MATH1180
Homework 9
Due Wednesday, May 3, 2017

Please read Sections 5.6, 6.1.1, and 6.3 of the textbook.

Please submit solutions to the following problems. Use 4 decimal places of accuracy for probabilities unless a problem says otherwise.

When submitting homework, please remember the following:

- Show all work leading to each solution.
- *Staple* all sheets together. A paper clip is not acceptable.
- Do not submit crossed-out or sloppy work.
- Do not submit ripped or torn pages.
- Be sure to submit your own work.

PROBLEM 1. A national survey reports that 66% of Americans favor restrictions on the use of firearms. A local advocacy group would like to determine if the proportion of Americans in their state who favor restrictions on the use of firearms (p) is the same as the national proportion or if it is different.

The group decides to test $p = 0.66$ versus $p \neq 0.66$ using information from a simple random sample of 88 individuals who live in their state, the z-statistic for large sample analyses of population proportions, and the following rejection region: $|z| \geqslant 1.645$.

(*a*) Find the probability of making a type I error in the test described in the paragraph above.

(*b*) If the null hypothesis is true, how many individuals who favor restrictions on firearms are expected to be in the sample? If the null hypothesis is true, how many individuals who do not favor restrictions on firearms are expected to be in the sample?

(*c*) Find the probability of making a type II error in the test described in the paragraph above if the true population proportion is 0.74.

(*d*) Find the power of the test described in the paragraph above if the true population proportion is 0.60.

PROBLEM 2 (BR). Research reports on a new antihypertensive drug state that 20% of hypertensive patients who use the drug will experience side effects. An independent medical advocacy group would like to verify this claim for themselves. They decide to test the null hypothesis that $p = 0.20$ versus the alternative hypothesis that $p \neq 0.20$ at the 5% significance level.

Suppose that the true population proportion is 0.30 (that is, suppose that 30% of hypertensive patients who use the drug will experience side effects).

(*a*) Find the power of the test proposed by the advocacy group if 80 subjects are used in the study.

(*b*) Find the power of the test proposed by the advocacy group if 170 subjects are used in the study.

PROBLEM 3 (WD, RG). The Poisson distribution can be used to model the number of patients referred to a specialist practice in a fixed period of time.

(*a*) Suppose that, for a certain practice, the mean rate of new referrals is 2.79 patients per week. Let X be the total number of new referrals in 52 weeks, and $\widehat{\lambda} = X/52$ be the sample mean rate per week. Find the interval containing the central 90% of the $\widehat{\lambda}$ distribution.

(*b*) If the true mean rate is actually 2.25 referrals per week, find the probability that a sample mean rate lies in the interval you calculated part (*a*).

(*c*) Consider testing the null hypothesis that $\lambda = 2.79$ versus the alternative hypothesis that $\lambda \neq 2.79$ at the 10% significance level using information gathered over 52 weeks. Does the probability you calculated in part (*b*) represent the probability of making a type I error when the true parameter is 2.25, the probability of making a type II error when the true parameter is 2.25, the power when the true parameter is 2.25, or none of these? Why? Be specific.

PROBLEM 4 (DSB). The following table appeared in a recent report summarizing body mass index (BMI) scores from simple random samples of men and women considered to be at high risk for coronary heart disease:

	Sample Size	*Mean (SD)*	*95% CI*
Men	20	31.6 (1.4)	????
Women	10	28.1 (2.0)	????

The column labeled "*Mean (SD)*" gives the sample mean BMI score for each sample of high risk individuals, with the sample standard deviation in parentheses, and the column labeled "95% *CI*" gives the 95% confidence interval for the mean BMI score constructed using small sample methods.

(*a*) Complete the table. Be sure to show all work.

(*b*) Do your confidence intervals suggest a difference between mean BMI scores for the men and women in these high risk populations? Why?

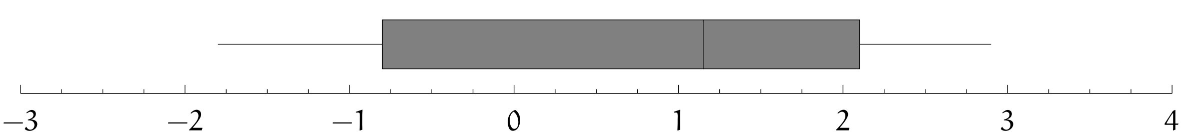
(*c*) What, if any, assumptions have you made in calculating the confidence intervals?

PROBLEM 5 (WT). The Activity Measure for Post-Acute Care (AMPAC) is a patient reported health survey measuring 3 physical domains. Scale scores are generated based on a patient's own responses to the survey items. However, when patients cannot provide their own data due to cognitive limitations, proxy reports from family members or clinicians are used. An interesting question is to compare the proxy reported scores to the patient recorded scores.

