p$
29. What are *order statistics*?

- A. statistics that show trends      B. the data sorted from smallest to largest  
C. the distribution in list form      D. the mean and the median

30. Why does it make sense that there a “-.5” in the equation  $\hat{y}_{(i-.5)/n} = y_{(i)}$ ?

- A. Because the average of two numbers is  $.5(y_1 + y_2)$ .  
B. Because there is a “-.5” in the exponent of the normal density function.  
C. Because the largest number in the data set is smaller than the largest possible number.

31. If the data generating process is a normal distribution, what explains the deviations from a perfect straight line in the  $q$ - $q$  plot?

- A. Imperfections in the design and measurement system.  
B. Invalid statistical assumptions.  
C. Discreteness.  
D. Chance alone.

32. When is a model “good”?

- A. When it is constructed from good DATA.  
B. When the root mean square error is low and  $R^2$  is high.  
C. When it produces DATA\* that look like DATA.

33. The \_\_\_\_\_ example was one where the form of the model determined by theory alone.

- A. real estate  
B. stock return  
C. stoplight  
D. human weight

34. A histogram is an estimate of a

- A. mean function  
B. quantile function  
C. cumulative distribution function  
D. probability distribution function

35. Suppose you have  $n = 100$  observations. Suggest how many intervals could be used for the histogram.
- A. 1     B. 10     C. 100     D. 1,000
36. Why is the q-q plot better than the histogram for diagnosing deviations from normality?
- A. The histogram is just an estimate; the q-q plot is the true distribution.  
B. The q-q plot shows tail behavior more clearly.  
C. The q-q plot shows the “bell shape” more clearly.
37. When are quantiles **not** completely defined?
- A. When the distribution is discrete.  
B. When the distribution is normal.  
C. When the distribution is continuous.
38. If there are  $n=5$  observations in the data set, then the smallest one is an estimate of which quantile?
- A. 0.00     B. 0.10     C. 0.20     D. 0.50
39. When data are sampled from the normal distribution (say, using R or EXCEL’s random number generators), the normal q-q plot does not look like a perfect straight line. What explains that?
- A. Chance alone.  
B. The distribution is truly non-normal.  
C. The q-q plot is expected to look like a bell curve when the distribution is normal.
40. Here are  $n = 4$  data values produced by a continuous pdf  $p(y)$ : 0.45, 2.12, -4.89, -3.16.
- 40A. Draw a graph of a continuous pdf  $p(y)$  that could have produced these data. Label and number both the horizontal and vertical axes.
- 40B. Draw a graph of a continuous pdf  $p(y)$  that could not have produced these data. Label and number both the horizontal and vertical axes.
41. What is another name for *quantile*?
- A. cumulative probability     B. random data value     C. average     D. percentile

42. The largest of  $n = 6$  data values is an estimate of which quantile of the distribution that produced the data?  
 A. 1    B. 11/12    C. 5/6    D. 2/3
43. What is the expected appearance of the normal q-q plot when the data are produced by a normal distribution?  
 A. A parabola    B. A needle plot    C. A straight line    D. A bell curve
44. What distribution produced the real stock return data?  
 A. non-normal    B. normal    C. poisson    D. uniform

## Chapter 5:

1. Give definition, ***in one sentence***, for the expression  $p(y | X=x)$ . ***Do not*** use mathematical expressions other than " $p(y | X=x)$ " in your definition.
2. Let  $X$  and  $Y$  be random variables. Give the mathematical definition of independence of  $X$  and  $Y$ , and explain what it means using a real example.
3. Explain the formula  $p(y) = \sum_{all\ x} p(x, y)$  in words.
4. Give an example showing how to use the formula  $p(y) = \sum_{all\ x} p(x, y)$ .
5. Nature favors  
 A. continuity over discontinuity    B. linearity over curvature  
 C. normal distributions over non-normal distributions
6. Use what you know to  
 A. predict what you know    B. predict what you don't know
7. Using a data set containing stock returns, the book showed that today's stock return is  
 A. dependent on yesterday's return    B. independent of yesterday's return  
 C. reasonably modeled as being independent of yesterday's return
8. Using a data set containing opinions about Barbara and George H.W. Bush, the book showed  
 A. there is dependence between opinions  
 B. the opinions are statistically independent

9. Let  $Y$  = a student's time to get to class (minutes), and let  $X$  be their distance to class, which is classified simply as either "far" or "near." Why are  $X$  and  $Y$  dependent? Answer in terms of distributions. Also draw graphs of these distributions.
10. Which mantra applies when deciding whether to use  $p(y|x)$  versus  $p(x|y)$ ?
- model produces data
  - nature favors continuity over discontinuity
  - use what you know to predict what you don't know
11. Suppose  $p(x,y)$  is a continuous joint distribution. Then \_\_\_\_\_ = 1
- $\iint p(x,y) \, dx \, dy$
  - $\iint p(y) \, dy$
  - $\iint p(y|x) \, dx \, dy$
12. Which formula gives you the conditional distribution  $p(y|x)$ ?
- $\iint p(y|x) \, dx \, dy$
  - $\int p(x,y) \, dx$
  - $p(x,y)/p(x)$
13. Suppose  $X \sim p(x)$  and  $Y \sim p(y)$ , independent of  $X$ . Then the joint distribution is given by  $p(x,y) = \underline{\hspace{2cm}}$ .
- $p(x)p(y)$
  - $\iint p(x,y) \, dx \, dy$
  - $p(x,y)/p(x)$
14. Suppose  $Y$  is smoking behavior (number of cigarettes per day), and  $X$  is an indicator of whether a person wears a nicotine patch (yes or no). What is the meaning of  $p(y|X=\text{yes})$ ? Two sentences maximum, with proper punctuation. No graphs.
15. In Chapter 5, the phrase " $X$  affects  $Y$ " meant that the \_\_\_\_\_ of  $Y$  depends on the \_\_\_\_\_ of  $X$ .
- value, value
  - value, conditional distribution
  - conditional distribution, value
  - conditional distribution, conditional distribution
16. If  $Y$  = a person's housing expense and  $X$  = the person's annual income, what happens to  $Y$  when  $X$  increases by \$1.00?
- Its value gets slightly bigger.
  - Its value stays the same.
  - Its conditional distribution shifts slightly to the right.
  - Its conditional distribution stays the same.

17. A die, when tossed, has  $\Pr(Y=1) = 1/6$ . Let  $X$  and  $Y$  be successive independent tosses.

Someone told you that  $X=1$ . What is  $\Pr(Y=1 \mid X=1)$  ?

- A.  $1/6$     B.  $2/6$     C.  $1/12$     D.  $1/36$

18. What was the conclusion about stock returns?

- A. Today's return is independent of yesterday's return.  
B. Today's return is nearly independent of yesterday's return.  
C. Today's return is strongly dependent on yesterday's return.

19. Here is a joint distribution of  $(X, Y)$ .

		$Y$		
		1	2	3
$X$	A	0.10	0.10	0.20
	B	0.05	0.25	0.10
	C	0.05	0.05	0.10

Give the marginal distribution of  $Y$  in list form

20.(20) What does the expression " $X \in A$ " mean?

- A.  $X$  is an element of  $A$     B.  $X$  is smaller than  $A$     C.  $X$  is larger than  $A$     D.  $X$  times  $\in$  equals  $A$

21.(20) What is the definition of independence of  $X$  and  $Y$ ?

- A. The conditional distribution of  $Y$  is the same for all  $X$ .  
B.  $Y$  is functionally unrelated to  $X$ .  
C.  $Y$  is the same number, no matter what is  $X$ .  
D. A statistical test of the hypothesis that  $Y$  and  $X$  are independent shows  $p > 0.05$ .

22. Let  $Y$  denote the number of minutes it takes to go from their home to school. Let  $X$  denote distance from campus. Draw a graphs of the continuous pdfs  $p(y \mid X = \text{far})$  and  $p(y \mid X = \text{near})$  on the same axes. Use your own personal judgment, aided by your own personal commute experience, when constructing the graph. Number and label both the horizontal and vertical axes. No words are needed, just the graph.

## Chapter 6:

1. The symbol  $p(x, y)$  denotes *joint distribution*. In the car color choice example, suppose  $p(\text{young, red}) = 0.25$ , or 25%. What does 25% mean, specifically, in this case? Answer by completing the following sentence.

“Approximately 25 out of 100 \_\_\_\_\_”

2. Pick the larger number.

A.  $\text{Pr}(\text{musician} \mid \text{mathematician})$     B.  $\text{Pr}(\text{mathematician} \mid \text{musician})$

3. Pick the larger number.

A.  $\text{Pr}(\text{death} \mid \text{drunk driving})$     B.  $\text{Pr}(\text{drunk driving} \mid \text{death})$

4. In the book, psychometric evaluation of employees was used to

A. predict mental illness    B. predict stealing behavior    C. predict employee fit with company

5. Which of the following is Bayes' theorem?

A.  $p(x \mid y) \propto 1/p(y \mid x)$     B.  $p(x \mid y) \propto p(y \mid x) p(x)$     C.  $p(x \mid y) \propto p(y \mid x) / p(x)$

6. Here is a joint pdf of  $(X, Y) = (\text{Sex}, \text{Beer Purchase})$  for supermarket customers.

		Y	
		No	Yes
X	Male	0.20	0.10
	Female	0.60	0.10

A. Give the marginal pdf of X.

B. Give the conditional pdf of Y, given that X = Male.

7. It was stated in the book that  $\text{Pr}(\text{Drunk} \mid \text{Fatality}) = 0.40$ . The number “0.40” can be interpreted as “40%”, or “40 out of 100.”

A. What does the number “100” refer to in this case? One phrase or sentence maximum.

B. What does the number “40” refer to in this case? One phrase or sentence maximum.

8. What does the “ $\propto$ ” symbol mean?

A. “is approximately equal to”

B. “is proportional to”

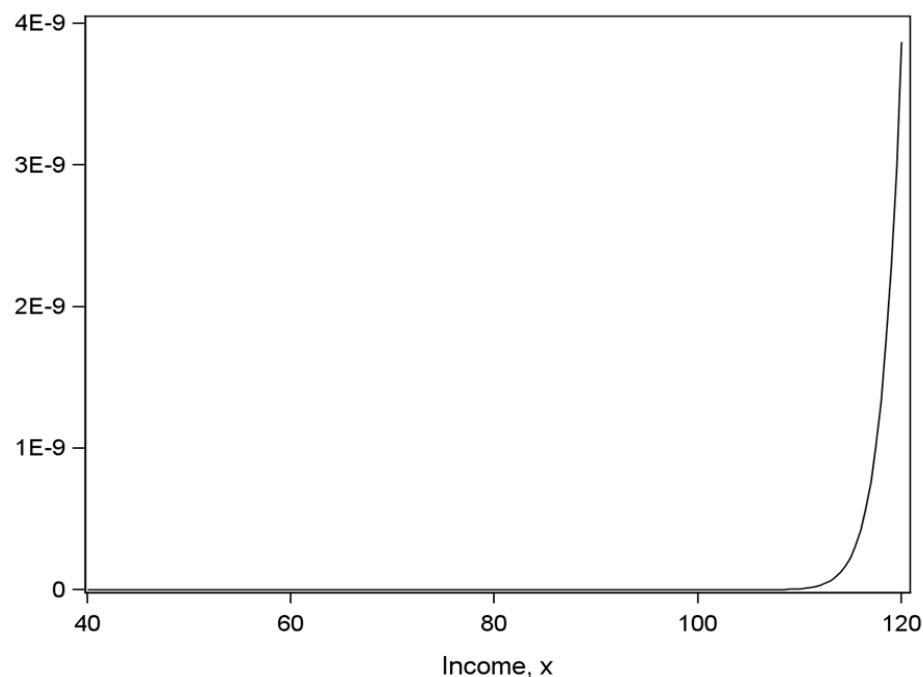
C. “is an estimate of”

9. Let X = drinking (Yes/No) and Y = killed while driving (Yes/No). If you are planning to drive drunk, then you want to know

A.  $p(y \mid x)$     B.  $p(x \mid y)$     C.  $p(x)$     D.  $p(y)$



10 . The following graph **not** a pdf because \_\_\_\_\_ .



- A. the area is less than 1      B. the area is greater than 1  
C. the density values are negative      D. the y axis is not labeled

11. What distribution was assumed for “employee fitness” in the psychometric evaluation case study?

- A. Normal    B. Uniform    C. Poisson    D. Generic

12. The symbol  $p(x, y)$  denotes joint distribution. In the car color choice example, suppose  $p(\text{young, red}) = 0.30$ , or 30%. What does 30% mean, specifically, in this case? Answer by completing the following sentence.

“About 30 out of 100

13.(20) If you drive somewhere while you are intoxicated, the probability that you will die is closest to  
A. 40%    B. 50%    C. 100%    D. 1 in a million

14.(20) What does the symbol “ $\propto$ ” that is used in Bayes’ Theorem mean?

- A. “significance level”    B. “ $p$ -value”    C. “is proportional to”    D. “is distributed as”

15. (40) In the age/car color choice example, it was found that  $p(\text{older} \mid \text{red}) = 17\%$ . Complete the sentence:

“About 17 out of 100 ...

16. Here is a joint distribution of  $(X, Y)$ .

		Y		
		1	2	3
X	A	0.10	0.10	0.20
	B	0.05	0.25	0.10
	C	0.05	0.05	0.10

Give the conditional distribution of  $Y \mid X = \text{“A”}$  in list form.

17.(20) Pick Bayes’ Theorem.

A.  $p(y|x) = p(y)/p(x)$     B.  $p(y|x) \propto p(y)/p(x)$     C.  $p(y|x) = p(x|y) p(y)$

D.  $p(y|x) \propto p(x|y) p(y)$

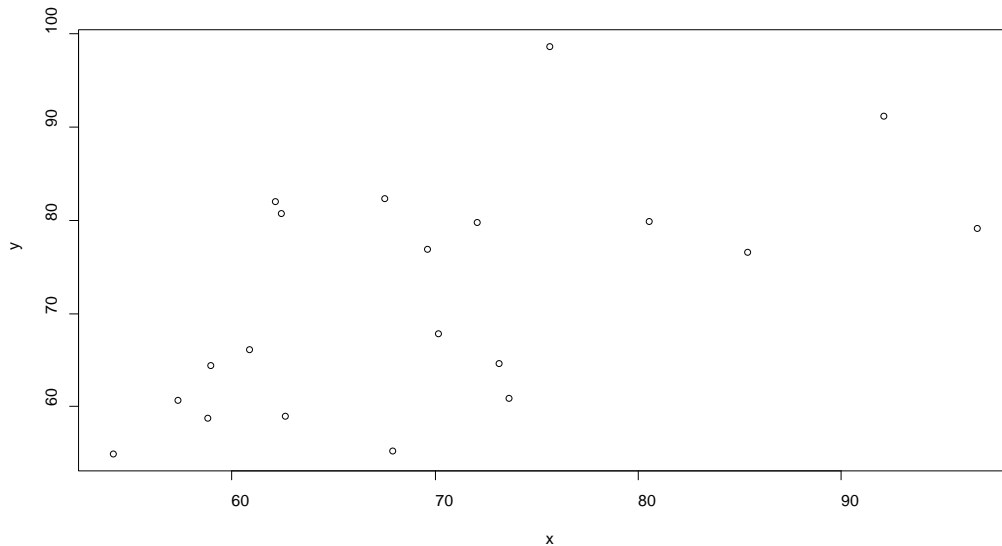
18.(20) In the “psychometric evaluation of employees” example, what distribution was assumed for employee fitness?

A. Normal    B. Bernoulli    C. Poisson    D. Generic

19. Which formula shows how the marginal distribution is related to the joint distribution?

A.  $\sum_{\text{all } x} \sum_{\text{all } y} p(x, y) = 1.0$     B.  $p(y|x) = \frac{p(x, y)}{p(x)}$     C.  $p(y) = \int_{\text{all } x} p(x, y) dx$     D.  $p(y) = p(y|x)$

20. Here is a graph.



This graph is called a

- A. joint distribution   B. marginal distribution   C. conditional distribution   D. scatterplot

21. The following table was given in the readings:

		Y			Total
		Red	Gray	Green	
X	Younger	0.50	0.20	0.30	1.00
	Older	0.20	0.40	0.40	1.00

This table displays

- A. a joint distribution   B. conditional distributions   C. marginal distributions

22. Which is an estimate of a joint distribution  $p(x,y)$ ?

- A. a bivariate histogram   B. a contingency table   C. a q-q plot   D. a regression curve

23. In the equation  $p(x|y) \propto p(y|x)p(x)$ , what is fixed (or constant)?

- A.  $p$    B.  $y$    C.  $x$    D.  $p(y|x)$

24. What is Bayes' Theorem used for?

- A. To find the marginal distribution of  $X$   
 B. To find the marginal distribution of  $Y$   
 C. To find the joint distribution of  $X,Y$

D. To find the conditional distribution of  $X|Y=y$

25. Which one is a prior distribution?

- A.  $p(x)$    B.  $p(x|y)$    C.  $p(y|x)$    D.  $p(x, y)$

26. The model relating  $Y$  = housing expense to  $X$  = income was called a

- A. Poisson distribution model  
B. uniform distribution model  
C. heteroscedastic regression model  
D. logistic regression model

## Chapter 7:

1. When sampling from a population, the process point of view states that

- A. The probability distribution  $p(y)$  is obtained from the population.  
B. The mean of the sample is equal to the process mean.  
C. The population is randomly selected from the process.  
D. The conditional distributions of  $Y|X=x$  are discrete distributions.

