Chapter 1: no key formulas.
Chapter 2: Relative Frequency=freq. of the class$/n$.
Approx. Class Width:
=(largest value-smallest value) /number of classes.
Chapter 3: sample and population means

$$\bar{x} = \sum x_i/n \text{ and } \mu = \sum x_i/N$$

Weighted mean and geometric mean

$$\bar{x} = \sum w_i x_i/w_i \text{ and } \bar{x}_g = [(x_1)(x_2)\ldots(x_n)]^{1/n}.$$

Interquartile Range: IQR $= Q_3 - Q_1$.
Population and sample variance

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{N} \text{ and } s^2 = \frac{\sum(x_i-\bar{x})^2}{n-1}.$$

Population and sample standard deviation

$$\sigma = \sqrt{\sigma^2} \text{ and } s = \sqrt{s^2}.$$

Coefficient of Variation

$$\left(\frac{\text{Standard deviation}}{\text{Mean}} \times 100\right)\%$$

z-Score: $z_i = \frac{x_i - \bar{x}}{s}$.
Population and Sample Covariance

$$\sigma_{xy} = \frac{\sum(x_i-\mu_x)(y_i-\mu_y)}{N} \text{ and } s_{xy} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n-1}$$

Population and Sample Pearson Correlation

$$\rho_{xy} = \sigma_{xy}/(\sigma_x\sigma_y) \text{ and } r_{xy} = s_{xy}/(s_x s_y).$$

Chapter 4: Counting Rule for Combinations

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

Counting Rule for Permutations

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!}.$$

Probability Rules: $P(A) = 1 - P(A^C)$

Chapter 4 continued:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

Multiplication Law for Independent Events

$$P(A \cap B) = P(B)P(A).$$

Bayes' Theorem

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

Chapter 5:
Discrete Uniform Probability Mass Function: $f(x) = 1/n$.
Expected Value of a Discrete R. V.: $E(x) = \mu = \sum x f(x)$.
Variance of a Discrete R. V.:

$$Var(x) = \sigma^2 = \sum(x-\mu)^2 f(x).$$

Number of Experimental Outcomes Providing Exactly x Successes in n Trials

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

Binomial Probability Mass Function

$$P(X=x) = f(x) = \binom{n}{x} p^x (1-p)^{(n-x)}.$$

Expected Value for Binomial Distribution: $E(x) = \mu = np$.
Variance for Binomial Distr.: $Var(x) = \sigma^2 = np(1-p)$.
Poisson Probability Mass Function:

$$P(X = x|\mu) = f(x) = \frac{\mu^x e^{-\mu}}{x!}.$$

Hypergeometric Probability Mass Function and Expected Value:

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \text{ and } E(x) = \mu = \frac{nr}{N}.$$

Chapter 5 continued: Variance for the Hypergeometric Distribution:

$$Var(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Chapter 6: Uniform PDF

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

Normal PDF The density function is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Converting to the Standard Normal rv:

$$z = \frac{x-\mu}{\sigma}.$$

Exponential PDF and CDF for $x \ge 0$

$$f(x) = \mu^{-1} e^{-x/\mu} \text{ and } P(x \le x_0) = 1 - e^{-x_0/\mu}.$$

Chapter 7: expected value of $\bar{x}$

$$E(\bar{x}) = \mu.$$

Standard Deviation of $\bar{x}$ (Standard Error)

$$\sigma_x = \frac{\sigma}{\sqrt{n}}.$$

Expected Value and Std Dev (Standard Error) of $\bar{p}$

$$E(\bar{p}) = p \text{ and } \sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Finite Pop. Correction Factor: $\sqrt{(N-n)/(N-1)}$.
Chapter 8: Interval Estimate of Population Mean, σ known and unknown

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ and } \bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

Necessary Sample Size for Interval Estimate of μ

$$n = \frac{(z_{\alpha/2})^2\sigma^2}{E^2}$$

Chapter 8 continued: Interval Estimate of p

$$\bar{p} \pm z_{\alpha/2} \frac{p(1-p)}{\sqrt{n}}$$

Necessary Sample Size for Interval Estimate of p

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

Chapter 9: Test Statistic for Hypothesis Tests About μ, σ known and unknown

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \text{ and } t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Test Stat for Hypothesis About p

$$z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Chapter 10: Point Estimate and Standard Error for Difference in Two Population Means

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

Interval Estimate and Test Statistic for Difference in Two Means with Known Variances

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \text{ and } z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Interval Estimate and Test Statistic for Difference in Two Means with Unknown Variances

$$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \text{ and } t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Degrees of Freedom for t, Two Independent Random Samples

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{s_2^2}{n_2}\right)^2}$$

Chapter 10 continued: Test Statistic (Matched Samples)

$$t = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}}$$

ANOVA Related:

