

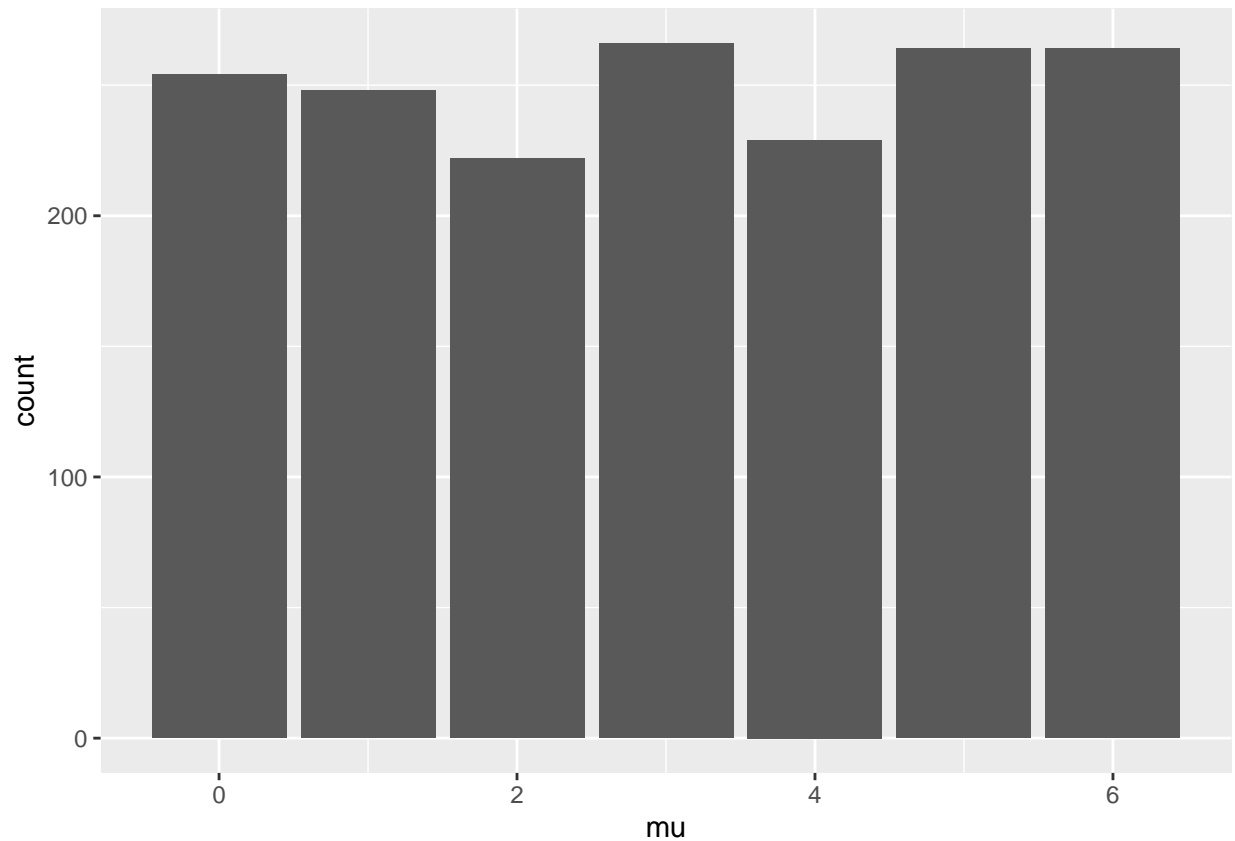
p8105_hw5

Cameron Chesbrough

2024-11-12

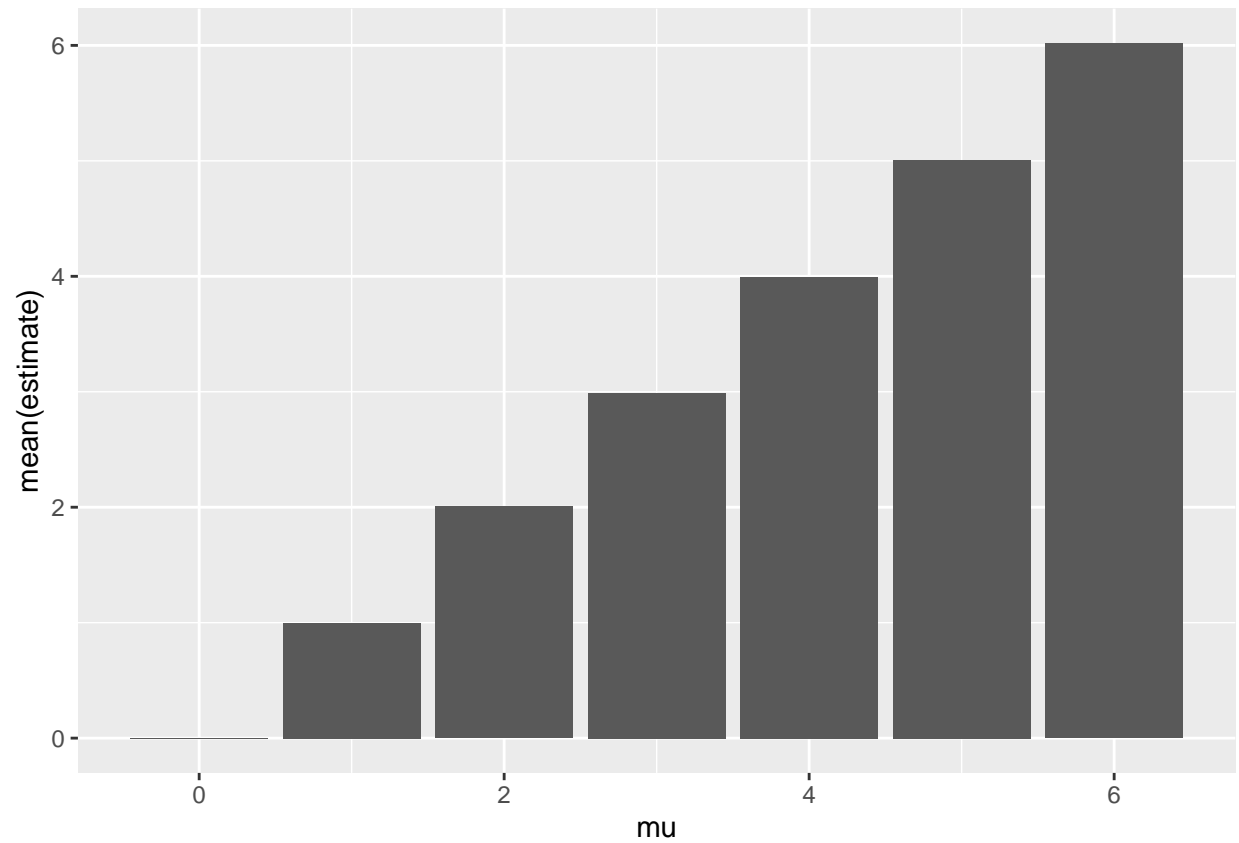
Question 2

```
sim_mean_pval = function(mu) {  
  n = 30  
  sigma = 5  
  sim_data = tibble(  
    x = rnorm(n, mu, sigma),  
  )  
  sim_test = t.test(sim_data, mu = mu)  
  sims = broom::tidy(sim_test)  
  sims[, c("estimate", "p.value")]  
}  
  
sim_results_df =  
  expand_grid(  
    mu = c(0,1,2,3,4,5,6),  
    iter = 1:5000  
  ) %>%  
  mutate(  
    estimate_df = map(mu, sim_mean_pval)  
  ) %>%  
  unnest(estimate_df)  
  
rejects = sim_results_df %>%  
  filter(p.value <= 0.05)  
ggplot(rejects, aes(x=mu)) + geom_bar()
```

