Chriss Jordan Oboa

Professor Tysse

Math213

May 6th, 2022

# MATH 213 – Homework 18

1. Our textbook authors have provided a data set called “ncbirths” with data about a random sample of 1000 births from the state of North Carolina in 2004. You can find a full description of the data set at <https://www.openintro.org/data/index.php?data=ncbirths>.

Here are 10 rows of the data matrix (excluding some of the variables). There is a blank column on the right that you will use later in this assignment.

1. Is the data from an observational study or an experiment?

It is an observational study.

1. Based on the information you have about the dataset, I claim we can generalize findings to a larger population. Tell me what the larger population is.

The state of North Carolina’s births.

1. Based on the information you have about the dataset, can we draw causal conclusions?

There is no causal conclusion because it is an observational study.

We will focus on two variables: birth weight and smoking habit of mother. We are interested in the effect of maternal smoking on birth weights of babies. In particular, we will examine the difference in mean birth weights for babies of mothers who are smokers vs non-smokers.

The following table summarizes the distribution of birth weights grouped by the smoking habit of the mother. One of the 1000 cases is excluded, since it did not have a value for the “smoking habit of mother” variable.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| smoking habit | N | Mean | std. dev. | Q1 | median | Q3 |
| nonsmoker | 873 | 7.14 | 1.52 | 6.44 | 7.31 | 8.06 |
| smoker | 127 | 6.83 | 1.39 | 6.08 | 7.06 | 7.74 |

1. Before performing any calculations, do you think the sample size here is large enough to glean meaningful information about the population of interest? Give reasons. (You are asked to revisit this question later.)

The sample size here is large enough to glean meaningful information about the population of interest because the current population is at 10.49 million, which man using a sample of size 1000 will result of having a lot of variability in which will make the generalization non considerable.

1. Compute the difference in sample mean birth weights in the order (nonsmoker – smoker).

Diff in mean= 7.14- 6.83= 0.31

1. What parameter can you estimate using this difference of sample means?

You can estimate using this difference of sample means, the population mean.

1. Go ahead and estimate the parameter of interest using a 95% confidence interval, showing your work.

Find the standard error both variable first:

Nonsmoker: 1.52/sqrt (873) = 0.05144

Smoker: 1.39/ sqrt (127) = 0.12334

Diff in standard error: sqrt ((0.05144)^2 + (0.12334)^2) = 0.13363

Then,

0.31 + 2\*0.13363= 0.57726

0.31 - 2\*0.13363= 0.04274

1. Write a sentence of interpretation for the parameter.

We are 95% confident that the birth weights for babies of mothers who are non-smokers would be 0.04274

and 0.57726 more than the birth weights for babies of mothers who are smokers.

1. Can we be confident that, in the population of interest, babies of nonsmokers tend to be heavier than babies of smokers, on average? What about the work you have done helps you to answer this question?

We can be confident that, in the population of interest, babies of nonsmokers tend to be heavier than babies of smokers, on average based on common knowledge. However, I also think he can be an overall statement but there will always have some outliers.

1. Now revisit the question about sample size in part d. Do you think you have learned something meaningful about the population?

Yes, I learned something meaning about the population. There is less mothers who are smokers and babies coming from a mother who smokes tend to have less weight compared to a baby who comes from a nonsmoker mother.

1. The *margin of sampling error* is half the width of a confidence interval. In other words, it is the value that we add and subtract to the statistic to create a confidence interval. In the case of a 95% confidence interval, the margin of sampling error is equal to 2SE. For example, if I say something like “the mean height for women in the US aged 20 to 29 is 64.1 inches with margin of sampling error 5.2 inches at 95% confidence”, another way of saying this is, “we are 95% confident that the mean height for women in the US aged 20 to 29 lies between 58.9 and 69.3 inches.” To find the confidence interval here, I added and subtracted the margin of error from 64.1 inches.

Sample size is an important consideration in designing a study. As we have learned, larger sample sizes lead to smaller standard errors, which means narrower confidence intervals, and therefore, better estimates for the population parameter of interest. But how large should the sample size be? In this question, we will figure out what sample size to use in a particular situation so that the margin of error is of a certain small size.

I think the sample size should about 25% of the population to allow less variability. The less variability the better it is.

In a typical competitive election, a major candidate will have the support of about 50% of the voters. Imagine that you are working for one of the two major candidates in an election. You are going to conduct a poll of a random sample of likely voters to determine the proportion of voters planning to vote for your candidate. How large a sample would you recommend to make the margin of sampling error (at 95% confidence) less than 1 percentage point? Show some of your work so that I can understand your reasoning.

This is the step:

Se =(sqrt(p(1-p)/n)

2Se= (sqrt(p(1-p)/n)

2(0.01) =sqrt ((0.5(1-0.5))/n)

N= 625

A sample we would you recommend to make the margin of sampling error (at 95% confidence) less than 1 percentage point is a sample size of 625.

1. In Quiz 7, we examined the 1993 study in which researchers tested the drug AZT to see if it might be effective in preventing the transmission of HIV infection from mothers to babies before and during birth.

In 1993, in one of the first studies about preventing transmission of HIV infection from mothers to babies before and during birth, researchers gave the drug AZT to pregnant, HIV-infected women (Connor et al., 1994). Roughly half of the women in the experiment were randomly assigned to receive the drug AZT, and the others received a placebo. The HIV-infection status was then determined for the babies. The results of the experiment are summarized in the table and graph below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Placebo treatment | AZT treatment | Total |
| HIV-infected baby (HIV+) | 40 | 13 | 53 |
| Non-HIV-infected baby (HIV-) | 143 | 167 | 310 |
| Total | 183 | 180 | 363 |

1. Compute the difference of sample proportions of mothers with HIV-positive babies between the placebo group and the AZT group.

Placebo treatment: 40/183= 0.218579

AZT treatment: 13/180= 0.072222

Diff in prop: 0.218579 - 0.07222= 0.146359

What symbol(s) do we use to represent this statistic?

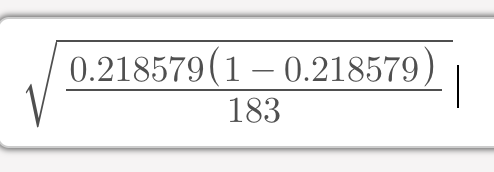
We use P\_hat to represent this statistic.

1. What parameter can we estimate using this statistic? Be specific and detailed in your wording, remembering the study type and what impact that has in the type of parameter we can estimate.

We can estimate the population proportion using this statistic.

1. Compute the 95% confidence interval that would estimate this parameter and write a sentence of interpretation for it. You can compute the confidence interval using the TBI applet or you can do it by hand. Tell which method you used. If you work by hand, show your work.

Placebo treatment= 0.03055

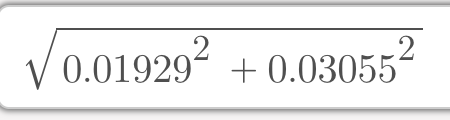


AZT treatment= 0.01929

A picture containing text

Description automatically generated

Diff\_Standard error= 0.03613



95% CI:

0.146359 - 2 (0.03613) = 0.074099

0.146359 + 2 (0.03613) = 0.218619

Method I used to find the CI: by hand

Interpretation: We are 95% confident that if using the AZT treatment, they would be 0.074099

and 0.218619 less HIV infected babies compared to not using anything at all (Placebo).

We are 95% confident that if the drug AZT were given to pregnant, HIV-infected women, there would be 0.074099 and 0.218619 less HIV infected babies compared to if the if the Placebo were given to pregnant, HIV-infected women.