Notes: MS 204 Chapter 2 part III normal distribution

Overview

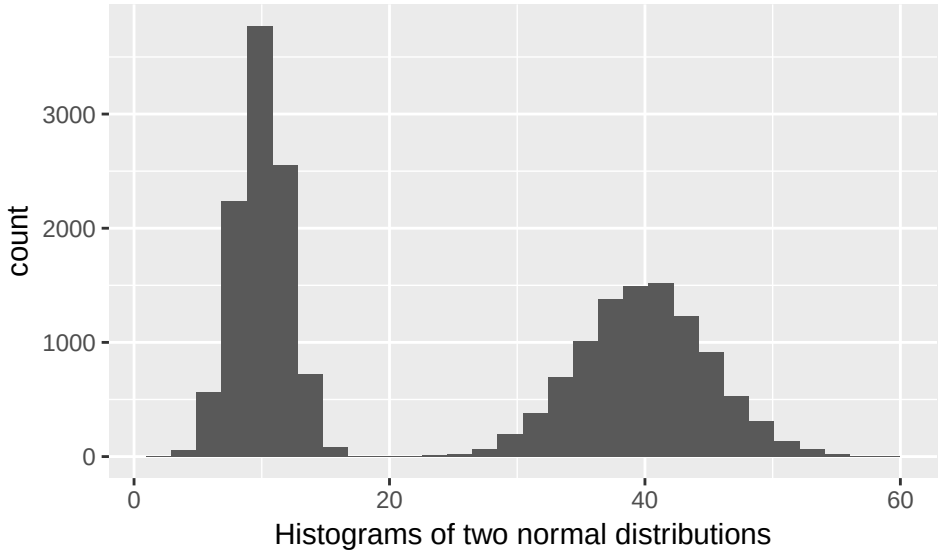
- Hypothesis testing review – single versus two population proportions
- Error types
- Normal distribution

Review of simulations & error types

Normal distribution

Normal distribution with different parameters

```
library(tidyverse)
x1 <- rnorm(10000, mean = 10, sd = 2)
x2 <- rnorm(10000, mean = 40, sd = 5)
qplot(c(x1, x2)) + xlab("Histograms of two normal distributions")
```



Standardizing with z-scores

SAT scores are normally distributed with mean 500 and standard deviation 100. (i) Draw a curve, (ii) find the percentage of scores below a 350, and (iii) identify what score would need to be obtained to ensure that 75 percent of the population lies below.

Rstats functions

```
pnorm(350, mean = 500, sd = 100)
```

```
## [1] 0.0668072
```

```
pnorm(-1.5)
```

```
## [1] 0.0668072
```

```
qnorm(.9, mean = 500, sd = 100)
```

```
## [1] 628.1552
```

```
qnorm(.9)
```

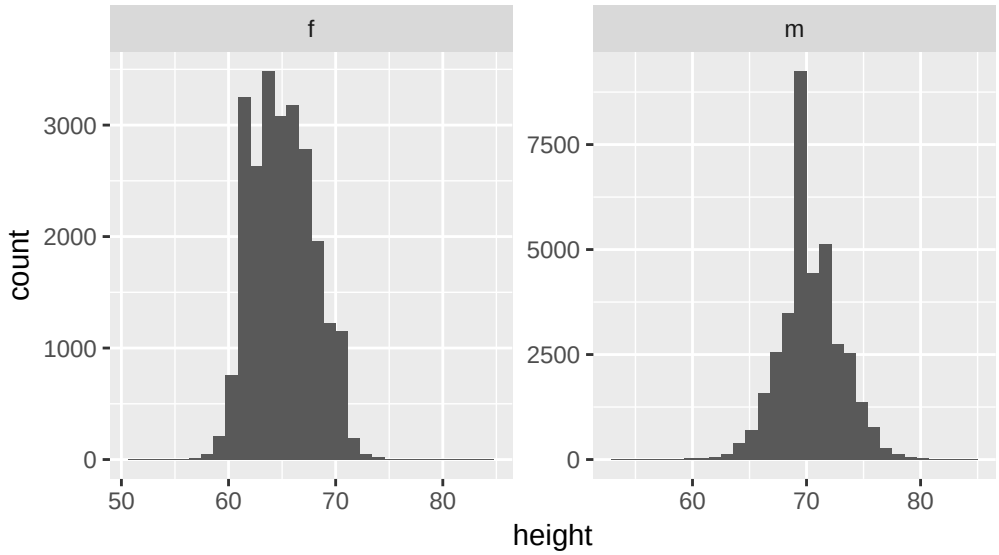
```
## [1] 1.281552
```

Follow up example

At the Heinz ketchup factory, the amount that goes in ketchup bottles is normally distributed with mean 36 oz and standard deviation 0.11 oz. A bottle is randomly selected. (i) How likely is it to have more than 35.9 ounces of ketchup? (ii) What percent of bottles will lie between 35.9 and 36.1 ounces of ketchup? Solve using both diagrams, equations, and code.

68-95-99.7 rule

```
library(okcupiddata)
#?profiles ### gives you varible names
profiles <- profiles %>% filter(height > 50, height < 85)
qplot(height, data = profiles) + facet_wrap(~sex, scales = "free")
```




```
profiles.sum <- profiles %>%
  group_by(sex) %>%
  summarise(ave.height = mean(height), sd.height = sd(height))
profiles.sum
```

```
## # A tibble: 2 x 3
##     sex ave.height sd.height
##   <chr>      <dbl>     <dbl>
## 1     f   65.10645  2.763380
## 2     m   70.44739  2.879944
```

```
profiles %>%
  filter(sex == "m") %>%
  mutate(one.sd = height > 70.44739 + (-1)*2.879944 & height < 70.44739 + (1)*2.879944,
         two.sd = height > 70.44739 + (-2)*2.879944 & height < 70.44739 + (2)*2.879944,
         three.sd = height > 70.44739 + (-3)*2.879944 & height < 70.44739 + (3)*2.879944) %>%
  summarise(percent.one = mean(one.sd),
            percent.two = mean(two.sd),
            percent.three = mean(three.sd))
```

```
##   percent.one percent.two percent.three
## 1   0.7007518   0.9671352     0.9961155
```