ECN221 Study Guide for Exams

The chapter numbers below correspond to the chapters in *Business Statistics, 2e* by Donnelly.

Note that you must have your student ID to be able to take the exam. We will begin seating a few minutes before the exam and the first 5 minutes will be dedicated to seating you. If your student ID ends 0-4 come in on the side furthest from the MU and if your student ID ends 5-9 come in on the door closest to the MU. Have your ID out to show to the proctor before you come in the room. Exams will be distributed when all papers, notebooks, phones, etc. are put away. Each exam will include the formula sheet attached here. The in semester exams will have 25 questions and the final exam will have 30.

Choose the best answer. Do not write letters in the margin or communicate with other students in any way. If you have a question note it on your exam and ask for clarification when your exam is returned. In the meantime choose the best answer. Neither the proctors nor I will answer questions during the exam. I will post a key and the scores to Bb after the scores are returned from the testing center.

Please check each question and possible answers thoroughly as questions at the bottom of a page sometimes run onto the next page. Below is a list of topics that may appear on an exam as well as some sample questions. Please note that what is below is only a guide and that any topic we have covered in class or which is part of the assigned reading is a potential topic for an exam question and that the purpose of the exam is specifically to test your knowledge. In addition the WP Carey School requires students to exhibit critical thinking skills in order to receive outstanding marks. The exam will have 70-80% of the points (what you need to pass) from standard questions such as you have seen on homework assignments, quizzes, and that we have worked in class. The remainder of the exam will involve questions that require you to bring together multiple pieces of information or exhibit a superior knowledge of the material. *Examples* of this type of question are marked with a **CT**.

Concepts to Study

Syllabus: basic questions about the course.

Chapter 1: vocabulary and definitions and ability to identify parts of data sets.

1. data and its components
2. scales of measurement
3. distinguishing types of data, e.g. categorical
4. difference between population and sample
5. difference between descriptive statistics and inferential statistics

Chapter 2: understand tabular and graphical displays

1. bar and pie charts
2. frequency, relative frequency, and percent frequency
3. cumulative frequency etc.
4. histograms
5. stem-and-leaf displays
6. crosstabulations and scatter plots
7. box and whisker plots

Chapter 3: numerical measures of central tendency, variability and association

1. mean, median, mode, percentiles and quartiles, know what they are and how to compute them from data
2. range, interquartile range, variance, standard deviation, and coefficient of variation: know what they are, how to compute them, and how they are different.
3. skewness versus symmetry
4. z-scores, how to compute and interpret them
5. understand the empirical rule
6. understand ways to detect outliers
7. understand measures of association such as covariance and correlation and be able to compute correlation from other summary statistics.

Chapter 4: Probability

1. understand counting rules and be able to calculate combinations possible
2. understand basic probability concepts and definitions along with the underlying rules
3. be able to compute joint, marginal, and conditional probabilities from data
4. understand rules about conditional probability including Bayes' Rule.

Chapter 5: vocabulary and definitions and ability to identify discrete random variables.

1. understand and compute expected value and variance
2. find probabilities for Binomial random variables
3. Find probabilities for random variables that follow the Poisson distribution.

Chapter 6: continuous random variables

1. know definitions and be able to identify continuous random variables
2. understand the uniform distribution
3. understand the exponential distribution.
4. understand the normal and standard normal distributions
5. be able to find probabilities, expected values and variances for random variables for all of the above distributions.

Chapter 7: sampling and sampling distributions

1. understand types of sampling and basic definitions from the chapter
2. understand the implications of the central limit theorem
3. understand the difference between standard deviation and standard error
4. know the distributions of $\bar{x}$ and $\bar{p}$

Chapter 8: Confidence Intervals

1. understand definitions and basic concepts such as confidence coefficient, margin of error etc.
2. be able to construct confidence intervals under various assumptions, for example σ known and unknown.
3. you will need to know how to read both the t and the z table.
4. be able to calculate the necessary sample size.

