

Assignment 9

Vishal

2022-11-30

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(ggplot2)
library(Stat2Data)
library(palmerpenguins)

## Warning: package 'palmerpenguins' was built under R version 4.2.2

data(package = 'palmerpenguins')
data("Hawks")
```

1. Basic concepts in classification

http://ianmadd.github.io/pages/Confidence_Intervals_Part4.html

2. A chi-squared test of populatinn variance

2 (Q1)

```
chi_square_test_one_sample_var <- function(sample, sigma_square_null) {
  sample <- sample[!is.na(sample)]
  n <- length(sample)
  chi_squared_statistic <- (n - 1) * var(sample) / sigma_square_null
  p_value <- 2 * min(pchisq(chi_squared_statistic, df = n - 1),
                    1 - pchisq(chi_squared_statistic, df = n - 1))
  return(p_value)
}
```

2 (Q2)

```
sample_size <- 100
mu_0 <- 1
sigma_0 <- 4

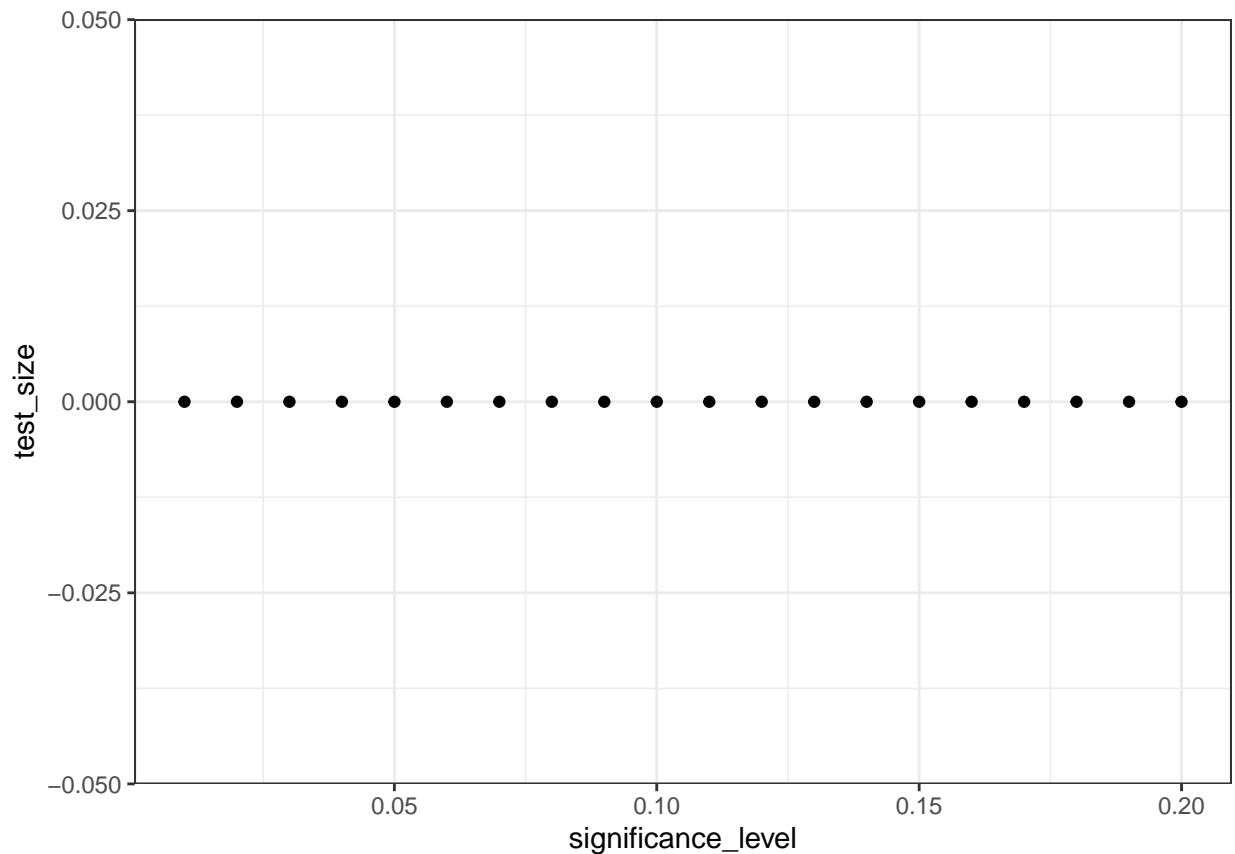
df <- data.frame(trial = seq(0.01, 0.20, 0.01)) %>%
  mutate(samples = map(.x = trial, .f = ~rnorm(n = sample_size, mean = mu_0, sd = sigma_0))) %>%
  mutate(p_val = map_dbl(.x = samples, ~chi_square_test_one_sample_var(.x, sigma_0)))