2. Select the true response.

- A. If  $Y_1, Y_2, \dots$  are iid, then they are a random sample from a population.  
B. If  $Y_1, Y_2, \dots$  are iid, then they are all produced by the same distribution.  
C. If  $Y_1, Y_2, \dots$  are iid, then they are normally distributed.  
D. If  $Y_1, Y_2, \dots$  are iid, then the conditional distribution of  $Y_2$  given  $Y_1$  is a histogram in the continuous case.

3. If  $Y$  randomly selected from the numbers 2.12, 3.55, 3.55, 5.13, then

- A.  $\Pr(Y = 3.55) = 0$    B.  $\Pr(2.12 < Y < 3.55) = 0$    C.  $\Pr(Y = 3.55) = 3.55$    D.  $\Pr(Y = 3.55) = 1/3$

4. If  $Y_1$  and  $Y_2$  are a simple random sample from the population of the numbers 2.12, 3.55, 3.55, 5.13, then  $Y_1$  and  $Y_2$  are

- A. independent but not identically distributed.  
B. identically distributed but not independent.  
C. independent and identically distributed.  
D. independent, and their distribution is a discrete distribution on the numbers 2.12, 3.55, 5.13.

5. People have different blood types, including O+, O-, A+, B+, and several others, some of which are very rare. What is the probability that a person in the US has blood type O+? Give the population definition. Two sentences maximum.

6. Which model is better for generalization?

- A. Population   B. Process

7. Which model has infinitely many sample sequences?

- A. Population   B. Process

8. What do the letters "iid" mean? (Be as brief as possible)

9. With cluster sampling, some observations are  
A. dependent    B. normally distributed    C. populations
10. How is  $\Pr(Y > 10)$  defined when using the population model?  
A.  $\int_{10}^{\infty} p(y)dy$     B.  $\#(y_i > 10)/N$     C.  $\#(y_i > 10)/n$
11. What is true about the conditional distributions  $p(y/x)$  in the population framework?  
A. They morph continuously    B. They allow generalization beyond the population  
C. They don't exist at all for some values of  $x$
12. Select the correct response.  
A. Nature produces the Population.    B. The Population produces Nature.
13. What problem is caused by nonresponse?  
A. Bias    B. Discreteness    C. Discontinuity
14. How come  $p(y|x)$  doesn't exist all for some cases in the population sampling framework? Give an example.
15. Roll a die 50 times; call the numbers  $Y_1, Y_2, \dots, Y_{50}$ . Roll it another 50 times and add ten to each outcome; call the numbers  $Y_{51}, Y_{52}, \dots, Y_{100}$ . Then the sequence  $Y_1, Y_2, \dots, Y_{100}$  is  
A. independent and identically distributed.  
B. independent but not identically distributed.  
C. identically distributed but not independent.  
D. neither independent nor identically distributed.
16. When  $Y_1, Y_2, \dots$  are sampled without replacement from a population, they are  
A. slightly dependent    B. independent
17. Cluster sampling produces DATA values that are  
A. dependent    B. independent
18. A lag scatterplot is a graph of the pairs  
A.  $(Y_{t-1}, Y_t)$     B.  $(X_t, Y_t)$     C.  $(t, Y_t)$
19. (20) The population size  $N$  is given as a capital letter because \_\_\_\_\_  
A. it is a random variable    B. it has a normal distribution    C. it is bigger than the sample size  $n$
20. (20) There are 500 people in a population, 50 of whom like tea better than all other drinks. Using the population definition, the probability that a person's favorite drink is tea is  
A. 50    B. 0.05    C. 10    D. 0.10
21. Five problems with the population definition of a probability model were given in the reading. Give me two of them (25 points for each).
22. The normal distribution is a \_\_\_\_\_ distribution.  
A. population    B. process    C. discrete    D. deterministic
23. Generalization is easier with a \_\_\_\_\_ distribution.  
A. population    B. process    C. discrete    D. deterministic

24. Roll a die three times and get  $X_1$ ,  $X_2$  and  $X_3$ . Let  $Y_1 = X_1 + X_2$  and let  $Y_2 = X_1 + X_3$ . (Note that  $X_1$  is involved in both  $Y_1$  and  $Y_2$ .) Then  $Y_1$  and  $Y_2$  are

- A. independent and identically distributed
- B. independent but not identically distributed
- C. identically distributed but not independent
- D. neither independent nor identically distributed

25. How do you prove that data are produced by an iid process?

- A. Draw the lag scatterplot.
- B. Draw histograms of separate groups of data.
- C. Use the iid test statistic.
- D. None of the above. It is impossible to prove that data are produced by an iid process.

26. Suppose 100 items are randomly sampled from 1000 items. The average value of the 100 items is \$24.50. Then the average value of the 1000 items is closest to

- A. \$2.45
- B. \$24.50
- C. \$245.00

27. What does the symbol " $U(0,1)$ " refer to?

- A. Up 1 from 0
- B. A random sample
- C. The uniform distribution

28. The population distribution  $p(y)$  is

- A. always discrete
- B. sometimes discrete, sometimes continuous
- C. always continuous
- D. always a normal distribution

29. Pick the correct answer.

- A. Nature defines the population.
- B. The population defines nature.

30.(40) You will flip a coin and observe whether the outcome is "heads" or "tails." Explain, using one sentence only, why the population model is wrong for this experiment.

31.(20) There are  $N = 100$  items in a population, 10 of which are the letter "A", and 90 of which are some other letter. Using the population definition of probability, what is  $\Pr(A)$ ?

- A. 0.10
- B. about 10%
- C. An unknown parameter

- 32.(20) How does process relate to population?
- A. The population produces the process data
  - B. The process produces the population data
- 33.(40) In the inventory example, there was exactly one item in the population having Value = \$321.20. That item was a furniture item, not an electronics item. Let  $Y$  be the type of item, either furniture or electronics. Using the population definition of probability distribution, give the conditional distribution of  $Y \mid \text{Value} = \$321.20$ , in list form.
- 34.(20) Which model is sometimes continuous? (Select one answer only.)
- A. Population
  - B. Process
  - C. Poisson
35. (20) Which tool can be used to check for dependence in time series data? (Select one answer only).
- A. A lag scatterplot
  - B. A q-q plot
  - C. A histogram
36. You roll a six-sided die having six possible outcomes, 1,2,...,6, three times, getting  $X_0$ ,  $X_1$ , and  $X_2$ . Let  $Y_1 = X_0 + X_1$  and  $Y_2 = X_0 + X_2$ . Then  $Y_1$  and  $Y_2$  are \_\_\_\_\_.  
(Select any, or all, that apply. If none applies select nothing. 40 points)
- A. independent
  - B. identically distributed
  - C. normally distributed
  - D. discrete random variables
  - E. sampled without replacement
37. There are  $N=1,000$  people in a population. Of these, 300 are 14 years old, 250 are 15 years old, 250 are 16 years old, and 200 are 17 years old. Letting  $Y = \text{Age}$ , give the population  $p(y)$  in list form.
38. Suppose you flip a bent coin ten times, in an ordinary way getting  $Y_1, Y_2, \dots, Y_{10}$ . Each  $Y_i$  is either "Heads" or "Tails". What is the appropriate model for how  $Y_1, Y_2, \dots, Y_{10}$  will appear?
- A. They are independent and identically distributed.
  - B. They are independent but not identically distributed.
  - C. They are identically distributed but not independent.
  - D. They are neither independent nor identically distributed.

39. Suppose you roll a die, getting  $X$ , either 1,2,3,4,5, or 6. Then you roll a different die repeatedly, getting  $Z_1, Z_2, Z_3, \dots$ . Then you add the same  $X$  to each  $Z_i$ , getting  $Y_1 = X + Z_1, Y_2 = X + Z_2, Y_3 = X + Z_3, \dots$ . What is the appropriate model for how  $Y_1, Y_2, Y_3, \dots$ , will appear?
- They are independent and identically distributed.
  - They are independent but not identically distributed.
  - They are identically distributed but not independent.
  - They are neither independent nor identically distributed.
40. Suppose you roll a die 5 times in an ordinary way, getting  $Y_1, Y_2, Y_3, Y_4, Y_5$ . Each of these  $Y$ 's is some number 1,2,3,4,5 or 6. Then you flip a bent coin 5 times in an ordinary way, getting  $Y_6, Y_7, Y_8, Y_9, Y_{10}$ . Each of these  $Y$ 's is either "Heads" or "Tails". What is the appropriate model for how  $Y_1, Y_2, \dots, Y_{10}$  will appear?
- They are independent and identically distributed.
  - They are independent but not identically distributed.
  - They are identically distributed but not independent.
  - They are neither independent nor identically distributed.
41. Suppose you sample  $Y_1, Y_2, \dots, Y_{10}$  as a with replacement sample from the numbers  $\{1,4,5,10,20\}$ . What is the appropriate model for how  $Y_1, Y_2, \dots, Y_{10}$  will appear?
- They are independent and identically distributed.
  - They are independent but not identically distributed.
  - They are identically distributed but not independent.
  - They are neither independent nor identically distributed.

## Chapter 8:

- The expression " $p(y | X=x)$ " refers to
  - The distribution of  $Y$  when the value of  $X$  is known.
  - The distribution of  $Y$  when the value of  $X$  is unknown.
  - The distribution of  $y$  when the value of  $X$  is known.
  - The distribution of  $y$  when the value of  $X$  is unknown.
- The "Law of Large Numbers" states what?
  - For large sample sizes, the average has an approximately a normal distribution.
  - For large mean values, the average has an approximately a normal distribution.
  - For large sample sizes, the average is approximately equal to the expected value.
  - For large mean values, the average is approximately equal to the expected value.
- What is the "Flaw" in the "Flaw of Averages"?
  - The assumption that the average has a normal distribution.
  - The assumption that the average converges to the expected value.
  - The assumption that average is a random variable.
  - The assumption that average can substitute for individual observations.



4. A distribution is  $p(y) = (1/16)y^2 e^{-y/2}$ , for  $y \geq 0$ . Write down the mathematical formula for  $E(Y)$  for this distribution. Don't compute its actual value.
5.  $E(Y)$  is determined by  
 A. the data    B. the population    C. the model  $p(y)$
6. The expected value of a  $U(0,1)$  random variable is  
 A. 0.0    B. 0.5    C. 1.0
7. The Law of Large Numbers states that  
 A. with larger  $n$ , averages are approximately normally distributed  
 B. with larger  $n$ , averages are closer to the expected value  
 C. with larger  $n$ , expected values are approximately normally distributed
8. What are the assumptions of the Law of Large Numbers? Select all that apply.  
 A. The  $Y_i$  are produced by a distribution  $p(y)$ .  
 B. The  $Y_i$  are independent.  
 C.  $E(Y_i) < \infty$   
 D. The  $Y_i$  are normally distributed  
 E. The  $Y_i$  are a random sample from a population
9. What is the preferred strategy for winning at roulette?  
 A. Bet the amounts 1, 2, 4, 8, 16, ... in succession  
 B. Bet on Red after Black has come up 10 times in a row  
 C. Chant "only winners no losers" before the marble is rolled (This is the correct answer! It's a joke, but one that underscores the fact that there is no winning strategy for roulette)
10. What is a main justification for needing calculus in statistics, according to this chapter?  
 A. The area under the curve is an integral  
 B. Averages are approximately integrals  
 C. The normal distribution is a continuous distribution
11. Which formula gives the mean,  $\mu$ , of a discrete distribution?  
 A.  $\mu = (y_1 + y_2 + \dots + y_n)/n$     B.  $\mu = (y_1 + y_2 + \dots + y_N)/N$     C.  $\mu = \sum yp(y)$     D.  $\mu = \int yp(y)dy$
12. The mean of a continuous distribution  $p(y)$  is the value  $\mu$  where  
 A. the graph  $p(y)$  balances    B. half the  $Y$  values are lower    C. the curve  $p(y)$  has a maximum
13. The average of Bernoulli (0/1) random variables is always  
 A. a proportion    B. a population mean    C. normally distributed    D. an expected value
14. The bootstrap distribution is most similar to a  
 A. normal distribution    B. exponential distribution    C. population distribution

15. Which of the following are random? Circle letters for all that are random. (Five points per correct selection and non-selection.)

A.  $Y_1$     B.  $\mu$     C.  $p(y)$     D.  $P(y)$     E.  $\bar{y}$     F.  $\bar{Y}$     G. Expected value    H. Process Mean

16. (20) What is the expected value of the  $U(0,1)$  random variable? (Circle one letter only).

A. 0.0    B. 0.5    C. 1.0    D.  $\infty$

17. (20) How large does  $n$  have to be to ensure that  $\bar{y}$  is equal to  $\mu$ ? Give a *very brief* answer.

18. State the Law of Large Numbers.

19. Suppose  $Y$  is a continuous RV. What is  $E(Y)$ ?

- A. the average of a sample of data values    B. the sum of a sample of data values  
C. the probability distribution that produces  $Y$     D. an integral

20. Suppose  $Y_1, Y_2, \dots, Y_{1000}$  are produced as iid from the following pdf.

$y$	$p(y)$
0	0.3
1	<u>0.7</u>
	1.0

What is  $(Y_1 + Y_2 + \dots + Y_{1000})/1000$ ?

- A. exactly 0.3    B. approximately 0.3    C. exactly 0.7    D. approximately 0.7

21. What does the symbol " $\mu$ " refer to?

- A. the sample mean    B. a random variable    C. the expected value    D. the population mean

22. Here is a data set:  $\{3,3,4,1\}$ . Give the bootstrap distribution in list form.

## Chapter 9:

1. How does Jensen's inequality explain the "flaw" in "the law of averages"?

2. Suppose  $X \sim U(0, 1)$ . Let  $Y = X^2$ . Find  $\Pr(Y < 0.5)$ .

3. Suppose  $Y \sim U(0, 2)$ . Then  $\text{Var}(Y) =$

- A.  $1/3$     B. 1    C. 2    D. 4

4. Suppose  $Y \sim N(70, 10^2)$ . Then  $\text{Var}(Y) =$   
 A. 70 B. 10 C. 100
5. Give the formula for mean absolute deviation (M.A.D.) of a random variable  $Y$ .
6. Suppose a non-normally distributed  $Y$  has parameters  $\mu = 70$  and  $\sigma = 10$ . Then  $\Pr(50 < Y < 90)$  is  
 A. = 0.95 B. > 0.95 C. = 0.75 D. > 0.75
7. Suppose  $Y \sim U(0,1)$ . What is the distribution of  $T = -\ln(Y)$ ?  
 A. Normal B. Uniform C. Exponential D. Poisson
8. Suppose  $Y$  is the outcome of a fair die roll, having the discrete uniform distribution on the numbers 1, 2, ..., 6. What is the expected value of  $Y + 1$ ?  
 A. 3 B. 3.5 C. 4 D. 4.5 E. 5
9.  $E(Y^2)$  \_\_\_\_\_  $\{E(Y)\}^2$   
 A. < B. > C. =
10.  $E(|Y - \mu|)$  \_\_\_\_\_  $\sigma$   
 A. < B. > C. =
11. What is the variance?  
 A.  $\int (y - \mu)^2 p(y) dy$  B.  $\int (y - \mu) p(y) dy$   
 C.  $\int (y - \mu)^2 dy$  D.  $\int (y - \mu)^2 p(\mu) d\mu$
12. Kurtosis is a measure of the \_\_\_\_\_ of the distribution  
 A. Peak B. Tail C. Spread D. Symmetry
13. Suppose  $Y \sim p(y)$ , where  $p(y)$  is given below. Find the distribution of  $T$ , where  $T = (Y-1)^2$ . Show work briefly.
- |          |             |
|----------|-------------|
| <u>y</u> | <u>p(y)</u> |
| 0        | 0.4         |
| 1        | 0.4         |
| <u>2</u> | <u>0.2</u>  |
|          | 1.0         |
14. Expected commute time is  $E(X) = 15$  minutes. Expected time from waking to leaving the house is  $E(Y) = 90$  minutes. What famous property allows you to say that  $E(X+Y) = 105$  minutes?
15. Suppose  $E(Y) = 10$ . Then  $E(Y^2)$  \_\_\_\_\_  
 A. = 100 B. > 100 C. < 100
16. The variance of a random variable  $Y$  is given by  $\sigma^2 =$  \_\_\_\_\_  
 A.  $\{(y_1 - \mu)^2 + \dots + (y_n - \mu)^2\} / (n-1)$  B.  $\{(y_1 - \mu)^2 + \dots + (y_n - \mu)^2\} / N$  C.  $E\{(Y - \mu)^2\}$

17. What is the ugly rule of thumb for identifying outliers?

- A.  $|z| > 3$       B.  $n > 30$       C.  $\rho > 0.7$       D.  $p < .05$

18. What does large kurtosis indicate about a distribution?

- A. It is peaked      B. It is skewed      C. It is bell-shaped      D. It occasionally produces outliers

19. (20) What is the maximum value of  $Y = -\ln(X)$ , where  $X$  is a  $U(0,1)$  random variable?

