

Stat 704, Homework 1

1. Let Y_1 , Y_2 , and Y_3 be independent random variables with means $E\{Y_i\} = \mu_i$ for $i = 1, 2, 3$ and common variance $\sigma\{Y_i\} = \sigma^2$. Define $\bar{Y} = \frac{1}{3}(Y_1 + Y_2 + Y_3)$.
 - (a) Find $\sigma\{Y_1 - \bar{Y}, \bar{Y}\}$.
 - (b) Find $E\{(Y_1 + 2Y_2 - Y_3)^2\}$.
 - (c) Now assume common covariance $\sigma\{Y_i, Y_{i'}\} = \omega$, $i, i' = 1, 2, 3$. Re-derive $\sigma\{Y_1 - \bar{Y}, \bar{Y}\}$.

2. Work from the merged data set created in Class Exercise 1 and Class Exercise 2.
 - (a) Take a look at the data and note all the missing values. Let's explore—construct a two-by-two frequency table (PROC FREQ) of Fall 2008 GPA after classifying GPA by missing/not missing and comment on the missingness pattern. How might this affect analysis? We could remove records with at least one missing value (e.g., use the subsetting *IF* statement `IF GPA08 NE . AND GPA09 NE .;`) if needed.
 - (b) Construct side-by-side boxplots of GPA in Fall 2008 and Fall 2009 and comment. Note that these displays do not reflect data pairing.
 - (c) Run a paired t-test in SAS on the differences. Comment on the boxplot and the test results. Do students improve their second year?
 - (d) Run the same analysis on the relative change in GPAs (the difference divided by Fall 2008 GPA). Discuss any differences in the two analyses. Which analysis do you think is more appropriate?

3. Examine the productivity data in Problem 16.7 on page 723 of your textbook. We will only deal with the data on the first two lines (“Low” and “Moderate”).

Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
High	8.5	9.7	10.1	7.8	9.6	9.5						

- (a) Use R to prepare side-by-side box plots for the two samples. Do the spreads seem to differ across samples?
- (b) Using a p-value from a t-test, test at $\alpha = 0.05$ whether the firms rated “Moderate” have a significantly higher mean productivity improvement than those rated “Low”.
- (c) From R, obtain the p-value from the F-test of $H_0 : \sigma_1 = \sigma_2$. What do you conclude here?
- (d) Obtain and interpret a 95% CI for the difference in mean productivity improvement between firms rated “Moderate” and those rated “Low”.
- (e) Construct normal quantile plots for the two samples and comment on whether the normality assumption appears satisfied.