Exam 4 Study Guide

Exam 4 will cover chapters 9 and 10

- Homework chapter 10 are due at the time of the exam.
- Your exam will come with a formula sheet identical to the one attached here (2 pages).
- Don’t forget your calculator! Unless otherwise stated on the exam, you are expected to do all hypotheses testing and finding all confidence intervals by hand. However, it is a good idea to check your solutions with that of your calculator programs before you turn in your exam.

Exam 4 Review Problems

Always show work where possible, and state any calculator programs that you are using. For all hypothesis tests, you are expected to use and show the six-step method specified in class and on worksheets. For confidence intervals you need to use and show the five-step method. For multiple choice questions, select the choice that best answers or finishes the statement, and no work needs to be shown.

1. Which type of sampling must be used to select the samples used for constructing confidence intervals and performing hypothesis tests?

2. Which condition(s) needs to be checked to determine if
   A) a sampling distribution of sample proportions is approximately normally distributed?
   B) a sampling distribution of sample means is approximately normally distributed?

3. When we want to calculate a confidence interval or perform a hypothesis test for a population mean, when will you use the $t$-distribution rather than the $z$-distribution in the formulas and procedures?

4. The null hypothesis is a claim about a:
   A) parameter, where the claim is assumed to be false until it is declared true
   B) parameter, where the claim is assumed to be true until it is declared false
   C) statistic, where the claim is assumed to be false until it is declared true
   D) statistic, where the claim is assumed to be true until it is declared false

5. You can decrease the width of a confidence interval by:
   A) lowering the confidence level or decreasing the sample size
   B) lowering the confidence level or increasing the sample size
   C) increasing the confidence level or decreasing the sample size
   D) increasing the confidence level or increasing the sample size

6. The mean federal income tax paid last year by a random sample of 19 persons selected from a city was $4275 with a standard deviation of $766. If we want to use this information to test at a 5% significance level, that the mean income tax of all persons in this city is more than $4000, we
   A) could use the ZInterval program on our TI83/84 calculators
   B) could use the Z-Test program on our TI83/84 calculators
   C) could use the TInterval program on our TI83/84 calculators
   D) could use the T-Test program on our TI83/84 calculators

7. In a hypothesis test, if we REJECT the null hypothesis at a 5% significance level, then it must be that
   A) P-value > 0.05
   B) P-value < 0.05
   C) P-value = 0.05
   D) P-value < 0.025
8. A two-tailed hypothesis test using the normal distribution reveals that the area under the sampling distribution curve of the mean and located in the tail to the right of the sample mean equals 0.028. Consequently, the \( p \)-value for this test equals:

9. We want to know if there is a difference in pay among females and males at a large cooperation. We draw two random samples, one from the population of female employees and one from the population of male employees at this cooperation. The two samples are
   A) independent      B) dependent      C) matched samples      D) paired samples

10. Match the following symbols with the correct meaning
   \( \mu_d \) the mean of the differences \( d \) for the population of all matched pairs of data
   \( \bar{d} \) or \( \bar{X}_d \) paired difference, the difference between the two data values in a single matched pair
   \( d \) the mean of the differences \( d \) for the paired sample data
   \( s_d \) the standard deviation of the difference \( d \) for the paired sample data

11. In 2008, the General Social Survey asked 1493 U.S. adults to rate their level of interest in environmental issues. Of these, 751 said that they were “very interested.”
   A) Does the survey provide convincing evidence that more than half of U.S. adults are very interested in environmental issues? Use a 5% level of significance.
   B) Based on your decision, which type of error could the sample data have led you to make?

12. A soft-drink manufacturer claims that its 12-ounce cans do not contain, on average, more than 30 calories. A random sample of 64 cans of this soft drink, which were checked for calories, contained a mean of 32 calories with a standard deviation of 3 calories.
   A) Does the sample information support the alternative hypothesis that the manufacturer’s claim is false? Use a significance level of 5%.
   B) Based on your decision, which type of error could the sample data have led you to make?

13. Maine Mountain Dairy claims that its 8-ounce low-fat yogurt cups contain, on average fewer calories than the 8-ounce low-fat yogurt cups produced by a competitor. A consumer agency wanted to check this claim. A sample of 27 such yogurt cups produced by this company showed that they contained an average of 141 calories per cup. A sample of 25 such cups produced by its competitor showed that they contained an average of 144 calories per cup. Assume that the two populations are normally distributed with population standard deviations of 5.5 and 6.4 calories, respectively.
   A) Make a 98% confidence interval for the difference between the mean number of calories in the 8-ounce low-fat yogurt cups produced by the two companies.
   B) Test at the 1% significant level whether Maine Mountain Dairy’s claim is true.