Chapter 9: Hypothesis Testing

1. understand definitions and basic concepts such as reject, type I error, level of significance etc.
2. be able to correctly state a null and alternative hypothesis given an actual problem.
3. be able to test a hypothesis for one mean under various assumptions, for example σ known and unknown.
4. be able to test a hypothesis, this involves being able to find a test statistic and/or p-value and to know what a p-value is.
5. you will need to know how to read both the t and the z table.
6. be able to conduct tests for both one tail and two tail scenarios.

Chapter 10: Comparisons involving means from two populations.

1. know terms and definitions
2. be able to test hypotheses concerning the means from two populations when the variance is known.
3. be able to test hypotheses concerning the means from two populations when the variance is *not* known.

Chapter 11: ANOVA.

4. Understand the concept of ANOVA.
5. Be able to test hypotheses concerning the equality of three or more means.
6. Be able to find F values from the F table.
7. Be able to read an ANOVA table and relate the values in the table to one another, for example be able to find missing values in the table given values appearing in the table.

Chapter 14: Simple Linear Regression

1. know definitions and concept concerning simple regression, for example the regression models and estimated equation.
2. understand the least squares method
3. know how to calculate and interpret the coefficient of determination or R^2.
4. Know what the model assumptions are and be able to draw conclusions about the residuals by looking at residual plots.
5. be able to read output from excel concerning regression and correctly interpret the output.
6. an example of interpreting the output involves being able to test hypotheses concerning the population parameters but you should be able to do this outside of excel.
7. be able to make predictions based on regression results and also be able to construct confidence intervals and prediction intervals.
8. be able to examine residuals to draw conclusions about the model assumptions.

Chapter 15: Multiple Regression

1. similar to chapter 14 you should be able to make all the extensions from simple regression to multiple regression
2. understand the F test in multiple regression and how it is different from the F test in simple regression
3. understand what adjusted R^2 is and how to calculate it using the formula on the formula sheet.

4. Understand the concept of multicollinearity
5. Be able to find predicted values for multiple regression
6. understand how to use categorical or dummy variables and how to interpret their estimated parameters.

Chapter 3:

sample mean: $\bar{x} = \sum_{i=1}^{n} x_i/n$

sample variance: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$.

sample standard deviation: $s = \sqrt{s^2}$.

Coefficient of Variation: $CV = \frac{s}{\bar{x}}\left(100\%\right)$

sample z-Score: $z = \frac{x_i - \bar{x}}{s}$

Interquartile Range: IQR $= Q_3 - Q_1$.

Sample Covariance: $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Sample Correlation Coefficient: $r_{xy} = s_{xy}/(s_x s_y)$

Chapter 4:

The complement rule: $P(A) + P(A') = 1$

addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

conditional probability: $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

Bayes' Theorem

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

Combinations: ${}_nC_x = \frac{n!}{(n-x)!x!}$

Chapter 5:

Expected Value and mean of a Discrete Probability Distribution:

$$E(x) = \mu = \sum_{i=1}^{n} x_i P(x_i)$$

Variance of a Discrete Probability Distribution:

$$\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2 P(x_i)$$

Chapter 5 continued:

Binomial Probability Dist.: $P(x,n) = \frac{n!}{(n-x)!x!}p^x(q)^{(n-x)}$

Mean of a Binomial Distribution: $\mu = np$

Standard Dev. of a Binomial Distribution: $\sigma = \sqrt{npq}$

Poisson Probability Distribution: $P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

Chapter 6:

Normal PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(1/2)[(x-\mu)/\sigma]^2}$

the z-score: $z = \frac{x-\mu}{\sigma}$

Exponential PDF: $f(x) = \lambda e^{-\lambda x}$

Exponential CDF: $P(x \leq a) = 1 - e^{-a\lambda}$

Standard Dev. of Exponential Dist.: $\sigma = \mu = \frac{1}{\lambda}$

Continuous Uniform PDF

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Uniform CDF: $P(x_1 \leq x \leq x_2) = \frac{x_2 - x_1}{b-a}$

mean of the continuous uniform dist.: $\mu = \frac{a+b}{2}$

standard dev. of the continuous uniform dist.: $\sigma = \frac{b-a}{\sqrt{12}}$

Chapter 7:

standard error of the mean: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

z-score for the mean: $z_{\bar{x}} = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$

Chapter 7 continued:

sample proportion: $\bar{p} = \frac{x}{n}$

standard error of the proportion: $\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$

Chapter 8:

Confidence Interval for the mean (σ known):

$$\bar{x} \pm z_{\alpha/2}\sigma_{\bar{x}}$$

margin of error for a CI for the mean: $ME_{\bar{x}} = z_{\alpha/2}\sigma_{\bar{x}}$

approximate standard error of the mean: $\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}}$.