- A. 0      B. 1      C. 1000      D. infinity

20. (60) Suppose  $Y$  has the distribution as shown below. Find  $E(Y^2)$ .

$y$	$p(y)$
1	0.5
2	0.5

21. (40) Show that  $f(x) = x^2$  is a convex function using calculus.

22. (40) If  $\mu = 100$  and  $\sigma = 10$ , then what does Chebychev's Inequality tell you about the percentage of DATA that will be between 80 and 120? (i.e., what percent of the DATA will be within plus or minus two standard deviations of the mean?)

23. The distribution of  $Y$  is as follows:

$y$	$p(y)$
1	0.2
2	0.2
3	0.3
4	0.3

Let  $X = (Y - 2)^2$ . Give the distribution of  $X$  in list form.

24. Suppose  $Y_1, Y_2, \dots, Y_n$  is a sequence of iid random variables produced by any probability distribution with finite variance. Then the Central Limit Theorem tells us that ...

25. According to Jensen's inequality,  $E(Y^2)$  \_\_\_\_\_  $\{E(Y)\}^2$ .

- A.  $>$       B.  $<$       C.  $=$       D.  $\sim$

26. The variance,  $\sigma^2$ , is defined as  $\sigma^2 = \underline{\hspace{2cm}}$ .

- A.  $E\{(Y - \mu)^2\}$     B.  $E\{(Y - \mu)\}^2$     C.  $\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$     D.  $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$

27. According to Chebyshev's inequality, \_\_\_\_\_% of the DATA will be within plus or minus two standard deviations of the mean.

- A. < 75%    B. > 75%    C. < 95%    D. > 95%

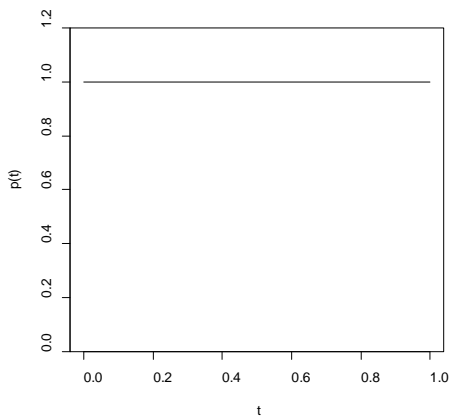
28. Kurtosis is a measure of the \_\_\_\_\_ of a distribution.

- A. outlier-producing potential    B. flatness    C. symmetry    D. pointiness

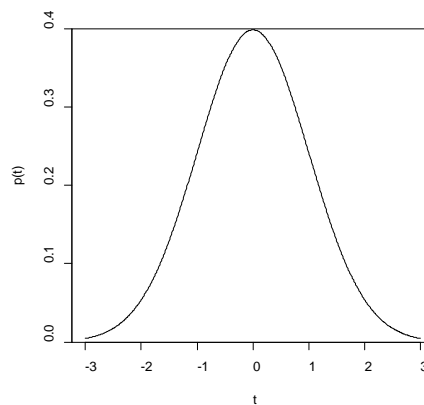
29. Which of the following is a function of the data, i.e., which is of the form  $f(Y_1, \dots, Y_n)$ ? Select all that apply. 10 points per correct selection/non-selection.

- A.  $\bar{Y}$     B.  $E(Y)$     C.  $(Y_1 + \dots + Y_n)^2$     D.  $\mu$

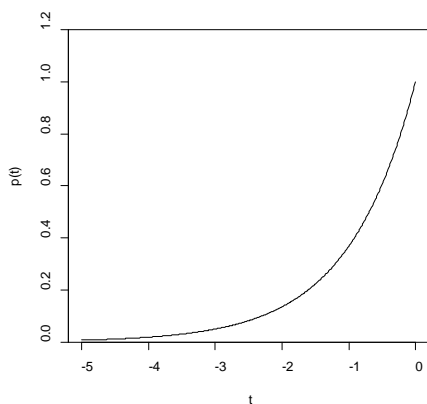
30. Suppose  $Y \sim U(0,1)$ , the uniform distribution between 0 and 1. Which is the pdf of  $T = \ln(Y)$ ?



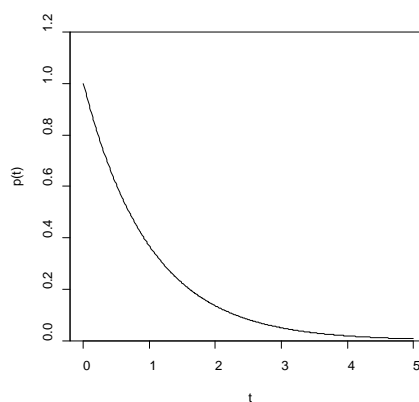
A.



B.



C.



D.

31. Suppose  $E(Y) = 3$ . Find  $E(2Y + 3)$ .

- A. 3    B. 6    C. 9    D.  $2\bar{Y} + 3$

### Chapter 10:

1. All theorems have the form “**If** certain assumptions are true, **then** certain conclusions are true.” State the Central Limit Theorem in that form as follows.

**If** (fill in the assumptions): \_\_\_\_\_

**Then** (fill in the conclusions): \_\_\_\_\_

2. Give the equation that defines the linearity property of expectation.

3. Give the equation that defines the additivity property of expectation.

4. Give the equation that defines the linearity property of variance.

5. Give the equation that defines the additivity property of variance, assuming independence.

6. Suppose  $X$  and  $Y$  are independent. *Select all that are true.* Ten points per correct selection.

- A.  $E(XY) = E(X)E(Y)$     B.  $E(X + Y) = E(X) + E(Y)$     C.  $\text{Var}(XY) = \text{Var}(X)\text{Var}(Y)$     D.  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$   
 E.  $\text{Stdev}(X + Y) = \text{Stdev}(X) + \text{Stdev}(Y)$

7. State the Central Limit Theorem.

8. Suppose that  $\text{Var}(Y_1) = 5$  and  $\text{Var}(Y_2) = 10$ . Let  $T = Y_1 - Y_2$ , with  $Y_1$  and  $Y_2$  independent. Note that  $T = Y_1 + (-1)Y_2$ , by simple algebra. Then  $\text{Var}(T) =$

- A. 5    B. 10    C. -5    D. 15

9. Suppose  $Y_1, Y_2, \dots, Y_{25} \sim_{\text{iid}} p(y)$ , with  $E(Y_1) = 0$  and  $\text{Var}(Y_1) = 1$ . Let  $T = Y_1 + Y_2 + \dots + Y_{25}$ . Pick a typical, but large value of  $T$ .  
A. 2   B. 10   C. 50   D. 200
10. How large does  $n$  have to be to make the distribution of the sample average approximately a normal distribution?  
A. 5   B. 30   C. 10,000   D. It depends on the distribution that produces the data
11. The expected value of the sum of two dice is 7.0. This is true because of  
A. The additivity property of expectation   B. The linearity property of expectation  
C. The central limit theorem   D. The law of large numbers
12. What assumption is required for the equation  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ ?  
A.  $X$  and  $Y$  are independent   B.  $X$  and  $Y$  are identically distributed   C.  $X$  and  $Y$  are normally distributed
13. What assumption is required for the equation  $E(XY) = E(X)E(Y)$ ?  
A.  $X$  and  $Y$  are independent   B.  $X$  and  $Y$  are identically distributed   C.  $X$  and  $Y$  are normally distributed
14. Suppose  $E\{(X - \mu_X)^2\} = 100$ ,  $E\{(Y - \mu_Y)^2\} = 4$ , and  $E\{(X - \mu_X)(Y - \mu_Y)\} = -10$ . Then the correlation between  $X$  and  $Y$  is  
A. 0.9   B. -2.0   C. -0.5   D. -0.025
15. If  $Y_1, Y_2, \dots, Y_n \sim_{\text{iid}} p(y)$ , where  $\text{Var}(Y_1) < \infty$ , then when  $n$  is large  
A. The distribution of  $p(y)$  is approximately a normal distribution  
B. The distribution of  $\mu$  is approximately a normal distribution  
C. The distribution of  $Y_1 + Y_2 + \dots + Y_n$  is approximately a normal distribution
- 16.(20) The symbol  $\sigma_{xy}$  in the text means  
A. Variance   B. Standard deviation   C. Covariance   D. Correlation
- 17.(20) Suppose  $\text{Var}(X) = 1$  and  $\text{Var}(Y) = 2$ , with  $X$  and  $Y$  independent. Then  $\text{Var}(X + Y) =$   
A. 1   B. 2   C. 3   D. 4   E. 5
18. (40) Finish the second sentence that follows. (Complete the sentence that is started, with no additional sentences.)
19. Suppose  $\text{Var}(Y) = 10$ . Let  $T = 5 - Y$ . Then  $\text{Var}(T) =$   
A. -10   B. -5   C. 0   D. 5   E. 10
20. Suppose  $\text{Var}(X) = 9$  and  $\text{Var}(Y) = 16$ , and  $X$  and  $Y$  are independent. Then  $\text{Var}(X+Y) =$   
A. 5   B. 10   C. 0   D. 25
21. Let  $\rho$  denote the correlation between two random variables,  $U$  and  $V$ . Pick the correct statement.  
A.  $|\rho| = 0$    B.  $|\rho| < 0$    C.  $|\rho| = 1$    D.  $|\rho| < 1$

22. The Central Limit Theorem states that the sample average of iid data has approximately a \_\_\_\_\_ distribution when  $n$  is large.  
 A. Poisson    B. Uniform    C. Discrete    D. Normal    E. Exponential    F. Generic
23. Suppose  $Y_1, Y_2 \sim_{\text{iid}} p(y)$ , with  $\text{Var}(Y_1) = 10$ . Then  $\text{Var}(Y_1 + Y_2) =$   
 A. 10    B. 20    C. 40    D. unknown
24. Suppose  $Y_1, Y_2, \dots, Y_n$  are produced as an iid sample from  $p(y)$ . What is the distribution  $p(y)$  when  $n$  is large?  
 A. normal    B. approximately normal    C. uniform    D. unknown
25. When is the sum of iid Bernoulli random variables approximately normally distributed?  
 A. when  $n > 30$     B. when  $n < 30$   
 C. when Bernoulli is approximately normal    D. when  $n\pi > 5$  and  $n(1 - \pi) > 5$
26. How does the Central Limit Theorem (CLT) differ from the Law of Large Numbers (LLN)?  
 A. CLT refers to small  $n$ ; LLN refers to large  $n$ .  
 B. CLT refers to large  $n$ ; LLN refers to small  $n$ .  
 C. CLT refers to the distribution of  $\bar{Y}$ ; LLN refers to the approximate value of  $\bar{Y}$ .  
 D. CLT refers to the approximate value of  $\bar{Y}$ ; LLN refers to the distribution of  $\bar{Y}$ .

#### Chapter 11:

1. If an estimator is unbiased, then  
 A. The estimator is equal to the estimand    B. The estimator is greater than the estimand  
 C. The estimator is less than the estimand    D. None of the above
2. The estimated standard deviation reported by most statistical software is  
 A. An unbiased estimator    B. A biased estimator  
 C. A normally distributed estimator    D. A linear estimator
3. Assuming positive data, the expected value of the ratio of sample means is  
 A. Greater than the ratio of true means    B. Less than the ratio of true means  
 C. Equal to the ratio of true means    D. Sometimes greater than, sometimes less than the ratio of true means



4. Suppose that  $Y_1$  and  $Y_2$  are unbiased estimators of the same parameter  $\theta$ . Which of the following is a *biased* estimator of  $\theta$ ?  
 A.  $0.5Y_1 + 0.5Y_2$    B.  $3Y_1 - 2Y_2$    C.  $(2Y_1 + Y_2)/3$    D.  $Y_1 - Y_2$
5. In what sense is a “best linear unbiased estimator” the “best”?  
 A. It has the most bias   B. It has the least bias   C. It has the most variance   D. It has the least variance
6. If  $\hat{\theta}_1$  is more efficient than  $\hat{\theta}_2$ , then  
 A.  $\hat{\theta}_1$  is closer to  $\theta$  than is  $\hat{\theta}_2$    B.  $\hat{\theta}_1$  tends to be closer to  $\theta$  than  $\hat{\theta}_2$   
 C.  $\text{Var}(\hat{\theta}_1)$  is larger than  $\text{Var}(\hat{\theta}_2)$    D.  $E(\hat{\theta}_1)$  is larger than  $E(\hat{\theta}_2)$
7. Suppose  $Y$  is an unbiased estimator of  $\theta$ . Then  $Y^2$  is a biased estimator of  $\theta^2$ . Explain why in no more than two sentences.
8. Suppose  $Y_1$  and  $Y_2$  are iid with  $E(Y_1) = \mu$ . Explain why  $Y_2$  is an unbiased estimator of  $\mu$  in no more than two sentences.
9. Suppose  $Y_1, Y_2, \dots, Y_n \sim_{\text{iid}} p(y)$ . Let  $\hat{\theta} = (1/n)(Y_1 + Y_2 + \dots + Y_n)$ . Explain why  $\hat{\theta}$  is (i) biased, and is (ii) consistent.
10. In a class exercise, students sampled dice from a bag. Each one got a percentage of white dice based on a sample of  $n = 5$  dice. If a student's percentage was 60%, then 60% is called  
 A. an estimator   B. an estimate   C. an estimand   D. a population parameter   E. A process parameter
11. Both  $\bar{Y}$  and  $\bar{X}$  are unbiased estimators of  $\mu_Y$  and  $\mu_X$ , respectively, and they are independent. Then  $E(\bar{Y}/\bar{X})$  \_\_\_\_\_  $\mu_Y/\mu_X$ .  
 A. >   B. <   C. =
12. A research paper that you are reading states that an estimator  $\hat{\theta}$  is a consistent estimator of  $\theta$ . What does this mean? (One sentence maximum.)
13. Suppose that the bias of the estimator  $\hat{\theta}$  is 3.0 and that its variance is 100.0. Find  $E\{(\hat{\theta} - \theta)^2\}$ .
14. Which of the following is random?  
 A.  $\theta$    B.  $\hat{\theta}$    C.  $E(\theta)$    D.  $E(\hat{\theta})$
15. Suppose that a single random number  $D$  will be produced by  $p(y)$ , a continuous distribution with  $\mu = \int y p(y) dy$ . Then  
 A.  $D$  is an unbiased estimator of  $\mu$   
 B.  $D$  is a biased estimator of  $\mu$   
 C.  $\mu$  is an unbiased estimator of  $D$
16. What explains the fact that  $\bar{Y} / \bar{X}$  is a biased estimator of  $\mu_Y/\mu_X$ ?  
 A. Jensen's inequality   B. Chebyshev's inequality   C. The Central Limit Theorem

17. What explains the fact that the “usual” standard deviation (the one with the  $n-1$  in the denominator) is a biased estimator of  $\sigma$ ?
- A. Jensen’s inequality    B. Chebyshev’s inequality    C. The Central Limit Theorem
18. Consistency of estimates is most closely related to what?
- A. The law of large numbers    B. The central limit theorem  
C. Jensen’s inequality    D. Chebyshev’s inequality
19. Suppose that a single random number  $D$  will be produced by  $p(y)$ , a continuous distribution with  $\mu = \int y p(y) dy$ . Then  $D$  is \_\_\_\_\_ estimator of  $\mu$ .
- A. a biased and inconsistent  
B. an unbiased and inconsistent  
C. a biased and consistent  
D. an unbiased and consistent
20. Suppose  $Y_1, Y_2, \dots, Y_{10}$  will be produced as an iid sample from  $p(y)$ , with  $\mu = \int y p(y) dy$ . Pick the most efficient estimator of  $\mu$ .
- A.  $Y_1$     B.  $Y_2$     C.  $Y_{10}$     D.  $(Y_1 + Y_2 + \dots + Y_{10})/10$
21. Suppose  $Y_1, Y_2, \dots, Y_{10}$  will be produced as an iid sample from a Poisson distribution. Pick the most efficient estimator of  $\sigma$ , the standard deviation of the Poisson distribution.
- A.  $\sqrt{\sum_{i=1}^{10} (Y_i - \bar{Y})^2 / 10}$     B.  $\sqrt{\sum_{i=1}^{10} (Y_i - \bar{Y})^2 / 9}$   
C.  $\sqrt{\sum_{i=1}^{10} (Y_i - \mu)^2 / 10}$     D.  $\sqrt{\bar{Y}}$

22. You saw the following equation in the book:

$$E\{(1/n)(Y_1 + Y_2 + \dots + Y_n)\} = (1/n)E\{(Y_1 + Y_2 + \dots + Y_n)\}$$

Why is this equation true?

