**MATH 8050**

# Homework Assignment 1
# (Due on 5/16 9:45am in class)

**Problem 1.** Let

$$Y_{11}, Y_{12}, ..., Y_{1n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2),$$

independent of

$$Y_{21}, Y_{22}, ..., Y_{2n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2).$$

Let $\overline{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}$ and $\overline{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}$ be the sample means for the two populations.

(a) Find the $E(\overline{Y}_1 - \overline{Y}_2)$.

(b) Find the $V(\overline{Y}_1 - \overline{Y}_2)$.

(c) What is the distribution of $\overline{Y}_1 - \overline{Y}_2$?

**Problem 2.** 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 $V(Y_i) = \sigma^2$. Define $\overline{Y} = \frac{1}{3}(Y_1 + Y_2 + Y_3)$.

(a) Find the $Cov(Y_1 - \overline{Y}, \overline{Y})$.

(b) Find the $V\{(Y_1 + 2Y_2 - Y_3)^2\}$

**Problem 3.** This problem will assess the validity of t-procedures in the situation in which model assumptions are not met.

(a) Consider the situation in which $Y_1, Y_2, ..., Y_n \overset{iid}{\sim} t_5$, and note that $E(Y_i) = 0$. Now say we are oblivious to the fact that the data are non-normal, find a value of $n$ such that a 95% confidence interval for the true mean (the CI that we discussed in class) is at its nominal level; i.e., we want to find the value of $n$ such that if we

1. Generate $n$ observations from a $t_5$ distribution (this can be done in R with the following function rt(n,5))
2. Calculate the 95% CI associated with the $n$ observations in step 1.
3. Record whether or not the 95% CI contains the true mean
4. Repeat the above process a large number of times (say 10,000)

then the percentage of the CIs that contains 0 will be approximately 95%.

(b) Repeat (a) under the assumption that $Y_1, Y_2, ..., Y_n \overset{iid}{\sim} \chi^2_1$, and note that $E(Y_i) = 1$. (Note: In R you can generate $\chi^2_{df}$ random variables using the following command rchisq(n,df))

(c) Why do you believe that when $n$ is large enough the t-based CI procedure becomes valid regardless of the distribution?

(d) Comment on how the shape of the true distribution effects how large $n$ needs to be, if the assumption of normality is not valid.

**Problem 4.** A random sample of 796 teenagers revealed that in this sample, the mean number of hours per week of TV watching was $\overline{y} = 13.2$, with a standard deviation of $s = 1.6$. Find and interpret a 95% confidence interval for the true mean weekly TV-watching time for teenagers. Why can we use a t CI procedure in this problem?

**Problem 5.** Suppose a sample of 10 types of compact cars reveals the following one-day rental prices (in dollars) for Hertz and Thrifty, respectively:

| | Car Type | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Renter | A | B | C | D | E | F | G | H | I | J |
| Hertz | 37.16 | 14.36 | 17.59 | 19.73 | 30.77 | 26.29 | 30.03 | 29.02 | 22.63 | 39.21 |
| Thrifty | 29.49 | 12.19 | 15.07 | 15.17 | 24.52 | 22.32 | 25.30 | 22.74 | 19.35 | 34.44 |

(a) Explain why this is a paired-sample problem.

(b) Use a graph to determine whether the assumption of normality is reasonable.

(c) Using a p-value, test at $\alpha = 0.05$ whether Thrifty has a lower true mean rental rate than Hertz via a t-test.