1: Hypothesis testing (think about)

(a) Solve:

For the following two points, is the statement a null hypothesis, or an alternative hypothesis?

- In Canada, the proportion of adults who favor legalized gambling equals 0.50.

- The proportion of all Canadian college students who are regular smokers is less than 0.24, the value it was 10 years ago.

- Introducing notation for a parameter, state the hypotheses in the previous two points in terms of the parameter values.

(b) Solve:

For each of the following, is the statement a null hypothesis or an alternative hypothesis? Why?

- The mean IQ of all students at Lake Wobegon High School is larger than 100.

- The probability of rolling a 6 with a particular die equals 1/6.

- The proportion of all new business enterprises that remain in business for at least five years is less than 0.50.

(c) Solve:

For a new pesticide, should the Environmental Protection Agency (EPA) have the burden of proof to show that it is harmful to the environment, or should the producer of the pesticide have the burden of proof to show that it is not harmful to the environment? Give the analog of the null hypothesis and the alternative hypothesis if the burden of proof is on the EPA to show the new pesticide is harmful.

(d) Solve:

Suppose the mean GPA of all students graduating from the University of Alabama in 1985 was 3.05. The registrar plans to look at records of students graduating in 2011 to see if mean GPA has changed. Define notation and state the null and alternative hypotheses for this investigation.

(e) Solve:

A study plans to have a sample of obese adults follow a proposed low-carbohydrate diet for three months. The diet imposes limited eating of starches (such as bread and pasta) and sweets, but otherwise no limit on calorie intake. Consider the hypothesis, The population mean of the values of weight change (= weight at start of study - weight at end of study) is a positive number.

- Is this a null or an alternative hypothesis? Explain your reasoning.

● Define a relevant parameter, and express the hypothesis that the diet has no effect in terms of that parameter. Is it a null or alternative hypothesis?

(f) Solve:

Give an example of a null hypothesis and an alternative hypothesis about a:

● population proportion,

● population mean.

2: Hypothesis testing (one sample tests)

(a) Solve:

A person who claims to be psychic says that the probability p that he can correctly predict the outcome of the roll of a die in another room is greater than 1/6, the value that applies with random guessing. If we want to test this claim, we could use the data from an experiment in which he predicts the outcomes for n rolls of the die. State hypotheses for a significance test, letting the alternative hypothesis reflect the psychic's claim.

(b) Solve:

For a test of $H_0 : p = 0.50$, the z test statistic equals 1.04.

● Find the P–value for $H_a : p > 0.50$.

● Find the P–value for $H_a : p?0.50$.

● Find the P–value for $H_a : p < 0.50$.
  (Hint: The P–values for the two possible one-sided tests must sum to 1.)

● Do any of the P–values in previous points give strong evidence against H_0? Explain.

NOTE: try to answer using both the Z–table and the proper R function

(c) Solve:

For a test of $H_0 : p = 0.50$, the sample proportion is 0.35 based on a sample size of 100.

● Show that the test statistic is z = -3.0.

● Find the P-value for Ha: p ¡ 0.50.

● Does the P-value in second point give much evidence against H_0? Explain.

(d) Solve:

According to an exit poll in the 2008 Vermont gubernatorial election, 54.5% of the sample size of 837 reported voting for the Republican candidate Douglas. Is this enough evidence to predict who won? Test that the population proportion who voted for Douglas was 0.50 against the alternative that it differed from 0.50. Answer by:

- Identifying the variable and parameter, and defining notation.

- Stating hypotheses and checking assumptions for a large-sample test.

- Reporting the P–value and interpreting it. (The test statistic equals 2.662.)

- Explaining how to make a decision for the significance level of 0.05.

(e) Solve:

A study has a random sample of 20 subjects. The test statistic for testing $H_0 : \mu = 100$ is $t = 2.40$. Find the approximate P–value for the alternative:

- $H_a : \mu \neq 100$,

- $H_a : \mu > 100$,

- $H_a : \mu < 100$.

NOTE: try to answer using both the t–table and the proper R function

(f) Solve:

When the 636 male workers in the 2008 GSS were asked how many hours they worked in the previous week, the mean was 45.5 with a standard deviation of 15.16. Does this suggest that the population mean work week for men exceeds 40 hours? Answer by:

- Identifying the relevant variable and parameter.

- Stating null and alternative hypotheses.

- Reporting and interpreting the P–value for the test statistic value of $t = 9.15$.

- Explaining how to make a decision for the significance level of 0.01.

(g) Solve:

A significance test about a proportion is conducted using a significance level of 0.05. The test statistic equals 2.58. The P–value is 0.01.

- If H_0 were true, for what probability of a Type I error was the test designed?