- A. Because of the linearity property of expectation  
B. Because of the additivity property of expectation  
C. Because of the Law of Large Numbers  
D. Because of the Central Limit Theorem

23. Why is it true that  $E(1/\bar{X}) > 1/E(\bar{X})$ ?

- A. Because of Chebyshev's inequality
- B. Because of Jensen's inequality
- C. Because  $1/\bar{X}$  is an unbiased estimator
- D. Because  $\bar{X}$  is approximately normally distributed

24. Fill in the blank:  $\sum_{i=1}^n (Y_i - \bar{Y})^2$  \_\_\_\_\_  $\sum_{i=1}^n (Y_i - \mu)^2$

- A. <      B. =      C. >

25. The plug-in estimator of variance is

- A. biased      B. unbiased

26. Suppose  $Y_1, Y_2, \dots, Y_{10} \sim_{\text{iid}} N(70, 10^2)$ . Then  $\theta = 70$  is both the mean and the median of the data-generating process. Let  $\hat{\theta}_1$  denote the average of the  $n = 10$  observations and let  $\hat{\theta}_2$  denote the median of the  $n = 10$  observations; both are estimators of  $\theta = 70$ .

Then  $\text{ESD}(\hat{\theta}_1)$  \_\_\_\_\_  $\text{ESD}(\hat{\theta}_2)$ .

- A. <      B. =      C. >

27. What is the best estimator of the standard deviation of a Poisson distribution?

- A. The bootstrap plug-in estimator of the standard deviation.
- B. The usual  $(n-1)$  bias-corrected estimator of the standard deviation.
- C. The square root of the sample average.
- D. The square root of the sample median.

28. Pick the true answer.

- A. An estimator must be unbiased to be consistent.
- B. An estimator must be efficient to be consistent.
- C. The bias of a consistent estimator must get closer to zero as  $n$  increases.
- D. The variance of a consistent estimator must get closer to one as  $n$  increases.

29. When is a biased estimator preferred over an unbiased estimator?

- A. Never.
- B. When  $n$  is large.
- C. When the distribution of the biased estimator is approximately a normal distribution.
- D. When the biased estimator tends to be more accurate than the unbiased estimator.

30. Suppose  $\hat{\theta}$  is an unbiased estimator of  $\theta$ . Then (pick one answer only) \_\_\_\_\_.

- A.  $\hat{\theta} = \theta$     B.  $E(\hat{\theta}) = \theta$     C.  $E(\theta) = \hat{\theta}$     D.  $\text{Var}(\hat{\theta}) = \sigma^2$

31. Why is  $E(Y_1 + Y_2)$  equal to  $E(Y_1) + E(Y_2)$ ? Because of the (pick one answer only) \_\_\_\_\_

- A. linearity property of expectation
- B. linearity property of variance
- C. additivity property of expectation
- D. additivity property of variance

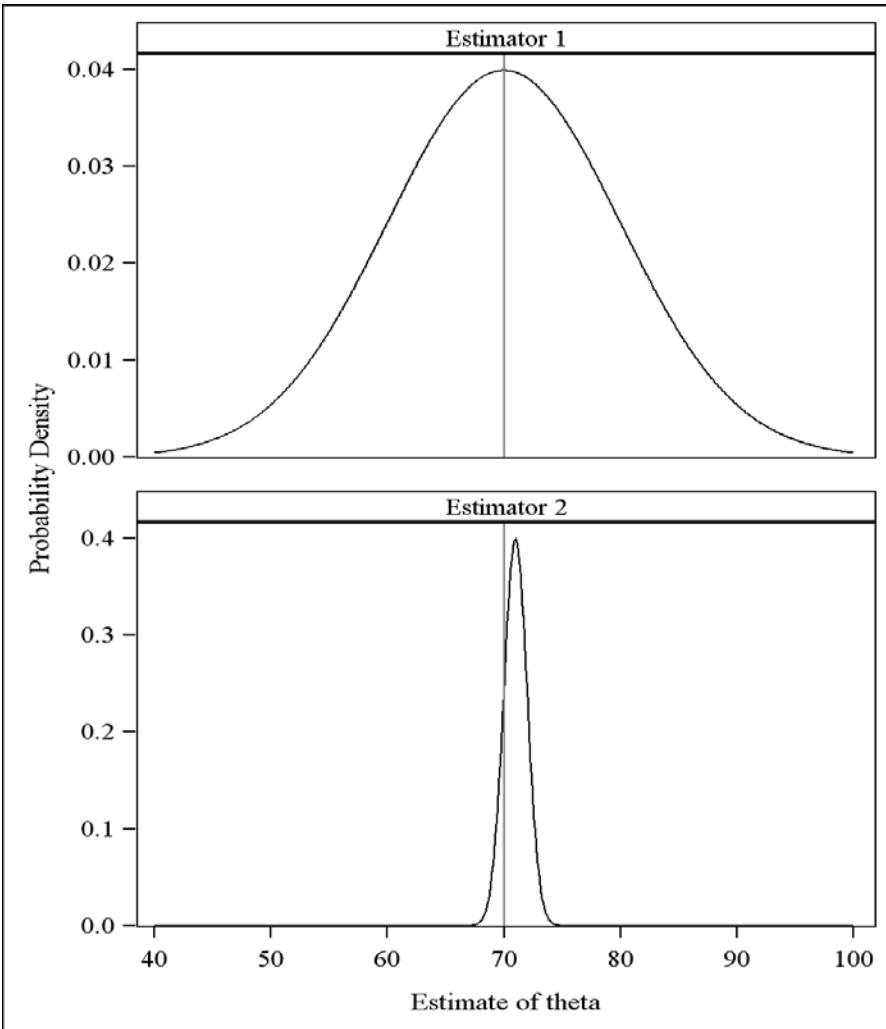
32. Suppose  $Y_1, Y_2, Y_3$  are produced as iid random variables from distribution whose mean is  $\mu$ . Which of the following are unbiased estimators of  $\mu$ ? Select all that apply. Ten points per correct selection/nonselection.

- A.  $Y_1$     B.  $(Y_1 + Y_2)/2$     C.  $(Y_1 + Y_2 + Y_3)/3$     D.  $Y_1 - Y_2$

33. Suppose  $Y_1, Y_2, \dots, Y_n$  are produced as iid random variables from distribution whose mean is  $\theta$ . Which of the following are consistent estimators of  $\theta$ ? Select all that apply. Ten points per correct selection/nonselection.

- A.  $\hat{\theta} = Y_1$     B.  $\hat{\theta} = (Y_1 + Y_2)/2$     C.  $\hat{\theta} = (Y_1 + Y_2 + \dots + Y_n)/n$     D.  $\hat{\theta} = (Y_1 + Y_2 + \dots + Y_n)/n + 1/n$

34. Consider the following graphs. The parameter is  $\theta = 70$ .



Select all that apply. 10 points per correct selection or nonselection.

- A. Estimator 1 is biased.
- B. Estimator 2 is biased.
- C. Estimator 1 is better than Estimator 2.
- D. Estimator 2 is better than Estimator 1.

35. You will roll a die one time getting  $Y$ . This  $Y$  might be 1, 2, 3, 4, 5, or 6. The mean of the distribution of  $Y$  is  $\mu$  (some unknown number close to 3.5). Then

- A.  $Y$  is a biased estimate of  $\mu$  because  $\mu$  is unknown
- B.  $Y$  is a biased estimate of  $\mu$  because  $Y \neq \mu$
- C.  $Y$  is an unbiased estimate of  $\mu$  because  $Y = \mu$

D.  $Y$  is an unbiased estimate of  $\mu$  because  $E(Y) = \mu$

36. Suppose  $\hat{\theta}$  is an unbiased estimator of  $\theta$ . Then

- A.  $2 \times \hat{\theta}$  is an unbiased estimator of  $\theta$ .
- B.  $2 \times \hat{\theta}$  is an unbiased estimator of  $2 \times \theta$ .
- C.  $2 \times \hat{\theta}$  is a biased estimator of  $2\theta$ .
- D.  $\theta$  is an unbiased estimator of  $\hat{\theta}$ .

37. Suppose  $Y_1, Y_2, \dots, Y_{10} \sim \text{iid } p(y)$ , with  $E(Y_1) = \mu$ . When is  $(1/10)(Y_1 + Y_2 + \dots + Y_{10})$  a biased estimator of  $\mu$ ?

- A. When  $n$  is small ( $< 30$  or so).
- B. When  $p(y)$  is not a normal distribution.
- C. When  $(1/10)(Y_1 + Y_2 + \dots + Y_{10}) \neq \mu$ .
- D. Never; it is always unbiased under the stated conditions.

38. Let  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2$ , the plug-in estimator of variance. Then

- A.  $E(\hat{\sigma}^2) > \sigma^2$
- B.  $E(\hat{\sigma}^2) < \sigma^2$
- C.  $E(\hat{\sigma}^2) = \sigma^2$

39. An estimator  $\hat{\theta}$  gets closer to its estimand  $\theta$  as  $n$  gets larger. Thus  $\hat{\theta}$  is \_\_\_\_\_ estimator of  $\theta$ .

- A. an approximately normally distributed
- B. an unbiased
- C. a consistent
- D. an efficient

40. Which of the following is impossible about an estimator  $\hat{\theta}$  and its estimand  $\theta$ ?

- A.  $\hat{\theta}$  is biased but gets closer to  $\theta$  with larger  $n$
- B.  $\hat{\theta}$  is unbiased but does not get closer to  $\theta$  with larger  $n$
- C.  $\hat{\theta}$  is efficient but does not have a normal distribution with larger  $n$
- D.  $\hat{\theta}$  is consistent but does not get closer to  $\theta$  with larger  $n$

41. When is  $\hat{\theta}_1$  more efficient than  $\hat{\theta}_2$ ?

- A. When  $\hat{\theta}_2$  is biased and  $\hat{\theta}_1$  is unbiased.
- B. When  $\text{Var}(\hat{\theta}_2) > \text{Var}(\hat{\theta}_1)$

C. When  $E(\hat{\theta}_2 - \theta)^2 > E(\hat{\theta}_1 - \theta)^2$

D. When  $\hat{\theta}_2$  is nonnormally distributed and  $\hat{\theta}_1$  is normally distributed.

42. When  $Y_1, \dots, Y_{10} \sim_{\text{iid}} N(\mu, \sigma^2)$ , what is the most efficient estimate of  $\mu$ ?

A.  $(1/10)(Y_1 + \dots + Y_{10})$       B. the median of  $Y_1, \dots, Y_{10}$

C.  $\frac{1}{10} \sum_{i=1}^{10} (Y_i - \bar{Y})^2$       D.  $\frac{1}{9} \sum_{i=1}^{10} (Y_i - \bar{Y})^2$

## Chapter 12:

1. Explain what the following expression means. Draw a graph and use it in your explanation.

$$L(\theta | y) \doteq C \times \exp \left\{ -\frac{1}{2} \frac{(\theta - \hat{\theta})^2}{\hat{\sigma}^2} \right\}.$$

2. Draw a graph that illustrates the meaning of the equation

$$\left. \frac{\partial L(\theta | y_1, y_2, \dots, y_n)}{\partial \theta} \right|_{\theta = \hat{\theta}} = 0.$$

Label both axes.

3. Name a specific software tool that will find the maximum likelihood estimate of the parameter of an exponential distribution.

4. What is noteworthy about the maximum likelihood estimate of the standard deviation of the normal distribution  $N(\mu, \sigma^2)$ ?

5. What is the “approximation” that is used to obtain the “Wald standard error?”

6. If  $\hat{\theta}$  is the maximum likelihood estimate of the parameter  $\theta$ , then

A.  $\hat{\theta}$  maximizes the likelihood function      B.  $\hat{\theta}$  is unbiased

C.  $\hat{\theta}$  is a convex function of the data      D.  $\hat{\theta}$  is the estimated mean of the normal pdf

7. The likelihood function is a function of

A. the data  $y$       B. the parameter  $\theta$       C. the average  $\bar{y}$

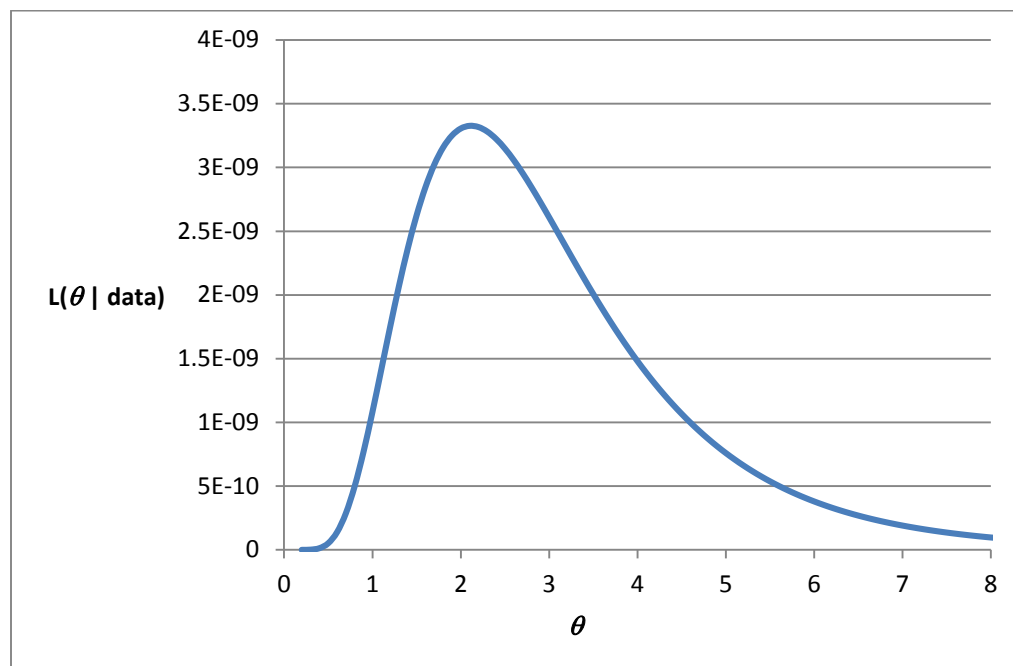
8. What happens to the likelihood function when  $n$  is larger?

A. It gets wider      B. It gets more skewed      C. It gets narrower

9. The data  $y_1 = 1, y_2 = 0, y_3 = 0$  are the result of an iid sample from a Bernoulli( $\pi$ ) process. Give the likelihood function:

$L(\pi | \text{data}) = \underline{\hspace{2cm}}$

10. See the following graph/ Using the inspection method, guess the MLE. (Just give a number).



11. From the graph above, which number is closest to the Wald standard error?

- A. 1E-09    B. 4E-09    C. 1.0    D. 4.0

12. Which method is best to find the true MLE?

- A. Inspection    B. Calculus    C. Numerical approximation    D. Simulation

13. What is the likelihood function used for?

- A. To reduce uncertainty about the data  
B. To reduce uncertainty about the estimates  
C. To reduce uncertainty about the unknown parameters

14. What is the area under the curve of a likelihood function?

- A. Always 1.0    B. Usually not 1.0

15. Suppose  $p(y | \theta) = y\theta^2 \exp(-\theta y)$ . Suppose  $y = 2$ . What is the likelihood function for  $\theta$ ?

- A.  $2y \exp(-2y)$     B.  $4y \exp(-2y)$     C.  $2\theta^2 \exp(-2\theta)$     D.  $4\theta^2 \exp(-2\theta)$

16. When is the likelihood function for the sample equal to the product of the likelihood functions for the individual observations?

- A. When the DATA are exponentially distributed  
B. When the DATA are normally distributed  
C. When the DATA are identically distributed  
D. When the DATA are independently distributed

17. Why is the log likelihood function used rather than the likelihood function? Pick the relevant answer.

- A. Because it gives better parameter estimates  
B. Because it allows the computer to deal with extremely small numbers  
C. Because it is approximately normally distributed

18. The data 1, 0, 0, 1, 0 are from an iid sampling of the Bernoulli( $\pi$ ) distribution



$y$	$p(y)$
0	$1 - \pi$
1	$\pi$
	1.0

Then  $L(\pi | \text{data}) =$

- A.  $\pi$     B.  $1 - \pi$     C.  $\pi^2(1 - \pi)^3$     D. 0.40    E. 0.60

19. Draw a graph of a likelihood function for a (one-dimensional) parameter  $\theta$ . Label the axes and show both the maximum likelihood estimate and the Wald standard error on this graph.

20. Where is the maximum likelihood estimate on the graph of the likelihood function? Describe in words.

21. Where is the Wald standard error on the graph of the likelihood function? Describe in words.

22. What assumption is needed for the equation

$$p(y_1, y_2, \dots, y_n | \theta) = p(y_1 | \theta) \times p(y_2 | \theta) \times \dots \times p(y_n | \theta)$$

to be true?

- A. The distributions are normal  
 B.  $\theta$  is an unknown parameter  
 C. The DATA are independent

23. Suppose  $Y_1, Y_2 \sim_{\text{iid}} p(y | \theta)$ , with  $p(y | \theta) = \theta y^{\theta-1}$ , for  $0 < y < 1$ , and  $0 < \theta < \infty$ . (Comment: This is an example of the famous *beta* distribution.) A sample gives  $y_1 = 0.4$  and  $y_2 = 0.5$ . Write down the likelihood function for  $\theta$ .

