VERSION A

Choose the best answer. Do not write letters in the margin or communicate with other students in any way; if you do you will receive a 0. 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 Dr. Cox will answer statistics questions during the exam.

Please check each question and possible answers thoroughly as questions at the bottom of a page sometimes run onto the next page. Please verify that your test version and scantron version are the same.

This exam has 25 questions. Cost for turning in exam late: 1st minute is 10 points; 2nd minute is 20 additional points (30 total); 3rd minute is 30 points (60 total) ; 4th minute is 40 points (100 total); 5th minute is 50 points (150 total); 6th minute is 80 points (230 total) and no exams are graded past that point.

As noted in the syllabus entering your student ID incorrectly on your scantron results in a 10 point deduction.

1. You are making a presentation to a group of mid level managers and want to show the differences in sales by geographic region. An appropriate graph to show this would be

   (a) a histogram
   (b) a scatter plot
   (c) a bar chart
   (d) any of the above
   (e) none of the above

2. You create a boxplot to show the distances of retailers from your dristibution centers. From the boxplot your audience can quickly tell or get an idea about

   (a) the prescence of outliers
   (b) the median distance
   (c) the interquartile range
   (d) a and b
   (e) all of the above

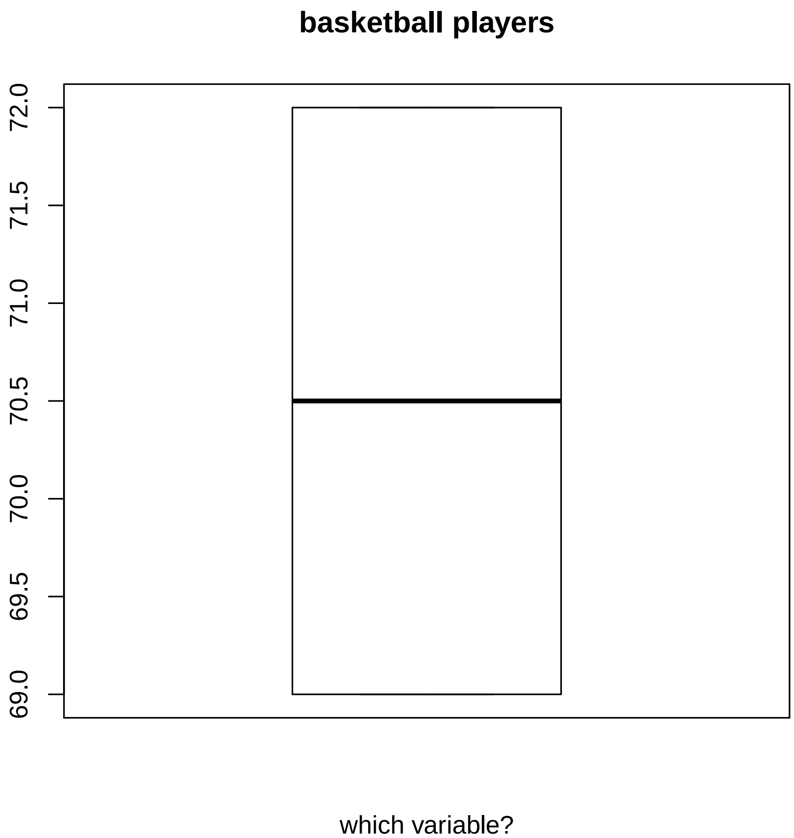
3. You will use the data below to answer multiple questions. The data are for a few basketball players in the NBA during 2016-2017. What is the average number of points scored for these players?

2016-2017 basketball players

name	height in inches	games played	shots blocked	points scored
Kay Felder	69	42	7	166
Isaiah Thomas	69	76	13	2199
Tyler Ulis	70	61	5	444
Ty Lawson	71	69	6	681
DJ Augustin	72	78	1	616
JJ Barea	72	35	1	381

   (a) 60.1666666666667
   (b) 650.615
   (c) 785.225
   (d) 747.833333333333
   (e) 751.833333333333

4. The graph here (below) is for which variable in the basketball players data set?



(a) height
(b) games
(c) blocks
(d) points
(e) the plot is not consistent with any of the above variables

5. Using the 2016-2017 basketball players table, what was the standard deviation for the players' height?

(a) 1.67840487520902
(b) 0.918404875209022
(c) 1.37840487520902

(d) 1.50840487520902

(e) 1.56840487520902

6. Using the 2016-2017 basketball players table, what was the range for shots blocked?

   (a) 7
   (b) 12
   (c) 15
   (d) 11
   (e) 8

7. Using the 2016-2017 basketball players table, what was the interquartile range for games played?

   (a) 36
   (b) 34
   (c) 37.5
   (d) 43
   (e) 39

8. Calculate the coefficient of variation for points scored.

   (a) 104.517612382521%.
   (b) 101.896701974232%.
   (c) 9.81386031760759%.
   (d) 98.1386031760759%.
   (e) 28.6388505641329%.

9. Using the basketball players table above, what would be the appropriate number of classes to use when creating a histogram for points scored?