14. Suppose you want to test if the average study time of all Cabrillo students differs from 10 hours per week.
   A) Explain what a Type I Error would be in this case
   B) Explain what a Type II Error would be in this case

15. Listed below are temperatures (in °F) of six subjects measured at 8:00 am and then again at 12:00am. It’s assumed that these differences are approximately normally distributed.
   
<table>
<thead>
<tr>
<th>Time</th>
<th>97.0</th>
<th>96.2</th>
<th>97.6</th>
<th>96.4</th>
<th>97.8</th>
<th>99.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00am</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:00am</td>
<td>98.0</td>
<td>98.6</td>
<td>98.8</td>
<td>98.0</td>
<td>98.6</td>
<td>97.6</td>
</tr>
</tbody>
</table>

   A) Construct a 95% confidence interval estimate of the difference between the 8:00 am temperatures and the 12:00 am temperatures.
   B) Test at 3% significance level the claim that the body temperature is the same at both times.
16. Lipitor is a drug used to control cholesterol. In clinical trials of Lipitor, 94 subjects were treated with Lipitor and 270 subjects were given a placebo. Among those treated with Lipitor, 7 developed infections. Among those given a placebo, 27 developed infections.

A) Use a 0.05 significance level to test the claim that the rate of infection was the same for those treated with Lipitor and those given a placebo.

B) If you were to create a 95% Confidence Interval for the difference between the proportion of individuals who developed infections when being treated with Lipitor and those who were not treated with Lipitor, do you expect that 0 would be included in your interval? (You don’t need to find the interval to answer this question.) What if you created a 99% confidence interval, or a 90% confidence interval?

17. True or false: When we reject $H_0$ we are certain that $H_1$ is true.

18. True or false: When we reject $H_0$ it indicates that $H_1$ is likely to be true.

19. True or false: When we fail to reject $H_0$ we are certain that $H_0$ is true.

20. True or false: When we fail to reject $H_0$ it is an indication that $H_0$ may be true.

21. Write the letter of the hypothesis test or confidence interval from the list on the right that would be used for the situation on the left:

- Tori wants to test if the percentage of females that use Facebook is different from the percentage of males. She selects two, random, independent samples of males and females respectively and counts the number of Facebook users in each sample.

- Daniel wants to test if the average speeds of cars on Highway 1 and Highway 17 are different. He selects two, random, independent samples and measures their speeds.

- Kaylee wants to estimate the difference between the average grades among students who prefer taking notes with a pencil vs. the students who prefer taking notes with a pen. She selects two, random, independent samples of students who write with pencil and pen respectively, and ask them about their GPA.

- Stephan wants to estimate the weekly average amount of hours spent on statistics homework by all statistics students at Cabrillo. He selects a random sample of statistics students, and records their weekly hours of doing statistics homework.

| A. Test about $\mu$ | B. Test about $p$ | C. Test about $\mu_d$ | D. Test about $\mu_1 - \mu_2$ | E. Test about $p_1 - p_2$ | F. Confidence Interval about $\mu$ | G. Confidence Interval about $p$ | H. Confidence Interval about $\mu_d$ | I. Confidence Interval about $\mu_1 - \mu_2$ | J. Confidence Interval about $p_1 - p_2$ |
Answers to some of the Review Problems

11. A) There is not sufficient evidence to show that more than half of U.S. adults are very interested in the environment.
    B) Type II Error

12. A) There is sufficient evidence to show that the manufacturer’s claim is false.
    B) Type I Error

13. A) -6.86 to 0.86 calories
    B) There is not sufficient evidence to show that Maine Mountain Dairy yogurt cups have fewer calories than their competitor’s yogurt cups.

14. A) Your conclusion is that the average study time of all Cabrillo students does differ from 10 hours per week, when in fact it does not.
    B) Your conclusion is that there is not enough evidence to show that average study time of all Cabrillo students is different than 10 hours per week, when in fact it is.

