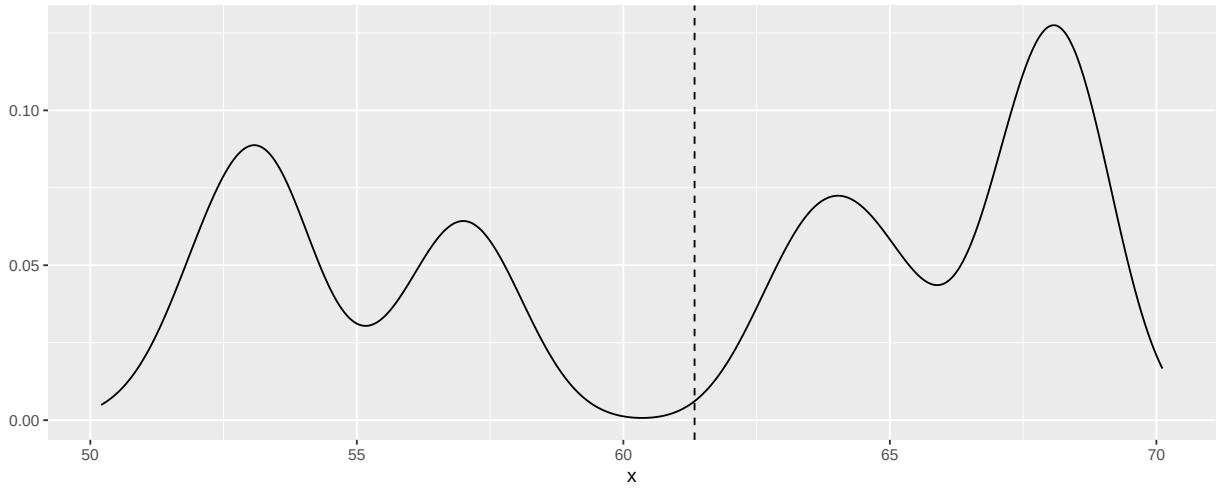
Sampling Distributions from Multimodal Populations

Consider the following multimodal probability distribution.



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

MLB Batting Averages

In 1941, Ted Williams of the Boston Red Sox hit .406, famously getting 6 hits in 8 at-bats on the last day of the season. No player in Major League Baseball has hit .400 since. Among the closest attempts was made by George Brett of the Kansas City Royals in 1980, when Brett hit .390. When viewed in relation to his peers, whose performance was more impressive?

```
library(tidyverse)
library(Lahman)
mlb <- Batting |>
  mutate(BAvg = H / AB) |>
  filter(yearID %in% c(1941, 1980) & AB > 400)
mlb |>
  filter(BAvg > .36) |>
  select(playerID, yearID, BAvg)
```

```
   playerID yearID      BAvg
1 brettge01   1980 0.3897550
2 willite01   1941 0.4057018
```

```
mlb |>
  group_by(yearID) |>
  summarize(N = n(), mean_BAvg = mean(BAvg), sd_BAvg = sd(BAvg))
```

```
# A tibble: 2 x 4
  yearID     N mean_BAvg sd_BAvg
   <int> <int>     <dbl>   <dbl>
1   1941    98     0.281  0.0328
2   1980   148     0.279  0.0276
```

1. Use the information above to calculate a z-score for both Williams in 1941 and Brett in 1980.

2. Whose performance do you think was more remarkable in the context of his peers? Why? What assumptions are you making?

Confidence intervals

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 μ. 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 known 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. Ima Bitlost 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.