Solutions to Homework 5

Statistics 302 Professor Larget

Textbook Exercises

4.79 Divorce Opinions and Gender In Data 4.4 on page 227, we introduce the results of a May 2010 Gallup poll of 1029 US adults. When asked if they view divorce as “morally acceptable”, 71% of the men and 67% of the women in the sample responded yes. In the test for a difference in proportions, a randomization distribution gives a p-value of 0.165. Does this indicate a significant difference between men and women in how they view divorce?

Solution

If we use a 5% significance level, the p-value of 0.165 is not less than $\alpha = 0.05$ so we would not reject $H_0 : p_f = p_m$. This means the data do not show significant evidence of a difference in the proportions of men and women that view divorce as “morally acceptable”.

4.82 Sleep or Caffeine for Memory? The consumption of caffeine to benefit alternateness is a common activity practiced by 90% of adults in North America. Often caffeine is used in order to replace the need for sleep. One recent study compares students’ ability to recall memorized information after either the consumption of caffeine or a brief sleep. A random sample of 35 adults (between the ages of 18 and 39) were randomly divided into three groups and verbally given a list of 24 words to memorize. During a break, one of the groups takes a nap for an hour and a half, another group is kept awake and then given a caffeine pill an hour prior to testing, and the third group is given a placebo. The response variable of interest is the number of words participants are able to recall following the break. The summary statistics for the three groups are shown below in the table. We are interested in testing whether there is evidence of difference in average recall ability between any two of the treatments. Thus we have three possible tests between different pairs of groups: Sleep vs Caffeine, Sleep vs Placebo, and Caffeine vs Placebo.

Group	Sample Size	Mean	Standard Deviation
Sleep	12	15.25	3.3
Caffeine	12	12.25	3.5
Placebo	11	13.70	3.0

(a) In the test comparing the sleep group to the caffeine group, the p-value is 0.003. What is the conclusion of the test? In the sample, which group had better recall ability? According to the rest results, do you think sleep is really better than caffeine for recall ability?
(b) In the test comparing the sleep group to the placebo group, the p-value is 0.06. What is the conclusion of the test using a 5% significance level? Using a 10% significance level? How strong is the evidence of a difference in mean recall ability between these two treatments?
(c) In the test comparing the caffeine group to the placebo group, the p-value is 0.22. What is the conclusion of the test? In the sample, which group had better recall ability? According to the test results, would we be justified in concluding that caffeine impairs recall ability?
(d) According to this study, what should you do before an exam that asks you to recall information?

Solution

(a) The p-value (0.003) is small so the decision is to reject H_0 and conclude that the mean recall for sleep ($\bar{x}_s = 15.25$) is different from the mean recall for caffeine ($\bar{x}_c = 12.25$). Since the mean for the sleep group is higher than the mean for the caffeine group, we have sufficient evidence to conclude that mean recall after sleep is in fact better than after caffeine. Yes, sleep is really better for you than caffeine for enhancing recall ability.

(b) The p-value (0.06) is not less than 0.05 so we would not reject H_0 at a 5% level, but it is less than 0.10 so we would reject H_0 at a 10% level. There is some moderate evidence of a difference in mean recall ability between sleep and a placebo, but not very strong evidence.

(c) The p-value (0.22) is larger than any common significance level, so do not reject H_0. The placebo group had a better mean recall in this sample ($\bar{x}_p = 13.70$ compared to $\bar{x}_c = 12.25$), but there is not enough evidence to conclude that the mean for the population would be different for a placebo than the mean recall for caffeine.

(d) Get a good night's sleep!

4.86 Radiation from Cell Phones and Brain Activity Does heavy cell phone use affect brain activity? There is some concern about possible negative effects of radiofrequency signals delivered to the brain. In a randomized matched-pairs study, 47 healthy participants had cell phones placed on the left and right ears. Brain glucose metabolism (a measure of brain activity) was measured for all participants under two conditions: with one cell phone turned on for 50 minutes (the "on" condition) and with both cell phones off (the "off" condition). The amplitude of radio frequency waves emitted by the cell phones during the "on" condition was also measured.

(a) Is this an experiment or an observational study? Explain what it means to say that this was a "matched-pairs" study.
(b) How was randomization likely used in the study? Why did participants have cell phones on their ears during the "off" condition?
(c) The investigators were interested in seeing whether average brain glucose metabolism was different based on whether the cell phones were turned on or off. State the null and alternative hypotheses for this test.
(d) The p-value for the test in part (c) is 0.004. State the conclusion of this test in context.
(e) The investigators were also interested in seeing if brain glucose metabolism was significantly correlated with the amplitude of the radio frequency waves. What graph might we use to visualize this relationship?
(f) State the null and alternative hypotheses for the test in part (e).
(g) The article states that the p-value for the test in part (e) satisfies $p < 0.001$. State the conclusion of this test in context.