24. Simplify the expression  $\ln(5\lambda^{10})$ . No reasons needed, just give the result.

25. Suppose  $y_1, y_2, \dots, y_{10}$  are produced as iid from  $N(\mu, \sigma^2)$ . What is the MLE of  $\sigma$ ?

- A.  $\sqrt{\sum_{i=1}^{10} (y_i - \bar{y})^2 / 10}$     B.  $\sqrt{\sum_{i=1}^{10} (y_i - \bar{y})^2 / 9}$   
 C.  $\sqrt{\sum_{i=1}^{10} (y_i - \mu)^2 / 10}$     D.  $\sqrt{\bar{y}}$

26. What does “convergence” mean, when speaking of computer solutions for MLEs?

- A. The difference between the estimate and the estimand is small  
 B. The difference between successive estimates is small  
 C. The difference between the maximum and the minimum likelihood is small

27. What assumption is made by the Wald standard error?

- A. The likelihood function is approximately proportional to a normal density function  
 B. The data are produced as iid from a normal distribution  
 C. The sample size is greater than 30  
 D. The MLE,  $\hat{\theta}$ , is a consistent estimator of the true value,  $\theta$

28.(40) The model that produces the data is  $p(y|\theta) = \theta y^{-\theta}$ . Suppose a single observation produced by this model is  $y = 2$ . Give the likelihood function for  $\theta$  by filling in the blank.

$L(\theta | y = 2) =$  \_\_\_\_\_

29. What does the subscript “ $T$ ” in the expression  $\theta_T$  refer to?

- A. Transformation    B. True value    C. Time    D. Triangular distribution

30. What does “vector” mean in the phrase “parameter vector”?

- A. A list    B. A direction    C. A magnitude    D. A distribution

31.  $\ln(e^y) =$

- A.  $y$     B.  $\ln(e) + \ln(y)$     C.  $y + \ln(e)$     D.  $e - \ln(y)$

32. The log-likelihood function (called  $LL$  in the book) for an iid sample is equal to

- A. the pdf for the sample.  
B. the sum of the pdfs for each observation in the sample.  
C. the sum of the logarithms of the pdfs for each observation in the sample.  
D. the product of the pdfs for each observation in the sample.

33. Suppose  $y_1, y_2, \dots, y_n$  are produced as an iid sample from  $N(\mu, \sigma^2)$ , where  $\mu$  and  $\sigma$  are unknown parameters. Then the MLE of  $\sigma$  is

- A.  $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$     B.  $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}$     C.  $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \mu)^2}$     D.  $\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \mu)^2}$

34. The Wald standard error assumes what?

- A. The distribution of the data is exponential
- B. The distribution of the data is approximately a bell-shaped curve
- C. The likelihood function of the parameter is exponential
- D. The likelihood function of the parameter is approximately a bell-shaped curve

35. Suppose that  $y_1 = 0.2$  and  $y_2 = 0.4$  are realizations from an iid process having

$$p(y | \theta) = \frac{1}{\sqrt{2 \times 3.1416}} \exp\{-(y - \theta)^2 / 2\}$$

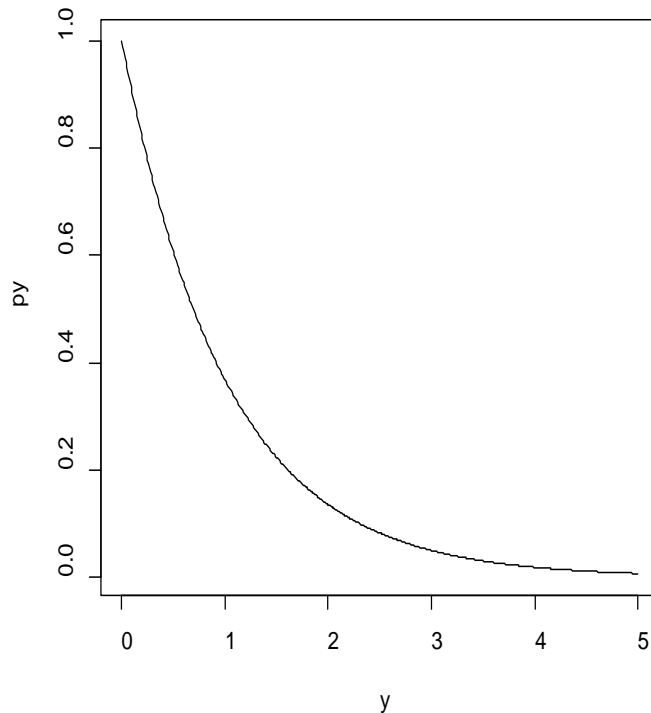
Give the likelihood function for  $\theta$ .  $L(\theta | 0.2, 0.4) =$

36. Suppose that the MLE of  $\theta$  is  $\hat{\theta} = 3.5$ , and the Wald standard error is 0.5. Give the 95% Wald confidence interval for  $\theta$ .

37. Suppose your data are produced by the  $N(0, \sigma^2)$  distribution. What is the parameter space?

- A. the range from 0 to 1
- B. the range from -1 to 1
- C. the range from  $-\infty$  to  $\infty$
- D. the range from 0 to  $\infty$

38. Consider the following probability distribution function.



The likelihood of  $y=2$  is

- A. 0     B. 0.135     C. 0.606     D. 1.0

39. Select the true statement about the likelihood function.

- A. data are fixed, parameters are fixed
- B. data are variable, parameters are fixed
- C. data are fixed, parameters are variable
- D. data are variable, parameters are variable

40. Data values  $y_1 = 2.0$  and  $y_2 = 4.5$  were produced independently by distributions  $p_1(y)$  and  $p_2(y)$  respectively. Then the likelihood function is

- A.  $p(2.0, 4.5)$      B.  $p_1(2.0) + p_2(4.5)$      C.  $p_1(2.0) \times p_2(4.5)$      D.  $p_1(2.0)/p_2(4.5)$

41. Suppose your data are produced by the exponential ( $\lambda$ ) distribution. Then the likelihood of a single observation  $y$  is

$$L(\lambda \mid y) = \lambda e^{-\lambda y}$$

Give the corresponding log likelihood, using properties of logarithms to simplify the expression.

## Chapter 13:

1. Where does the prior distribution come from?  
A. The likelihood function. B. The current data. C. The investigator's mind.
2. You flip a coin 10 times and there are 2 heads. Then  
A. the probability of heads is .5. B. the probability of heads is .2.  
C. your guess of the probability of heads depends on your prior knowledge of the coin and the coin-flipper.
3. Give an example of an "ignorance prior."  
A. A uniform distribution B. A normal distribution C.  $\theta = \mu$  D.  $\theta = \sigma$
4. What is "InternetSavvy"?  
A. A Likert scale B. A fictitious company C. A construct measuring familiarity with the internet
5. What is "value at risk"?  
A. The maximum attainable loss  
B. A quantile of the attainable loss distribution  
C. The expected loss  
D. The volatility of the loss
6. What do you need to perform Bayesian analysis?  
A. A likelihood function. B. A prior distribution.  
C. A likelihood function and a prior distribution. D. None of the above.
7. Where does prior information come from?  
A. Other similar studies. B. Life experience and common sense.  
C. Parameter constraints. D. All of the above.
8. A credible interval is  
A. A range of values for the observed data.  
B. A range of values for the unknown parameter.  
C. The time between first and last observations.  
D. The permissible time between first and last observations.
9. The beta distribution is used to model  
A. The prior probability of success. B. The age of a student.  
C. The standard deviation of students' ages. D. The expected value of student ages.
10. What do you need to perform Bayesian analysis?  
A. Normally distributed data B. A large sample size  
C. Jensen's inequality D. None of the above.
11. A credible interval is  
A. A range of values for the observed data.  
B. A range of values for the unknown parameter.  
C. Usually, a three standard deviation range.  
D. All of the above
12. The "kernel" of a distribution is  
A. a prior probability of success. B. a random variable.  
C. proportional to the distribution. D. the mean of the distribution.

13. Suppose you have 1,000,000 values of  $\theta^*$ , which you have obtained by simulating from the posterior distribution  $p(\theta | \text{data})$ . Describe how to obtain the 90% equal-tail credible interval for  $\theta$  using these 1,000,000 values. Two sentences maximum.

14. Hans got two “successes” and 8 “failures” in both his thumbtack toss experiment and in his coin toss experiment. His likelihood function for the thumbtack toss data was \_\_\_\_\_ his likelihood function for the coin toss data.

- A. identical to
- B. shifted to the left of
- C. shifted to the right of

15. The \_\_\_\_\_ is what you use to express your uncertainty about the parameters before collecting your data.

- A. prior distribution
- B. posterior distribution
- C. likelihood function

16. Hans gives his prior distribution for  $\theta$  = mean driving time as follows:

$\theta$	$p(\theta)$
20.0 min	0.4
20.5 min	0.6
Total	1.0

Hans' prior is an example of a \_\_\_\_\_ prior.

- A. non-informative
- B. vague
- C. uniform
- D. dogmatic

17. The function  $\pi(1 - \pi)^2$  is the kernel of the beta(\_\_\_\_\_) distribution.

- A. 1, 2
- B. 0, 1
- C. 1/3, 2/3
- D. 2, 3

18. Suppose you have 10,000 values  $\theta^*$  which are simulated from  $p(\theta | \text{data})$ . What is the best way to estimate  $p(\theta | \text{data})$ ?

- A. Use the normal distribution with the mean and standard deviation from the 10,000 values.
- B. Use the scatterplot of the 10,000 values.
- C. Use the histogram of the 10,000 values.

19. In the logistic regression example, the posterior correlation between  $\beta_0$  and  $\beta_1$  was \_\_\_\_\_ 0.

- A. >
- B. <
- C. =
- D.  $\cong$

20. What is the 95<sup>th</sup> percentile of the  $N(\mu, \sigma^2)$  distribution?

- A. 1.6445
- B.  $1.6445\sigma$
- C.  $\mu + 1.6445\sigma$

21. In the drug company example, 68.4% of the treated patients were cured while only 65.3% of the placebo patients were cured.

In the analysis of the book, what reason was given that the company might choose **not** to continue product development?

- A. The true proportion of cures might actually be smaller for the treated group.
- B. The treated group might suffer too many side effects.
- C. The  $p$  value might be greater than 0.05.
- D. The US Food and Drug administration might decline the application.

22. Give the posterior distribution for  $\theta$ .

- A.  $L(\theta | \text{data})$
- B.  $\int p(\theta) d\theta$
- C.  $(\text{constant}) \times L(\theta | \text{data}) \times p(\theta)$
- D.  $(\text{constant}) \times \exp\{-0.5(\theta - \hat{\theta})^2\}$

23. Which distribution is used as non-informative prior?

- A. normal
- B. Poisson
- C. Bernoulli
- D. uniform

24. The parameter space is usually
- A. the same as the sample space for the DATA
  - B. a continuum
  - C. the derivative of the likelihood function
  - D. a list of values
25. What was the standard deviation of the prior distribution for  $\pi$ , the probability of Heads for Hans' toss of a fair coin?
- A. 0.003      B. 1.0      C. 10.0      D. 1000.0
26. What is the most commonly-used method for Bayesian analysis?
- A. Central limit theorem
  - B. Law of large numbers
  - C. Simulation from the posterior distribution
  - D. Maximum likelihood
27. How many parameters were there in the logistic regression model in Section 13.7?
- A. 1    B. 2    C. 100    D. 10,000
28. "Value at risk" is a \_\_\_\_\_
- A. mean    B. median    C. standard deviation    D. quantile
29. What is the benefit of Bayesian methods?
- A. They allow you to determine the true value of the parameter  $\theta$
  - B. They allow you to select different values of the parameter  $\theta$  that are consistent with your data
  - C. They allow you to determine the true value of your DATA
  - D. They free you from having to specify a model for how your DATA are produced
30. A prior is called "dogmatic" when
- A. it is a uniform distribution
  - B. it is a normal distribution
  - C. it assigns 0% probability to some regions of the parameter space
  - D. it assigns 50% probability to some regions of the parameter space
31. The posterior distribution function is proportional to the
- A. likelihood function
  - B. log likelihood function
  - C. likelihood function times the prior distribution function
  - D. log likelihood function plus the logarithm of the prior distribution function
32. Comparing two interval ranges for a parameter  $\theta$ , the shorter interval range shows
- A. less bias
  - B. more variability

- C. more precision
- D. less prior information

33. When you use a beta prior distribution for the Bernoulli parameter  $\pi$ , the posterior distribution for  $\pi$  has the \_\_\_\_\_ distribution.

- A. beta
- B. uniform
- C. Bernoulli
- D. normal

34. Which prior distribution can be called “vague”?

- A.  $N(0, 1)$
- B.  $\text{Exponential}(1)$
- C.  $U(0, 1)$
- D.  $\text{beta}(5, 5)$

35. Which prior distribution is “improper”?

- A.  $p(\theta) = 1$ , for  $-\infty < \theta < \infty$
- B.  $p(\theta) = 0.5$ , for  $0 < \theta < 2$
- C.  $p(\theta) = \exp(-\theta)$ , for  $0 < \theta < \infty$
- D.  $p(\theta) = (1/\sqrt{2\pi})\exp\{-0.5\theta^2\}$ , for  $-\infty < \theta < \infty$

36. When do you need to use Bayesian methods?

- A. When you need to select values of your parameters that are consistent with the data.
- B. When your data are produced by a normal distribution.
- C. When the parameters of your data generating process are known.
- D. When you can assume that the data produces the model.

37. In the example of forecasting stock returns, how were the potential values of the future stock returns produced?

- A. From a normal distribution with known mean and variance.
- B. From a normal distribution where the mean and variance were randomly sampled.
- C. From extrapolating the current time series into the future using an exponentially weighted moving average.
- D. By using autoregressive integrated moving average (ARIMA) methods.

38. A certain proportion of TTU students ride a bicycle to class every day. Call this proportion  $\theta$ , a number between 0 and 1. Draw a graph of  $p(\theta)$ , which is your own personal prior distribution for  $\theta$ . (A uniform distribution is not allowed).



39. An improper prior  $p(\theta)$  is called “improper” because \_\_\_\_\_
- $p(\theta) < 0$  for some values of  $\theta$
  - $\int p(\theta) d\theta \neq 1$
  - $\int \theta p(\theta) d\theta \neq 1$
  - normal distributions have infinite means
40. What happens when you use a vague prior?
- You bias the results of the data analysis toward your prior thoughts
  - You let the data “talk for themselves”
  - You get smaller standard errors of parameter estimates
  - Your credible intervals become narrower
41. You need to use Bayesian analysis when \_\_\_\_\_
- the data are produced by a normal distribution
  - you need to maximize the likelihood function
  - you perform logistic regression
  - you want to select values of the parameters that are consistent with your data and prior
42. In the student age example, what was  $\mu^*$ ?
- the true mean of the age-producing process
  - the sample average of the  $n = 16$  age values
  - a value produced by the prior distribution
  - a value produced by the posterior distribution
43. Consider the following simplified version of the Bayesian data analysis in the book. Again, you observe a survey with  $y=4$ .

Satisfaction Rating, $y$	Company Name, $\theta$				
	BankTen	DoggyTreats	CraftyCrafts	AgBus	InternetSavvy
1	25%	35%	15%	20%	10%
2	20%	25%	30%	20%	25%
3	20%	5%	10%	25%	5%
4	25%	5%	10%	25%	5%
5	10%	30%	35%	10%	55%
Totals:	100%	100%	100%	100%	100%

Give the posterior distribution  $p(\theta \mid y=4)$  when the uniform prior is assumed for  $\theta$ , by entering the correct numbers in the table below.

Company Name, $\theta$				
BankTen	DoggyTreats	CraftyCrafts	AgBus	InternetSavvy

44. What does computer software need in order to perform simulation-based posterior inference?
- A. The likelihood function
  - B. The prior distribution
  - C. Both the likelihood function and the prior distribution
  - D. A data set
45. What is a “vague” prior probability distribution function?
- A. One that imposes little, if any, prior knowledge.
  - B. A distribution with infinite variance.
  - C. A distribution with unknown mean and unknown variance.
  - D. A function whose area under the curve is not equal to 1.0.
46. What is an “improper” prior probability distribution function?
- A. One that imposes little, if any, prior knowledge.
  - B. A distribution with infinite variance.
  - C. A distribution with unknown mean and unknown variance.
  - D. A function whose area under the curve is not equal to 1.0.
47. When do you use logistic regression?
- A. When the “Y” variable is normally distributed.
  - B. Whenever the “Y” variable has a continuous distribution.
  - C. When the “Y” variable has a Bernoulli distribution.
  - D. When the “Y” variable has a Poisson distribution.