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \quad s_j^2 = \frac{\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2}{n_j - 1} \quad \bar{\bar{x}} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n_T}$$

$$MSTR = \frac{SSTR}{k-1} \quad \text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 \quad MSE = \frac{SSE}{n_T - k}$$

$$SSE = \sum_{j=1}^{k}(n_j - 1)s_j^2 \quad F = MSTR/MSE$$

$$\text{SST} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{\bar{x}})^2 \quad \text{SST=SSTR+SSE}$$

Chapter 11: not covered in this course
Chapter 12: $y = \beta_0 + \beta_1 x + \epsilon$

$$E(y) = \beta_0 + \beta_1 x \quad \hat{y} = b_0 + b_1 x \quad b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad SSE = \sum(y_i - \hat{y}_i)^2$$

$$SST = \sum(y_i - \bar{y})^2 \quad SSR = \sum(\hat{y}_i - \bar{y})^2 \quad \text{SST=SSR+SSE}$$

$$r^2 = \frac{SSR}{SST} \quad r_{xy} = (\text{sign of } b_1)\sqrt{r^2} \quad s^2 = MSE = \frac{SSE}{n-2}$$

Standard Error of the Estimate, $s = \sqrt{MSE}$.

$$\sigma_{b_1} = \frac{\sigma}{\sqrt{\sum(x_i - \bar{x})^2}} \quad s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}} \quad t = \frac{b_1}{s_{b_1}}$$

For simple regression, $MSR = SSR$ because there is only one independent variable.

$$F = \frac{MSR}{MSE} \quad s_{\hat{y}*} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_1 - \bar{x})^2}}$$

Confidence Interval for $E(y^*)$: $\hat{y}^* \pm t_{\alpha/2}s_{\hat{y}*}$

$$s_{\text{pred}} = s\sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_1 - \bar{x})^2}}$$

Chapter 12 continued: Prediction Interval for y^*:

$$\hat{y}^* \pm t_{\alpha/2}s_{\text{pred}}$$

Residual for Observation i: $y_i - \hat{y}_i$
Chapter 13:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \epsilon$$

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p$$

$$SST = SSR + SSE \quad R^2 = \frac{SSR}{SST}$$

$$R_a^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

$$MSR = \frac{SSR}{p} \quad MSE = \frac{SSE}{n-p-1} \quad F = \frac{MSR}{MSE}$$

$$t = \frac{b_i}{s_{bi}}$$

Other Math Rule Reminders:

$$e^x = \exp(x)$$

$$\ln 1 = 0 \quad \ln e = 1$$

$$x! = (x)(x-1)(x-2)\cdots(2)(1)$$

$$0! = 1 \quad x^0 = 1$$

VERSION A

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.

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 20. What are the total number of possible combinations?
   (a) 2147
   (b) 816
   (c) 15504
   (d) 1140
   (e) 253

3. The table here is from the classes' responses to the survey for homework 1. How many male students are there that have not taken a statistics class before?

Gender	first stats class	took stats class before	total
Female	4	4	8
Male	27	5	32
Total	31	9	40

   (a) 27
   (b) 5
   (c) 4
   (d) 31
   (e) 0.775

4. 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	first stats class	took stats class before	total
Female	4	4	8
Male	27	5	32
Total	31	9	40

(a) 0.1
(b) 0.5
(c) 0.225
(d) 0.2580645
(e) 0.2

5. 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	first stats class	took stats class before	total
Female	4	4	8
Male	27	5	32
Total	31	9	40

(a) 0.1290323
(b) 0.2
(c) 0.775
(d) 0.1
(e) 0.5

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

Gender	first stats class	took stats class before	total
Female	4	4	8
Male	27	5	32
Total	31	9	40

(a) 0.8709677
(b) 0.1290323
(c) 0.2903226
(d) 0.225
(e) 0.8709677

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

Gender	first stats class	took stats class before	total
Female	4	4	8
Male	27	5	32
Total	31	9	40

(a) 0.9

(b) 0.775

(c) 1.575

(d) 0.7

(e) 0.2

8. You will use the data below to answer multiple questions. The data are for the top votes getters in the 2016 NBA MVP voting and are per game averages. How many total steals per game did these players have? That is, if they were all playing together in a game what would be the total number of steals we might expect?