alpha_list <- seq(0.01, 0.20, 0.01)

compute_test_size <- function(alpha) {
  reject_null <- df$p_val[df$p_val < alpha]
  return (mean(reject_null))
}

multiple_alpha_test_size_simulation_df <- data.frame(significance_level = alpha_list) %>%
  mutate(test_size = map_dbl(significance_level, compute_test_size))

multiple_alpha_test_size_simulation_df %>%
  ggplot(aes(x = significance_level, y = test_size)) +
  geom_point() +
  theme_bw()
```



2 (Q4)

```
sample_size <- 100
mu_0 <- 1
sigma <- 6
sigma_0 <- 4

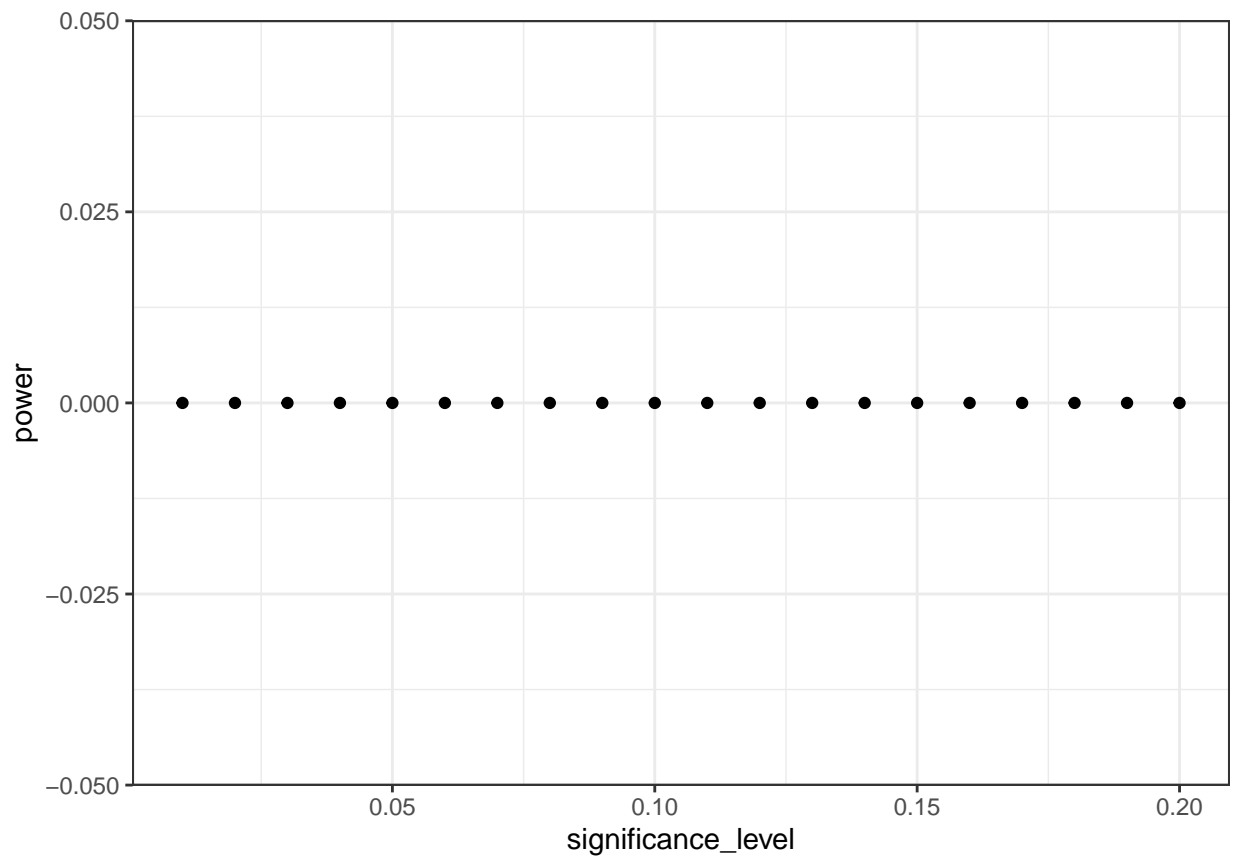
df <- data.frame(trial = seq(0.01, 0.20, 0.01)) %>%
  mutate(samples = map(.x = trial, .f = ~rnorm(n = sample_size, mean = mu_0, sd = sigma))) %>%
  mutate(p_val = map_dbl(.x = samples, ~chi_square_test_one_sample_var(.x, sigma_0)))