15. A) There is not sufficient evidence to show that the body temperatures are different at both times (p-value=0.1632, t=-1.6341)
    B) -2.32°F< µ_d <0.52°F

16. A) There is not sufficient evidence to show that there is a difference in the infection rate among those treated with Lipitor and those not. (p-value=0.4638, z=-0.73)
    B) We would expect 0 to fall in our 95% interval, since this corresponds to the same significance/confidence level as in our two-tailed hypothesis test in part A). A 99% confidence interval will be wider, so it would be included here as well. A 90% confidence interval would be narrower, so we don’t know if 0 would be included in this interval.

17. False
18. True
19. False
20. True
21. E, D, I, A
# Formula Sheet for Ch 8, 9, and 10

## Confidence Intervals

<table>
<thead>
<tr>
<th>Desired Estimate:</th>
<th>Requirements:</th>
<th>Distribution:</th>
<th>Formula:</th>
<th>Program:</th>
</tr>
</thead>
</table>
| Conf. Int. for $\mu$ | 1. SRS  
2. $n > 30$ or population is normal  
3. $\sigma$ known | $z$ distribution | $\bar{x} \pm z \frac{\sigma}{\sqrt{n}}$ | ZInterval |
| Conf. Int. for $\mu$ | 1. SRS  
2. $n > 30$ or population is normal  
3. $\sigma$ unknown, but $s$ known | $t$ distribution  
$df = n - 1$ | $\bar{x} \pm t \frac{s}{\sqrt{n}}$ | TInterval |
| Conf. Int. for $p$ | 1. SRS  
2. $np > 10$ and $n(1-p) > 10$ | $z$ distribution | $\hat{p} \pm z \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ | 1-PropZInt |
| Conf. Int. for $\mu_1 - \mu_2$ | 1. Independent SRSs  
2. $n_1, n_2 > 30$ or pops. normal  
3. $\sigma_1, \sigma_2$ unknown, and $s_1, s_2$ known | $t$ distribution  
$df = n - 1$  
for smallest $n$ or use calculator to find $df$ (it will be different) | $(\bar{x}_1 - \bar{x}_2) \pm t \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ | 2-SampTInt (not pooled) |
| Conf. Int. for $\mu_d$ | 1. Paired SRSs  
2. $n \geq 30$ or pop. of diff. normal  
3. $\sigma_d$ unknown, $s_d$ known | $t$ distribution  
$df = n - 1$ | $\bar{x}_d \pm t \frac{s_d}{\sqrt{n}}$ | TInterval |
| Conf. Int. for $p_1 - p_2$ | 1. Independent SRSs  
2. Each pop. size $\geq 20 \cdot n$  
3. Two categories with at least 10 in each. | $z$ distribution | $(\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ | 2-PropZInt |

SRS = Simple Random Sample
### Hypothesis Tests

<table>
<thead>
<tr>
<th>Type of Test:</th>
<th>Requirements:</th>
<th>Distribution:</th>
<th>Details:</th>
</tr>
</thead>
</table>
| Test about $\mu$: | 1. SRS  
2. $n > 30$ or pop. is normal  
3. $\sigma$ known | $z$ distribution | Test Stat: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$  
P-Value: Normal83, or ZTest |
| Test about $\mu$: | 1. SRS  
2. $n > 30$ or pop. is normal  
3. $\sigma$ not known, $s$ known | $t$ distribution | Test Stat: $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$  
$df = n - 1$  
P-Value: T83, or TTest |
| Test about $p$: | 1. SRS  
2. $np > 10$ and $n(1 - p) > 10$ | $z$ distribution | Test Stat: $z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$  
P-Value: Normal83, or 1-PropZTest |
| Test about $\mu_1 - \mu_2$: | 1. Independent SRSs  
2. $n_1, n_2 > 30$ or pops. Normal  
3. $\sigma_1, \sigma_2$ unknown, $s_1, s_2$ known | $t$ distribution  
$df = n - 1$ for smallest $n$ or use calculator to find df (it will be different) | Test Stat: $t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$  
P-Value: T83, or 2-SampTTest (not pooled) |
| Test about $\mu_d$: | 1. Paired SRSs  
2. $n > 30$ or pop. of diff. normal  
3. $\sigma_d$ unknown, $s_d$ known | $t$ distribution | Test Stat: $t = \frac{\bar{x}_d - \mu_d}{s_d/\sqrt{n}}$  
P-Value: TTest |
| Test about $p_1 - p_2$: | 1. Independent SRSs  
2. Each pop. size $\geq 20 \cdot n$  
3. Two categories with at least 10 in each. | $z$ distribution | Test Stat: $z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$,  
where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$  
P-Value: NORMAL83, or 2-PropZTest |