# Chriss Jordan Oboa

Professor Tysse

Math213

April 22nd, 2022

# MATH 213 – Homework 16

1. Recall a study that we considered in Homework 13: an article “Change in Prevalence of Hearing Loss in US Adolescents” from about 10 years ago reported on a study of hearing loss in US residents aged 12 to 19. The authors used data from the National Health and Nutrition Examination Survey (NHANES), which is an ongoing program run by the US government to assess all aspects of the health of US residents. The NHANES data come from physical exams and interviews of nationally representative samples of US residents.

Part of the study focused on data from the 2005-2006 NHANES. In this data set, there were 1771 people aged 12 to 19. Of those 1771 people, there were 333 who had experienced any sort of hearing loss.

Although the NHANES did not use an actual random sample from the population of US residents aged 12 to 19 in 2005-2006, we will analyze it using our methods for random samples. (This approach is reasonable because of the careful steps used to avoid sampling bias.)

1. Look up the solutions for HW 13. Notice that the sample statistic of interest is the proportion of people in the sample who experienced any sort of hearing loss. What symbol do we use to represent this sample statistic? *(Look up our table of symbols if necessary, to get this right.)* What is its value?

**We used P\_hat to represent this statistic and the value was 0.19**

1. Again, looking at the solutions for HW 13, we said the parameter of interest is the proportion of people aged 12-19 in all of the United States in 2005-2006 who experienced any sort of hearing loss. What symbol do we use to represent this parameter?

**The parameter is simply P.**

1. Compute the estimated standard error for the sample proportion. *(You may need to look at the slides titled “Formulas for Standard Error” from class on 4/18. Pay close attention to the last slide. The activity Using the formulas for SE from that day may also be helpful.)*

**p\_hat = 0.19**

**SE = sqrt((p\_hat(1-p\_hat))/n) = sqrt((0.188(1-0.19))/1771) = 0.00932**

1. Use your standard error to calculate a 95% confidence interval for the population parameter, showing your work.

**95% CI = statistic ± 2(SE) = 0.19 ± 2(0.00932) = 0.17136 to 0.20864**

1. Write a sentence of interpretation for your confidence interval.

**We are 95% confident that the proportion of people aged 12 to 19 who had experienced any sort of hearing loss lies between 0.4090 and 0.4328.**

1. Check your confidence interval using either prop.test in R or the theory-based inference (TBI) applet in the Rossman/Chance applet collection (see new link to these applets in left-hand menu of Blackboard). The result from prop.test especially may not match your confidence interval exactly, but it should be close. If it’s not close, check your work!

Method used to check: **The theory-based inference (TBI) applet in the Rossman/Chance applet collection**

Confidence interval from this method: **95%**

Chart, bar chart

Description automatically generated

1. In the National Health and Nutrition Examination Survey (NHANES), subjects were whether they had smoked at least 100 cigarettes in their lives. The researchers then asked the 2328 people who answered “yes” (our sample) the age at which they started smoking.
2. Identify some features of this study. (I filled out part of this for you!)

