

## Unit 7 – STATA for Analysis of 1, 2, and 3+ Samples Homework

**Due: Wednesday November 4, 2020**

### Preliminary

Exercises 1 -4 give you practice with “immediate” commands. You are also asked to provide interpretations of your findings. You will need to refresh your memory of introductory biostatistics!

Exercises 5-8 ask you to work with the data set [sepsis.dta](#). Download it from the course website.

### Description of [sepsis.dta](#)

This data set contains 455 observations. Data are from a study conducted by Bernard et al (1997). This was a randomized clinical trial to assess the effect of intravenous ibuprofen on mortality in patients with sepsis. Study investigators used a specific measure of mortality risk called the APACHE score; higher scores indicate greater risk. This data is used for exercises 6-9.

### Information for Exercises 1-2

**In the general population the population mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of serum creatinine are  $\mu = 1.0$  mg/dL and  $\sigma = 0.4$  mg/dL, respectively.**

- \_\_\_\_ 1. A sample of  $n=12$  patients were administered a new antibiotic. One day later, their serum creatinine levels were measured. The mean of this sample was 1.2 mg/dL. Using an appropriate *immediate* command, test the null hypothesis that the mean in the sample is different from the mean in the general population. Do NOT assume that the population variance is known. Instead, suppose you are given the sample standard deviation ( $s$ ) and it has value  $s = 0.6$  mg/dL. Provide a 1 sentence interpretation of your output.

- \_\_\_ 2. Using an appropriate *immediate* command, obtain a 95% confidence interval estimate of the true mean serum creatinine among patients who have received the new antibiotic. Again, do NOT assume that the population variance is known. Again, use the sample standard deviation  $s = 0.6$  mg/dL. Provide a 1-sentence interpretation of your output.

### Information for Exercise 3

**Suppose that in the year 1990, the incidence rate of myocardial infarction (MI) was 5 per 1000 among 45- to 54-year old men. It was of interest to study change over time in this incidence. Beginning in the year 2000, a sample of 5000 men aged 45- to 54 were followed for one year. In that one year, 15 new cases of MI occurred.**

- \_\_\_ 3. Using an appropriate *immediate* command, test the null hypothesis that the 2000 incidence of MI among 45- to 54-year old men is equal to the 1990 incidence of MI. Provide a 1 sentence interpretation of your output.

### Information for Exercise 4

**The drug erythromycin has been proposed to possibly lower the risk of premature delivery. A related area of interest is its association with the incidence of side effects during pregnancy. Assume 30% of all pregnant women complain of nausea between weeks 24 and 28 of pregnancy.**

- \_\_\_ 4. Consider now a sample of  $n=300$  women who are taking erythromycin regularly between weeks 24 and 28 of their pregnancies. Suppose that, of these, 110 complaints of nausea occurred. Using an appropriate *immediate* command, test the null hypothesis that the incidence rate of nausea for the erythromycin group is the same as that for a typical pregnant woman. Provide a 1 sentence interpretation of your output.

**Information for Exercises 5-8**

**Download from the course website**  
**sepsis.dta**

- \_\_\_ 5. Consider treated patients whose race is recorded as “other”. For this subset of the data, test whether the baseline temperature (**temp0**) is significantly different from their temperature after 2 hours (**temp1**). Provide a 1 sentence interpretation of your output.
  
- \_\_\_ 6. Consider, still, ONLY the treated patients whose race is recorded as “other”. For this subset of the data, obtain a 90% confidence interval for the true change in temperature between baseline and 2 hours. Provide a 1 sentence statement and interpretation of your confidence interval.
  
- \_\_\_ 7. Test whether the baseline APACHE score (**apache**) is different between treated and untreated patients. The treatment variable is **treat**. Provide a 1 sentence interpretation of your output.
  
- \_\_\_ 8. Exercise 7, continued – Obtain a 95% confidence interval for the true difference in APACHE scores between treated and untreated patients. Provide a 1 sentence statement and interpretation of your confidence interval.