MATH1180
Homework 8
Due Wednesday, April 19, 2017

Please (re)read Chapter 5 through Section 5.5.6 of the Baglivo text.

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 (DS). Glaucoma is a disease of the optic nerve. An important risk factor for glaucoma is elevated eye pressure, where researchers define *elevated eye pressure* as values greater than 20.1 millimeters of mercury (mmHg). Assume that the following intersection probability table shows the proportions of individuals in a certain population cross-classified by whether or not they have glaucoma, and whether or not they have elevated eye pressure:

	Has Elevated Pressure	*Does Not Have Elevated Pressure*
Has Glaucoma	0.1479	0.0221
Does Not Have Glaucoma	0.1245	0.7055

Suppose that researchers would like to use elevated eye pressure as a diagnostic test for glaucoma. Specifically, the test will give a positive result (POS) if a person's eye pressure is greater than 20.1 mmHg, and will give a negative result (NEG) if a person's eye pressure is 20.1 mmHg or less.

(a) Find the prevalence of glaucoma in this population.

(b) Find the sensitivity and specificity of the elevated eye pressure test for glaucoma.

(c) Find the proportion of individuals for whom the test would give the correct result.

(d) Find the predictive value of a positive test and the predictive value of a negative test.

(e) Suppose that this test is used in a population where 8% of individuals have glaucoma. Use your answer to part (b) to find the predictive value of a positive test and the predictive value of a negative test in this population.

PROBLEM 2 (RG). For purposes of this problem, assume there are 4 major blood types: O, A, B, and AB. Blood type distributions vary significantly in different subpopulations in the United States.

Considering Caucasians and non-Caucasians in the United States as two separate populations, the blood type distributions are as follows:

	O	*A*	*B*	*AB*
Caucasians:	0.45	0.40	0.11	0.04
Non-Caucasians:	0.51	0.26	0.18	0.05

Notice that each row separately sums to 1.00. THIS IS NOT A TABLE OF JOINT PROBABILITIES.

Suppose that 77% of U.S. citizens are Caucasian and 23% are non-Caucasian.

(*a*) If a random individual is chosen from the population of U.S. citizens, find the probability that the chosen individual has each of the 4 possible blood types. Display your answers in a chart.

	O	*A*	*B*	*AB*
US Citizen:	????	????	????	????

(*b*) Suppose it is known that the randomly chosen citizen has type *AB* blood. Find the probability that the individual is non-Caucasian.

PROBLEM 3 (AF). In some research reports, the investigators fail to give all the important summary information. Suppose, for example, that the following statement is in a report on the results of a survey of a simple random sample of adults in the United States concerning their attitudes toward whether the use of marijuana should be legalized:

> "With 95% confidence, we believe that between 31.8% and 38.4% of adults in the United States believe that the use of marijuana should be legalized."

Using this information, find each of the following four quantities:

(*a*) the value of the sample proportion, $\widehat{p}$;

(*b*) the margin of error;

(*c*) the estimated standard error (use 4 decimal places of accuracy);

(*d*) the sample size (give a good approximation).

PROBLEM 4 (DSB). The following table appeared in a recent report summarizing a study investigating the relationship between body mass index (BMI) and dietary habits:

Characteristics of Study Sample ($n = 100$)	*Mean* (SE)
Body mass index	25 (0.67)
Total calories per day	1875 (36.4)
Total fat grams per day	21 (0.31)

The study involved subjects 18 years of age or older. The column labelled "*Mean* (SE)" gives the sample mean for each characteristic, with the estimated standard error of the sample mean based on a simple random sample of size 100 from the population in parentheses.

Suppose that we would like to design a study to investigate the relationship between body mass index and dietary habits among individuals aged 15–17 years.

(a) Use the table above to fill in the following table of sample standard deviations:

Sample Standard Deviation:		
Body mass index	Total calories per day	Total fat grams per day
???	???	???

(b) Using the sample standard deviation for body mass index values for individuals 18 years of age or older as your best guess for the standard deviation, find the sample size needed to estimate the mean body mass index in individuals aged 15–17 to within 2 units with 99% confidence.

