MAS 170 Elementary Statistics Spring 2020
Vocabulary and Facts, Exam 2, Chapters 8,9,10,11,12

This document is an inventory of vocabulary terminology, facts, and problem types from Chapters 8--12 in the textbook. This is **not** a complete study guide. It is intended to be a *checklist* to help you form your own vocabulary cards or lists.

Prerequisite vocabulary and facts from Ch 3,4,5

average of a list of data

SD of a list of data

conversion of data from original units to ***standard units***

conversion from standard units to ***original units***

68% estimate rule for many kinds of data

median, ***percentiles*** for a list of data

specific skills for normal data
- read normal table "+/- z balanced" area from a z value
- read normal table z value from an area
- cut and combine "non +/- z" areas to use normal table areas

Vocabulary and Facts Ch 8--12

scatter diagram
- ***linear*** association (versus ***nonlinear*** association)
- ***homoscedastic*** versus ***heteroscedastic***

correlation
- formula, procedure to calculate
- range of values and interpretation in the scatter diagram
- how r behaves under rescaling of variables
- ***ecological correlation***
- association is not causation

point of averages

SD line
- formula
- rms error for the SD line

regression line
- formula
- rms error for the regression line
  - fact: regression line has least rms error among all possible lines
- regression line for y on x versus regression line for x on y

graph of averages

regression effect

regression fallacy

residuals
- what the graph of residuals tells you about data

"new average" and ***"new SD"*** (for a homoscedastic scatter diagram)

Procedural Problems Ch 8--12
==========

calculate r:
  Given paired X,Y data in a table, calculate r.

"Basic regression":
  Given summary statistics AVE(X), SD(X), AVE(Y), SD(Y), and r, for data called X and Y with a scatter diagram that shows linear association, estimate the Y-coordinate of the center of a given thin vertical strip of the scatter diagram.

"Basic regression" plus normal X,Y data:
  Given r for data X and Y that are both normal and linearly associated, estimate the percentile rank (among all Y data) of the Y-coordinate of the center of a thin vertical strip whose location has been specified as a percentile rank among X data.

Compare errors for estimates using various lines:
  Given data X,Y with linear association, compare the error in using
   (i) the horizontal line through the point of averages
   (ii) the SD line
   (iii) the regression line
  to estimate the Y-coordinate of the center of a given thin vertical strip in the scatter diagram.

normal data in thin vertical strips:
  Given X,Y data, both normal, with linear association and homoscedastic scatter diagram, use the "new average" and the "new SD" to estimate a percentage of Y data in a given range within a thin vertical strip of the scatter diagram.