Solution
(a) This is an experiment since the explanatory factor (cell phone "on" or "off") was controlled. The design is matched pairs, since all 47 participants were tested under both conditions. For each participant, we find the difference in brain activity between the two conditions.

(b) Randomization in this case means that the order of the conditions ("on" and "off") was randomized for all the participants. Cell phones were on the ears for both conditions to control for any lurking variables and to make the treatments as similar as possible except for the variable of interest (the radiofrequency waves).

(c) Using μ_{on} to represent average brain glucose metabolism when the cell phones are on and μ_{off} to represent average brain glucose metabolism when the cell phones are off, the hypotheses are:

$$H_0 : \mu_{on} = \mu_{off}$$

$$H_a : \mu_{on} \neq \mu_{off}$$

Notice that since this is a matched pairs study, we could also write the hypotheses in terms of the average difference μ_D between the two conditions, with $H_0 : \mu_D = 0$ vs $H_a : \mu \neq 0$.

(d) Since the p-value is quite small (less than a significance level of 0.01), we reject the null hypothesis. There is significant evidence that brain activity is affected by cell phones.

(e) Both of these variables (brain glucose metabolism and amplitude of radiofrequency) are quantitative, so we use a scatterplot to graph the relationship.

(f) We are testing to see if the correlation ρ between these two variables is significantly different from zero, so the hypotheses are

$$H_0 : \rho = 0$$

$$H_a : \rho \neq 0$$

where ρ is the correlation between brain glucose metabolism and amplitude of radiofrequency.

(g) This p-value is very small so we reject H_0. There is strong evidence that brain activity is correlated with the amplitude of the radiofrequency waves emitted by the cell phone.

For 4.94 and 4.96, indicate whether it makes more sense to use a relatively large significance level (such as $\alpha = 0.10$) or a relatively small significance level (such as $\alpha = 0.01$).

4.94 Using your statistics class as a sample to see if there is evidence of a difference between male and female students in how many hours are spent watching television per week.

Solution

A Type I error (saying there's a difference in TV habits by gender for the class, when actually there isn't) is not very serious, so a large significance level such as $\alpha = 0.10$ will make it easier to see any difference.

4.96 Testing to see if a well-known company is lying in its advertising. If there is evidence that the company is lying, the Federal Trade Commission will file a lawsuit against them.

Solution

A Type I error (suing the company when they are not lying) is quite serious so it makes sense to use a small significance level such as $\alpha = 0.01$.

For 4.100 and 4.102, describe what it means in that context to make a Type I and Type II error. Personally, which do you feel is a worse error to make in the given situation?

4.100 The situation described in Exercise 4.94.

Solution

Type I error: Conclude there's a difference in TV habits by gender for the class, when actually there is no difference. Type II error: Find no significant difference in TV habits by gender, when actually there is a difference. Personal opinions will vary on which is worse.

4.102 The situation described in Exercise 4.96.

Solution

Type I error: Sue the company when they are not lying. Type II error: Let the company off the hook, when they are actually lying in their advertising. Personal opinions will vary on which is worse.

4.123 Paul the Octopus In the 2010 World Cup, Paul the Octopus (in a German aquarium) became famous for being correct in all eight of the predictions it made, including predicting Spain over Germany in a semifinal match. Before each game, two containers of food (mussels) were lowered into the octopus's tank. The containers were identical, expect for country flags of the opposing teams, one on each container. Whichever container Paul opened was deemed his predicted winner. Does Paul have psychic powers? In other words, is an 8-for-8 record significantly better than just guessing?

(a) State the null and alternative hypotheses.

(b) Simulate one point in the randomization distribution by flipping a coin eight times and counting the number of heads. Do this five times. Did you get any results as extreme as Paul the Octopus?

(c) Why is flipping a coin consistent with assuming the null hypothesis is true?

Solution

(a) The hypotheses are $H_0 : p = 0.5$ vs $H_a : p > 0.5$, where p is the proportion of all games Paul the Octopus picks correctly.

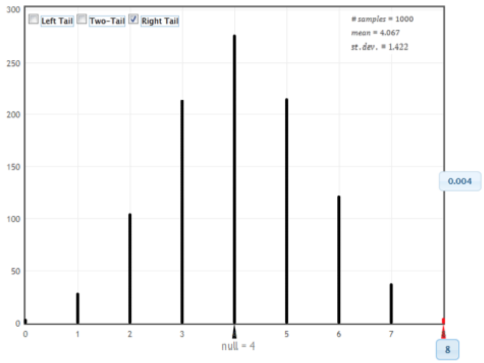
(b) Answers vary, but 8 out of 8 heads should rarely occur.

(c) The proportion of heads in flipping a coin is $p = 0.5$ which matches the null hypothesis.

