

Worksheet 11

Author:

Discussants:

=====

The purpose of this worksheet is to practice running parametric hypothesis tests and computing parametric confidence intervals for proportions. Please answers the following questions and submit a compiled pdf with your answers to Moodle by 11:59pm on Sunday December 2nd. Two functions you will find useful are: `pnorm()` and `qnorm()`

```
library(knitr)
# makes sure the code is wrapped to fit when it creates a pdf
opts_chunk$set(tidy.opts = list(width.cutoff = 60))
set.seed(1) # set the random number generator to always give the same sequence of random numbers
```

Please answer the following Lock5 questions:

Lock5 exercise 5.2 first edition, P.130 second edition: In figure below (Fig. 5.11/P.13), the percentage of the population that is more than 30 is closet to which value: 4%, 27% 50%, 73% or 95%.

Answer

Lock5 exercise 5.6 first edition, P.130 second edition: Use the R chunk below and the `pnorm()` function to find the areas on the standard normal density curve $N(0, 1)$ listed below and then check your answers using the apps at: (https://asterius.hampshire.edu:3939/intro_stats/normal_area/) and (https://asterius.hampshire.edu:3939/intro_stats/normal_area_range/)

- a) The area below $z = 0.8$
- b) The area above $z = -1.5$
- c) The area between $z = -1.75$ and $z = 0.5$

```
# a) The area below z = 0.8

# b) The area above z = -1.5

# c) The area between z = -1.75 and z = 0.5
```

Answers

- a)

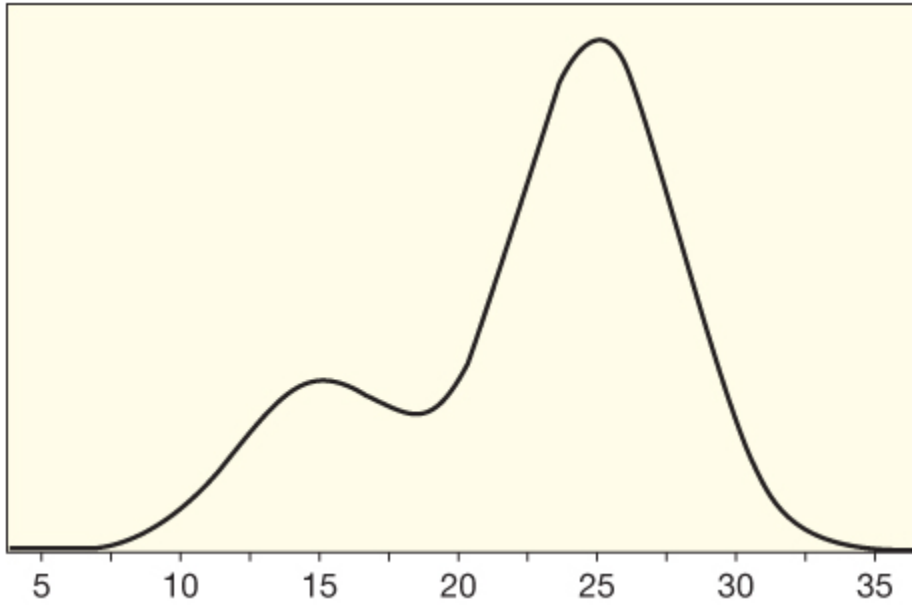


Figure 1: Figure 5.11

- b)
- c)

Lock5 exercise 5.18 first edition, P.146 second edition: Find the ‘endpoints’ (quantile function values) for the specified area under the specified normal distribution. To do this use the `qnorm()` function and then check your results using this app: (https://asterius.hampshire.edu:3939/intro_stats/normal_area_quantiles/)

- a) For a normal curve of $N(25, 8)$, what is the endpoint value such that the area under the curve to the right of the endpoint is 0.25?
- b) For a normal curve of $N(500, 80)$, what is the endpoint value such that the area under the curve to the left of the endpoint is 0.02?
- c) For a normal curve of $N(10, 3)$, what are the endpoint values such that the symmetric area under the middle of the curve is 0.95?

a)

b)

c)

Answers:

- a)
- b)
- c)

Curving Grades on an Exam Lock5 exercise 5.40 first edition, P.168 second edition: A statistics instructor designed an exam so that the grades would be roughly normally distributed with mean $\mu = 75$ and standard deviation $\sigma = 10$. Unfortunately, a fire alarm went off with ten minutes to go in the exam which made it difficult for some students to finish. When the instructor graded the exams, he found that the grades were roughly normally distributed, but the mean grade was 62 and the standard deviation was 18. To be fair, he decides to “curve” the scores to match the desired $N(75, 10)$ distribution. To do this, he standardizes the actual scores to z-scores using the $N(62, 18)$ distribution and then “unstandardizes” those z-scores to shift to $N(75, 10)$. What is the new grade assigned for a student whose original score was 47? Now about a student who originally scores a 90? Hint: you can use the R chunks below to convert the two original scores to z-scores and then convert them back to the new scores from the $N(75, 10)$ distribution.

```
# calculate the z-scores
```

```
# transform the z-scores to a N(75, 10) distribution
```

Answers:

Lock5 exercise 5.56 first edition, 5.13 second edition: Find the p-value based on the standard normal distribution for each of the following standardized test statistics. By an upper-tailed test, we mean a test where the alternative hypothesis is framed as $H_A : param > x$ while a two-tailed test is framed as $H_A : param \neq x$. Please show the code you used to get your answers below, or explain how you got your answers.

