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

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

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

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

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:

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

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

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

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:

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

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

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

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

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.

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

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:

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

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

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

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

ECN221 Exam 1 A (NL) Spring 2016 (Chapters 1-4), ASU-COX

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.

You can do a great job so relax and show me what you know.

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 7 numbers from a total of 21. What are the total number of possible combinations?
   (a) 117287
   (b) 50388
   (c) 293930
   (d) 116280
   (e) 134596

3. A baseball team has 10 pitchers of which 5 are in the starting rotation. How many different possible combinations are possible for the starting rotation?
   (a) 252
   (b) 504
   (c) 462
   (d) 56
   (e) 210

4. You will use the data below to answer multiple questions. The data are for the top votes getters in the 2015 NL MVP voting. How many total HR (home runs) did these players have?

2015 NL MVP results

name	mvp rank	HR	hits	salary	position
Bryce Harper	1	42	172	$2,500,000	RF
Paul Goldschmidt	2	33	182	$3,100,000	1B
Joey Votto	3	29	171	$14,000,000	1B
Anthony Rizzo	4	31	163	$5,000,000	1B
Andrew McCutchen	5	23	165	$10,000,000	CF
Jake Arrieta	6	2	12	$3,630,000	P

(a) 865

(b) 160

(c) 180

(d) 145

(e) 112

5. Using the 2015 NL MVP table above, what was the average number of hits?

(a) 144.166666666667

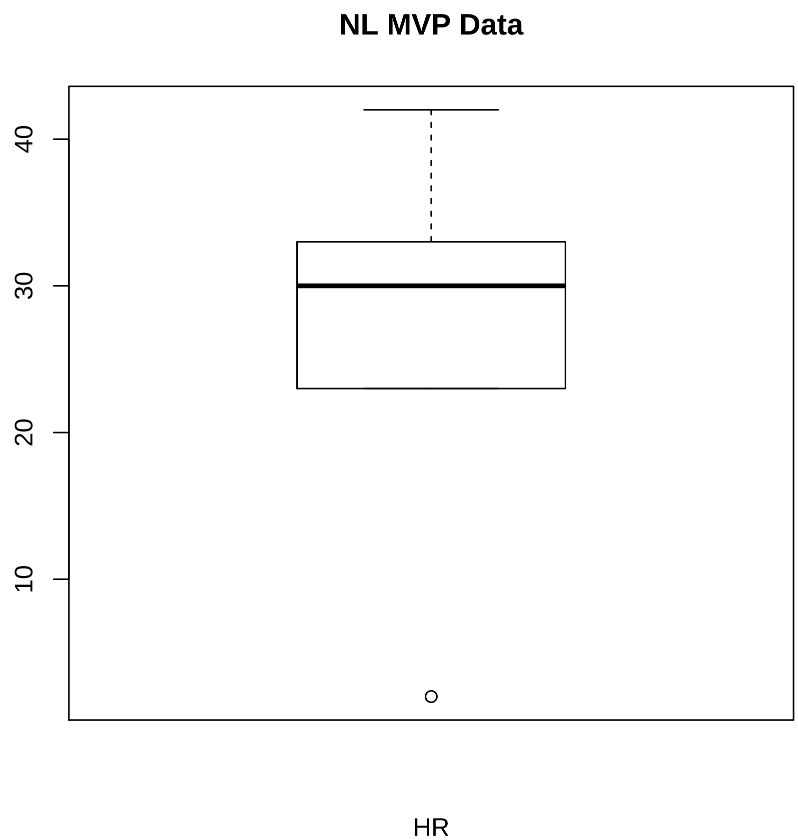
(b) 26.6666666666667

(c) 28.6666666666667

(d) 85.4166666666667

(e) 173

6. The graph here is a



(a) histogram

(b) box and whiskers plot

(c) scatter plot

(d) stem and leaf display

(e) ogive

7. Using the 2015 NL MVP table above, what is the range for hits?

(a) 166

(b) 181

(c) 204

(d) 40

(e) 170

8. Using the 2015 NL MVP table above, what is the median for salary?
   (a) $4365000
   (b) $6371666.66666667
   (c) $4315000
   (d) $4415000
   (e) $4794444.44444444
9. What type of variable is “position”?
   (a) quantitative
   (b) numeric
   (c) qualitative
   (d) time series
   (e) alphabetical
10. Using the 2015 NL MVP table above, what is the standard deviation for hits?
    (a) 61.0889135465224
    (b) 76.0889135465224
    (c) 78.1066962558269
    (d) 13.5744858711727
    (e) 65.0889135465224
11. Using the 2015 NL MVP table above, what is the coefficient of variation for HR (home runs)?
    (a) 55.9947542185873%
    (b) 50.9043220168975%
    (c) 45.1483793386283%
    (d) 13.5744858711727%
    (e) 43.2686737143629%
12. Using the 2015 NL MVP table above, what is the z value for HRs for Paul Goldschmidt?
    (a) 0.606530030785031