4.124 How Unlikely Is Paul the Octopus's Success? For the Paul the Octopus data in Exercise 4.123, use *StatKey* or other technology to create a randomization distribution. Calculate a p-value. How unlikely is his success rate if Paul the Octopus is really not psychic?

Solution

We use technology to simulate many samples of size 8 from a population that has an equal number of "successes" and "failures", i.e. one where $p = 0.5$. For each sample we count the number of successes out of the 8 trials to obtain a randomization distribution such as the one shown below (or find the proportion of successes in each sample). We then count the number of samples for which all 8 trials are successes, and divide by the total number of samples to get a p-value. For the distribution below, only 4 of the 1000 samples gave 8 correct guesses in 8 trials, so we estimate the p-value = 0.004. Answers will vary for other randomizations but the p-value will always be small, indicating that it is very unlikely to predict all eight games correctly when just guessing at random.



4.126 Finger Tapping and Caffeine In Data 4.6 we look at finger-tapping rates to see if ingesting caffeine increases average tap rate. The sample data for the 20 subjects (10 randomly getting caffeine and 10 with no-caffeine) are given in Table 4.4 on page 241. To create a randomization distribution for this test, we assume the null hypothesis $\mu_c = \mu_n$ is true, that is, there is no difference in average tap rate between the caffeine and no-caffeine groups.

(a) Create one randomization sample by randomly separating the 20 data values into two groups. (One way to do this is to write the 20 tap rate values on index cards, shuffle, and deal them into two groups of 10.)

(b) Find the sample mean of each group and calculate the difference $\bar{x}_c - \bar{x}_n$, in the simulated sample means.

(c) The difference in sample means found in part (b) is one data point in a randomization distribution. Make a rough sketch of the randomization distribution shown in Figure 4.11 on page 242 and locate your randomization statistic on the sketch.

Solution

(a) Answers vary. For example, one possible randomization sample is shown below.

caffeine	244	250	248	246	248	245	246	247	248	246	mean = 246.8
no caffeine	250	244	252	248	242	250	242	245	242	248	mean = 246.3

(b) Answers vary. For the randomization sample above, $\bar{x}_c - \bar{x}_{nc} = 246.8 - 246.3 = 0.5$.

(c) The sample difference of 0.5 for the randomization above would fall a bit to the right of the center of the randomization distribution.

4.130 Effect of Sleep and Caffeine on Memory Exercise 4.82 on page 261 describes a study in which a sample of 24 adult are randomly divided equally into two groups and given a list of 24 words to memorize. During a break, one group takes a 90-minute nap while another group is given a caffeine pill. The response variable of interest is the number of words participants are able to recall following the break. We are testing to see if there is a difference in the average number of words a person can recall depending on whether the person slept or ingested caffeine. The data are shown in the table below and are available in **SleepCaffeine**.

Sleep	14	18	11	13	18	17	21	9	16	17	14	15	Mean=15.25
Caffeine	12	12	14	13	6	18	14	16	10	7	15	10	Mean=12.25

(a) Define any relevant parameter(s) and state the null and alternative hypotheses.
(b) What assumption do we make in creating the randomization distribution?
(c) What statistic will we record for each of the simulated samples to create the randomization distribution? What is the value of the statistic for the observed sample? sample means
(d) Where will the randomization distribution be centered?
(e) Find one point on the randomization distribution by randomly dividing the 24 data values into two groups. Describe how you divide the data into two groups and show the values in each group for the simulated sample. Compute the sample mean in each group and compute the difference in the sample means for this simulated result.
(f) Use *StatKey* or other technology to create a randomization distribution. Estimate the p-value for the observed difference in means given in part (c).
(g) At a significance level of $\alpha = 0.01$, what is the conclusion of the test? Interpret the results in context.

Solution

(a) The hypotheses are $H_0 : \mu_s = \mu_c$ vs $H_a : \mu_s \neq \mu_c$, where μ_s and μ_c are the mean number of words recalled after sleep and caffeine, respectively.

(b) The number of words recalled would be the same regardless of whether the subject was put in the sleep or the caffeine group.

(c) The sample statistic is $\bar{x}_s - \bar{x}_c$. For the original sample the value is $15.25 - 12.25 = 3.0$.

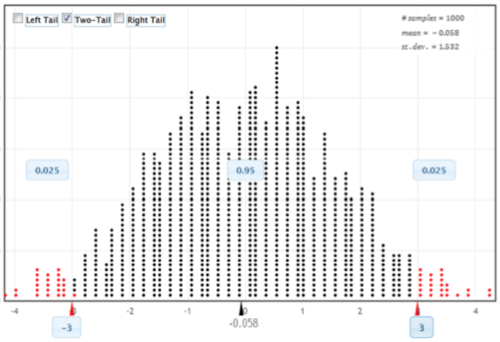
(d) Under H_0 we have $\mu_s - \mu_c = 0$ so the randomization distribution should be centered at zero.

