

## STA 3032 - Practice set 2

### True or False

1. In conducting a hypothesis test regarding a mean, the larger the p-value, the stronger the evidence against the null hypothesis.
2. The (large-sample) sampling distribution for the sample mean is normally distributed only if the underlying population of measurements is normal.
3. Nonparametric/Distribution free procedures should ONLY be used for non-normal data.
4. All else being equal, as we increase the sample size, we increase the width of a Confidence Interval for  $\mu$ .

### Open Ended Problems

1. A method of spraying for rust mites has been conducted on a random sample of 100 10-acre plots in citrus groves in Florida. From the sample, the mean and standard deviation for the yield of fruit is 750 and 200 tons respectively. Construct a 95% confidence interval for  $\mu$ , the population mean yield per 10-acre plot.
2. A scientist wishes to estimate the mean breaking strength of a certain type of steel rod within 3.0 psi with 95% confidence. Based on experience with a similar type of steel, she believes the standard deviation is approximately 15.0 psi in individual breaking strengths. How many steel rods should she test?
3. A textile engineer is interested whether the variation in breaking strength of Yarn Type A is larger than the variation in Type B. Random samples of  $n_A = n_B = 7$  measurements were obtained, and the sample variances were  $s_A^2 = 5.0$  and  $s_B^2 = 1.25$ . Assuming the two types are normally distributed, do the data provide sufficient evidence to conclude  $\sigma_A^2 > \sigma_B^2$ ? Infer by using  $\alpha = 0.05$ .
4. A forensic researcher samples 100 adult males and 100 adult females and measures each subjects right foot length (cm). The following table gives the summary results:

Gender	$\bar{x}$	$s$
Female	23.80	1.1
Male	26.30	1.6

- (a) Test  $H_0 : \mu_F = \mu_M$  versus  $H_A : \mu_F \neq \mu_M$  at the  $\alpha = 0.05$  significance level.
- (b) Compute a 95% CI for  $\mu_F - \mu_M$
- (c) Assuming equal variances, recompute the 95% CI. How does it differ from the previous interval?

5. Suppose a sample of 20 students were given a diagnostic test before studying a particular module and then again after completing the module. We want to find out if, in general, our teaching leads to improvements in students knowledge/skills (i.e. test scores). We can use the results from our sample of students to draw conclusions about the impact of this module in general.

Student	Pre-module score	Post-module score
1	18	22
2	21	25
3	16	17
4	22	24
5	19	16
6	24	29
7	17	20
8	21	23
9	23	19
10	18	20
11	14	15
12	16	15
13	16	18
14	19	26
15	18	18
16	20	24
17	12	18
18	22	25
19	15	19
20	17	26

- (a) Is the data paired?
- (b) Check the normality assumption.
- (c) Test the null hypothesis (using both parametric and nonparametric/distribution free approaches) that the true mean difference is zero. What is your final conclusion?
6. Is cocaine deadlier than heroin? A study was conducted where rats were randomly assigned to receive unlimited supplies of cocaine or heroin. Of the 200 rats assigned to cocaine, 80% died in 30 days. Of the 200 rats assigned to heroin, 50% died in 30 days.
- (a) Compute a 95% Confidence Interval for  $p_C - p_H$ . Difference in population mean 30-day death rates for rats assigned to cocaine versus heroin.
- (b) Do you feel this is evidence to conclude that cocaine is deadlier than heroin in this setting?

7. Recent studies of the private practices of physicians who saw no Medicaid patients suggested that the 40th percentile length of each patient visit was 22 minutes. It is believed that the 40th percentile visit length in practices with a large Medicaid load is shorter than 22 minutes. A random sample of 20 visits in practices with a large Medicaid load yielded, in order, the following visit lengths:

9.4, 13.4, 15.6, 16.2, 16.4, 16.8, 18.1, 18.7, 18.9, 19.1,

19.3, 20.1, 20.4, 21.6, 21.9, 23.4, 23.5, 24.8, 24.9, 26.8

Based on these data, is there sufficient evidence to conclude that the 40th percentile visit length in practices with a large Medicaid load is shorter than 22 minutes?

8. A built in dataset in R named [InsectSprays](#) contains the counts of insects in agricultural experimental units treated with (6) different insecticides. Test whether the variance in counts of the insecticides is equal. (If using R just type `InsectSprays` and you will be able to see the data.
9. Students were asked to classify themselves according to smoking (Heavy, Regularly, Occasionally, Never) and according to exercise (Frequently, Some, None)

Smoking/Exercise	Frequently	Some	None
Heavy	7	3	1
Regularly	9	7	1
Occasionally	12	4	3
Never	87	84	18

Test whether the variables Smoking and Exercise are dependent. If so, interpret using standardized residuals.