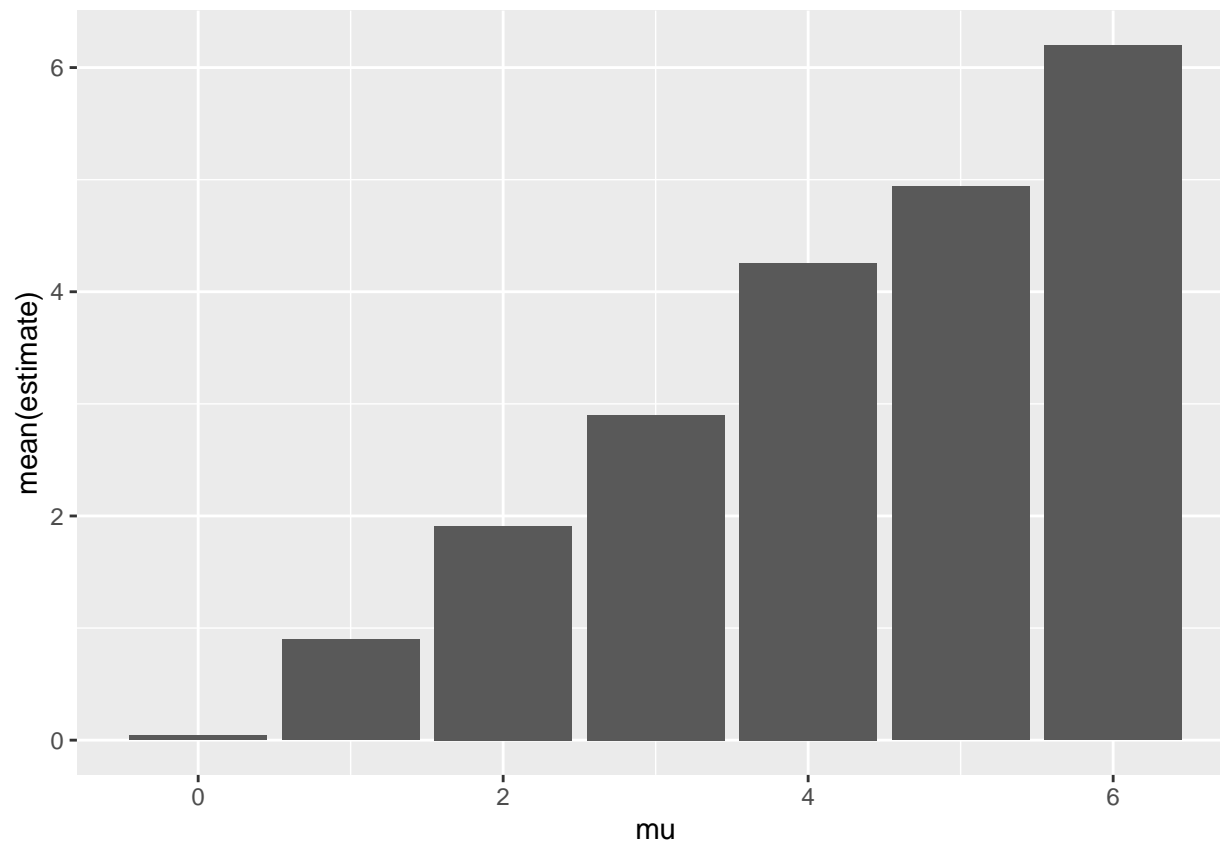


As effect size grows larger so will the power. Power is the measure of rejecting the null hypothesis when it should be rejected. Increasing the effect size will have a smaller likelihood of type 2 error and an increase in power.

```
avgs = sim_results_df %>%  
  group_by(mu) %>%  
  summarise(mean_estimate = mean(estimate))  
ggplot(avgs, aes(x = mu, y = mean_estimate)) + geom_col()
```



```
avgs_rejects = sim_results_df %>%  
  filter(p.value <= 0.05) %>%  
  group_by(mu) %>%  
  summarise(mean(estimate))  
ggplot(avgs_rejects, aes(x = mu, y = `mean(estimate)`)) + geom_col()
```



In the scenarios where the null hypothesis was rejected, the average mean is approximately equal to the true mean. Following an alpha of 0.05 we are rejecting all which are less than 5% likely to have occurred, but this will lead to rejections for those both significantly larger and smaller. Because of this, when looking across all tests, the mean will be approximately equal.

Question 3

```
homicide_df = read_csv(file = "./data/homicide-data.csv")

## Rows: 52179 Columns: 12
## -- Column specification -----
## Delimiter: ","
## chr (9): uid, victim_last, victim_first, victim_race, victim_age, victim_sex...
## dbl (3): reported_date, lat, lon
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

homicide_df = homicide_df %>%
  mutate(city_state = paste(city, state, sep = ", "))
```

This dataset describes homicides in large American cities in the past decade. It includes information on where the homicide took place, information describing the victim, and what action from law enforcement followed. The dataset has 52,179 rows and 12 columns.

```

total = homicide_df %>%
  group_by(city_state) %>%
  summarise(murders = n())

unsolved = homicide_df %>%
  filter(disposition == c("Closed without arrest", "Open/No arrest")) %>%
  group_by(city_state) %>%
  summarise(unsolved_murders = n())

## Warning: There was 1 warning in 'filter()'.
## i In argument: 'disposition == c("Closed without arrest", "Open/No arrest")'.
## Caused by warning in 'disposition == c("Closed without arrest", "Open/No arrest")':
## ! longer object length is not a multiple of shorter object length

both = full_join(total, unsolved) %>%
  drop_na()

## Joining with 'by = join_by(city_state)'

prop.test(both$unsolved_murders[3], both$murders[3])

##
## 1-sample proportions test with continuity correction
##
## data:  both$unsolved_murders[3] out of both$murders[3], null probability 0.5
## X-squared = 345.29, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3078863 0.3427525
## sample estimates:
##           p
## 0.3250796

calc_prop_murders = function(dfrow) {
  unsolved = dfrow[[3]]
  murders = dfrow[[2]]
  mur_prop = prop.test(unsolved, murders)
  results = broom::tidy(mur_prop)
  results[, c("estimate", "conf.low", "conf.high")]
}

testing = split(both, both$city_state)

listcol_df = both %>%
  mutate(summary = map(testing, calc_prop_murders)) %>%
  unnest(summary) %>%
  arrange(estimate)

ggplot(listcol_df, aes(reorder(x = city_state, estimate), y = estimate)) +
  geom_col(fill = "white", color = "gray") +
  geom_errorbar(aes(ymin = conf.low, ymax = conf.high)) +

```

```
coord_flip() +
labs(
  title = "Proportion of Murders Unsolved in 50 Large Cities",
  y = "Proportion of Murders Unsolved",
  x = "City and State"
) +
theme_bw(base_size = 5)
```