## Chapter 14:

1. For a frequentist, what does “95%” refer to, in the 95% confidence interval  $2.1 \leq \theta \leq 3.4$ ? You can assume 95% means “95 out of 100”, but what does “95” refer to, and what does “100” refer to? Be specific.
2. You are a frequentist. You have already collected a data set, which you believe to be produced by a model  $p(y | \theta)$ . What is random at this point in time?  
A. Your data    B. The parameter(s) of your model    C. Neither the data nor the parameter(s) are random
3. You are a frequentist. What is the probability that your parameter  $\theta$  is greater than 0?  
A. 50%    B. 100%    C. 0%    D. Either 0% or 100%
4. The variance of  $Y$  is 4. What is the standard deviation of  $\bar{Y}$ , assuming an iid sample?  
A. 4    B.  $4/n$     C. 2    D.  $2/n$     E.  $4/\sqrt{n}$     F.  $2/\sqrt{n}$
5. What is the true confidence level of an approximate 95% confidence interval?  
A. 95%    B. 90%    C. 80%    D. It depends on the distribution that produced the data
6. Suppose  $Y_1, \dots, Y_{10}$  are iid, with  $\text{Var}(Y_1) = \sigma^2$ . Then  $\text{Var}(\bar{Y}) =$   
A.  $\sigma$     B.  $\sigma^2$     C.  $\sigma/10$     D.  $\sigma^2/10$     E.  $\sigma/\sqrt{10}$
7. What is an advantage of frequentist methods over Bayesian methods?  
A. They are more accurate  
B. You can assume normal distributions  
C. You don’t need a prior distribution
8. For a frequentist, the interval  $9.3 < \theta < 17.9$  is called a \_\_\_\_\_ interval.  
A. likely    B. confidence    C. credible
9. For a 95% frequentist interval, the true percentage is not 95%. How can you estimate the true percentage?  
A. Use simulation    B. Use Jensen’s inequality    C. Use Bayesian methods
10. A frequentist is  
A. A non-Bayesian  
B. A cross-tabulation table  
C. A person who uses statistics frequently  
D. A posterior distribution
12. In the student age example of Chapter 14, the parameter  $\theta$  of the distribution  $p(y | \theta)$  that produces the age data is  
A. the mean  
B. the standard deviation  
C. the vector of regression coefficients  
D. the list of probabilities
13. What is the true confidence level of an approximate 95% confidence interval?  
A. 95%    B. 90%    C. 80%    D. It depends on the distribution that produced the data
14. If a lion is within 10 kilometers of a town, then  
A. the town is within 10 kilometers of the lion.

- B. the town is more than 10 kilometers from the lion.
- C. the town could be more than 10 kilometers from the lion, or it could be less than 10 kilometers from the lion.

15. Frequentist methods may be considered **better** than Bayesian methods because \_\_\_\_\_.

(Select all that apply; 10 points per correct selection).

- A. you do not have to assume a prior  $p(\theta)$ .
- B. you do not have to assume that the model produces the DATA.
- C. you do not have to assume independence of the DATA.
- D. it is easier to understand frequentist methods using simulation.
- E. frequentist methods allow you to select parameters from the posterior distribution.

16. What approximations that are applied in the approximate interval  $\bar{y} \pm 1.96\hat{\sigma}/\sqrt{n}$  ?

(Select all that apply; 10 points per correct selection.)

- A. The distribution of  $\bar{Y}$  is approximately normal.
- B.  $\hat{\sigma}$  is approximately equal to  $\sigma$ .
- C. The sample size,  $n$ , is approximately greater than 30.

17. When does the frequentist use the word “probability” rather than “confidence” to describe data?

- A. Before seeing the data.
- B. After seeing the data.
- C. Never.

18. How many unknown parameters were there in the generic distribution that was assumed for student age (in years)?

- A. 0
- B. 1
- C. 2
- D. more than 2

19. When do you use  $\sigma/\sqrt{n}$  rather than  $\sigma$  ?

- A. When  $n > 30$ .
- B. When  $n$  approaches infinity.
- C. When the distribution is normal.
- D. When considering the variability of  $\bar{Y}$  rather than  $Y$ .

20. Why is  $\sigma$  usually unknown?

- A. Because  $n$  is usually unknown.
- B. Because  $\hat{\sigma}$  is usually unknown.
- C. Because  $\bar{y}$  is usually unknown.
- D. Because models have unknown parameters.

21. If a mountain lion is within 20 kilometers of a town, then what must be true?

- A. The data are normally distributed
- B. The central limit theorem is true

- C. The iid assumption is true
- D. The town is within 20 kilometers of the mountain lion

22. In the town/mountain lion analogy,  $\mu$  is like the \_\_\_\_\_ and  $\bar{y}$  is like the \_\_\_\_\_.

- A. town, mountain lion
- B. mountain lion, town
- C. population mean, process mean
- D. random mean, fixed mean

23. There are two approximations in the 95% frequentist interval  $\bar{y} \pm 1.96\hat{\sigma}/\sqrt{n}$ . What are they?

- A.  $1.96 \cong 2.00$ ;  $n$  is approximately large.
- B. The distribution of the data is approximately normal; the distribution of the variance is approximately chi-squared.
- C.  $\bar{Y}$  has an approximate normal distribution;  $\hat{\sigma} \cong \sigma$ .
- D. The sample size is approximately greater than 30;  $t$  is approximately  $z$ .

24. Here is a 95% interval for a parameter:  $7.6 < \theta < 12.3$ . What does “95%” mean for the frequentist?

- A. About 95 out of 100  $\theta$ 's are between 7.6 and 12.3.
- B. About 95 out of 100  $\hat{\theta}$ 's are between 7.6 and 12.3.
- C. About 95 out of 100  $Y$ 's are between 7.6 and 12.3.
- D. In about 95 out of 100 samples from the same process,  $\theta$  will lie between the limits of similarly constructed intervals.

25.(20) In the book you read the following sentence:

If  $Y_1, Y_2, \dots, Y_{16} \sim_{\text{iid}} p(y | \theta)$ , then  $\bar{Y} \sim N(\mu, \sigma^2/16)$ .

This statement is true because of the

- A. law of large numbers
- B. central limit theorem
- C. bootstrap plug-in principle

26. From the frequentist standpoint, which of the following quantities are random variables.

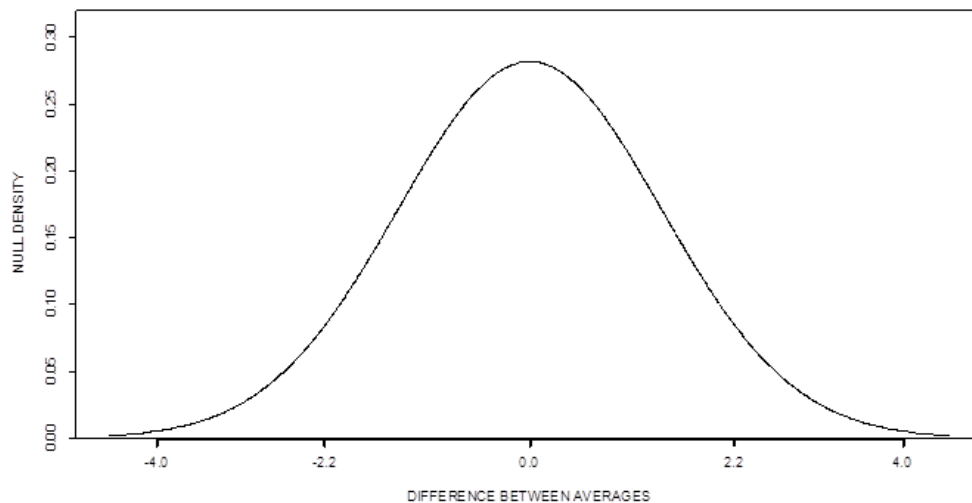
Circle all that apply. Ten points per correct selection/nonselection.

- A.  $\mu$     B.  $\sigma$     C.  $n$     D.  $\sigma^2$     E.  $\bar{Y}$     F.  $Y_1$     G.  $\bar{y}$     H.  $y_1$

## Chapter 15:

- What is the difference between results that are *explainable* by chance alone, versus results that are *explained* by chance alone?
- $\text{Var}(Y) = 4$  and  $\text{Var}(X) = 12$ . What is  $\text{Var}(Y - X)$ , assuming independence?  
 A. 16      B. -8      C.  $\sqrt{-8}$       D. 4      E.  $16/\sqrt{n}$       F.  $-8/\sqrt{n}$       G.  $\sqrt{-8} / \sqrt{n}$   
 H.  $4/\sqrt{n}$
- What is the distribution of the  $p$ -value, when a null model produces the data?  
 A. Normal      B. Bernoulli      C. Uniform
- If a null model produces the data, then the observed differences in the data are \_\_\_\_\_ by chance alone.  
 A. explained      B. explainable
- The randomization model is also called the \_\_\_\_\_ model.  
 A. permutation      B. significance
- If a  $p$ -value for testing a null hypothesis is  $p=0.0059$ , then  
 A. the null hypothesis is true      B. the null hypothesis is false      C. you may accept the null hypothesis  
 D. you may reject the null hypothesis
- If a two-sided test of  $H_0: \mu=250$  fails to reject the null hypothesis, then the confidence interval for  $\mu$   
 A. includes 250      B. excludes 250      C. includes the sample average      D. excludes the sample average
- Suppose  $Y_1, Y_2, \dots, Y_{16} \sim_{\text{iid}} p_0(y)$ , where  $E(Y_1) = \mu$  and  $\text{Var}(Y_1) = \sigma^2$ . Let  $\bar{Y}_1 = (Y_1 + Y_2 + \dots + Y_8)/8$  and let  $\bar{Y}_2 = (Y_9 + Y_{10} + \dots + Y_{16})/8$ . Then  $E(\bar{Y}_1 - \bar{Y}_2) =$  \_\_\_\_\_.  
 A. 0      B.  $\mu_1 - \mu_2$       C.  $\bar{Y}_1 - \bar{Y}_2$
- Suppose the same as in 8. Then  $\text{Var}(\bar{Y}_1 - \bar{Y}_2) =$   
 A. 0      B.  $\sigma_1^2 - \sigma_2^2$       C.  $\sigma_1^2 + \sigma_2^2$       D.  $\sigma^2/8 + \sigma^2/8$
- A large  $p$ -value indicates that  
 A. There is truly no difference  
 B. There is truly a difference  
 C. The difference can be easily explained by chance alone  
 D. The difference cannot be easily explained by chance alone
- What is the distribution of the (random)  $p$ -value when the chance-only model is the true data-generating process?  
 A. Uniform      B. Normal      C. Poisson      D. Bernoulli      E. Exponential
- You simulate 100 numbers from the normal distribution with mean 70 and standard deviation 10. The average of the 100 numbers is 71.8. The difference between 71.8 and 70.0 is  
 A. explained by chance alone.      B. explainable by chance alone.  
 C. statistically significant.      D. normally distributed.

13. The average commute time for this class is 18.9 minutes; last semester it was 17.2 minutes. State an appropriate **null model** to assess whether the difference is explainable by chance alone.
- A. The distribution of driving time depends on distance travelled.
  - B. Driving times are normally distributed.
  - C. Last semester's data are produced by  $p_1(y)$ ; this semester's data are produced by  $p_2(y)$ .
  - D. Last semester's data are produced by  $p(y)$ ; this semester's data are also produced by  $p(y)$ .
14. The following graph is the null distribution of the difference between average commute times under a null model. The test statistic is 2.2. Show how the two-sided  $p$ -value appears by drawing on the graph.



15. You simulate 10 random observations from the  $N(70, 10^2)$  distribution, and calculate their average to be 68.47. The difference between 68.47 and 70 is best characterized how?
- A. The difference is explained by chance alone.
  - B. The difference is explainable by chance alone.

16. The standard convention is that when the  $p$ -value is more than \_\_\_\_, then the differences found in study can be called “explainable by chance alone.”
- A 0.01    B. 0.05    C. 0.10    D. 0.50
17. An analysis in the chapter showed that daily stock returns are \_\_\_\_\_ the previous days return
- A. independent of  
B. not independent of  
C. positively skewed given  
D. normally distributed given
18. What statistic can be used to test whether the data were produced by a normal distribution?
- A. The difference between two sample averages  
B. The difference between two sample standard deviations  
C. The correlation coefficient calculated from a normal q-q plot  
D. The standardized Hoeffding-Wassily  $\tau$  statistic
19. State a null model for assessing whether the difference between mean age in the front and the back of the class are equal.
- A. All the age data are produced as iid from the same  $p(y)$ .  
B. The age data for the front of the class are produced as iid from  $p_1(y)$  and the age data from the back of the class are produced as iid from  $p_2(y)$ .  
C. The sample average of the age data from the front of the class is produced to be equal to the sample average of the age data in the back of the class.  
D. The ages in the front of the class are independent of ages in the back of the class.
20. State a null model for assessing whether the data are produced by a normal distribution.
- A. The data are produced by a normal distribution.  
B. The data are produced by a non-normal distribution.  
C. The data are produced by the bootstrap distribution.  
D. The data are produced by a discrete distribution.
21. What is a  $p$ -value?
- A. The probability that the null model is true, given the data.  
B. The probability that the data are true, given the null model.



C. The probability that the null model is more extreme than what was observed, given the data.

D. The probability that the data are more extreme than what was observed, given the null model.

22. Suppose the sample average of 100 numbers produced at random (using R, say) from the  $N(0,1)$  distribution is 0.168.

Then the difference between 0.168 and \_\_\_\_\_ is \_\_\_\_\_ by chance alone.

A. 0, explained

B. 0, explainable

C. 1, explained

D. 1, explainable

23. Suppose  $Y_1, Y_2, \dots, Y_n \sim_{\text{iid}} p(y)$ . You wish to test the hypothesis that  $p(y)$  is a normal distribution. What is the null model for how the data  $Y_1, Y_2, \dots, Y_n$  are produced?

#### Chapter 16:

1. Suppose  $X \sim N(10, 9)$  and  $Y \sim N(20, 16)$ . Then the distribution of  $X - Y$  is  
A.  $N(30, 25)$  B.  $N(-10, 25)$  C.  $N(10, 25)$  D.  $N(30, -7)$  E.  $N(-10, -7)$  F.  $N(10, -7)$

2. Assuming iid normal DATA, the true confidence level of the interval  $\bar{Y} \pm 1.96\hat{\sigma} / \sqrt{n}$  is  
A. = 95% B. > 95% C. < 95%

3. Suppose  $Y_1, Y_2, Y_3$  are iid  $N(0,1)$ . Then the distribution of  $Y_1^2 + Y_2^2 + Y_3^2$  is  
A.  $N(0, 3)$  B.  $N(3, 3)$  C.  $N(0, 1/\sqrt{3})$  D.  $\chi_1^2$  E.  $\chi_3^2$

4. When  $n$  increases,  $\hat{\sigma} / \sigma$  gets closer to  
A. 0 B. 1 C.  $\sigma$  D.  $1/\sqrt{n}$

5. The two-sample  $t$ -test uses  $\text{Var}(\bar{Y}_2 - \bar{Y}_1)$ , which is  
A.  $\sigma^2 - \sigma^2$  B.  $\sigma^2 + \sigma^2$  C.  $\sigma^2 / n_1 - \sigma^2 / n_2$  D.  $\sigma^2 / n_1 + \sigma^2 / n_2$

6. The degrees of freedom used in the two-sample  $t$ -test are  
A.  $n - 1$  B.  $n_1 + n_2 - 2$  C.  $n_1 - 1$  D.  $n_2 - 1$

7. The  $p$ -value for testing the ANOVA hypothesis  $\mu_1 = \mu_2 = \mu_3 = \mu_4$  is computed using which distribution?  
A.  $F$  B.  $\chi^2$  C.  $z$  D.  $t$

8. The confidence interval for the ratio of variances is computed using critical values from which distribution?  
 A.  $F$     B.  $\chi^2$     C.  $z$     D.  $t$
9. Suppose  $Y_1, Y_2, \dots, Y_{16} \sim_{\text{iid}} N(0,1)$ . Then the distribution of  $Y_1^2 + Y_2^2 + \dots + Y_{16}^2$  is  
 A.  $\chi_{15}^2$     B.  $\chi_{16}^2$     C.  $N(16, 32)$     D.  $T_{15}$     E.  $T_{16}$
10. The chi-squared distribution is a \_\_\_\_\_ distribution.  
 A. right-skewed    B. left-skewed    C. symmetric
11. Student's  $t$  distribution is a \_\_\_\_\_ distribution.  
 A. right-skewed    B. left-skewed    C. symmetric
12. Critical values from Student's  $t$  distribution are \_\_\_\_\_ the corresponding critical values from the standard normal distribution.  
 A. larger than    B. smaller than    C. equal to
13. Suppose  $Y_{ij} \sim_{\text{independent}} N(\mu_i, \sigma^2)$ , for  $i = 1, 2$ . Define the within-group averages as  $\bar{Y}_1, \bar{Y}_2$ . Then  $\text{Var}(\bar{Y}_1 - \bar{Y}_2) =$   
 A.  $\sigma_1^2 - \sigma_2^2$     B. 0    C.  $2\sigma^2$     D.  $\sigma^2(1/n_1 - 1/n_2)$     E.  $\sigma^2(1/n_1 + 1/n_2)$
14. With the same notations as in problem 13, what is the distribution of  $\bar{Y}_1 - \bar{Y}_2$ ?  
 A. Normal    B. Student's  $t$     C. Chi-squared
15. The  $F$  distribution was used in the book to test two different null hypotheses. What were they? (One line maximum for each).  
 Null Hypothesis 1: \_\_\_\_\_  
 Null Hypothesis 2: \_\_\_\_\_
16. Suppose  $Y_1, Y_2, \dots, Y_n \sim_{\text{iid}} N(\mu, \sigma^2)$ . How large does the sample size  $n$  have to be to ensure that the average  $\bar{Y} = (1/n) \sum_{i=1}^n Y_i$  is normally distributed?  
 A.  $n \geq 1$     B.  $n \geq 30$     C. It depends on the process that produces the data  
 D. The average  $\bar{Y}$  is never exactly normally distributed, for any  $n$
17. Suppose  $Y_1, Y_2 \sim_{\text{iid}} N(\mu, \sigma^2)$ . A confidence interval for  $\mu$  is  $\bar{Y} \pm 1.96\hat{\sigma}/\sqrt{2}$ , where  $\bar{Y}$  and  $\hat{\sigma}$  are the usual average and sample standard deviation calculated from the two values  $Y_1, Y_2$ . This confidence interval contains  $\mu$  in \_\_\_\_\_ of repeated samples.