Confidence Interval for the mean (σ unknown):

$$\bar{x} \pm t_{\alpha/2}\hat{\sigma}_{\bar{x}}$$

Sample Size needed to Estimate a population mean

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{(ME_{\bar{x}})^2}$$

Sample Size needed to Estimate the population proportion

$$n = \frac{(z_{\alpha/2})^2\bar{p}(1-\bar{p})}{(ME_p)^2}$$

Chapter 9:

the z-test statistic for a hypothesis test for the population mean (when σ is known)

$$z_{\bar{x}} = \frac{\bar{x} - \mu_{H_0}}{\sigma/\sqrt{n}}$$

the t-test statistic for a hypothesis test for the population mean (when σ is unknown)

$$t_{\bar{x}} = \frac{\bar{x} - \mu_{H_0}}{s/\sqrt{n}}$$

Chapter 10:

the mean of the sampling distribution for the difference in means:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2}$$

the standard error of the difference between two means:

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

the z-test statistic for a hypothesis test for the difference between two means (σ_1 and σ_2 known)

$$z_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sigma_{\bar{x}_1-\bar{x}_2}}$$

the t-test statistic for a hypothesis test for the difference between two means (σ_1 and σ_2 unknown but equal)

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

pooled variance: $s_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)}$

the t-test statistic for a hypothesis test for the difference between two means (σ_1 and σ_2 unknown and unequal)

$$t_{\bar{x}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_{H_0}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Confidence Interval for the difference between the means of two independent populations (σ_1 and σ_2 unknown but equal)

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

the matched-pair difference: $d = x_1 - x_2$

the mean of matched-pair difference: $\bar{d} = \dfrac{\sum_{i=1}^n d_i}{n}$

the standard deviation of the matched-pair differences

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$$

the t-Test Statistic for a Matched-Pair hypothesis test for the mean

$$t_{\bar{x}} = \frac{\bar{d} - (\mu_d)_{H_0}}{s_d/\sqrt{n}}$$

Chapter 11:

the total sum of squares (SST): $SST = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{\bar{x}})^2$

the mean square total (MST): $MST = \dfrac{SST}{n_T - 1}$

the partitioning of the Total Sum of Squares (SST) for a One-Way ANOVA: SST=SSB+SSW.

sum of squares between (SSB): $SSB = \sum_{j=1}^{k} n-j(\bar{x}_j - \bar{\bar{x}})^2$

the mean square between (MSB): $MSB = \dfrac{SSB}{k-1}$

sum of squares within (SSW): $SSW = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2$

the mean square within (MSW): $MSW = \dfrac{SSW}{n_T - k}$

the F-test statistic for One-Way ANOVA: $F_{\bar{x}} = \dfrac{MSB}{MSW}$

Tukey-Kramer critical range:

$$CR_{ij} = Q_\alpha\sqrt{\frac{MSW}{2}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Chapter 14: simple linear regression model for a population $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$$\hat{y} = b_0 + b_1 x \qquad e_i = y_i - \hat{y}_i$$

sum of squares error: $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

total sum of squares (SST): $SST = \sum(y - \bar{y})^2$

$$SST = SSR + SSE$$

sum of squares regression (SSR): $SSR = \sum(\hat{y} - \bar{y})^2$

Chapter 14 continued:

$$R^2 = \frac{SSR}{SST}$$

F-statistic for the coef. of determination: $F = \dfrac{SSR}{SSE/(n-2)}$

Standard Error of the Estimate, $s_e = \sqrt{SSE/(n-2)}$.
Confidence Interval (CI) for an average value of Y:

$$CI = \hat{y}^* \pm t_{\alpha/2}s_e\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x^2 - ((\sum x)^2/n)}}$$