   (a) 5
   (b) 8
   (c) 14
   (d) 20
   (e) a histogram would not be appropriate in this example

10. Using the basketball players data set, what is the minimum for shots blocked?

    (a) 13
    (b) 5.5
    (c) 3
    (d) 5.5
    (e) 1

11. In the basketball players data which variables use the ratio measurement scale?

    (a) games, blocks, points
    (b) blocks, points
    (c) games
    (d) height, games, blocks, points
    (e) none

12. Calculate the z-value for points scored for Isaiah Thomas. It is:

    (a) 1.97730016512065
    (b) -0.792782229980038
    (c) 1.68125939873333
    (d) 2.25412218823754
    (e) 2.35298719649357

13. Calculate the z value for an observation that has a value of 10 when the sample mean is 13 and the sample standard deviation is 2.22.

    (a) -1.39189189189189
    (b) -1.41891891891892
    (c) -1.35135135135135
    (d) -1.44594594594595
    (e) -1.30405405405405

14. The sample space is all the different sets of 4 people. There are 12 people from which to choose. How big is the sample space?

    (a) 220
    (b) 210

(c) 45
(d) 495
(e) none of the above

15. There are 20 major film actors and 5 will be nominated for an award as best actor. How many different possible combinations of nominees are possible?

(a) 4845
(b) 1140
(c) 125970
(d) 3003
(e) none of the above

16. You want to study how tuition at public universities has changed since 1997. You look at 42 public universities in assembling your data set. How would we best describe your tuition data set?

(a) the data are a times series and cross sectional
(b) the data are cross sectional
(c) the data are categorical
(d) the data are a times series
(e) the data are a categorical time series

17. The sample variance is never greater than the population variance?

(a) True
(b) False
(c) True if the variance is greater than the mean but false otherwise.

18. Suppose you have a frequency table which is made for the number of sales by the number of sales representatives as shown below. Choose the correct answer.

sales	frequency
10	1
14	2
18	3
19	5
20	8
22	3
25	1

(a) from this you could construct a relative frequency table

(b) from this you could construct a relative cumulative frequency table

(c) from this you could construct a stem and leaf display

(d) a and b are true

(e) a, b, and c are true

19. Consult the table in the previous question concerning sales. What is the empirical probability of observing a sales representative that makes more than 22 sales?

    (a) 0.173913043478261

    (b) 0.25

    (c) 0.0434782608695652

    (d) none of the above

    (e) cannot be calculated from the information in the table

20. The table here is from the classes' responses to the survey for homework 1. How many students got the question correct about the categorical variable?

	answer about categorical variable		total
took a stats class before	correct	wrong	total
NO	197	17	214
YES	72	5	77
Total	269	22	291

    (a) 269

    (b) 197

    (c) 72

    (d) 291

    (e) 0.676975945017182

21. The table here is from the classes' responses to the survey for homework 1. What is the probability that a student randomly drawn from the class has never taken a statistics class before?

	answer about categorical variable		
took a stats class before	correct	wrong	Total
NO	197	17	214
YES	72	5	77
Total	269	22	291

(a) 0.247422680412371
(b) 0.735395189003436
(c) 0.924398625429553
(d) 0.264604810996564
(e) 0.676975945017182

22. The table here is from the classes' responses to the survey for homework 1. What is the probability that a student randomly drawn from the class took a statistics class before and got the question about categorical variables wrong?

	answer about categorical variable		
took a stats class before	correct	wrong	Total
NO	197	17	214
YES	72	5	77
Total	269	22	291

(a) 0.0756013745704467
(b) 0.0584192439862543
(c) 0.0171821305841924
(d) 0.264604810996564
(e) 0.0649350649350649

23. The table here is from the classes' responses to the survey for homework 1. What is the probability that a student got the question about categorical variables wrong given that the student took a statistics class before?

	answer about categorical variable		
took a stats class before	correct	wrong	Total
NO	197	17	214
YES	72	5	77
Total	269	22	291

(a) 0.0756013745704467
(b) 0.220779220779221
(c) 0.0171821305841924
(d) 0.0649350649350649
(e) 0.0694444444444444

24. Apply your knowledge of probabilites rules to find the correct answer. Suppose you know that $P(A \cap B) = 0$.

(a) $P(A|B) = 0$.
(b) $P(A \cup B) = 0$.
(c) $P(A)P(B) = 0$.
(d) all of the above
(e) none of the above

25. Choose the correct statement.

(a) The z-score method is the best way to identify outliers.
(b) Outliers based on the IQR method always have |z scores| > 3.
(c) Outliers which make the mean increas by more than 150% or decrease by more than 50% should be excluded from the analysis.
(d) a and b are correct.
(e) none of the above.

Key

1. c
2. e
3. d
4. a
5. c
6. b
7. b
8. d
9. e
10. e
11. d
12. a
13. c
14. d
15. e
16. a
17. b
18. e
19. c
20. a
21. b
22. c
23. d
24. a
25. e

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: $\text{IQR} = Q_3 - Q_1$.

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

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

Continuous Uniform PDF

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

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

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:

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

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