- A. 0%      B. 69.96%      C. 95.44%      D. 100%

18. Suppose  $Y_1, Y_2 \sim_{\text{iid}} N(0,1)$ . Then  $Y_1^2 + Y_2^2 \sim$  \_\_\_\_\_

- A.  $N(0, 1/\sqrt{2})$       B.  $N(2, 2)$       C.  $T_2$       D.  $\chi^2_2$

19. The  $T$  distribution is a \_\_\_\_\_ distribution.

- A. normal      B. symmetric      C. right skewed      D. left skewed      E. discrete

20. The two-sample  $t$ -test uses  $\text{Var}(\bar{Y}_2 - \bar{Y}_1)$ , which, under the usual assumptions, is equal to

- A.  $\sigma^2 - \sigma^2$       B.  $\sigma^2 + \sigma^2$       C.  $\sigma^2/n_1 - \sigma^2/n_2$       D.  $\sigma^2/n_1 + \sigma^2/n_2$

21. Under the usual assumptions, the degrees of freedom used in the two-sample  $t$ -test are

- A.  $n - 1$       B.  $n_1 + n_2 - 2$       C.  $n_1 - 1$       D.  $n_2 - 1$

22. The  $p$ -value for testing the ANOVA hypothesis  $\mu_1 = \mu_2 = \mu_3 = \mu_4$  is computed using which distribution?

- A.  $F$       B.  $\chi^2$       C.  $z$       D.  $t$

23. The confidence interval for the ratio of variances is computed using critical values from which distribution?

- A.  $F$       B.  $\chi^2$       C.  $z$       D.  $t$

24. Let  $Y \sim N(70, 10^2)$  and let  $X = Y/100$ . Then  $X \sim$  \_\_\_\_\_

- A.  $N(70, 1^2)$       B.  $N(0.70, 1^2)$       C.  $N(70, 0.1^2)$       D.  $N(0.70, 0.1^2)$

25. Suppose  $Y_1, Y_2 \sim_{\text{iid}} N(\mu, \sigma^2)$ . Then  $(Y_1 + Y_2)/2 \sim$  \_\_\_\_\_

- A.  $N(\mu, \sigma^2)$       B.  $N(\mu/2, \sigma^2)$       C.  $N(\mu, \sigma^2/2)$       D.  $N(\mu/2, \sigma^2/2)$

26. The distribution of the sample variance  $\hat{\sigma}^2$  most directly involves which distribution?

- A. normal      B. chi-square      C. Student's t      D. F

27. How is the Student's t distribution different from the standard normal distribution?

- A. Student's t distribution is symmetric
- B. Student's t distribution is right skewed
- C. Student's t distribution is uniform (flat)
- D. Student's t distribution has a larger variance

28. What is the advantage of constructing a confidence interval for the difference  $\delta = \mu_1 - \mu_2$  versus testing the hypothesis that  $\delta = 0$ ?

- A. The confidence interval does not assume normality
- B. There is only a 5% chance of making a mistake with the confidence interval
- C. The confidence interval allows you to quantify the size of the difference between  $\mu_1$  and  $\mu_2$

29. See problem 28 above. The test for  $\delta = 0$  is called the \_\_\_\_\_ test.

- A. Lilliefors      B. Two-sample t      C. Chi-squared      D. Levene

30. The distribution of the ratio of sample variance  $\hat{\sigma}_1^2 / \hat{\sigma}_2^2$  most directly involves which distribution?

- A. normal      B. chi-square      C. Student's t      D. F

31. What is the null hypothesis in an ANOVA?

- A. The variances of all groups are equal
- B. The standard deviations of all groups are equal
- C. The means of all groups are equal
- D. The medians of all groups are equal

32. You wish to use a statistical method that is based on the assumption that your data are produced by a normal distribution. However, you also know that no real data you can ever analyze are produced by a normal distribution. Thus,

- A. this method should not be used at all.
- B. this method should be used only if the test for normality shows  $p > .05$ .
- C. this method should be used only if the distribution that produces the data is a normal distribution.
- D. the validity of this method depends on the degree of non-normality of the data generating process.

33. Suppose  $X \sim N(70, 12^2)$  and  $Y = 100 - X$ . Then  $Y$  \_\_\_\_\_

- A.  $\sim N(30, 112^2)$     B.  $\sim N(30, -12^2)$     C.  $\sim N(30, 12^2)$     D.  $\sim N(70, -12^2)$     E.  $\sim N(30, 88^2)$

34. Suppose  $Y_1$  and  $Y_2$  are iid from the  $N(\mu, \sigma^2)$  distribution. Then  $Y_1 + Y_2$  \_\_\_\_\_

- A.  $\sim N(\mu, \sigma^2)$     B.  $\sim N(2\mu, \sigma^2)$     C.  $\sim N(\mu, 2\sigma^2)$     D.  $\sim N(2\mu, 2\sigma^2)$

35. Suppose  $Y_1$  and  $Y_2$  are iid from the  $N(\mu, \sigma^2)$  distribution. Let  $\bar{Y}$  and  $\hat{\sigma}$  be the ordinary sample mean and standard deviation calculated from  $Y_1$  and  $Y_2$ . If 100 samples of size  $n = 2$  each are taken, about how many of the 100 intervals  $\bar{Y} \pm 1.96(\hat{\sigma}/\sqrt{2})$  will contain  $\mu$ ?

- A. 100    B. 95    C. 70    D. 0

36. Suppose the 95% confidence interval for  $\mu_1 - \mu_2$  is  $2.3 < \mu_1 - \mu_2 < 10.2$ . Then the  $p$ -value for testing  $H_0: \mu_1 = \mu_2$  will show

- A.  $p < 0.05$     B.  $p > 0.05$     C.  $p = 0.05$     D.  $p = 0.95$

37. Consider the  $F_{30,15}$  distribution, which has 30 numerator and 15 denominator degrees of freedom. The mean of this distribution is closest to \_\_\_\_\_.

- A. 1    B. 2    C. 15    D. 30

38. There are three groups in an analysis of variance (ANOVA) model. The null hypothesis is  $H_0$ : \_\_\_\_\_

- A.  $\bar{y}_1 = \bar{y}_2 = \bar{y}_3$     B.  $\mu_1 = \mu_2 = \mu_3$     C.  $\hat{\sigma}_1 = \hat{\sigma}_2 = \hat{\sigma}_3$     D.  $\sigma_1 = \sigma_2 = \sigma_3$

39. Suppose  $Y_{ij} \sim_{\text{ind}} N(\mu_i, \sigma^2)$ , for  $i=1,2$  (groups), and with  $n_1=10$  observations in group 1 and  $n_2=15$  observations in group 2. Let  $\hat{\sigma}_1^2$  and  $\hat{\sigma}_2^2$  be the ordinary estimated variances for each of the two groups. Then  $\hat{\sigma}_1^2 / \hat{\sigma}_2^2 \sim$  \_\_\_\_\_
- A.  $N(0,1)$     B.  $T_{23}$     C.  $F_{10,15}$     D.  $F_{9,14}$
40. Suppose  $Y_1, \dots, Y_{10} \sim_{\text{iid}} N(70, 10^2)$  and  $X_1, \dots, X_{20} \sim_{\text{iid}} N(75, 10^2)$ , independent of the  $Y$ 's. What is the distribution of  $\bar{X} - \bar{Y}$ ?
41. Suppose  $U \sim \chi_2^2$  and  $V \sim \chi_3^2$ , independent of  $U$ . What is the distribution of  $(U/2)/(V/3)$ ?

#### Chapter 17:

- The unrestricted (full) model is  $Y | (X_1, X_2) = (x_1, x_2) \sim N(\beta_0 + \beta_1 x_1 + \beta_2 x_2, \sigma^2)$ . The restricted (null) model is  $Y | (X_1, X_2) = (x_1, x_2) \sim N(\beta_0, \sigma^2)$ . The degrees of freedom for the likelihood ratio test statistic are  $df$ . What is  $df$ ?  
A. 1    B. 2    C. 3    D. 4    E. 5    F. 6    G. 7    H. 8    I. 9    J. 10
- The null distribution of the likelihood ratio test statistic is  
A. chi-squared    B. normal    C. approximately chi-squared    D. approximately normal
- Which is the most commonly observed form of the likelihood ratio chi-squared statistic?  
A.  $n \ln(\hat{\sigma}_0^2 / \hat{\sigma}_1^2)$     B.  $n \ln\{(\bar{Y} - m_0)/(\hat{\sigma} / \sqrt{n})\}$     C.  $n\{\hat{\rho} / s.e(\hat{\rho})\}$     D.  $n \ln\left\{\int yp(y)dy\right\}$
- Which is the Pearson chi-squared statistic?  
A.  $2\sum f_i \ln(f_i / e_i)$     B.  $\sum (f_i - e_i)^2 / e_i$     C.  $1 - \hat{\sigma}_1^2 / \hat{\sigma}_0^2$     D.  $2(LL_1 - LL_0)$     E.  $\frac{\exp(\beta_0)}{1 + \exp(\beta_0)}$
- Which is the best approximation to  $\ln(1+x)$  when  $x$  is near 0?  
A.  $1+x$     B.  $1+x^2/2$     C.  $x$     D.  $x^2/2$
- What is the ugly rule of thumb about the  $p$ -values from the likelihood ratio test?  
A. They tend to be too small    B. They tend to be too large    C. They tend to be around 0.05    D. They tend to be around 0.10
- Which model is the "chance-only" model?  
A. The restricted model    B. The unrestricted model
- When is the test based on the square of the  $t$  statistic optimal?  
A. When the data-generating process is iid normal

- B. When the restricted model has one parameter  
C. When the chi-square distribution is used to find the  $p$ -value
9. In many examples in the chapter, the likelihood ratio test rejected the null model  $p_0(y|\theta_0)$  when the variance ratio  $\hat{\sigma}_0^2 / \hat{\sigma}_1^2$  was  
A. Large      B. Small      C. Less than 0.05      D. More than 0.05
10. In multiple regression analysis, which statistic is optimal for testing the “chance-only” model?  
A. The  $R^2$  statistic      B. The least squares statistic      C. The error sum of squares
11. When are the likelihood ratio chi-squared statistic and Pearson’s chi-square statistic similar?  
A. When the restricted model is true  
B. When the unrestricted model is true
12. The cross-classification table has 2 rows and 3 columns. What are the degrees of freedom for the chi-squared test of independence? (Just give a single number, no words).
13. A statistic measuring model fit is  $G = LL - k$ . When comparing two different models, the one with \_\_\_\_  $G$  is preferred.  
A. lower      B. higher
14. Which is a better model for dice outcomes?  
A. Multinomial      B. Shifted Poisson
15. Which method for testing hypotheses is optimal?  
A. bootstrap test      B. t test      C. F test      D. likelihood ratio test
16. What assumption is needed to apply the likelihood ratio test?  
A. data are a random sample from a population  
B. the data are normally distributed  
C. the data are produced by a distribution within a specified family of distributions  
D. the sample size is greater than 30
17. Under the null (restricted) model,  $\hat{\sigma}_0^2$  is the estimated variance of  $Y$ . Under the alternative (unrestricted) model,  $\hat{\sigma}_1^2$  is the estimated variance. When should you reject the null (restricted) model?  
A. When  $\hat{\sigma}_0^2$  is much larger than  $\hat{\sigma}_1^2$ .  
B. When  $\hat{\sigma}_1^2 / \hat{\sigma}_0^2 < F_{1-\alpha, n-1}$ .  
C. When  $\chi^2 = 2\ln(\hat{\sigma}_1^2 / \hat{\sigma}_0^2)$ .  
D. When  $\hat{\sigma}_1^2 / \hat{\sigma}_0^2 \sim N(0,1)$ .
18. Suppose the data are produced as iid  $N(10, \sigma^2)$ . Then the MLE of  $\sigma^2$  is  
A.  $\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$       B.  $\frac{1}{n} \sum_{i=1}^n (y_i - 10)^2$       C.  $\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$       D.  $\frac{1}{n-1} \sum_{i=1}^n (y_i - 10)^2$
19. Here is a generic probability distribution:
- |          |   |         |         |         |         |         |         |       |
|----------|---|---------|---------|---------|---------|---------|---------|-------|
| $y$      | : | 1       | 2       | 3       | 4       | 5       | 6       | Total |
| $p(y)$ : |   | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\pi_4$ | $\pi_5$ | $\pi_6$ | 1.0   |

How many degrees of freedom are there in the unknown parameters of this distribution?

- A.  $n$     B.  $n - 1$     C. 5    D. 6

20. What is the null hypothesis that is tested in the analysis of cross-classification tables?

- A. The row and column variables are normally distributed  
B. The row means are equal  
C. The column variances are equal  
D. The column variable is independent of the row variable

21. Consider the following cross classification table.

	Masters	Ph.D.
Male	229	16
Female	224	12

How many degrees of freedom are there for the likelihood ratio chi-square test statistic?

- A. 1    B. 2    C. 3    D. 4    E.  $n - 1$

22. The AIC statistic is defined as  $AIC = -2LL + 2k$ , where  $LL$  is the maximized log likelihood and  $k$  is the number of parameters in the model. When comparing two models, you should pick the one for which AIC is \_\_\_\_\_.

- A. greater than  $\chi^2_{0.95,k}$   
B. less than 1.0  
C. smallest

23.  $\ln(L_1/L_0) =$

- A.  $\ln(L_1) + \ln(L_0)$     B.  $\ln(L_1) - \ln(L_0)$     C.  $L_1 + L_0$     D.  $L_1 - L_0$

24. When are test statistics equivalent?

- A. When one is a monotonic function of the other.  
B. When they are both based on the iid normal model  
C. When they are both linear functions of the data  
D. When they are both distributed as chi-squared random variables

25. Which assumptions are needed for the square of the t-statistic to be equivalent to a likelihood ratio statistic? (Circle all that apply, seven points per correct selection)

- A. The data are produced by a normal distribution



- B. The data are produced independently
- C. The distributions that produce the data are identical
- D. The data are produced by an exponential distribution
- E. The means of the distributions are known
- F. The variances of the distributions are known

26. Suppose 100 data values are produced by the following distribution:

$y$	$p(y)$
A	0.30
B	0.50
C	0.20

What is the expected number of data values that are "C"? \_\_\_\_\_

27. What is the statistic  $\sum (f_i - e_i)^2 / e_i$  called?

- A. The likelihood ratio
- B. The Pearson chi-square
- C. The multinomial p-value
- D. The coefficient of determination

28. When tested whether Bernoulli data are produced by the Bernoulli(0.5) model rather than the unrestricted Bernoulli( $\pi$ ) model, the degrees of freedom are

- A. 0.0
- B. 0.5
- C. 1.0
- D. 2.0

29. The AIC statistic is defined as  $AIC = -2 * (\log \text{likelihood}) + 2 * k$ , where  $k$  is the number of parameters in the model. The AIC for model 1 is 34.5, and the AIC for model 2 is 94.2. Which model is better?

