

Agenda

1. Finish simulation stuff from last time
2. Central Limit Theorem
3. Normal Distribution

Warmup: Distribution of Height on OkCupid Consider the distribution of reported male height for users of the online dating site OkCupid.

1. What observations can you make from this data graphic?

Example: 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?

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

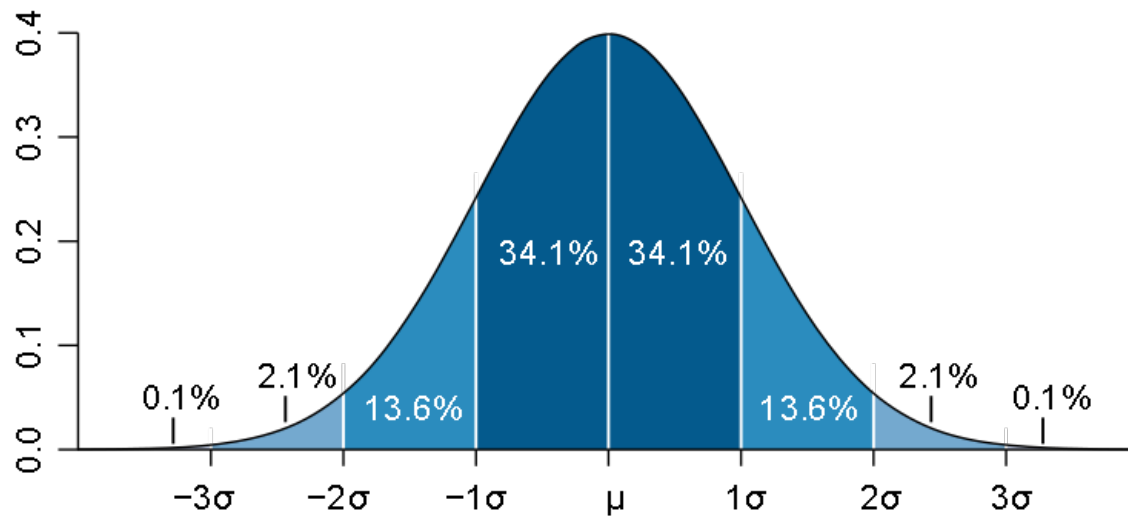
##	playerID	yearID	BAvg
## 1	willite01	1941	0.4057018
## 2	brettge01	1980	0.3897550

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

```
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	0.2806367	0.03279026
## 2	1980	148	0.2788247	0.02757441

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



The Empirical Rule for Normal Distributions For any normally distributed variable:

- About 68% of the distribution is contained within 1 standard deviation of the mean.
- About 95% of the distribution is contained within 2 standard deviations of the mean.
- About 99.7% of the distribution is contained within 3 standard deviations of the mean.
- About 38% of the distribution is contained within 0.5 standard deviations of the mean.
 - In other words, the middle 38% of the distribution is about 1 standard deviation wide.

Sample Calculations

1. What percentage of the distribution is less than 2 standard deviations above the mean?
 - By the rule, about 95% of the population is within two standard deviations of the mean. By symmetry, half of those are above the mean, and half below it. Thus, we estimate that about $95/2 = 47.5\%$ is less than 2 standard deviations above the mean.
 - From the picture, we can calculate the area as about $34.1\% + 13.6\% = 47.7\%$
2. Assume that the distribution of heights of adult women is approximately normal with mean 64 inches and standard deviation 2.5 inches.
 - (a) What percentage of women are taller than 5'9"?
 - (b) Between what heights do the middle 95% of women fall?
 - (c) What percentage of women are shorter than 61.5 inches?
 - (d) A professor claims that about 51% of women are between 61.5 and 65.25 inches tall. Is this claim accurate?