STAT 240 Homework 2

Rebecca Barter, Andrew Do, and Kellie Ottoboni

March 4, 2015

Chapter 26

Review exercise 2

Part (a)

The null hypothesis is that in a box of 38 tickets, there are 18 that are red. The alternative is that in this box of 38 tickets, there are more than 18 red tickets.

Part (b)

The null says that the percentage of reds in the box is $\frac{18}{38} \approx 48\%$. The alternative says that the percentage of reds in the box is greater than 48%.

Part (c)

$$z = \frac{1890 - (\frac{18}{38})(3800)}{\sqrt{(\frac{18}{38})(\frac{20}{38})(3800)}} = \frac{1890 - 1800}{\sqrt{\frac{36000}{38}}} \approx 2.924$$

Under the null hypothesis, z has approximately a standard normal distribution. Thus the p-value for the one-sided test is $P = P(z \geq 2.924) \approx 0.0017$.

Part (d)

At the significance level 0.05, we reject the null hypothesis that there were 18 red tickets in the box. In particular, this means that there were too many reds in the 3800 roulette spins to be due to chance alone.

Review exercise 5

Part (a)

Consider a box containing 3000 tickets, from which we take a random sample of 100 without replacement. The null hypothesis is that the average of the box is 7.5 and the alternative is that the average is something other than 7.5.

Part (b)

The null hypothesis is that the average of the box is 7.5 and the alternative is that the average is something other than 7.5.

Part (c)

$$T = \frac{7.5 - 6.6}{\frac{9}{\sqrt{100}}\sqrt{\frac{100}{99}}} = \frac{0.9}{0.904534} \approx 0.9950$$

Under the null hypothesis, T has a t distribution with 99 degrees of freedom. Thus the p-value for the two-sided test is $P(t_{99} \geq T) \approx 0.16$. We fail to reject the null hypothesis that the box average is different from 7.5, at the 0.05 significance level.

Review exercise 12

(Hard.) Does the psychological environment affect the anatomy of the brain? This question was studied experimentally by Mark Rosenzweig and his associates. The subjects for the study came from a genetically pure strain of rats. From each litter, one rat was selected at random for the treatment group, and one for the control group. Both groups got exactly the same kind of food and drink—as much as they wanted. But each animal in the treatment group lived with 11 others in a large cage, furnished with playthings which were changed daily. Animals in the control group lived in isolation, with no toys. After a month, the experimental animals were killed and dissected.

On the average, the control animals were heavier and had heavier brains, perhaps because they ate more and got less exercise. However, the treatment group had consistently heavier cortexes (the "grey matter," or thinking part of the brain). This experiment was repeated many times; results from the first 5 trials are shown in the table: "T" means the treatment, and "C" is for control. Each line refers to one pair of animals. In the first pair, the animal in treatment had a cortex weighing 689 milligrams; the one in control had a lighter cortex weighing only 657 milligrams. And so on.

Two methods of analyzing the data will be presented in the form of exercises. Both methods take into account the pairing, which is a crucial feature of the data. (The pairing comes from randomization within litter.)

Part (a)

First Analysis. How many pairs were there in all? In how many of these pairs did the treatment animal have a heavier cortex? Suppose treatment had no effect, so each animal of the pair had a 50-50 chance to have the heavier cortex, independently from pair to pair. Under this assumption, how likely is it that an investigator would get as many pairs as Rosenzweig did, or more, with the treatment animal having the heavier cortex? What do you infer?

Part (b)

Second Analysis. For each pair of animals, compute the difference in cortex weights "treatment - control." Find the average and SD of all these differences. The null hypothesis says that these differences are like draws made at random with replacement from a box whose average is 0—the treatment has no effect. Conduct an appropriate hypothesis test. What do you infer?

Part (c)

To ensure the validity of the analysis, the following precaution was taken. "The brain dissection and analysis of each set of littermates was done in immediate succession but in a random order and identified only by code number so that the person doing the dissection does not know which cage the rat comes from." Comment briefly on the following: What was the point of this precaution? Was it a good idea?