(e) We randomly divide the 24 sample word recall values into two groups of 12 (one for "sleep" group, the other for "caffeine") and find the difference in sample means. One such randomization is shown below where $x_s - x_c = 14.75 - 12.75 = 2.0$. Answers vary.

sleep	11	14	15	17	17	18	6	12	18	13	21	15	mean = 14.75
caffeine	16	14	10	13	7	10	14	18	9	14	12	16	mean = 12.75

ne razbiram prasanjeto

(f) A randomization distribution for 1000 differences in means is shown below. Since this is a two-tailed test so we double the count in one tail (25 out of 1000 values at or beyond $x_s - x_c = 3.0$) in order to account for both tails. We see that the p-value is $= 2 \times 0.025 = 0.05$.



(g) The p-value is more than $\alpha = 0.01$ so we do not reject H_0. There is not sufficient evidence (at a 1% level) to show a difference in mean number of words recalled after taking a nap or ingesting caffeine.

4.132 Does Massage Help Heal Muscles Strained by Exercise? After exercise, massage is often used to relieve pain, and a recent study shows that it also may relieve inflammation and help muscles heal. In the study, 11 male participants who had just strenuously exercised had 10 minutes of massage on one quadricep and no treatment on the other, with treatment randomly assigned. After 2.5 hours, muscle biopsies were taken and production of the inflammatory cytokine interleukin-6 was measured relate to the resting level. The differences (control minus massage) are given in the table below.

(a) Is this an experiment or an observational study? Why is it not double blind?

(b) What is the sample mean difference in inflammation between no massage and massage?

(c) We want to test to see if the population mean difference μ_D is greater than zero, meaning muscle with no treatment has more inflammation than muscle that has been massaged. State the null and alternative hypotheses.

(d) Use *StatKey* or other technology to find the p-value from a randomization distribution.

(e) Are the results significant at a 5% level? At a 1% level? State the conclusion of the test if we assume a 5% significance level (as the authors of the study did).

0.6	4.7	3.8	0.4	1.5	-1.2	2.8	-0.4	1.4	3.5	-2.8

Solution

(a) This is an experiment since a treatment was actively manipulated. In fact, it is a matched pairs experiment. The experiment cannot be blind to the subject since s/he will know which muscle is being massaged. However the person measuring the level of inflammation should not know which is which.

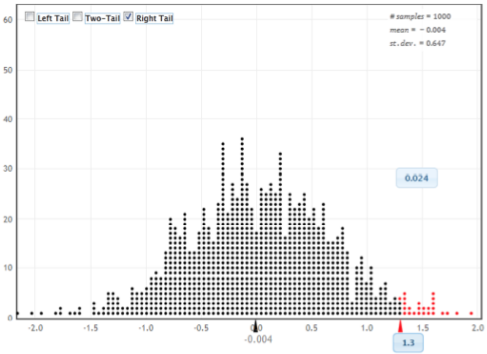
(b) The mean difference is $\bar{x}_D = 1.30$.

(c) We define μ_D to be the mean difference in inflammation in muscle between a muscle that has not been massaged and a muscle that has (using control level minus massage level). The null hypothesis is no difference from a massage and the alternative is that levels are lower in a muscle that has been massaged. The hypotheses are:

$$H_0 : \mu_D = 0$$

$$H_a : \mu_D > 0$$

(d) Using StatKey or other technology, we create a randomization distribution of mean differences under the null hypothesis that $\mu_D = 0$. One such distribution is shown below. Since the alternative hypothesis is $H_a : \mu_D > 0$, this is a right-tail test. Using our sample mean $\bar{x}_D = 1.30$, we see that 24 of the 1000 simulated samples were as extreme so the p-value is 0.024.



(e) Based on the randomization distribution and p-value in (d), the results are significant at a 5% level but not at a 1% level. Using a 5% level, we reject H_0 and find evidence that massage does reduce inflammation in muscles after exercise.

4.141 Quiz vs Lecture Pulse Rate Do you think that students undergo physiological changes when in potentially stressful situations such as taking a quiz or exam? A sample of statistics students were interrupted in the middle of quiz and asked to record their pulse rates (beats for 1-minute period). Ten of the students had also measured their pulse rate while siting in class listening to a lecture, and these values were matched with their quiz pulse rates. The data appear in the table below and are stored in **QuizPulse10**. Note that this is paired data since we have two values, a quiz and a lecture pulse rate, for each student in the sample. The question of interest is whether quiz pulse rates tend to be higher, on average, then lecture pulse rates. (Hint: Since this is paired data, we work with the differences in pulse rate for each student between quiz and lecture. If the difference are D = quiz pulse rate-lecture pulse rate, the question of interest is whether mu_D is greater than 0.)

(a) Define the parameter(s) of interest and state the null and alternative hypotheses.