Prediction Interval (PI) fpr a specific value of y:

$$PI = \hat{y}^* \pm t_{\alpha/2}s_e\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{\sum x^2 - ((\sum x)^2/n)}}$$

t-test statistic for the regression slope: $t = \dfrac{b_1 - \beta_1}{s_b}$

the standard error of a slope: $s_b = \dfrac{s_e}{\sqrt{\sum x^2 - n(\bar{x})^2}}$

confid. interval for the pop. slope: $CI = b_1 \pm t_{\alpha/2}s_b$

Chapter 15:

mean square regression (MSR): $MSR = SSR/k$

mean square error (MSE): $MSE = SSE/(n-k-1)$

F-test stat. for the overall regression model: $F = \dfrac{MSR}{MSE}$

adjusted multiple coef. of det.: $R_A^2 = 1 - (1-R^2)\dfrac{n-1}{n-k-1}$

variance inflation factor: $VIF_j = \dfrac{1}{1-R_j^2}$

Other Math Rule Reminders:

$$e^x = \exp(x) \quad \text{and} \quad \ln 1 = 0 \quad \text{and} \quad \ln e = 1$$

$$x! = (x)(x-1)(x-2)\cdots(2)(1) \quad \text{and} \quad 0! = 1 \quad \text{and} \quad x^0 = 1$$

Use these tables to answer the relevant questions.

Table 1

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1.086667	2	0.543333	0.091659	0.913257	4.256495
Within Groups	53.35	9	5.927778			
Total	54.43667	11				

Table 2

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Treatments	112.0467	2	56.02333		0.006684	4.256495
Error						
Total	166.8867	11				

Table 3

Regression Statistics	
Multiple R	0.982539264
R Square	0.965383404
Adj R Square	0.958460085
Standard Error	1.207604307
Observations	7

ANOVA

	df	SS	MS	F	Significance F
Regression	1	203.345602	203.3456	139.4394	7.66542E-05
Residual	5	7.29154081	1.458308		
Total	6	210.637143			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.479831327	1.14324078	0.419712	0.69212	-2.45896266	3.4186253
advertising	1.987183044	0.16828489	11.80845	7.67E-05	1.554592952	2.4197731

Chapter 1

1. Consider the data set below:

customer information			
customer	$ spent	days since billing	tenure
1	74	3	25
2	989	25	79
⋮	⋮	⋮	⋮
76	4553	15	74

   Which variables are quantitative?

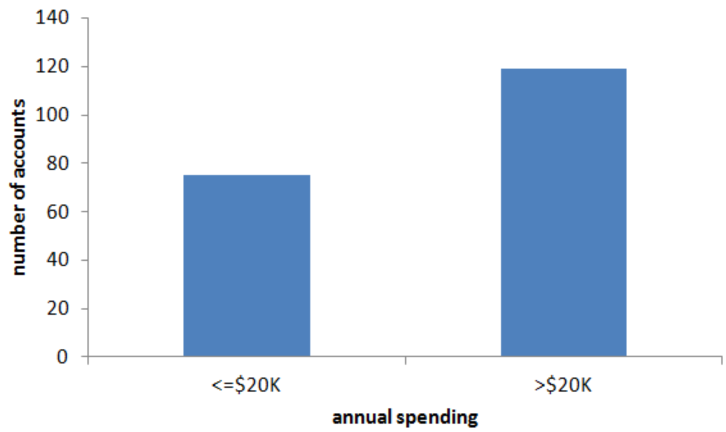
   (a) $ spent, days since billing, and tenure.
   (b) customer and $ spent.
   (c) customer, $ spent and tenure.
   (d) they are all quantitative.

2. Statistical inference

   (a) is the same as Data and Statistics.
   (b) refers to the process of drawing inferences about the sample based on the characteristics of the population.
   (c) is the same as a census.
   (d) is the process of drawing inferences about the population based on the information taken from the sample.

Chapter 2

3. The display below has



(a) too many classes.

(b) the right number of classes.

(c) too few classes.

(d) a class range that is too narrow.

