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

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:

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

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

VERSION B

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 Dr. Cox will answer 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.

1. I have checked that my ID is bubbled in correctly. If it is bubbled in incorrectly I will get this question wrong. I also understand that questions and their possible answers may run onto the next page and so I should always check the top of the next page for possible answers. I understand that if I have a question I should simply make a note on my exam and ask Dr. Cox afterwards. I should always choose the best answer.

   (a) False.
   (b) I didn't read the directions.
   (c) True.

2. A lottery allows you to choose 3 numbers from a total of 24. What are the total number of possible combinations?

   (a) 3031
   (b) 1540
   (c) 42504
   (d) 2024
   (e) 351

3. The table here is from the classes' responses to the survey for homework 1. How many female students are there that have taken a statistics class before?

	Gender		total
took a stats class before	female	male	total
NO	217	335	552
YES	78	101	179
Total	295	436	731

   (a) 217
   (b) 78
   (c) 295
   (d) 179
   (e) 0.7355932

4. The table here is from the classes' responses to the survey for homework 1. What is $P(\text{Female} \cap \text{NOT already taken stats class})$?

	Gender		total
took a stats class before	female	male	total
NO	217	335	552
YES	78	101	179
Total	295	436	731

(a) 0.0135593
(b) 0.4582763
(c) 0.4035568
(d) 0.3931159
(e) 0.2968536

5. The table here is from the classes' responses to the survey for homework 1. What is $P(\text{Male}|\text{already took stats class})$?

	Gender		total
took a stats class before	female	male	total
NO	217	335	552
YES	78	101	179
Total	295	436	731

(a) 0.5642458
(b) 0.4357542
(c) 0.1080711
(d) 0.3242754
(e) 0.8709677

6. The table here is from the classes' responses to the survey for homework 1. What is the probability of randomly drawing a female student from the class?

	Gender		total
took a stats class before	female	male	total
NO	217	335	552
YES	78	101	179
Total	295	436	731

(a) 0.4035568
(b) 0.6766055
(c) 0.6477612
(d) 0.3931159
(e) 0.4357542

7. The table here is from the classes' responses to the survey for homework 1. What is $P(\text{did not take stats class before} \cup \text{Male})$?

	Gender		total
took a stats class before	female	male	total
NO	217	335	552
YES	78	101	179
Total	295	436	731

(a) 0.8932969
(b) 0.4582763
(c) 1.3515732
(d) 0.75513
(e) 0.5964432

8. You will use the data below to answer multiple questions. The data are for the 2015-2016 PHX SUNS that played at least 60 games. Games is the total number of games played while the other variables are per game averages. How many total free.throws per game did these players have? That is, if they were all playing together in a game what would be the total number of free.throws we might expect?

2015-2016 PHX SUNS results

name	games	free.throws	rebounds	minutes
P.J. Tucker	82	1.7	6.2	31
Mirza Teletovic	79	2	3.8	21.3
Alex Len	78	3.1	7.6	23.3
Devin Booker	76	3.4	2.5	27.7
Jon Leuer	67	1.5	5.6	18.7
Tyson Chandler	66	2.6	8.7	24.5
Ronnie Price	62	0.7	1.6	19.5

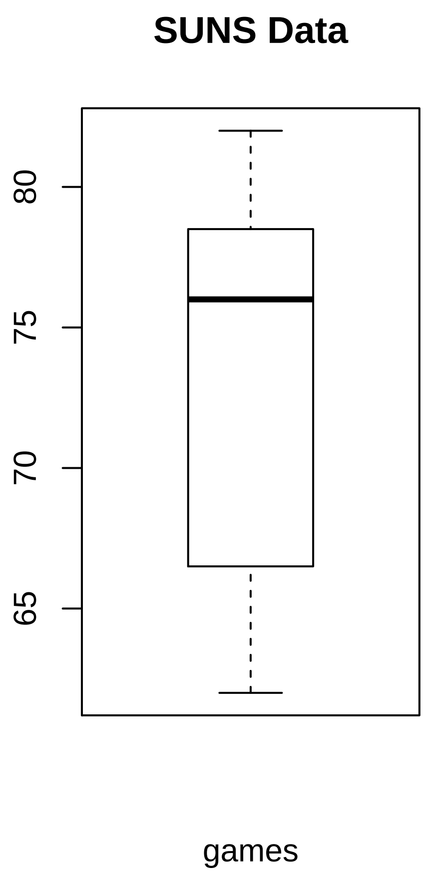
(a) 13.5
(b) 10
(c) 36
(d) 15
(e) 17

9. Using the 2015-2016 PHX SUNS table above, what was the average number of rebounds for the players?

(a) 4.6285714
(b) 74.8571429

(c) 5.1428571
(d) 39
(e) 6.1714286

10. The graph here is a



(a) histogram
(b) scatter plot
(c) stem and leaf display
(d) ogive
(e) box and whiskers plot