(b) Determine an appropriate statistic to measure and compute its value for the original sample.

(c) Describe a method to generate randomization samples that is consistent with the null hypothesis and reflects the paired nature of the data There are several viable methods. You might use shuffled index cards, a coin, or some other randomization procedure.

(d) Carry out your procedure to generate one randomization sample and compute the statistic you chose in part (b) for this sample.

(e) Is the statistic for your randomization sample more extreme (in the direction of the alternative) than the original sample?

Student	1	2	3	4	5	6	7	8	9	10
Quiz	75	52	52	80	56	90	76	71	70	66
Lecture	73	53	47	88	55	70	61	75	61	78

Solution

(a) We are interested in whether pulse rates are higher on average than lecture pulse rates, so our hypotheses are

$$H_0 : \mu_Q = \mu_L$$

$$H_a : \mu_Q > \mu_L$$

where μ_Q represents the mean pulse rate of students taking a quiz in a statistics class and μ_L represents the mean pulse rate of students sitting in a lecture in a statistics class. We could also word hypotheses in terms of the mean difference $D = Lecturepulse - Quiz$ pulse in which case the hypotheses would be $H_0 : \mu_D = 0$ vs $H_a : \mu_D > 0$.

(b) We are interested in the difference between the two pulse rates, so an appropriate statistic is $\bar{x}_D$, the differences ($D = Lecture - Quiz$) for the sample. For the original sample the differences are:

$$+2, -1, +5, -8, +1, +20, +15, -4, +9, -12$$

and the mean of the differences is $\bar{x}_D = 2.7$.

(c) Since the data were collected as pairs, our method of randomization needs to keep the data in pairs, so the first person is labeled with (75, 73), with a difference in pulse rate of 2. As long as we keep the data in pairs, there are many ways to conduct the randomization. In every case, of course, we need to make sure the null hypothesis (no difference) is met and we need to keep the data in pairs.

how to randomise

One way to do this is to sample from the pairs with replacement, but randomly determine the order of the pair (perhaps by flipping a coin), so that the first pair might be (75,73) with a difference of 2 or might be (73,75) with a difference of -2. Notice that we match the null hypothesis that the quiz/lecture situation has no effect by assuming that it doesn't matter - the two values could have come from either situation. Proceeding this way, we collect 10 differences with randomly assigned signs (positive/negative) and compute the average of these differences. That gives us one simulated statistic.

A second possible method is to focus exclusively on the differences as a single sample. Since the randomization distribution needs to assume the null hypothesis, that the mean difference is 0, we can subtract 2.7 (the mean of the original sample of differences) from each of the 10 differences, giving the values

$$-0.7, -3.7, 2.3, -10.7, -1.7, 17.3, 12.3, -6.7, 6.3, -14.7.$$

Notice that these values have a mean of zero, as required by the null hypothesis. We then select samples of size ten (with replacement) from the adjusted set of differences (perhaps by putting the 10 values on cards or using technology) and compute the average difference for each sample.

There are other possible methods, but be sure to use the paired data values and be sure to force the null hypothesis to be true in the method you create!

(d) Here is one sample if we randomly assign +/- signs to each difference:

$$+2, +1, -5, -8, +1, -20, -15, +4, +9, +12 \Rightarrow \bar{x}_D = -1.9$$

Here is one sample drawn with replacement after shifting the differences.

$$-6.7, -1.7, 6.3, 2.3, -3.7, -1.7, -0.7, 17.3, 2.3, -0.7 \Rightarrow \bar{x}_D = 1.3$$

(e) Neither of the statistics for the randomization samples in (d) exceed the value of $\bar{x}_D = 2.7$ from the original sample, but your answers will vary for other randomizations.

4.144 Exercise Hours Introductory statistics students fill out a survey on the first day of class. One of the questions asked is "How many hours of exercise do you typically get each week?" Responses for a sample of 50 students are introduced in Example 3.25 on page 207 and stored in the file **Exercise Hours**. The summary statistics are shown in the computer output. The mean hours of exercise for the combined sample of 50 students is 10.6 hours per week and the standard deviation is 8.04. We are interested in whether these sample data provide evidence that the mean number of hours of exercise per week is difference between male and female statistics students.

Variable	Gender	N	Mean	StDev	Minimum	Maximum
Exercise	F	30	9.40	7.41	0.00	34.00
	M	20	12.40	8.80	2.00	30.00

Discuss whether or not the methods described below would be appropriate ways to generate randomization samples that are consistent with $H_0 : \mu_f = \mu_m$ vs $H_A : \mu_f \neq \mu_m$. Explain your reasoning in each case.

(a) Randomly label 30 of the actual exercise values with "F" for the female group and the remaining 20 exercise value with "M" for the males. Compute the difference in the sample means $\bar{x}_f - \bar{x}_m$.