(c) Using the sample standard deviation for total daily calories values for individuals 18 years of age or older as your best guess for the standard deviation, find the sample size needed to estimate the mean daily caloric intake to within 100 calories with 99% confidence.

(d) Using the sample standard deviation for total daily fat values for individuals 18 years of age or older as your best guess for the standard deviation, find the sample size needed to estimate the mean daily fat intake to within 1 gram with 99% confidence.

(e) What is the smallest sample size that will ensure that all three criteria (that is, the three criteria given in parts (b), (c), and (d)) are satisfied?

PROBLEM 5. A 2014 report from the U.S. Census Bureau states that 13% of adults do not have health insurance of any kind. In order to determine if this percentage is the same in the subpopulation of adults aged 26 to 34, researchers decided to test

$$H_0 : \ p = 0.13 \quad \text{versus} \quad H_A : \ p \neq 0.13$$

at the 1% significance level, where p is the proportion of adults between the ages of 26 and 34 who do not have health insurance of any kind.

The following table summarizes information the investigators gathered from a simple random sample of 326 adults in this age range:

Has Insurance:	*Does Not Have Insurance*:
265	61

(a) Find the rejection region for this test.

(b) Find the value of the test statistic for the given sample. Is this value in the acceptance region or the rejection region?

(c) Clearly state your conclusion.

PROBLEM 6. A 2008 report published by the state traffic department states that the numbers of cars passing a certain interchange during midday (11AM to 2PM) have a mean hourly rate of 9.8 cars per hour.

In order to determine if the traffic patterns have remained the same, county officials decide to test

$$H_0 : \ \lambda = 9.8 \quad \text{versus} \quad H_A : \ \lambda \neq 9.8$$

at the 5% significance level, where λ is the mean hourly rate of cars passing the interchange during the midday hours of 11AM to 2PM.

The officials determined that a total of 275 cars passed the interchange during 30 hours of observation.

(a) Find the rejection region for this test.

(b) Find the value of the test statistic for the given sample. Is this value in the acceptance region or the rejection region?

(c) Clearly state your conclusion.

PROBLEM 7 (PG). For the population of male industrial workers in London who have *never* experienced a major coronary event, the mean systolic blood pressure (SBP) is 136 millimeters of mercury (mmHg).

In order to determine if the mean SBP is the same for the population of male workers who have suffered a coronary event, researchers decided to test

$$H_0 : \ \mu = 136 \quad \text{versus} \quad H_A : \ \mu \neq 136$$

at the 5% significance level, where μ is the mean SBP for male industrial workers in London who have experienced a major coronary event.

The following table gives summary measures for a simple random sample of workers who have experienced a major coronary event:

Sample Mean (mmHg):	*Sample Standard Deviation (mmHg)*:	*Sample Size*:
142.8	24.4	86

(a) Find the rejection region for this test.

(b) Find the value of the test statistic for the given sample. Is this value in the acceptance region or the rejection region?

(c) Clearly state your conclusion.

PROBLEM 8 (AF). In a recent General Social Survey (GSS), participants were asked

> "Would you be willing to pay much higher taxes in order to protect the environment?"

Of the 852 responders, 369 said yes and 483 said no.

Let p be the proportion of adults in the U.S. who would answer yes to this question. The following table summarizes analyses of whether a majority or minority of Americans would answer yes:

Large sample test of $p = 0.50$ *versus* $p \neq 0.50$:

x	n	$\widehat{p}$	95% CI	z-Statistic	p-Value
369	852	0.4431	[0.3998, 0.4664]	−3.91	0.0000

(The p-value is 0 to 4 decimal places of accuracy.)

(a) Interpret the results given in the table above. Use the 5% significance level to interpret the result of the test.

(b) Explain an advantage of the confidence interval shown over the hypothesis test.

(c) What, if any, assumptions were used in the analyses reported in the table?

PROBLEM 9 (AF). In a recent General Social Survey (GSS), participants in a simple random sample of adults in the United States were asked to respond to the statement

> "A preschool child is likely to suffer if his or her mother works"

using the ordinal response categories: Strongly Disagree, Disagree, Agree, Strongly Agree. The responses were then scored using

> -2 for strongly disagree, -1 for disagree, 1 for agree and 2 for strongly agree.