- If this test resulted in a decision error, what type of error was it?

(h) Solve:

Consider the test of H_0: The defendant is not guilty against H_a: The defendant is guilty.

- Explain in context the conclusion of the test if H_0 is rejected.

- Describe the consequence of a Type I error.

- Explain in context the conclusion of the test if you fail to reject H_0.

- Describe the consequence of a Type II error.

(i) Solve:

Consider the test of H_0: The new drug is safe against Ha: the new drug is not safe.

- Explain in context the conclusion of the test if H_0 is rejected.

- Describe the consequence of a Type I error.

- Explain in context the conclusion of the test if you fail to reject H_0.

- Describe the consequence of a Type II error.

3: Hypothesis testing (two sample tests)

(a) Solve:

The PACE project (pace.uhs.wisc.edu) at the University of Wisconsin in Madison deals with problems associated with high–risk drinking on college campuses. Based on random samples, the study states that the percentage of UW students who reported bingeing at least three times within the past two weeks was 42.2% in 1999 ($n = 334$) and 21.2% in 2009 ($n = 843$).

- Estimate the difference between the proportions in 1999 and 2009, and interpret.

- Find the standard error for this difference. Interpret it.

- Construct and interpret a 95% confidence interval to estimate the true change, explaining how your interpretation reflects whether the interval contains 0.

- State and check the assumptions for the confidence interval in previous point to be valid.

(b) Solve:

In the study for cancer death rates, consider the null hypothesis that the population proportion of cancer deaths π_1 for placebo is the same as the population proportion π_2 for aspirin. The sample proportions were $\hat{\pi}_1 = 347/11535 = 0.030$ and $\hat{\pi}_2 = 327/14035 = 0.023$.

- For testing $H_0 : \pi_1 = \pi_2$ against $H_a : \pi_1 \neq \pi_2$, show that the pooled estimate of the common value π under H_0 is $\hat{\pi} = 0.027$ and the standard error is 0.002.

- Show that the test statistic is $z = 3.5$.

- Find and interpret the P–value in context.

(c) Solve:

Do women tend to spend more time on housework than men? If so, how much more? Based on data from the National Survey of Families and Households, one study reported the results in the table for the number of hours spent in housework per week. (Source: Data from A. Lincoln, Journal of Marriage and Family, vol. 70, 2008, pp. 806–814.)

Housework Hours			
Gender	Sample size	Mean	Standard deviation
Women	476	33.0	21.9
Men	496	19.9	14.6

- Based on this study, calculate how many more hours, on the average, women spend on housework than men.

- Find the standard error for comparing the means. What factor causes the standard error to be small compared to the sample standard deviations for the two groups?

- Calculate the 95% confidence interval comparing the population means for women and men. Interpret the result including the relevance of 0 being within the interval or not.

- State the assumptions upon which the interval in part c is based.

(d) Solve:

Female college student participation in athletics has increased dramatically over the past few decades. Sports medicine providers are aware of some unique health concerns of athletic women, including disordered eating. A study (M. Reinking and L. Alexander, Journal of Athletic Training, vol. 40, 2005, p. 47–51) compared disordered–eating symptoms and their causes for collegiate female athletes (in lean and nonlean sports) and nonathletes. The sample mean of the body dissatisfaction assessment score was 13.2 ($s = 8.0$) for 16 lean sport athletes (those sports that place value on leanness, including distance running, swimming, and gymnastics) and 7.3 ($s = 6.0$) for the 68 nonlean sport athletes. Assuming equal population standard deviations,

- Find the standard error for comparing the means.

- Construct a 95% confidence interval for the difference between the mean body dissatisfaction for lean sport athletes and nonlean sport athletes. Interpret.

(e) Solve:

Anna's project for her introductory statistics course was to compare the selling prices of textbooks at two Internet bookstores. She first took a random sample of 10 textbooks used that term in courses at her college, based on the list of texts compiled by the college bookstore. The prices of those textbooks at the two Internet sites were:

Site A:	115	79	43	140	99	30	80	99	119	69
Site B:	110	79	40	129	99	30	69	99	109	66

- Are these independent samples or dependent samples? Justify your answer.

- Find the mean for each sample. Find the mean of the difference scores. Compare, and interpret.

- Construct a 90% confidence interval comparing the population mean prices of all textbooks used that term at her college. Interpret.

(f) Solve:

For the data in the previous exercise,

- perform a significance test comparing the population mean prices. Show all steps of the test, and indicate whether you would conclude that the mean price is lower at one of the two Internet bookstores.

- perform a significance test comparing the population variance prices. Show all steps of the test, and indicate whether you would conclude that the variance is lower at one of the two Internet bookstores.