(b) 0.319226331992122
(c) 0.466561562142331
(d) 0.358893509340255
(e) 0.166561562142331

13. Which player has the largest z-value for HR in terms of absolute value; the z-value furthest away from 0?

    (a) Bryce Harper
    (b) Paul Goldschmidt
    (c) Joey Votto
    (d) Anthony Rizzo
    (e) Andrew McCutchen

14. I asked each of you how many cousins you have. I got back some answers like "a lot" or "I'm [nationality], so more than I can keep track of" and the such. Of the acutal numeric values I got back the average was 10.15, the standard deviation was 10.65 and the maximum was 86 and the sample size was $n = 627$. What is the z-value for the student with 86 cousins in this sample?

    (a) 4.30516431924883
    (b) 7.12206572769953
    (c) 8.76014084507042
    (d) 5.93268075117371
    (e) 7.42364532019704

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

16. The covariance of hits and HR in the data set is 798.866666666667 and you should have already calculated the other pieces you need to calculate the correlation. What is the correlation between hits and HR? (0).

(a) 1.02815708128649

(b) 0.654157081286488

(c) 0.939444678217485

(d) 0.8175000276405

(e) 0.904157081286488.

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

Already took Stats class	Male	Female	total
YES	103	55	158
NO	324	182	506
Total	427	237	664

(a) 55

(b) 158

(c) 237

(d) 664

(e) 0.0828313253012048

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

Already took Stats class	Male	Female	total
YES	103	55	158
NO	324	182	506
Total	427	237	664

(a) 0.237951807228916

(b) 0.643072289156627

(c) 0.241217798594848

(d) 0.758782201405152

(e) 0.0828313253012048

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

Already took Stats class	Male	Female	total
YES	103	55	158
NO	324	182	506
Total	427	237	664

(a) 0.881024096385542
(b) 0.643072289156627
(c) 0.370023419203747
(d) 0.725903614457831
(e) 0.155120481927711

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

Already took Stats class	Male	Female	total
YES	103	55	158
NO	324	182	506
Total	427	237	664

(a) 0.881024096385542
(b) 0.356927710843373
(c) 0.274096385542169
(d) 1.1288056206089
(e) 0.155120481927711

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

Already took Stats class	Male	Female	total
YES	103	55	158
NO	324	182	506
Total	427	237	664

(a) 0.232067510548523
(b) 0.359683794466403
(c) 0.356927710843373
(d) 1.1288056206089
(e) 0.640316205533597

22. In which graph is it easiest to identify outliers? If you think this question is a matter of opinion then the opinion that matters is the opinion of the person that wrote the exam which is me, Dr. Cox.

(a) stem and leaf display

(b) scatter plot
(c) box plot
(d) histogram
(e) bar chart

23. Conditional probabilities are always greater than unconditional probabilities.

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

24. Suppose that you have a histogram where the data are skewed left.

    (a) The mean will exceed the median.
    (b) The median will exceed the mean.
    (c) The mean will exceed the median as long as the first quartile is within 3 standard deviations of the median.
    (d) The median will exceed the mean as long as the first quartile is within 3 standard deviations of the median.
    (e) none of the above.

25. Consider the formulas for the variance and standard deviation. We know that they must both be positive.

    (a) The variance is always greater than the mean.
    (b) The variance is always greater than the standard deviation.
    (c) The variance is always greater than the standard deviation as long as all data values are positive.
    (d) The variance is always greater than the mean as long as all data values are positive.
    (e) none of the above.

Key

Most questions required your knowledge of definitions and vocabulary discussed in class and in the notes or they required you to be able to make calculations similar to those done in class and required in the homework. There were a few questions where some critical thinking as opposed to simple memorization or plug and chug techniques were useful. .

1. c
2. d, use the $\binom{N}{r}$ formula
3. a, use the $\binom{N}{r}$ formula
4. b
5. a
6. b
7. e
8. c
9. c
10. e
11. b, check the formula for cv=$s/\bar{x} \times 100\%$.
12. c
13. a
14. b, check the formula for z values, $z = \frac{x_i - \bar{x}}{s}$ which is on the formula sheet.
15. d
16. e
17. a
18. b, 427/664=0.643072289156627
19. d, (158+427-103)/664=0.725903614457831
20. c, 182/664=0.274096385542169
21. b, 182/506=0.359683794466403

22. c

23. b

24. e, The data set (1, 1, 2, 2, 2) violates c as the mean of 1.6 is less than the median of 2 and Q1 is within 3 standard deviations of the median and the data are left skewed. The data (0, 1, 2, 4, 4) violates d though it may be more difficult to see that it is left skewed. You could think of lots of other data sets that violate a-d.

25. e, if the values are 100, 101, and 101.5 then $\bar{x} = 100.83 > .76376 = s > .5833 = s^2$.