compute_power <- function(alpha) {
  reject_null <- df$p_val[df$p_val < alpha]
  return (mean(reject_null))
}

multiple_alpha_power_df <- data.frame(significance_level = seq(0.01, 0.20, 0.01)) %>%
  mutate(power = map_dbl(significance_level, compute_power))

multiple_alpha_power_df %>%
  ggplot(aes(x = significance_level, y = power)) +
```

```
geom_point() +
theme_bw()
```



```
## 5
```

```
bill_adelie <- penguins %>%
  filter(species == "Adelie") %>%
  pull(bill_length_mm)
```

```
bill_adelie <- bill_adelie[!is.na(bill_adelie)]
bill_adelie
```

```
## [1] 39.1 39.5 40.3 36.7 39.3 38.9 39.2 34.1 42.0 37.8 37.8 41.1 38.6 34.6 36.6
## [16] 38.7 42.5 34.4 46.0 37.8 37.7 35.9 38.2 38.8 35.3 40.6 40.5 37.9 40.5 39.5
## [31] 37.2 39.5 40.9 36.4 39.2 38.8 42.2 37.6 39.8 36.5 40.8 36.0 44.1 37.0 39.6
## [46] 41.1 37.5 36.0 42.3 39.6 40.1 35.0 42.0 34.5 41.4 39.0 40.6 36.5 37.6 35.7
## [61] 41.3 37.6 41.1 36.4 41.6 35.5 41.1 35.9 41.8 33.5 39.7 39.6 45.8 35.5 42.8
## [76] 40.9 37.2 36.2 42.1 34.6 42.9 36.7 35.1 37.3 41.3 36.3 36.9 38.3 38.9 35.7
## [91] 41.1 34.0 39.6 36.2 40.8 38.1 40.3 33.1 43.2 35.0 41.0 37.7 37.8 37.9 39.7
## [106] 38.6 38.2 38.1 43.2 38.1 45.6 39.7 42.2 39.6 42.7 38.6 37.3 35.7 41.1 36.2
## [121] 37.7 40.2 41.4 35.2 40.6 38.8 41.5 39.0 44.1 38.5 43.1 36.8 37.5 38.1 41.1
## [136] 35.6 40.2 37.0 39.7 40.2 40.6 32.1 40.7 37.3 39.0 39.2 36.6 36.0 37.8 36.0
## [151] 41.5
```