(b) Add 1.2 to every female exercise value to give a new mean of 10.6 and subtract 1.8 from each male exercise value to move their mean to 10.6 (and match the females). Sample 30 values (with replacement) from the shifted female values and 20 values (with replacement) from the shifted male values. Compute the difference in the sample means, $\bar{x}_f - \bar{x}_m$.

(c) Combine all 50 sample values into one set of data having a mean amount of 10.6 hours. Select 30 values (with replacement) to represent a sample of female exercise hours and 20 values (also with replacement) for a sample of male exercise values. Compute the difference in the sample means, $\bar{x}_f - \bar{x}_m$.

Solution

(a) This method is appropriate. Under H_0 the exercise amounts should be unrelated to the genders, so each exercise value could have just as likely come from either gender.

(b) This method is appropriate. Adjusting the values so that the male and female exercise means are the same produces a "population" to sample from that agrees with the null hypothesis of equal means. Note that adding 3.0 to all of the female exercise values or subtracting 3.0 from all the male exercise values would also accomplish this goal. Adjusting both, as suggested in part (b) has the advantage of keeping the combined mean at 10.6. Any of these methods give equivalent results.

(c) This method is appropriate. As in part(b) the randomizations samples are taken from a "population" where the mean exercise levels are the same for males and females, so the distribution of $\bar{x}_F - \bar{x}_M$ should be centered around zero as specified by $H_0 : \mu_F = \mu_M$.

4.158 Testing for a Home Field Advantage in Soccer In Exercise 3.108 on page 215, we see that the home team was victorious in 70 games out of a sample of 120 games in the FA premier league, a football (soccer) league in Great Britain. We wish to investigate the proportion p of all games wine by the home team in this league.

(a) Use *StatKey* or other technology to find and interpret a 90% confidence interval for the proportion of games won by the home team.

(b) State the null and alternative hypotheses for a test to see if there is evidence that the proportion is different from 0.5.

(c) Use the confidence interval from part (a) to make a conclusion in the test from part (b). State the confidence level used.

(d) Use *StatKey* or other technology to create a randomization distribution and find the p-value for the test in part (b).

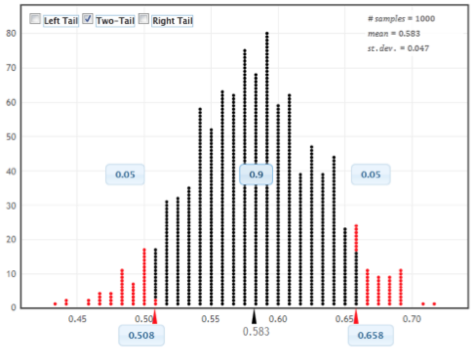
(e) Clearly interpret the results of the test using the p-value and using a 10% significance level. Does your answer match your answer from part (c)?

(f) What information does the confidence interval give that the p-value doesn't? What information does the p-value give that the confidence interval doesn't?

(g) What's the main different between the bootstrap distribution of part (a) and the randomization distribution of part (d)?

<u>Solution</u>

(a) The proportion of home wins in the sample is $\hat{p} = 70/120 = 0.583$. Using StatKey or other technology, we construct a bootstrap distribution of sample proportions such as the one below. We see that a 90% confidence interval in this case goes from 0.508 to 0.658. We are 90% confident that the home team will win between 50.8% and 65.8% of soccer games in this league.

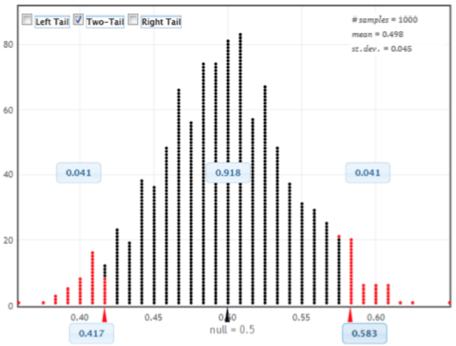


(b) To test if the proportion of home wins differs from 0.5, we use $H_0 : p = 0.5$ vs $H_a : p \neq 0.5$.

(c) Since 0.5 is not in the interval in part (a), we reject H_0 at the 10% level. The proportion of home team wins is not 0.5, at a 10% level.

(d) We create a randomization distribution (shown below) of sample proportions when $n = 120$

using $p = 0.5$. Since this is a two-tailed test, we have p-value = 2(0.041) = 0.082.



(e) At a 10% significance level, we reject H_0 and find that the proportion of home team wins is different from 0.5. Yes, this does match what we found in part (c).

(f) The confidence interval shows an estimate for the proportion of times the home team wins, which the p-value does not give us. The p-value gives a sense for how strong the evidence is that the proportion of home wins differs from 0.5 (only significant at a 10% level in this case).

