Math 10B with Professor Stankova
Worksheet, Discussion #15; Thursday, 3/8/2018
GSI name: Roy Zhao

1 Estimators and Confidence Intervals

1.1 Concepts

1. Often times, we are not given the distribution or parameters of the distribution (but we know what kind of distribution it is), and we want to figure out what the parameters are. One example is if you are given a biased coin and you want to figure out how biased it is (how likely flipping heads/tails is).

   The **estimator for the mean** is the **sample mean** which is given as
$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k.$$
   The **biased standard deviation estimator** is given by
$$x_* = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$
   The **unbiased standard deviation** or **sample standard deviation** is given by
$$s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$
   Given estimators for the mean and standard deviation (or the sample mean and sample standard deviation) $\hat{\mu}, \hat{\sigma}$ respectively, the 95% **confidence interval** for the expected value μ is
$$(\hat{\mu} - 2\hat{\sigma}/\sqrt{n}, \hat{\mu} + 2\hat{\sigma}/\sqrt{n}).$$
   You say that you are 95% confident that μ is in that interval.

1.2 Examples

2. I have a loaded die and I think that it is more likely to be a 1 than normal. Suppose I roll it 100 times and get 1 25 times. What is the 95% confidence interval for p, the probability of getting a 1?

3. Suppose that the amount of lightning strikes at Berkeley per thunderstorm is Poisson distributed. In the past 10 storms, I observe $2, 3, 5, 1, 3, 7, 9, 2, 0$, and 1 lightning strikes. What is the 95% confidence interval for λ?

1.3 Problems

4. True False A smaller 95% confidence interval means that we are less sure about what the mean μ could be.

5. Suppose that you measure the heights of everyone in a 400 person class to estimate the average height of a Berkeley student. Suppose that the sample mean is 65 inches with a standard deviation of 10 inches. What is the 95% confidence interval for the average height of a Berkeley student?

6. Assume the standard deviation of student heights is 5 inches. How large of a sample do you need to be 95% confident that the sample mean is within 1 inch of the population mean?

7. In a class of 25 students, the time that students spent on the midterm was 40 minutes with a standard deviation of 5 minutes. What is the 95% confidence interval for the average time taken on the midterm?

8. In 2012, you interview $10,000$ people and ask who they voted for. Out of these people, 51% of people voted for Obama. Are you 95% sure that a majority of people in America support Obama? (Hint: Formulate asking someone who they voted for as a Bernoulli trial)

1.4 Extra Problems

9. I have a loaded die and I think that it is more likely to be a 1 than normal. Suppose I roll it 100 times and get 1 50 times. What is the 95% confidence interval for p, the probability of getting a 1?

10. Suppose that the amount of lightning strikes at Berkeley per thunderstorm is Poisson distributed. In the past 9 storms, I observe $2, 6, 1, 1, 3, 4, 2, 2,$ and 6 lightning strikes. What is the 95% confidence interval for λ?

11. Suppose that you measure the heights of everyone in a 100 person class to estimate the average height of a Berkeley student. Suppose that the sample mean is 68 inches with a standard deviation of 10 inches. What is the 95% confidence interval for the average height of a Berkeley student?

12. Assume the standard deviation of student heights is 8 inches. How large of a sample do you need to be 95% confident that the sample mean is within 1 inch of the population mean?

13. In a class of 36 students, the time that students spent on the midterm was 45 minutes with a standard deviation of 10 minutes. What is the 95% confidence interval for the average time taken on the midterm?

14. In 2016, you interview 100 people and ask who they voted for. Out of these people, 54% of people voted for Hillary. Are you 95% sure that a majority of people in America support Hillary?

2 Hypothesis Testing

2.1 Concepts

15. In general, statistics does not allow you to prove anything is true, but instead allows you to show that things are probably false. So when we do hypothesis testing, the **null hypothesis** H_0 is something that we want to show is false and the **alternative hypothesis** H_1 is something that you want to show is true. For example, to show that a drug cures cancer, the null hypothesis would be that the drug does nothing and the alternative hypothesis would be that the drug does help cure cancer.

    A **type 1 error** is rejecting a true null which means that in our example, saying a drug cures cancer when it doesn't. A **type 2 error** is failing to reject a false null which means in our case as saying that the drug doesn't do anything when it does. The **significance level** is the probability of making a type 1 error. The **power** is 1 minus the probability of making a type 2 error.

    You use a χ^2 test to determine if a distribution is how you expect it to be. Suppose that you expect it to be distributed with a different values and for each of these values, you expect to get outcome k m_k times but actually get it n_k times. Then you compare the statistic

    $$r = \sum_{k=1}^{a} \frac{(n_k - m_k)^2}{m_k}$$

    with the $\chi^2(a-1)$ distribution.

2.2 Examples

16. Chip bags say that they have 14 ounces of chips inside with a standard deviation of 0.5 ounces. You weigh 100 bags and get an average of 13.8 ounces. What can you say with significance level $\alpha = 0.05$?

17. In a skittle bag, you get 11 red skittles, 12 blue, 5 green, 10 yellow, and 13 orange skittles. Is it possible that the colors are evenly distributed with a significance level of $\alpha = 0.05$?

2.3 Problems

18. True False The null hypothesis is something we want to be false.

19. True False If we get a value that is not smaller than α, then we have shown that the null hypothesis is true.

20. True False We want our test to have a high significance level and high power.

21. True False A type-2 error made by a road patrol may result in letting drunken drivers continue driving.

22. You flip a coin 100 times and get 55 heads. Can you say that it is biased towards heads? (use $\alpha = 0.05$)

23. An infomercial claims that a miracle drug will cause you to grow all your hair back. There are 25 brave participants and surprisingly 10 people regrew their hair. If normally 10% of people regrow their hair, can you say that this drug worked?

24. You take 400 cards and get 100 spades, 105 hearts, 107 diamonds, and 88 clubs. Can you say that the suits are not evenly distributed with $\alpha = 0.05$?

25. You expect to get a distribution of brown eyes brown hair to brown eyes blond hair to blue eyes brown hair to blue eyes blond hair as $9 : 3 : 3 : 1$. When looking around in class, you get a distribution of $61 : 19 : 11 : 9$ after looking at 100 people. Is this distribution accurate (use $\alpha = 0.05$)?

2.4 Extra Problems

26. Chip bags say that they have 2 ounces of chips inside with a standard deviation of 0.05 ounces. You weigh 100 bags and get an average of 2.01 ounces. What can you say with significance level $\alpha = 0.05$?

27. In a skittle bag, you get 15 red skittles, 8 blue, 5 green, 7 yellow, and 15 orange skittles. Is it possible that the colors are evenly distributed with a significance level of $\alpha = 0.05$?

28. You flip a coin 100 times and get 38 heads. Can you say that it is biased towards tails? (use $\alpha = 0.05$)

29. An infomercial claims that a miracle drug will cause you to grow all your hair back. There are 25 brave participants and surprisingly 7 people regrew their hair. If normally 10% of people regrow their hair, can you say that this drug worked?

30. You take 400 cards and get 95 spades, 110 hearts, 109 diamonds, and 86 clubs. Can you say that the suits are not evenly distributed with $\alpha = 0.05$?