Chapter 27

Review exercise 8

Part (a)

The data tell us nothing about the research questions (i) and (ii); the two proportions measure different things. The fraction 22 of 46 is the proportion of people who *predicted* that they would help whereas 2 of 46 is the proportion of people who actually agreed to help. A z-test is not applicable here.

Part (b)

This data can be used to answer question (i), can people predict how well they will behave? However, a two-sample z-test is inappropriate because the groups under comparison are comprised of the same individuals. This introduces correlation between the samples so the standard error of the difference in proportions cannot be computed with the usual formula. We should use a paired z-test instead of a two-sample z-test.

Part (c)

The proportion of volunteers in each group can be used to answer question (ii), do predicted responses influence behavior? We would conduct a two-sample z-test to compare the proportions.

The null hypothesis is that the rate of volunteers among those who are asked to predict their response is the same as the rate of volunteers among those who are not asked. The alternative hypothesis is that the rates are different. The observed proportion in the "prediction-request" group is $22/46$ and the observed proportion in the "request-only" group is $14/46$. The test statistic is

$$z = \frac{\frac{22}{46} - \frac{14}{46}}{\sqrt{\frac{22 \times 24}{46^2}\frac{1}{46} + \frac{14 \times 32}{46^2}\frac{1}{46}}} = \frac{8}{\sqrt{\frac{976}{46}}} \approx 1.7368$$

Under the null, z is approximately standard normally distributed, so the p-value for the two-sided test is

$$P(|Z| \geq |z|) = 2P(Z \geq 1.7368) \approx 0.0824$$

At the 0.05 level, we fail to reject the null hypothesis that the rates are the same.

Review exercise 9

The null hypothesis is that the rate of rejections is the same for the paper with positive results and the paper with negative results. The alternative hypothesis is that the rate is higher for the paper with negative results. The observed proportion of rejections for the positive paper is $28/53$ and the observed proportion of rejections for the negative paper is $8/54$. We will conduct a one-sided, two-sample z-test for the difference in proportions. The test statistic is

$$z = \frac{\frac{28}{53} - \frac{8}{54}}{\sqrt{\frac{28 \times 25}{53^2}\frac{1}{53} + \frac{8 \times 46}{54^2}\frac{1}{54}}} \approx \frac{0.5283 - 0.1481}{\sqrt{\frac{0.249199}{53} + \frac{0.1262}{54}}} = 4.5317$$

The p-value for this test is

$$P(Z \geq 4.5317) = 2.9255 \times 10^{-6}$$

There is strong evidence to reject the null hypothesis that the rejection rates are the same in favor of the alternative hypothesis that the rate of rejections is higher for the paper with negative results.

Review exercise 10

There is a natural pairing in this experiment: children from the same household are the same with respect to a variety of potential confounders for IQ, including parents' socioeconomic status, parents' education, level of parent involvement, etc. Thus, siblings in this study are highly correlated. It therefore doesn't make sense to use a test which assumes the sample of first-borns, $X_1, \ldots, X_{400}$ is independent of the sample of second-borns, $Y_1, \ldots, Y_{400}$.

In reality, the correlation between the IQ of the first-borns and the IQ of the second-borns is positive, so

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}\bar{X} + \mathrm{Var}\bar{Y} - 2\mathrm{Cov}(\bar{X}, \bar{Y})$$
$$< \mathrm{Var}\bar{X} + \mathrm{Var}\bar{Y}$$

Therefore the estimated standard error, given by $\sqrt{0.5^2 + 0.5^2}$ overestimates the true standard error of the difference in means. Consequently, the test statistic is biased towards 0. The z-test here is inappropriate.

If we had accounted for the covariance between $\bar{X}$ and $\bar{Y}$, the estimated standard error would have been smaller, thus the test statistic z would have been larger. In other words, using a better estimator of the standard error, accounting for the correlation between siblings in the sample, would have given more power to detect a difference in sample means.

Chapter 29

Exercise set C 6

Part (a)

The two-sample z-test is legitimate here. We have two randomly sampled groups, drawn without replacement from a population of reviewers. The randomization justifies the test.