- a) $z = 0.84$ for an upper tail test for a difference in two proportions.
- b) $z = -2.38$ for a lower tail test for a difference in two means.
- c) $z = 2.225$ for a two-tailed test for a correlation.

```
# a)
```

```
# b)
```

```
# c)
```

Answers:

- a)
- b)
- c)

Hearing Loss in Teenagers Lock5 exercise 5.58 first edition, 5.43 second edition: A recent study

found that, of the 1771 participants aged 12 to 19 in the National Health and Nutrition Examination Survey, 19.5% had some hearing loss (defined as loss of 15 decibels in at least on ear). This is a dramatic increase from a decade ago. The sample size is large enough to use the normal distribution, and a bootstrap distribution shows that the standard error for the proportion is $SE = 0.009$. Find and interpret a 90% confidence interval for the proportion of teenagers with some hearing loss. Use the code chunk below to show your work.

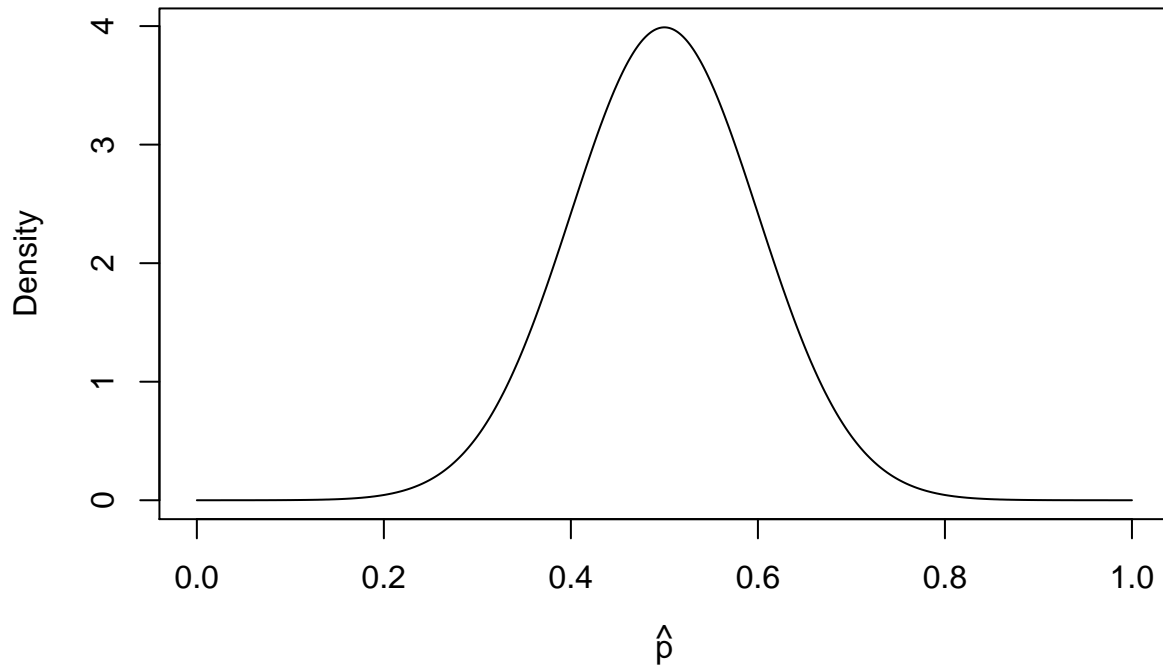
Answer:

Modified from Lock5 exercise 6.4 both editions: Suppose we have a population proportion $\pi = 0.27$. Let's create the sampling distribution for $n = 50$ observations of a categorical variable by doing the following:

- a) First, find mean and the standard error of the sampling distribution of proportions
- b) Second, the code below plots the density curve for $N(.5, .1)$. Modify the code below to plot the sampling distribution of proportions for $\pi = 0.27$, $n = 50$. Is the sample size large enough for this density function to be valid to use as the sampling distribution? Why or why not?

```
# a) report/calculate the mean and SE of the sampling  
# distribution (hint: use the equation from class)  
  
# b) plot the density curve  
  
# The x-values of the function - you do not need to change  
# this!  
x_vals <- seq(0, 1, by = 0.001)  
  
# change the parameters in the dnorm() function to create the  
# appropriate density function  
density_curve <- dnorm(x_vals, 0.5, 0.1)  
  
# Plotting the function - you do not need to change this  
# either  
plot(x_vals, density_curve, type = "l", xlab = expression(hat(p)),  
      ylab = "Density", main = expression(paste("Sampling distribution of ",  
                                                hat(p))))
```

Sampling distribution of \hat{p}



Answer:

- a)
- b)

**** What Percent of US Adults Say They Never Exercise? ** Lock5 exercise 6.34 first edition, 6.21 second edition:** In a survey of 1,000 US adults, twenty percent say they never exercise. This is the highest level seen in five years. Use the R chunk below to find a 99% confidence interval for the proportion of US adults who say they never exercise. Also report what the margin of error is and interpret what the 99% confidence interval tells you.

Answer:

Home Field Advantage in Baseball in 2014 6.64 first edition, 6.56 second edition There were 2428 Major League Baseball (MLB) games played in 2014, and the home team won 1288 of those games. If we consider the games played in 2014 as a sample of all MLB games, test to see if there is evidence, at the $\alpha = 1\%$ level, that the home team wins more than half the games. Show all details of the test in the code chunk below. Hints: Start by stating the null and alternative hypotheses. Then write down \hat{p} , n and compute the SE. Create the z-statistic and then use `pnorm()` to find the probability you would get a value as big or bigger than the observed \hat{p} .

Answer:

Reflection

How did this worksheet go? Is the concept that sample proportions (\hat{p}) can be described as coming from a normal distribution clear (when n is large enough)? Is it clear how to use the formulas to calculate the SE based on n and \hat{p} values, and then how to use the SE and \hat{p} values to do inference using normal distributions?