2016 NBA MVP results

name	mvp rank	points	rebounds	assists	steals
Stephen Curry	1	30.1	5.4	6.7	2.1
Kawhi Leonard	2	21.2	6.8	2.6	1.8
LeBron James	3	25.3	7.4	6.8	1.4
Russell Westbrook	4	23.5	7.8	10.4	2
Kevin Durant	5	28.2	8.2	5	1
Chris Paul	6	19.5	4.2	10	2.1
Draymond Green	7	14	9.5	7.4	1.5

(a) 10.4

(b) 8

(c) 49.3

(d) 11.9

(e) 13.9

9. Using the 2016 NBA MVP table above, what was the average number of rebounds?

(a) 7.0428571

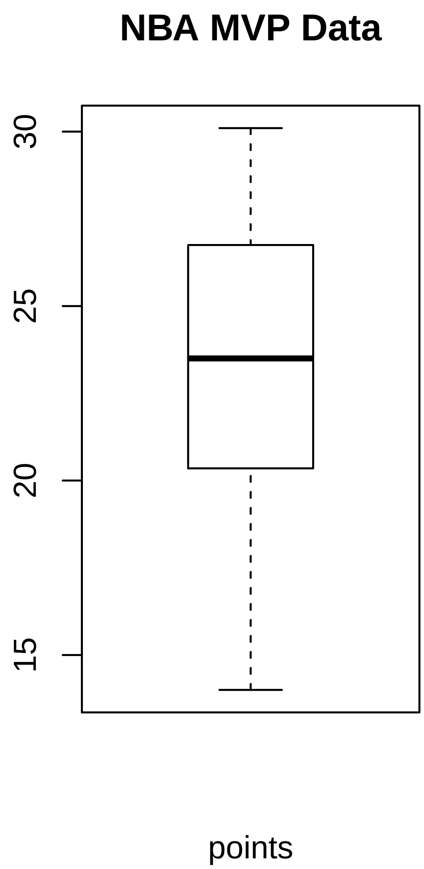
(b) 23.1142857

(c) 25.1142857

(d) 15.0785714

(e) 8.4514286

10. The graph here is a



(a) histogram

(b) box and whiskers plot

(c) scatter plot

(d) stem and leaf display

(e) ogive

11. Using the 2016 NBA MVP table above, what is the range for rebounds?

(a) 1.3
(b) 16.3
(c) 16.1
(d) 5.3
(e) 6.36

12. Using the 2016 NBA MVP table above, what is the median for assists?

    (a) 7.3
    (b) 7.55
    (c) 6.8
    (d) 7.8
    (e) 7.5555556

13. Qualitative data are always numeric and quantitative data are never numeric.

    (a) True
    (b) False

14. Using the 2016 NBA MVP table above, what is the standard deviation for rebounds?

    (a) -2.2243713
    (b) 12.7756287
    (c) 2.1307544
    (d) 5.4679151
    (e) 1.7756287

15. Using the 2016 NBA MVP table above, what is the coefficient of variation for points (per game)?

    (a) 26.0215983%
    (b) 23.6559985%
    (c) 25.211766%
    (d) 5.4679151%
    (e) 20.1075987%

16. Using the 2016 NBA MVP table above, what is the z value for points for Kawhi Leonard?

(a) -0.3500943
(b) -0.4551225
(c) -0.7158644
(d) -0.2693033
(e) -0.6500943

17. In any data set the observation with the minimum value for a variable will always have the smallest z-value.

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

18. A basketball team has 15 players of which 5 are starters. How many different possible combinations are possible for the starters?

    (a) 6006
    (b) 4368
    (c) 1287
    (d) 3003
    (e) 1365

19. I asked each of you how many hours you worked the previous week. The average for the class was 9.795. The z-value for a student working 0 hours a week was -.68. There were 40 students that took the survey. 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 40 or more hours a week?

    (a) 0
    (b) 1
    (c) 4
    (d) 6
    (e) We cannot tell without the standard deviation

20. What is the probability that two mutually exclusive events happen at the same time?

    (a) 1
    (b) .5

(c) It depends on the original probabilities.

(d) 0.

(e) $P(A) + P(B) - P(A \cap B)$.

21. The covariance of rebounds and points in the data set is -2.6957143. What is the correlation between rebounds and points? (0).

    (a) -0.1536515

    (b) -0.5276515

    (c) 0.4518247

    (d) 0.0770903

    (e) -0.2776515.

22. In which graph shows the relationship between two continuous variables?

    (a) stem and leaf display

    (b) scatter plot

    (c) box plot

    (d) histogram

    (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 first quartile.

    (a) True.

    (b) False.

25. The median must be positive.

    (a) True.

    (b) False.

Key

1. c
2. d
3. a
4. e
5. d
6. b
7. a
8. d
9. a
10. b
11. d
12. c
13. b, false, quantitative data can be numeric, e.g., zip code
14. e
15. b
16. a
17. a, true
18. d, use the formula $\binom{N}{r}$
19. b, you can find the standard deviation because you know $-9.795/s = -.68$ so that you find $s = 14.4044118$ and then you can use the empirical rule in approximating this.
20. d
21. e
22. b
23. b, false

24. b, false, consider the data set 0, 1, 1, 1, 1, 1, 1, 1, 1, 5, 5, 5 where Q1=1=Q2

25. b, false, consider the data set -10, -9, -8, -7.5, -7, -6.2, -3.