Part (b)

We aren't given the sample sizes, but we can recover the standard error using the z-score. The one-sided p-value is approximately 0.02, so the z-score is $z = \Phi^{-1}(1 - 0.02) \approx 2.0537$. Then

$$z = \frac{0.715 - 0.25}{\sqrt{\mathrm{Var}(\bar{X} - \bar{Y})}} = 2.0537$$

Solving gives $\mathrm{SE}(\bar{X} - \bar{Y}) = 0.2264$.

Part (c)

Yes, the difference between 71.5% and 25% is substantial. Just from a qualitative perspective, this means the rate of rejection with negative results is nearly three times higher than if one had reported positive results.

Part (d)

In this case, the z-test does not add a great deal of information. It confirms our belief that there is a significant qualitative difference in the rates and gives evidence that the difference is not simply due to chance sampling variability.

Part (e)

The data suggest that the review process is biased to favor positive results. Since reviewers were selected at random to receive either the paper with positive findings or the paper with negative findings, we expect that the two groups are roughly the same with respect to all other variables. Thus, the difference in acceptance rates can be attributed to the difference in the paper's results.

Exercise set D 3

It is inappropriate to do a statistical test to compare infant mortality rates in this scenario. There is no randomness in the data. Babies weren't randomized to be born in different seasons; their birth simply happened when it happened and we observed the date. Furthermore, we have data for every baby that was born, so there is no sampling variability. Since we observe the entire population of interest and the "treatment" (season of birth) is fixed by Nature, applying a probability model makes no sense.

Review exercise 9

Investigators are studying the relationship between income and education, for women age 25-54 who are working.

Part (a)

Investigator A computes the correlation between income and education for all these women. Investigator B computes the correlation only for women who have professional, technical, or managerial jobs. Who gets the higher correlation? Or should the correlations be about the same? Explain.

Part (b)

Investigator C computes the correlation between income and education for all the women. Investigator D looks at each state separately, computes the average income and average education for that state—and then computes the correlation coefficient for the 50 pairs of state averages. Which investigator gets the higher correlation? Or should the correlations be about the same? Explain.

Review exercise 11

A university made a study of all students who completed the first two years of undergraduate work. The average first-year GPA was 3.1, and the SD was 0.4. The correlation between first-year and second-year GPA was 0.4. The scatter diagram was football shaped.
Sally Davis was in the study. She had a first-year GPA of 3.5, and her second-year GPA was just about average—among those who had a first-year GPA of 3.5. What was her percentile rank on the second-year GPA, relative to all the students in the study? If this cannot be determined from the information given, say what else you need to know and why.

Special exercise 33

In the U.S., there are two sources of national statistics on crime rates:

  i The FBI's Uniform Crime Reporting Program, which publishes summaries on all crimes reported to police agencies in jurisdictions covering virtually 100% of the population.

  ii The National Crime Survey, based on interviews with a nationwide probability sample of households.

In 2001, 3% of the households in the sample told the interviewers they had experienced at least one burglary within the past 12 months. The same year, the FBI reported a burglary rate of 20 per 1,000 households, or 2%. Can this difference be explained as chance error? If not, how would you explain it? You may assume that the Survey is based on a simple random sample of 50,000 households out of 100 million households. We are interested in the probability that of the $N = 50,000$ interviewed, $1,500$ (3% of $50,000$) experienced burglary when the finite population of $M = 100,000,000$ has $K = 2,000,000$ cases on file. Let $F_{M,K,N}$ be the hypergeometric distribution function with parameters M, K, N as defined above, then the probability p of observing a burglary rate of 3% or higher in a sample size of $50,000$ is given by:

$$
\begin{aligned}
p &= P(\text{number of households burglarized} \geq 1500) \\
&= 1 - P(\text{number of households burglarized} < 1500) \\
&= 1 - F_{M,K,N}(1499) \\
&= 1 - \sum_{i=0}^{1499} \frac{\binom{2,000,000}{i}\binom{98,000,000}{50000-i}}{\binom{100,000,000}{50,000}} \\
&\approx 4.7 \times 10^{-55}
\end{aligned}
$$

In other words, the difference between the two rates is almost certainly not due to chance. The most likely explanation is that many burglaries go unreported because people don't want to deal with the hassle of police reports and/or increased insurance rates especially if the stolen items are not particularly valuable.