(g) The bootstrap and randomizations distributions are similar, except that the bootstrap proportions are centered at the original $\hat{p} = 0.583$, while the randomization proportions are centered at the null hypothesis, $p = 0.5$.

4.160 How Long Do Mammals Live? Data 2.2 on page 61 includes information on longevity (typical lifespan), in years, for 40 species of mammals.

(a) Use the data, available in **MammalLongevity**, and *StatKey* or other technology to test to see if the average lifespan of mammal species is difference from 10 years. Include all details of the test: the hypotheses, the p-value, and the conclusion in context.

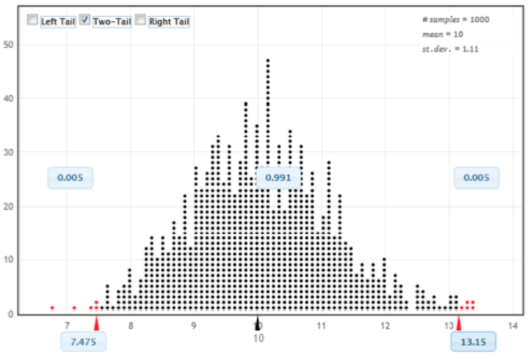
(b) Use the result of the test to determine whether $\mu = 10$ would be included as a plausible value in a 95% confidence interval of average mammal lifespan. Explain.

Solution

(a) We are testing $H_0 : \mu = 10$ vs $H_a : \mu \neq 10$, where μ represents the mean longevity, in years, of all mammal species. The mean longevity in the sample is $\bar{x} = 13.15$ years. We use StatKey or other technology to create a randomization distribution such as the one shown below. For this two-tailed test, the proportion of samples beyond $\bar{x} = 13.15$ in this distribution gives a p-value of 20.005 = 0.010. We have strong evidence that mean longevity of mammal species is different from 10 years.

(b) Since, for a 5% significance level, we reject $\mu = 10$ as a plausible value of the population mean in the hypothesis test in part (a), 10 would not be included in a 95% confidence interval for the mean longevity of mammals.

4.167 Does Massage Really Help Reduce Inflammation in Muscles? In Exercise 4.132 on page 279, we learn that massage helps reduce levels of the inflammatory cytosine interleukin-6 in muscles when muscle tissue is tested 2.5 hours after massage. The results were significant at the 5% level. However, the authors of this study actually performed 42 different tests. They tested



for significance with 21 different compounds in muscles and at two different times (right after the massage and 2.5 hours after).

(a) Given the new information, should we have less confidence in the one results described in the earlier exercise? Why?

(b) Sixteen of the tests done by the authors involved measuring the effect of massage on muscle metabolites. None of these tests were significant. Do you think massage affects muscle metabolites?

(c) Eight of the tests done by the authors (including the one described in the earlier exercise) involving measuring the effects of massage on inflammation in the muscle. Four of these tests were significant. Do you think it is safe to conclude that massage really does reduce inflammation?

Solution (a) We should definitely be less confident. If the authors conducted 42 tests, it is likely that some of them will show significance just by random chance even if massage does not have any significant effects. It is possible that the result reported earlier just happened to one of the random significant ones.

(b) Since none of the tests were significant, it seems unlikely that massage affects muscle metabolites.

(c) Now that we know that only eight tests were testing inflammation, and that four of those gave significant results, we can be quite confident that massage does reduce muscle inflammation after exercise. It would be very surprising to see four p-values (out of eight) less than 5% if there really were no effects at all.

Computer Exercises

For each R problem, turn in answers to questions with the written portion of the homework. Send the R code for the problem to Katherine Goode. The answers to questions in the written part should be well written, clear, and organized. The R code should be commented and well formatted.

R problem 1

1. Write a function in R that will compute a p-value by a randomization test for a single population proportion. You may use the following template to begin, but will need to add missing code to make the function work.

```
function for p-value from a proportion
# n is sample size
# x is number of successes in sample
# p0 is null success probability
# R is the number of samples to generate
pvalue.p = function(n,x,p0,R,alternative=c("not.equal","less","greater")) {
alternative = match.args(alternative)
p.hat = numeric(R)
for ( i in 1:R ) {
p.hat[i] = mean(sample(c(0,1),size=n,replace=TRUE,probs=c(1-p0,p0)))
}
if ( alternative == "not.equal" ) {
## do something
}
else if ( alternative == "less" ) {
## do something
}
else if ( alternative == "greater" ) {
## do something
}
## return something
}
```