4. The *cumulative* relative frequency of cards with a balance under $20,000 is:

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	
10,000-19,999	321	
20,000-29,999	397	
30,000+	422	

(a) .761.

(b) .421.

(c) .339.

(d) .941.

Chapter 3

5. A set of credit card accounts have the following estimated probabilities of default: .067, .743, .042, .022, .031, .194, .001. Find the 30th percentile.

(a) 2.1

(b) .036

(c) .031

(d) .022

6. The average of 22, 25, 26, 29 is

   (a) 25
   (b) 26
   (c) 26.5
   (d) 25.5

7. **{CT}** Suppose you observe the following values, 1, 2, 3, 4, 5, 6, 7, 8, 9, 20.

   (a) 20 is not an outlier.
   (b) 20 is an outlier based on both the interquartile range method and the z-score..
   (c) 20 is an outlier based on the interquartile range method and even more of an outlier based on the z-score.
   (d) 20 is not an outlier based on the interquartile range method but is based on the z-score.
   (e) 20 is an outlier based on the interquartile range method but not based on the z-score.

Chapter 4

8. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student did not complete their degree and is currently delinquent on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

   (a) .26.
   (b) .74.
   (c) .50.
   (d) .34.

9. A study by the Institute for Higher Education Policy found the values in the joint probability table below. The underlying data are for former college students that had taken out student loans. The table shows whether the student received a college degree versus whether they are successfully making their student loan payments. What is the probability that a former student did not complete their degree given that they are currently delinquent on their loan payments?

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

   (a) .34.
   (b) .68.
   (c) .50.
   (d) .59.

10. **{CT}** Suppose that P(econ)=.33, P(finance)=.4, and P(marketing)=.27 are the probabilities that a student has the respective major. Suppose further that some student work outside of school and others don't and that the conditional probabilities of working given the majors are, P(works|econ)=.4, P(works|finance)=.2, P(works|marketing)=.1. Find P(econ|works).

   (a) .671.
   (b) .552.
   (c) .433.
   (d) .132.
   (e) .388

Chapter 5

11. A random variable which can take on only two possible values has which distribution?

   (a) Poisson.
   (b) Binomial.
   (c) Uniform.
   (d) Discrete.
   (e) Taylor Series.

12. Suppose at ASU that 60% of students take out some form of loan to pay for their schooling. In a class of 220 students what is the expected number of students with a loan?

    (a) 156.
    (b) 110.
    (c) 132.
    (d) .6.
    (e) 88.

Chapter 6

13. Suppose you have a random variable that is uniformly distributed between 50 and 175. What is the expected value for this random variable?

    (a) 112.5.
    (b) 50.
    (c) 175.
    (d) 225.
    (e) 125.

14. Suppose you have data for several years of inflation adjusted bids for major road resurfacing projects in the greater Phoenix MSA. The bids are given in terms of per mile cost. You observe that bids are fairly evenly distributed on the interval \$110,000 to \$126,000. Given this information what it the probability of observing a bid >\$120,000?

    (a) .752.
    (b) .425.
    (c) .625.
    (d) .50.
    (e) .375.

Chapter 7

15. For a sample of 23 cars you find the standard deviation of the price is 5750 (measured in dollars). What is the standard error?

(a) 5750.
(b) 1437500.
(c) 15.8.
(d) 1199.
(e) 250.

16. The average wage for hourly wage workers at a RTE cereal company is \$17.65 with a known standard deviation of \$4.1. If the we take a survey of 40 employees what is the probability of getting an average wage of less than \$15? Note in this problem the standard deviation is **known** which means that you will find a variable that follows the z distribution.

    (a) .023.
    (b) .259.
    (c) .741.
    (d) 0.
    (e) 1.

Chapter 8

17. A sample of 15 recent business graduate had an average starting salary of \$51,317 with a standard deviation of \$7,127. What is the 99% confidence interval for the expected starting salary?

    (a) $[44266, 57918]$.
    (b) $[45839, 56795]$.
    (c) $[48076, 54558]$.
    (d) $[46577, 56057]$.
    (e) $[47710, 54924]$.

18. Suppose that you have a sample of 22 and you need to find the t value for sample with a mean of 56 and a standard deviation of 3.7. How many degrees of freedom are there in this case?