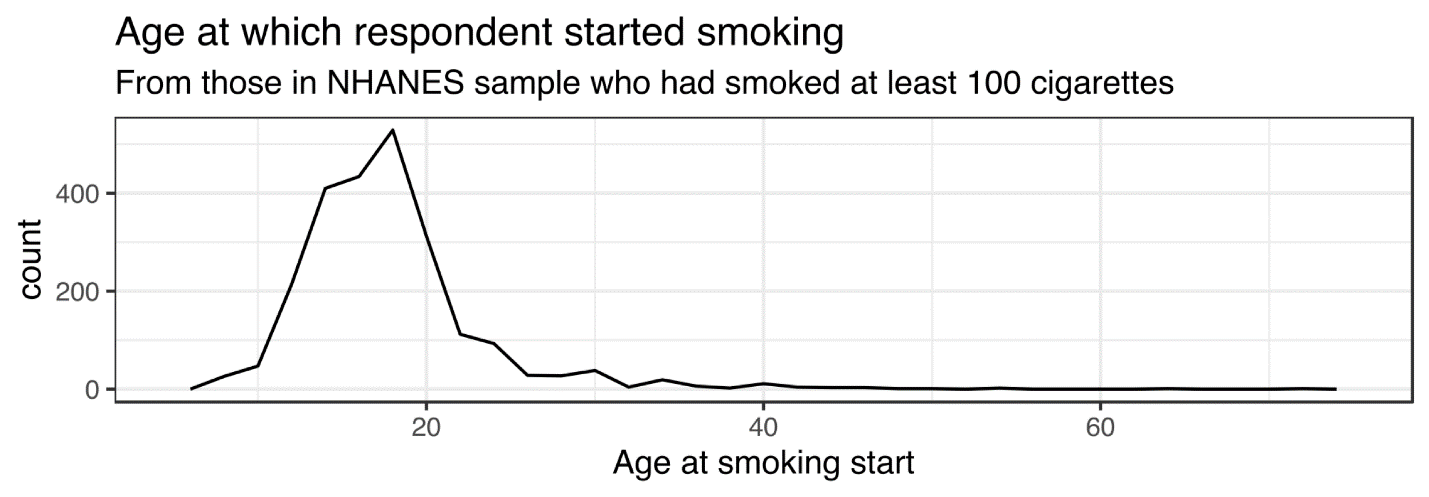
Population of interest: US residents who have smoked at least 100 cigarettes.

Variable of interest: **The age at which they started smoking**

Type of variable (categorical or **numerical**):

What population parameter might we use the study data to estimate: **All U.S citizen who have smoked at least 100 cigarettes**

We can summarize the distribution of ages at which people in the sample started smoking with our usual tools: a frequency polygon and some summary statistics.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 0.025 quantile | Q1 | median | Q3 | 0.975 quantile |
| 11 years | 15 years | 18 years | 20 years | 31 years |

|  |  |
| --- | --- |
| mean | standard deviation |
| 18.2 years | * 1. years |

1. Describe the distribution of smoking-start ages in this sample of 2328 people who have smoked at least 100 cigarettes. As always, address the center, spread, and shape of the distribution. Include the context in your description.

**The frequency polygon above shows the distribution of ages at which people in the sample started smoking. We see that the graph is right skewed and is centered around 18 years old. Most of the people in the sample started smoking were between 15 and 20 years old, with very few who started smoking as low as 11 years old and as high as 31 years old. There are uncommon outliers where who started smoking around 40 and above.**

1. I have provided a CSV file containing the data from the study.

**ok**

1. Load this data file into the StatKey bootstrap applet “CI for Single Mean, Median, St. Dev” > Upload File as we have done in class.

**done**

1. Use the applet to compute 50% and 95% bootstrap confidence intervals. Report them here.

Chart

Description automatically generated

Chart, histogram, scatter chart

Description automatically generated

**95% confidence is 17.998 to 18.421**

**50% confidence is 18.127 to 18,284**

1. Write sentences of interpretation.

I am 95% confident that **U.S citizen who have smoked at least 100 cigarettes started smoking when they were between 17 to 18 years old, (17.998 to 18.421).**

I am 50% confident that **U.S citizen who have smoked at least 100 cigarettes started smoking when they were at 18 years old, (18.127 to 18,284).**

1. Compute the 95% confidence interval in TWO more ways:
2. Estimate the standard error using our formula and the “plug-in” principle. Then use it to compute a 95% CI. *(You may need to look at the slides titled “Formulas for Standard Error” from class on 4/18. Pay close attention to the last slide. The activity Using the formulas for SE from that day may also be helpful.)*

*Estimated SE:* **0.111**

*95% CI:* **statistic ± 2(SE) = 18.203 ± 2(0.111) = 17.981 to 18.425**

1. Use the TBI applet from the Rossman/Chance applet collection. The confidence interval from the applet will not exactly match yours, but it should be quite close.

95% confidence interval from TBI applet: **(17.9917, 18.4143)**