- A. Model 1
- B. Model 2

30. A model is  $Y_{ij} \sim_{\text{ind}} N(\mu_i, \sigma^2)$ . Relative to this model, the model  $Y_{ij} \sim_{\text{ind}} N(\mu, \sigma^2)$  is called a

- A. restricted model
- B. unrestricted model
- C. generic model
- D. likelihood ratio model

31. What is the approximate distribution of the likelihood ratio test statistic  $2(LL_1 - LL_0)$ ?

- A. Normal
- B. Student's T
- C. Chi-squared
- D. F

32. When does  $2(LL_1 - LL_0)$  suggest that the unrestricted model is true?

- A. when it is too large

- B. when it is too small
- C. when it is approximately 0
- D. when it is less than 0

33. The ANOVA F statistic is \_\_\_\_\_ a likelihood ratio statistic.

- A. equal to
- B. more than
- C. less than
- D. monotonically related to

34. Consider the following model:

$$y \sim p(y)$$

- A.  $\pi_1$
- B.  $\pi_2$
- C.  $\pi_3$
- 1.0

The observed data are B,B,C,A,B. Then the likelihood function is

- A.  $(3/5) \times (1/3) \times (1/3)$
- B.  $(0.2, .6, .2)$
- C.  $\pi_1 \times \pi_2^3 \times \pi_3$
- D.  $A \times B^3 \times C$

35. Which distribution is shown in problem 34?

- A. Normal
- B. Poisson
- C. Bernoulli
- D. Multinomial

36. The likelihood ratio test statistic for testing independence in two-way cross-classification tables is  $\chi^2 =$

- A.  $2 \sum_i f_i \ln(f_i / e_i)$
- B.  $\sum_i (f_i - e_i)^2 / e_i$
- C.  $2 \sum_i \sum_j f_{ij} \ln(f_{ij} / e_{ij})$
- D.  $\sum_i \sum_j (f_{ij} - e_{ij})^2 / e_{ij}$

37. Suppose the AIC statistic for model is given as  $AIC = -2LL + 2k$ . Model 1 gives  $AIC_1$  and model 2 gives  $AIC_2$ . When is model 1 preferred over model 2?

- A. When  $AIC_1 > AIC_2$
- B. When  $AIC_1 < AIC_2$
- C. When  $AIC_1 = AIC_2$
- D. When  $AIC_1 \cong AIC_2$

38. Suppose  $Y_1, \dots, Y_n \sim_{\text{iid}} N(\mu, \sigma^2)$ . Then the likelihood ratio test statistic for testing  $H_0: \mu = m_0$  is equivalent to the
- ordinary t-statistic
  - squared t-statistic
  - estimated standard deviation
  - estimated variance
39. Why is it important that a test statistic be a likelihood ratio statistic?
- likelihood ratio test statistics are normally distributed
  - likelihood ratio test statistics are optimal
  - likelihood ratio test statistics have the chi square distribution
  - likelihood ratio test statistics are nonparametric
40. Suppose the restricted (null) parameter space is  $\Theta_0$  and the unrestricted parameter space is  $\Theta_1$ . Then
- $\max_{\theta \in \Theta_0} L(\theta \mid \text{data}) < \max_{\theta \in \Theta_1} L(\theta \mid \text{data})$
  - $\max_{\theta \in \Theta_0} L(\theta \mid \text{data}) > \max_{\theta \in \Theta_1} L(\theta \mid \text{data})$
  - $\max_{\theta \in \Theta_0} L(\theta \mid \text{data}) = \max_{\theta \in \Theta_1} L(\theta \mid \text{data})$
41. For which model is the ANOVA F statistic equivalent to a likelihood ratio test statistic?
- $Y_{ij} \sim_{\text{independent}} p_i(y)$
  - $Y_{ij} \sim_{\text{iid}} p(y)$
  - $Y_{ij} \sim_{\text{independent}} N(\mu_i, \sigma^2)$
  - $Y_{ij} \sim_{\text{independent}} \text{Bernoulli}(\pi_i)$
42. The Pearson chi-square test statistic for testing independence in a contingency table is given by
- $n \{ \ln(\hat{\sigma}_0^2 / \hat{\sigma}_1^2) \}$
  - $n \times \ln \left\{ 1 + \frac{g-1}{n-g} f \right\}$
  - $2 \sum \sum f_{ij} \ln(f_{ij} / e_{ij})$
  - $\sum \sum (f_{ij} - e_{ij})^2 / e_{ij}$

43. Why is it important that a test statistic be a likelihood ratio statistic?

- A. likelihood ratio test statistics are normally distributed
- B. likelihood ratio test statistics are optimal
- C. likelihood ratio test statistics have the chi square distribution
- D. likelihood ratio test statistics are nonparametric

44. The restricted and full models for dice rolls are, respectively, given as follows:

$y$	$p_0(y)$	$y$	$p_1(y)$
1	1/6	1	$\pi_1$
2	1/6	2	$\pi_2$
3	1/6	3	$\pi_3$
4	1/6	4	$\pi_4$
5	1/6	5	$\pi_5$
6	1/6	6	$\pi_6$

How many parameters must be estimated from the data in the restricted model?

- A. 0   B. 1   C. 2   D. 3   E. 4   F. 5   G. 6

45. Consider the logistic regression model which states that

$$\Pr(Y=1 | X=x) = \exp(\alpha + \beta x) / \{1 + \exp(\alpha + \beta x)\}.$$

After estimation, the maximized logarithm of the likelihood function is  $\ln L = -5.0$ . What is the AIC statistic?

- A. -3.0   B. -1.0   C. 12.0   D. 14.0

## Chapter 18:

1. In the manufacturing example, the number 250 is the

- A. Hypothesized expected value   B. Expected value   C. Hypothesized sample average   D. Sample average

2. A food vendor sells various food and beverage items. Each sale results in a ticket amount, for examples, \$0.85 for a cup of coffee, \$6.54 for a bagel and tea, etc. They want to estimate the mean sale amount per ticket. There is a formula

$$n = \frac{Z_{1-\alpha/2}^2 \sigma^2}{(a.m.)^2}$$

that they can use to determine a sample size. Give a value of *a.m.* that is clearly too large and explain why it is too large. (one sentence)

3. When the alternative hypothesis is true, the distribution of the t-statistic is  
A. centered at 0    B. shifted away from zero    C. standard normal    D. uniform
4. Suppose  $Y_1, \dots, Y_8 \sim_{\text{iid}} p(y)$ , where  $\text{Var}(Y_1) = 4$ . Then  $\text{Var}\{(Y_1 + \dots + Y_8)/8\} =$   
A. 0.25    B. 0.50    C. 1.0    D. 2.0    E. 4.0
5. What is the difference between standard error (s.e.) and accuracy margin (a.m.)?  
A. s.e. is calculated from data; a.m. is from the mind of the analyst  
B. a.m. is calculated from data; s.e. is from the mind of the analyst
6. Suppose  $Y \sim \text{Bernoulli}(\pi)$ . What is the *maximum* value of  $\text{Var}(Y)$ ?  
A. 0.25    B. 0.50    C. 1.00
7. Which statistical software does Hans prefer?  
A. SAS    B. Minitab    C. SPSS    D. Stata    E. R    F. HappyStat
8. Power = \_\_\_\_\_  
  
A. 1- p-value    B. 1 -  $\alpha$     C.  $\text{Pr}(\text{reject } H_0 \mid H_0 \text{ is false})$
9. When  $H_0$  is false and all else is held fixed, what happens to power when the sample size increases?  
A. It converges to  $\mu$   
B. It becomes closer to normally distributed  
C. It gets higher  
D. It becomes closer to uniformly distributed
10. What is the difference between a standard error (s.e.) and accuracy margin (a.m.)?  
A. s.e. is calculated *after* observing data, a.m. *before*  
B. s.e. has a square root of n in the denominator, a.m. does not  
C. s.e. applies to proportions, a.m. applies to means
11. When performing power calculations, how do you pick the parameters of the model?  
A. using “what-if” scenarios  
B. using the data collected from the study  
C. using the defaults of the software
12. In group 1, 30% of 100,000 patients are cured, and in group 2, 32% of 100,000 patients are cured. The difference between 30% and 32% is  
A. not easily explained by chance alone.  
B. the power of the test.

- C. a chi-squared random variable.
- D. the  $p$ -value for testing that the two true proportions are equal.

13. Based on your desired accuracy margin (a.m.) of 20.0, you found that you need to sample  $n = 100$  observations. After discussion with the boss, you decide on a.m. = 10.0 instead. Keeping all else fixed, what sample size do you now need?

- A.  $n = 25$     B.  $n = 50$     C.  $n = 200$     D.  $n = 400$

14. Suppose the alternative hypothesis is true. What happens to the power of the test when you increase the sample size?

- A. It becomes approximately normally distributed.
- B. It converges to the true mean.
- C. It becomes unbiased.
- D. It gets closer to 1.0.

15. A statistic is distributed as chi-squared when the null hypothesis is true. What is its distribution when the alternative hypothesis is true?

- A. standard normal
- B. Student's  $t$
- C. Exponential
- D. noncentral chi-squared

16. Suppose 32% of 100,000 people treated with drug A are cured and 30% of 100,000 different people treated with drug B are cured. The difference between 32% and 30% is

\_\_\_\_\_.

- A. Explainable by chance alone.
- B. Explained by chance alone.
- C. Not easily explained by chance alone.
- D. Statistically insignificant.

17. What is the main problem with choosing a larger sample size?

- A. Results become less significant.
- B. The  $p$ -values become larger.

- C. The distribution of the sample average becomes non-normal.
- D. The study becomes more costly.

18. What is power?

- A. The probability of observing a difference as large as what you observed, when the null model is true.
- B. The probability of observing a difference as large as what you observed, when the null model is false.
- C. The probability of rejecting the null model, when the null model is true.
- D. The probability of rejecting the null model, when the null model is false.

19. What is “post hoc power”?

- A. The power of the study, assuming the standard deviation  $\sigma$  is some pre-specified value.
- B. The power of the study, assuming the mean  $\mu$  is some pre-specified value.
- C. The power of the study, assuming the sample size  $n$  is some pre-specified value.
- D. A useless statistic

20. How large a sample size,  $n$ , do you need for a statistical study?

- A.  $n \geq 30$
- B.  $n < 30$
- C. an  $n$  that gives you sufficient confidence
- D. an  $n$  that gives you sufficient power

21. Suppose  $Y_1, \dots, Y_{10} \sim_{\text{iid}} N(310, 4.5^2)$ . What is the distribution of  $(1/10)(Y_1 + \dots + Y_{10})$ ?

- A.  $N(310, 4.5^2)$
- B.  $N(310/10, 4.5^2)$
- C.  $N(310, 4.5^2/10)$
- D.  $N(310/10, 4.5^2/10)$

22. Suppose you want to estimate a mean  $\mu$  to within  $\pm 10$ , with 95% confidence, using the confidence interval  $\bar{y} \pm 1.96\sigma / \sqrt{n}$ . How large should  $n$  be?

- A.  $1.96^2 \times \sigma^2 / 10^2$
- B.  $1.96 \times \sigma^2 / 10^2$
- C.  $1.96^2 \times \sigma / 10^2$
- D.  $1.96 \times \sigma / 10^2$
- E.  $1.96 \times \sigma / 10$

23. Consider the following table.

		Action based on data	
		Fail to Reject $H_0$	Reject $H_0$
State of Nature	$H_0$ true		
	$H_0$ false	??????	

Fill in the ?????? cell with the appropriate term.

- A. Power   B. Type I error   C. Type II error   D. correct decision   E. significance level

#### Chapter 19:

- Nonparametric methods usually assume what?  
A. iid DATA   B. normally distributed DATA   C. Poisson distributed DATA   D. Simulated DATA
- When should you use permutation (or randomization) tests for contingency tables?  
A. When the expected frequencies in the table are all more than 5.0  
B. When the expected frequencies in the table are all less than 5.0  
C. When the expected frequencies in the table are non-normally distributed  
D. When the degrees of freedom are large
- When is the rank-transformed two-sample  $t$ -test likely to be best?  
A. When the distributions are normal  
B. When the distributions are Cauchy  
C. When the data are dependent  
D. When the variances are identical
- What is an advantage of the bootstrap percentile  $t$  confidence interval compared to the ordinary confidence interval for the mean when the distribution is skewed?  
A. It is wider   B. It is asymmetric   C. It has lower confidence   D. It is a Bayesian method
- What did George Box say?  
A. All models are useful.   B. All models are useless.   C. All models are wrong.   D. All models are wrong, but some are useful.   E. All models are wrong, and most are useless.
- If a procedure is *level robust*, then  
A. it is more powerful than competing procedures   B. its error rate is approximately  $\alpha$   
C. its estimates are consistent   D. it is a simulation-based method



7. What assumptions do you need to make about the data  $Y_1, Y_2, \dots, Y_n$  in order to know that  $T = \frac{\bar{Y} - \mu}{\hat{\sigma} / \sqrt{n}}$  has the  $T_{n-1}$

distribution? Select all answers that apply.

- A. They are independent
  - B. They are identically distributed
  - C. They are produced by a normal distribution
  - D. They are fixed (non-random) data values
  - E. Their mean,  $\mu$ , is known
  - F. Their variance,  $\sigma^2$ , is known
  - G. The sample size,  $n$ , is greater than 30 ( $n > 30$ )
  - H. Their  $p$ -value is greater than 0.05
8. What was the application in the example with the percentile- $t$  bootstrap interval?
- A. Burn patient mortality    B. Cancer survival    C. Back injuries    D. Public opinion polls
9. Which transformation of the data tends to make your procedures more robust?
- A. Rank    B. Linear    C. Quadratic    D. Exponential
10. Another name for “permutation test” is “\_\_\_\_\_ test.”
- A. bootstrap    B. randomization    C. Student’s  $t$     D. chi-square
11. For which data-producing distribution is the two-sample  $t$  test not robust?
- A. Discrete ordinal    B. Normal    C. Cauchy
12. The bootstrap percentile- $t$  method gives a confidence interval for the \_\_\_\_\_.
- A. mean    B. median    C. standard deviation    D. significance level
13. Suppose a hypothesis testing method is level robust. Then when the null hypothesis is true, the  $p$ -value will be less than 0.05 approximately \_\_\_\_\_ of the time.
- A. 0%    B. 5%    C. 50%    D. 95%    E. 100%
14. What is used to construct the percentile  $t$  confidence interval?
- A. bootstrap
- B. rank transformation
- C. permutation test
- D. Student’s  $t$  distribution
15. What is a “nonparametric” method?
- A. one without assumptions
- B. one where generic distributions are assumed to produce the data
- C. one where the parameters are known
- D. one that you should use whenever the distribution that produces the data is not the normal distribution
16. Which test has the highest power when the distributions are Cauchy?
- A. rank-transformed test
- B. two-sample  $t$  test
- C. log-transformed test
- D. chi-square test

17. The shifted Cauchy distribution is \_\_\_\_\_.
- A. right skewed
  - B. left skewed
  - C. normally distributed
  - D. outlier-prone
18. A “level robust” test will
- A. result in a “reject  $H_0$ ” conclusion  $\cong 5\%$  of the time when  $H_0$  is true.
  - B. result in a “reject  $H_0$ ” conclusion  $\cong 5\%$  of the time when  $H_0$  is false.
  - C. never result in a “reject  $H_0$ ” conclusion when  $H_0$  is true.
  - D. never result in a “reject  $H_0$ ” conclusion when  $H_0$  is false.
19. When using the permutation test, how many random permutation samples are ideal?
- A. 2    B. 5    C. 30    D. Infinitely many
20. What assumption is needed for the bootstrap percentile-t confidence interval as given in the chapter?
- A. The data are produced by normal distributions.
  - B. The data are produced as an iid sample.
  - C. The data are produced from shifted Cauchy distributions.
  - D. The sample size is less than 100.
21. When is a supposed 95% confidence interval for  $\mu$  non-robust?
- A. When the true confidence level is close to 95%.
  - B. When the true confidence level is 30%.
  - C. When  $\mu$  is not in the interval.
  - D. When the sample size,  $n$ , is less than 30.
22. The p-value for a randomization test is 0.15. This means “about 15 out of 100.” What does the “100” in “15 out of 100” refer to?
- A. A table of critical values.
  - B. Randomly shuffled data sets.
  - C. The sample size in the study.
  - D. The number of p-values that are less than .15.

23. If two-sample data are produced by Cauchy distributions, then which method is most powerful test?
- The Wilcoxon rank sum test.
  - The two-sample t test.
  - The F test.
  - The normal test.
24. What do you assume about the distribution that produces the data  $Y_1, Y_2, \dots, Y_n$  when you use the bootstrap percentile t interval?
- It is the rank distribution.
  - It is the permutation distribution.
  - It is a generic distribution  $p(y)$ .
  - It is a normal distribution  $N(\mu, \sigma^2)$  having unknown mean and
25. Someone says the 95% confidence interval  $\bar{Y} \pm T_{n-1, 0.975} \hat{\sigma} / \sqrt{n}$  is non-robust. What does that mean?
- It means that  $\mu$  is not inside the interval.
  - It means that the percentage of samples for which  $\mu$  is not inside the interval is different from 5%.
  - It means that  $\bar{y}$  is not inside the interval.
  - It means that the percentage of samples for which  $\bar{y}$  is not inside the interval is different from 5%.
26. A data set is 4.5, 5.4, 1.2. The rank-transformed data values are, respectively,
- 10, 5, 15
  - 2, 3, 1
  - 50, 100, 0
  - medium, high, low
27. The Kruskal-Wallis test is used to test which null hypothesis?
- The data are all produced by the same distribution
  - The means of the groups are all different
  - The standard deviations of the groups are all the same
  - The data are independent in each group
28. What is an advantage of the bootstrap confidence interval over the standard  $\bar{Y} \pm T_{n-1, 0.975} \hat{\sigma} / \sqrt{n}$  interval?
- The bootstrap interval is better when the data are produced by a normal distribution
  - The bootstrap interval requires less computation
  - The bootstrap interval is less robust

D. The bootstrap interval is not symmetric