In one study, researchers examined the agreement between patient reported scores and proxy reported scores in the AMPAC-*Movement* domain from a simple random sample of 14 patients. Each patient was scored twice: the first score was based on the patient's own report and the second was based on the proxy report. The variable of interest is the following difference:

$$\text{Difference} = \text{Patient Recorded Score Minus Proxy Reported Score.}$$

The following box plot shows the distribution of differences:



The following table gives numerical summaries of the differences:

Sample Mean:	*Sample SD:*	*Sample Size:*
0.736	1.538	14

Let μ be the population mean difference.

(*a*) Use small sample methods to test the null hypothesis that $\mu = 0$ versus the alternative hypothesis that $\mu \neq 0$ at the 5% significance level. Clearly state your conclusion. Be sure to show all work.

(*b*) What, if anything, does your conclusion from part (*a*) say about self-reported scores and proxy reported scores?

PROBLEM 6 (BR). High serum cholesterol levels have been associated with heart disease, and recent studies have pointed to the possible link between diet and serum cholesterol levels.

For women aged 20–39, serum cholesterol levels are approximately normally distributed with mean 215mg/dL. A research team is interested in determining if women who follow a macrobiotic diet have the same mean serum cholesterol level or if it is different. A simple random sample of 24 women in this age range who eat a primarily macrobiotic diet was chosen.

The following table summarizes analyses concerning this small sample:

Small sample test of $\mu = 215$ *versus* $\mu \neq 215$:

$\bar{x}$	n	s	se	95% CI	t-Statistic	p-Value
195.131	24	34.933	????	????	????	0.0106

(*a*) Fill in the missing information. Be sure to show all work.
(*b*) Interpret the resulting table. Use the 5% significance level when interpreting the test result.

PROBLEM 7 (AF). A recent General Social Survey asked participants to respond to the question:

"What is the ideal number of children for a family to have?"

The following table summarizes the responses of the 772 women and 530 men who participated:

	Sample Size:	*Sample Mean*:	*Sample SD*:
Women	772	2.48	0.81
Men	530	2.45	0.90

(*a*) Let μ_1 be the mean response to this question for adult women in the United States and let μ_2 be the mean response to this question for adult men in the United States. Use the information above to construct a 95% confidence interval for the difference in means $\mu_1 - \mu_2$. Interpret your interval.
(*b*) What, if any, assumptions have you made in constructing the interval in part (*a*)?

PROBLEM 8 (WD). Researchers questioned 326 women over 45 years old who were receiving free bone mineral density screening. The questions focused on past smoking history. Subjects undergoing hormone replacement therapy (HRT), and subjects not undergoing HRT, were asked if they had been a regular smoker.

Sixty-five of the 220 women in the HRT group stated that they were at some point in their life a regular smoker, and 18 of the 106 women in the non-HRT group responded positively to being at some point in their life a regular smoker.

(*a*) Let p_1 be the proportion of women over 45 years old on HRT who at some point in their life were regular smokers, and let p_2 be the proportion of women over 45 years old not on HRT who at some point in their life were regular regular smokers. Use the information above to construct a 95% confidence interval for the difference in proportions $p_1 - p_2$. Interpret your interval.
(*b*) What, if any, assumptions have you made in constructing the interval in part (*a*)?

PROBLEM 9 (DSB). A simple random sample of 200 individuals with gastroesophageal reflux disease (GERD) agreed to participate in a randomized clinical trial comparing two treatments for the disease; 100 individuals were randomly assigned to each treatment group. The following table compares background characteristics for subjects in each treatment group:

Background Characteristic	*Treatment 1* ($n_1 = 100$)	*Treatment 2* ($n_2 = 100$)
Mean age in years, with Standard Deviation in parentheses	43.0 (7.2)	45.0 (8.1)
Mean annual income in dollars, with Standard Deviation in parentheses	\$41,352 (\$8754)	\$39,459 (\$9687)
Gender: % Male	65%	48%
Race: % Nonwhite	36%	39%

(*a*) In each case, find the observed value of the test statistic, and the p value for the two-sided test:

(*i*) A test of the null hypothesis that the difference in mean ages for the two groups equals 0 versus the alternative that the difference is not 0.

(*ii*) A test of the null hypothesis that the difference in mean annual incomes equals 0 versus the alternative that the difference is not 0.

(*iii*) A test of the null hypothesis that the difference in proportions of men equals 0 versus the alternative that the difference is not 0.

(*iv*) A test of the null hypothesis that the difference in proportions of nonwhites is 0 versus the alternative that the difference is not 0.

Be sure to show all steps.

(*b*) Using the 5% significance level, are there any statistically significant differences in background characteristics? Explain.

Footnote: This is a typical application of testing methods in randomized experiments. After individuals are randomly assigned to receive a particular treatment, researchers will test whether the randomization process has "evenly split" individuals on important background characteristics. Because we have large samples, p-values are calculated in exactly the same way as in the third workbook.