Solution
The function which computes a p-value by a randomization test for a single population proportion is included below.

```
# n is sample size
# x is number of successes in sample
# p0 is null success probability
# R is the number of samples to generate
pvalue.p = function(n,x,p0,R,alternative=c("not.equal","less","greater")) {
  # This is fancy code that will set the alternative to one of these three possibilities
  #   if the passed argument matches the beginning of ay alternative.
  alternative = match.arg(alternative)
  # Create an empty array of size R to store the computed sample proportions.
  p.hat = numeric(R)
  # Do the sampling R times.
  # Here the first argument is the population to sample from, the two numbers 0 and 1.
  # The argument probs=... places probability 1-p0 onto 0 and p0 onto 1
  for ( i in 1:R ) {
    p.hat[i] = mean(sample(c(0,1),size=n,replace=TRUE,prob=c(1-p0,p0)))
  }
  # Compute the sample phat
  p.sample = x/n

  if ( alternative == "not.equal" ) {
  # Need to add up left and right tail probabilities.
  # This code will just find the tail area on the side that p.sample falls
  # and double it.
    if ( p.sample == p0 ) {
      p.value = 1
    }
    else if ( p.sample < p0 ) {
      p.value = 2*sum( p.hat <= p.sample ) / R
    }
    else if ( p.sample > p0 ) {
      p.value = 2*sum( p.hat >= p.sample ) / R
    }
  }
  else if ( alternative == "less" ) {
    p.value = sum( p.hat <= p.sample ) / R
  }
  else if ( alternative == "greater" ) {
    p.value = sum( p.hat >= p.sample ) / R
  }
  return( p.value )
}
```

2. Use the function to compute p-values in the following situations.

(a) Katherine's first ten tries at the ESP test: $n = 10, x = 6, H_0 : p = 0.2, H_a : p > 0.2$.

Solution
Using our function, we find that the p-value is 0.0077.

```
R Code
print( pvalue.p(n=10,x=6,p0=0.2,R=10000,alternative="greater") )
```

(b) Paul the Octopus for all Euro 2008 and World Cup 2010 matches: n = 13, x = 11, $H_0 : p = 0.5$, $H_a : p > 0.5$.

Solution
Using our function, we find the the p-value is 0.01.

```
R Code
print( pvalue.p( n=13, x=11, p0=0.5, R=10000, alternative="greater") )
```

(c) A genetics hypothesis where if true, $p = 11/16 : n = 230, x = 161, H_0 : p = 11/16, H_a : p \neq 11/16$.

Solution
Using our function, we find the the p-value is 0.7386.

```
R Code
print( pvalue.p( n=230, x=161, p0=11/16, R=10000, alternative="not.equal") )
```

R problem 2 (a) Write a function that returns a p-value using a randomization distribution for testing a difference in two population means. Do this by repeatedly assigning individuals to the two groups at random. (This is also called a permutation test.) Here is a very brief template for the function.

```
# function for p-value from differences in sample means
# x is the first sample
# y is the second sample
# R is the number of samples to generate
pvalue.mudiff = function(x,y,R,alternative=c("not.equal","less","greater")) {
alternative = match.args(alternative)
stat = numeric(R)
total = c(x,y)
nx = length(x)
ny = length(y)
for ( i in 1:R ) {
newTotal = sample(total)
stat[i] = mean(newTotal[1:nx]) - mean(newTotal[(nx+1):(nx+ny)])
}
if ( alternative == "not.equal" ) {
## do something
}
else if ( alternative == "less" ) {
## do something
}
else if ( alternative == "greater" ) {
## do something
}
## return something
}
```

Solution

Below is the code for the function in r, which returns a p-value using a randomization distribution for testing a difference in two population means.

```
# function for p-value from differences in sample means
# x is the first sample
# y is the second sample
# R is the number of samples to generate
pvalue.mudiff = function(x,y,R,alternative=c("not.equal","less","greater")) {
  alternative = match.arg(alternative)
  stat = numeric(R)
  total = c(x,y)
  nx = length(x)
  ny = length(y)
  for ( i in 1:R ) {
    newTotal = sample(total)
    stat[i] = mean(newTotal[1:nx]) - mean(newTotal[(nx+1):(nx+ny)])
  }
  d = mean(x) - mean(y)
  if ( alternative == "not.equal" ) {
    if ( d == 0 ) {
      p.value = 1
    }
    else if ( d < 0 ) {
      p.value = 2 * sum( stat <= d ) / R
    }
    else if ( d > 0 ) {
      p.value = 2 * sum( stat >= d ) / R
    }
  }
  else if ( alternative == "less" ) {
    p.value = sum( stat <= d ) / R
  }
  else if ( alternative == "greater" ) {
    p.value = sum( stat >= d ) / R
  }
  return ( p.value )
}
```

(b) Use the function to find a p-value for the data in Table 4.11 on page 278.

Solution

Using our function, we find that the p-value for the data in Table 4.11 on page 278 is 4e-05.

```
R Code
ll = c(10,10,11,9,12,9,11,9,17)
ld = c(5,6,7,8,3,8,6,6,4)

print("Problem 2")
print( pvalue.mudiff(ll,ld,R=100000) )
```