11. Using the 2015-2016 PHX SUNS table above, what is the median for minutes?

(a) 23.3
(b) 23.8
(c) 24.05
(d) 24.3
(e) 25.8888889

12. For a variable with observations that can be negative it is possible to get a standard deviation that is negative.

    (a) True
    (b) False

13. Using the 2015-2016 PHX SUNS table above, what is the standard deviation for rebounds?

    (a) -1.3745749
    (b) 2.6254251
    (c) 13.6254251
    (d) 3.1505102
    (e) 7.7120808

14. Using the 2015-2016 PHX SUNS table above, what is the coefficient of variation for games ?

    (a) 11.6437299%
    (b) 10.585209%
    (c) 51.0499332%
    (d) 7.7120808%
    (e) 8.9974276%

15. Using the 2015-2016 PHX SUNS table above, what is the range for rebounds?

    (a) 7.1
    (b) 3.1
    (c) 18.1
    (d) 20
    (e) 8.52

16. Using the 2015-2016 PHX SUNS table above, what is the z value for rebounds for PJ Tucker ?

   (a) 0.4026559
   (b) −11.9441546
   (c) 0.4550011
   (d) −7.0675471
   (e) 0.3543372

17. A baseball team has 10 pitchers of which 3 will be used in one game. How many different possible combinations are possible for 3 pitchers in a game?

   (a) 120
   (b) 240
   (c) 210
   (d) 45
   (e) 165

18. We can use z-values to help identify outliers. The empirical rule also can give us some hints about the potential distribution of z-values. True or false: the maximum number of observations with a z-value greater than 4 in a data set is 5.

   (a) True
   (b) False
   (c) depends on whether or not the values can be negative

19. I asked each of you how many hours you worked the previous week. The average for the class was 13.89 as mentioned in one lecture. The z-value for a student working 0 hours a week was −0.7648678. There were 696 students that took the survey (different from the other table in this test for a couple of reasons). Suppose (contrary to fact) that the distribution of the data is roughly bell shaped. Based on the empirical rule we will see *approximately* how many students working 30 or more hours a week?

   (a) 139
   (b) 223
   (c) 14
   (d) 104
   (e) We cannot tell without the standard deviation

20. Bar charts are useful for?

    (a) observing the relationship between two variables.
    (b) providing a different picture from what we would see in a frequency table.
    (c) organizing interval data.
    (d) showing an empirical distribution.
    (e) translating pivot tables into a picture.

21. The covariance of rebounds and games in the data set is 2.8571429. What is the correlation between rebounds and games? (0).

    (a) 0.265111
    (b) -0.108889
    (c) 0.2969803
    (d) 0.0199123
    (e) 0.141111.

22. Which graph is useful for finding empirical probabilities?

    (a) stem and leaf display
    (b) relative frequency table
    (c) histogram
    (d) scatter plot
    (e) bar chart

23. $P(A \cap B) \geq P(A \cup B)$.

    (a) True
    (b) False
    (c) It depends.

24. The median is always greater than the mean.

    (a) True.
    (b) False.
    (c) It dependes on whether or not the data can have negative values.

25. If the data are right skew this *suggests* that many values will be to the left of the mean while a few observations will be fairly large and far to the right of the mean.

   (a) True.
   (b) False.

Key

1. c
2. d
3. b
4. e
5. a
6. a
7. a
8. d
9. c
10. e
11. a
12. b
13. b
14. b
15. a
16. a
17. a, use the formula $\binom{N}{r}$
18. b,
19. a or d, you can find the standard deviation because you know $-13.89/s = -.7648678$ so that you find $s = 18.160001$ and then you can use the empirical rule in approximating this. Neither a nor d are exactly right but both are conceivable approximations.
20. b
21. e
22. b
23. b, false

24. b, false,

25. a, true,