    (a) Degrees of freedom are not used in this type of problem.
    (b) 22.
    (c) 21.

(d) 23.

(e) 34.

19. Construct a 90% confidence interval for the mean when $\bar{x} = 50$ and there are 31 observations. Suppose $\sigma = 7$. What is the confidence interval? Don't round until the final step. Note that in this example σ is known so you will work with a z value in constructing the interval.

    (a) $[48.24, 51.76]$.

    (b) $[47.93, 52.07]$.

    (c) $[47.54, 52.46]$.

    (d) $[47.87, 52.13]$.

    (e) $[46.54, 53.45]$.

20. **{CT}** You are given a 95% confidence interval for the mean balance on a new loan product by the statisticians at your firm. The interval is $[5009, 5791]$. You know from working with the statisticians that they were going to use a sample of 40 to provide this interval estimate. Your boss says she doesn't want the interval, she wants the point estimate and the standard deviation and she wants them yesterday. Can you provide them while the statisticians are out to lunch?

    (a) 5400 and 1111.

    (b) 5200 and 1435.

    (c) 5400 and 1223

    (d) 5500 and 1317

    (e) 5300 and 1009

Chapter 9

21. Set the maximum acceptable probability of a Type I error at .05. Test the hypothesis that the average exam scores for students that previously took a statistics class is equal to the average exam score for student that did not previously take a statistics class. For those that took a class before the mean score was 16.65 with a standard deviation of 3.44 and for those that did not take a class before the mean was 17.28 with a standard deviation of 3.45. There were 171 students that took a statistics class before and 293 that did not. What is the value of the test statistic?

    (a) 0.88

(b) 3.47
(c) -1.90
(d) -1.28
(e) cannot be computed without the degrees of freedom

Chapter 10

22. Set the maximum acceptable probability of a Type I error at .05. Test the hypothesis that the average salary for men is at least \$1,000 more than the average salary for women; $H_0 : \mu_{\mathtt{M}} \geq \mu_{\mathtt{W}} + 1,000$. You have a sample of 40 men and 40 women and the average salaries are \$56,500 and \$53,800. The estimated standard deviations are \$1,566 and \$2,022. What is the value of the test statistic and do you reject or fail to reject H_0? Note that there are 73 degrees of freedom.

(a) -1.88, fail to reject
(b) .47, fail to reject
(c) 2.03, reject
(d) 4.2, fail to reject
(e) 6.6, reject

23. You have a paired sample with 9 employees that have a pre training program efficiency rating and a post training program rating. The pre training average rating is 8.93 and the post training rating is 9.28. The standard deviation of the differences is .464. Test the hypothesis that the pre and post training population efficiency ratings are the same at the .01 level of significance. What is the test statistic and conclusion?

(a) -1.57, fail to reject
(b) -2.26, fail to reject
(c) -2.26, reject
(d) 1.57, reject
(e) 3.76, reject

Chapter 11

24. See the excel output in ANOVA Table 1. What is the test statistic and conclusion?

(a) 4.256, fail to reject

(b) 4.256, reject
(c) .091, fail to reject
(d) .091, reject
(e) .913, reject

25. See the excel output in ANOVA Table 2. What is MSE?
    (a) 6.09
    (b) 56.02
    (c) 112
    (d) 9.19
    (e) not enough information

26. For an alpha value of .05, 10 numerator df and 13 denominator df the critical F value is?
    (a) 3.35
    (b) 1.81
    (c) 1.77
    (d) 2.95
    (e) 2.67

27. **{CT}** Consider a situation in which you are conducting ANOVA. You have a set number of treatments defined by your research problem. As you increase the number of observations suppose that the increases in SSE and SSTR are proportional as in if n increases by 10% then so do SSE and SSTR. What happens to the F statistic as n increases in this situation?
    (a) nothing, the increases are proportional so F stays the same.
    (b) it depends on the original value of F
    (c) F gets larger
    (d) F gets smaller
    (e) it depends on the number of treatments

Chapter 14

28. Using the least squares method for simple regression the sum of the residuals will be

(a) between -1 and 1
(b) 0
(c) between 0 and 1
(d) minimized
(e) depends on the values of x and y

29. Table 3 shows the regression results for a regression of sales on advertising expenditures. What is the lowest value of α for which you would reject the null hypothesis that $\beta_0 = 0$?

