Midterm and Finals Preparation for Math 147B

General Remarks:

> **a.** To prepare for a midterm you may ignore the comments for chapters that are not in scope for that exam.
>
> **b.** Not all chapters are covered in the order of the book. For example, I will cover ch.9 (Multiple Regression) after ch.23 (Inferences for Regression).
>
> **c.** Except for the chapters that have been skipped (#24: Multiple Regression Wisdom, #26: Multifactor ANOVA, #27: Intro to Statistical Learning and Data Science:) you are advised to read their "What Have We Learned" summaries, in particular the "Review of Terms" subsection.
>
> **d.** Unless specified otherwise, skip the starred (optional) material. Note that I have tried to be explicit about all that you may skip.
>
> **e.** All problems you are given on an exam can be solved with a basic scientiufic calculator. For example, if you have to do a simple or multiple regression, then you will not be asked to compute the regression equation. Rather, the computer output will be provided and your task will be to interpret it correctly and answer questions such as "What is the interpretation of the regression equation?", "Is the xyz coefficient relevant?", and the like.

Chapter by Chapter Comments

Ch.01: Skip the following: ● nominal variable ● metadata
Skip for now (will be covered in detail in later chapters): ● case ● respondent ● subject ● participant ● experimental unit

Ch.02: Skip the following: ● stem-and-leaf display ● dot plot (will use in lecture) ● density plot (will use in lecture)

Ch.03: Skip the following: ● segmented bar charts ● mosaic plots

Ch.04: ● Boxplots: Understand the difference between fences and whiskers!
● Skip the optional Smoothing Timeplots chapter.

Ch.05: ● An alternate name for "normal probability plot" is "qq–plot" ● Skip the subchapter "How Does a Normal Probability Plot Work?" (will not cover in lecture).

Ch.06: ● Understand that independent variable = predictor variable = explanatory variable ● Understand that dependent variable = response variable
● Skip optional subchapters on Kendall's tau, Spearman's rho ● Skip ch.6.4: Straightening Scatterplots. ● Skip terminology: ● re-expression (of scatterplots) ● ladder of powers.

Ch.07: Skip nothing!

Ch.08: ● Skip optional ch.8.6 and 8.7 on re-expressing scatterplots

Ch.09: Skip this chapter for now: moved between ch.23 and ch.24!

Ch.10: Not quantitative, but lots of concepts to remember!

Ch.11: Not quantitative, but lots of concepts to remember!

Ch.12: • Review the link in HW1 about Venn diagrams. • Not in the book but taught in class: You may use "iid" as a short for "independent and identically distributed" (random variables) in your answers if it comes in handy. If you do not feel comfortable with that concept, you can use alternate terminology from the book.

Ch.13: In ch.13.5 you will not be asked to write down the Bayes rule formula

$$P(B|A) \;=\; \frac{P(A|B)P(B)}{P(A|B)P(B) \;+\; P(A|B^{\complement}P(B^{\complement})}$$

(approx. p.411), but knowing how to use it will certainly not hurt. You **must understand reverse conditioning** by use of a tree diagram!

Ch.14: • Skip the subchapter on correlation and covariance.

Ch.15: • **Study** ch.15.5 (Continuity Correction), even though it is optional!
• Skip Poisson and exponential probability (ch.15.6 and part of ch.15.7).

Ch.16: Don't confuse • sample distribution = distribution of the actual sample that was collected (the SPECIFIC list of data) as was done in the beginning chapters: It is represented by a relative frequency diagram. • sampling distribution = distribution of a random item which is a function of the sample picks $\vec{X}(\omega) \;=\; X_1(\omega), X_2(\omega), ; X_n(\omega)$. Example: sampling distribution of the mean = distribution of $\bar{Y}(\omega) \;=\; (X_1(\omega) + X_2(\omega) + X_n(\omega))/n$.
• Note that "sampling variability" is also known as "sampling error".
• STUDY ch.16.6: Choosing the Sample Size (marked as optional).
• STUDY the What Can Go Wrong section more carefully than usual.

Ch.17: • STUDY the optional ch.17.4 (Picking Our Interval up by Our Bootstraps). You need not know how to do bootstraps with the computer but you must understand the theory of creating bootstrap samples and how you use them to obtain confidence intervals, P-values (defined in ch.18), ...

Ch.18: • ch.18.1: Understand why we use $\mathrm{SD}(\hat{p}) = \sqrt{(p_0 q_0)/n}$ and not $\mathrm{SE}(\hat{p}) = \sqrt{(\hat{p}\hat{q})/n}$.
• STUDY the optional "Random Matters" section of ch.18.5.

Ch.19: • SKIP the "Random Matters" section of ch.19.4.
• We write α for Alpha and β for Beta, so "Alpha level" becomes "α–level". Understand how a given α–level is used a decision criterion for rejecting the Null.

Ch.20: • Pooling **for means**: You need not know about the formulas since there will not be a computational problem on the exam, but you must understand **WHY** doing a pooled t–test for two means is strongly discouraged.
• Ch.20.4: No need to know that formula about degrees of freedom for a confidence interval or a test concerning the difference of two means in the footnote on p.615. If you need it, we will give it to you.
• Ch.20.5: Skip Tukey's quick test and the Wilcoxon Rank–Sum test.
• Skip the optional ch.20.6.
• Ch.20.7 (Pooling): • STUDY the first part about pooled **proportions** • Skip the second part about pooled **means** (starts with "For means, the assumption that ...") • STUDY the subsection "Is the Pool All Wet?"
• STUDY the optional ch.20.8.

Ch.21: • Skip most of ch.23.1 (Blocking), EXCEPT: understand that the advantage of paired sampling is that it reduces variability. • In particular, skip the Random Matters section.

Ch.22: • Understand how to compute expected counts and the subtle difference between testing for proportions and testing for independence.

Ch.23: • Chapter 23.5 (Multiple Regression Inference): Notes are further down since you need to understand ch.09!
• Terminology: Skip • logistic regression Variance • Inflation factor (VIF)
• STUDY LATER (belongs to ch.23.5): • collinearity • adjusted R^2

Ch.09: Moved to this spot!!
• Skip ch.9.4: Partial Regression Plots. • **NOT SURE YET** about (optional) ch.9.5: Indicator Variables

Ch.23.5: • Chapter Multiple Regression Inference: Ignore the material about the Variance Inflation factor (VIF)
• Terminology: Skip • Inflation factor (VIF) • But REVIEW • collinearity • adjusted R^2

Ch.24: • SKIP THE ENTIRE CHAPTER!

Ch.25: • No F–table lookups in the exams, but do understand that the F–distributions come with TWO degrees of freedom $F_{\mathrm{df}_T, \mathrm{df}_E}$ where $\mathrm{df}_T = k - 1$ and $\mathrm{df}_E = N - k$.
• Ch.25.4 (Comparing Means): Ignore all formulas and numerical calculations. Only understand that you need to increase the confidence level for multiple comparisons and that Bonferroni says that you must decrease α to α/J.
• Terminology: Skip • LSD (least significant difference) • MSD (minimum significant difference)

Ch.26: • SKIP THE ENTIRE CHAPTER!

Ch.27: • SKIP THE ENTIRE CHAPTER!