

**Agenda**

1. Confidence intervals
2. Central limit theorem
3. Sampling Distributions from Multimodal Populations

**Reminder** Confidence intervals for test statistics that are normally distributed are of the form:

$$\text{point estimate} \pm z_{\alpha/2}^* \cdot SE$$

Computing the point estimate is usually easy. Once you've chosen a confidence level, finding  $z_{\alpha/2}^*$  is trivial (use `qnorm()` or the table in the back of the book). The difficult part is usually computing the  $SE$ , since that depends on the sampling distribution of the test statistic!

**Warmup**

1. A recent study estimated the mean U.S. per capita consumption of sugar-sweetened beverages among adults 20 to 44 years of age to be 289 kcal/day with a standard error of 7 kcal/day.
  - (a) The 68-95-99.7 rule says that the probability is about 0.95 that  $\bar{x}$  is within  $y$  kcal/day of the population mean  $\mu$ . What is  $y$ ?
  - (b) About 99% of all samples will capture the true mean of kcals consumed per day in the interval  $\bar{x}$  plus or minus 7 kcal/day times what? Draw a labeled picture and indicate where the missing quantity is. Estimate it. What does the computer need to know in order to compute it?
2. Suppose 400 randomly selected alumni of the University of Okoboji were asked to rate the university's counseling services on a 1 to 10 scale. The sample mean was found to be 8.6. Assume that the standard error was computed to be 0.4.
  - (a) A researcher computes the 99% confidence interval for the average satisfaction score as  $8.6 \pm 1.96 \cdot 0.4$ . What is her mistake?
  - (b) After correcting her mistake in part (a), she states: "I am 95% confident that the sample mean falls between 7.82 and 9.38." What is wrong with this statement?
  - (c) She quickly realizes her mistake in part (b) and instead states: "The probability the true mean is between 7.82 and 9.38 is 0.95." What misinterpretation is she making now?
  - (d) Finally in her defense for using the Normal distribution to determine the confidence interval she says "Because the sample size is quite large, the population of alumni ratings will be approximately Normal." Explain to Ima her misunderstanding and correct this statement.
3. Explain whether a test of significance can answer each of the following questions.
  - (a) Is the sample or experiment properly designed?

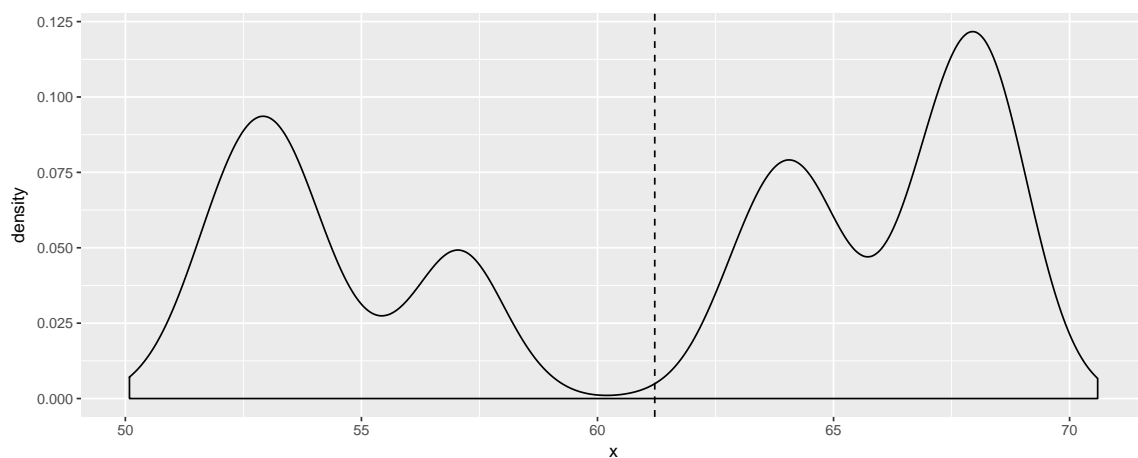
- (b) Is the observed effect compatible with the null hypothesis?
  - (c) Is the observed effect important?
4. Justify whether or not you agree with each of the following statements.
- (a) If the p-value is larger than 0.05, the null hypothesis is true.
  - (b) Practical significance is not the same as statistical significance.
  - (c) We can perform a statistical analysis using any set of data.
  - (d) If you find an interesting pattern in a set of data, it is appropriate to then use a significance test to determine its significance.

**Central limit theorem** See <http://45.55.32.181/shiny/SamplingDist/>

1. What happens to the sampling distribution as the size of samples increases?
2. What happens to the sampling distribution as the number of samples increases?

**Sampling Distributions from Multimodal Populations** Consider the following multimodal probability distribution.

```
require(mosaic)
n <- 1000
ds <- data.frame(a = rnorm(n, mean = 53), b = rnorm(n, mean = 57, sd = 0.8),
                 c = rnorm(n, mean = 64), d = rnorm(n, mean = 68, sd = 0.8), p = runif(n)) %>%
  mutate(x = ifelse(p < 0.25, a, ifelse(p < 0.4, b, ifelse(p < 0.65, c, d))))
pop_plot <- qplot(data = ds, x = x, geom = "density", adjust = 0.5) +
  geom_vline(aes(xintercept = mean(ds$x)), linetype = 2)
pop_plot
```



1. Sketch what you think the sampling distribution of the mean looks like for samples of size 1.
2. Sketch what you think the sampling distribution of the mean looks like for samples of size 2.
3. Sketch what you think the sampling distribution of the mean looks like for samples of size 10,000.
4. Sketch what you think the sampling distribution of the mean looks like for samples of size 4.

```
sim <- do(1000) * mean(~x, data = sample_n(ds, 4))  
pop_plot + geom_density(data = sim, aes(x = mean), adjust = 0.5, color = "red")
```