(a) .001
(b) .01
(c) .05
(d) .70
(e) 1.14

Key

1. a

2. d

3. c

4. a The value is found from noting that there are a total of 422 observations and the values are found in the table:

credit card balance		
card	cum. freq.	cum. relative freq.
<= 9,999	143	.339
10,000-19,999	321	.761
20,000-29,999	397	.941
30,000+	422	1

5. c. $n = 7$ and $p = 30$ so the index point is $i = (7)(30)/100 = 2.1$ which we round up to 3. The third value is .031.

6. d

7. e The first quartile is 3 and the third is 8 so the interquartile range is 5. 20 is (20-8)/5=2.4 interquartile ranges above the Q_3 and so it is considered an outlier.

   Note that $\bar{x} = 6.5$ and $s = 5.4$. The z-score is:

$$z = \frac{20 - 6.5}{5.4} = 2.5 < 3.$$

   So based on the z-score 20 is not an outlier.

8. d. Take the value from the table.

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

9. b. Take the values from the table to make the calculations.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.34}{.5} = .68.$$

	holds a college degree		
loan status	yes	no	total
satisfactory	.26	.24	.50
delinquent	.16	.34	.50
total	.42	.58	1

10. b b. You can use Bayes' Rule to find this,

$$P(\text{econ}|\text{works}) = \frac{(.33)(.4)}{(.33)(.4) + (.4)(.2) + (.27)(.1)} = .552.$$

11. b. Binomial.

12. c. (.6)(220)=132.

13. a.
$$\frac{50 + 175}{2} = 112.5.$$

14. e.
$$1 - \frac{120 - 110}{126 - 110} = .375.$$

15. d.
$$\frac{5750}{\sqrt{23}} = 1198.96$$
so the closest answer is 1199.

16. d. Find the z value
$$z = \frac{15 - 17.65}{4.1/\sqrt{40}} = \frac{-2.65}{.6482} = -4.08.$$
Find the probability using the z table. In this case it is 0.

17. b. You should have found a t value of 2.977 and a standard error of 1840 which would give you an interval of $[45839, 56795]$.

18. c.

19. b. Find the margin of error
$$z\frac{\sigma}{\sqrt{n}} = 1.645\frac{7}{\sqrt{31}} = 2.068.$$
Then 50 ± 2.068 gives approximately $[47.93, 52.07]$.

20. c. You know that the interval is symmetric around $\bar{x}$ so $\bar{x} = (5009 + 5791)/2 = 5400$. You know that with df=39 and $\alpha = .05$ that your t value was about 2.022. Your margin of error is $5791 - 5400 = 391$. Then you can solve for the standard deviation:

$$s = \frac{(391)\sqrt{40}}{2.022} = 1223.$$

21. c. The point estimate for the difference is 16.65-17.28=-.63. The standard error is

$$\sqrt{\frac{11.834}{171} + \frac{11.9}{293}} = .331.$$

Then $-.63/.331 = -1.90$.

22. d. The standard error is

$$\sqrt{\frac{(1566)^2}{40} + \frac{(2022)^2}{40}} = 404.377$$

so that the test statistic is

$$\frac{56500 - 53800 - 1000}{404.377} = 4.20.$$

The critical value is about -1.665 and $4.2 > -1.665$ so we fail to reject the null hypothesis.

23. b. The point estimate of the difference is -.35 and the test statistic is

$$t = \frac{-.35}{.464/\sqrt{9}} = \frac{-.35}{.1547} = -2.26.$$

which is not more extreme than the critical values of -3.355 and 3.355 so we fail to reject.

24. c.

25. a. First find SSE as 166.88-112.04=54.84 and then divide by 9 degrees of freedom to get 6.09.

26. e.

27. F gets larger. Notice that the df for the SSE increases and so does the denominator for MSE which means that MSE gets relatively smaller and F gets larger.

28. b.

29. d.