The following table shows the frequency distribution for a simple random sample of 996 individuals:

Agreement Score	-2	-1	1	2	*Total*
Frequency	99	421	385	91	996

(*a*) Find the proportion of individuals who either agreed or strongly agreed with the statement.

(*b*) Let X be the agreement score of a randomly chosen adult in the United States and $\mu = E(X)$ be the mean of the distribution of agreement scores.

The table below gives the sample mean ($\bar{x}$) score for the sample chosen by the GSS, the number of individuals asked (n), the sample standard deviation (s), the estimated standard error of the sample mean (se), a 95% confidence interval for the population mean, the large sample z statistic for the test of $\mu = 0$ versus $\mu \neq 0$, and the observed significance level for the test.

Large sample test of $\mu = 0$ *versus* $\mu \neq 0$*:*

$\bar{x}$	n	s	se	95% CI	z-Statistic	p-Value
-0.0522	996	1.2535	???	???	???	???

Please fill in the missing information. (Interpretation will be done in part (*d*).)

(*c*) Based on the scores developed by the researchers, explain why the researchers would be interested in testing $\mu = 0$ versus $\mu \neq 0$.

(*d*) Interpret the results in the table you completed in part (*b*). Use the 5% significance level when interpreting the test result.

PROBLEM 10 (WMS). The U.S. Fire Administration reports that there were nearly 1.4 million fires nationwide in 2011, and that there were, on average, 2.16 deaths per thousand fires.

Government officials are interested in determining if the mean death rate per thousand fires is the same today as it was in 2011. They decide to test the null hypothesis that $\lambda = 2.16$ versus the alternative hypothesis that $\lambda \neq 2.16$ at the 5% significance level.

Assume that 80 thousand fires were reported during the study period, and that a total of 154 deaths occurred in these fires.

(a) Complete the following table:

Large sample test of $\lambda = 2.16$ *versus* $\lambda \neq 2.16$*:*

x	n	$\widehat{\lambda}$	95% CI	z-Statistic	p-Value
154	80	????	????	????	????

(b) Interpret the results in your table. Use the 5% significance level when interpreting the test result.

PROBLEM 11 (AF). Suppose that the p-value for a properly conducted test of the null hypothesis that $\mu = 100$ versus the alternative hypothesis that $\mu \neq 100$ is 0.043.

(a) What conclusion would you draw using a 5% significance level?

(b) If the conclusion in part (a) is incorrect, what type of error has been made?

(c) Does a 95% confidence interval for μ contain 100? Explain.

PROBLEM 12 (AF). A medical study in 1996 concluded that women who take birth control pills have a higher chance of getting breast cancer. A later study by scientists at the National Institutes of Health (NIH) and the Centers for Disease Control (CDC) reported no effect (*NEJM*, 6/27/2002).

Discuss the factors that can cause different medical studies to come to different conclusions. Include in your discussion the words "chance," "type I error," and "type II error."

PROBLEM 13 (RG). On March 23, the *New York Times* reported in an article:

> About 215,000 children under 5 die each year of rotavirus, almost half of them in just four countries: India, Pakistan, Nigeria and the Democratic Republic of Congo, according to the W.H.O. A major 2013 study sponsored by the Bill and Melinda Gates Foundation found that rotavirus was the leading cause of fatal diarrhea in children under age 2—and the only major one not caused by bacteria or parasites, which are treatable with antibiotic and antiparasitic drugs.

A test of a new vaccine had this result:

> There were only 31 cases among the 1780 children who got three doses of the vaccine, while there were 87 among the 1728 children who got a placebo.

One test of the effectiveness of the vaccine is to test the null hypothesis that $p = \frac{87}{1728} \approx 0.0503$. Complete the following table:

Large sample test of $p = 0.0503$ *versus* $p \neq 0.0503$*:*

x	n	$\widehat{p}$	95% CI	z-Statistic	p-Value
31	1780	???	???	???	???

Interpret your results.