

pset5_gonzalez

Nicholas R. Gonzalez

Problem 1

a

A *Type II* error in this hypothesis test would be if we failed to reject the null hypothesis, despite the alternative hypothesis being true. This would mean that the test's results are incorrect, and there is no real effect that exists.

b

i

```
effect <- seq(0.1, 3, by = 0.1)

variance <- seq(1, 10, by = 1)

sample <- seq(10, 200, by = 10)

effect
```

```
[1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9
[20] 2.0 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3.0
```

```
variance
```

```
[1] 1 2 3 4 5 6 7 8 9 10
```

```
sample
```

```
[1] 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190  
[20] 200
```

ii

```
library(tidyverse)
```

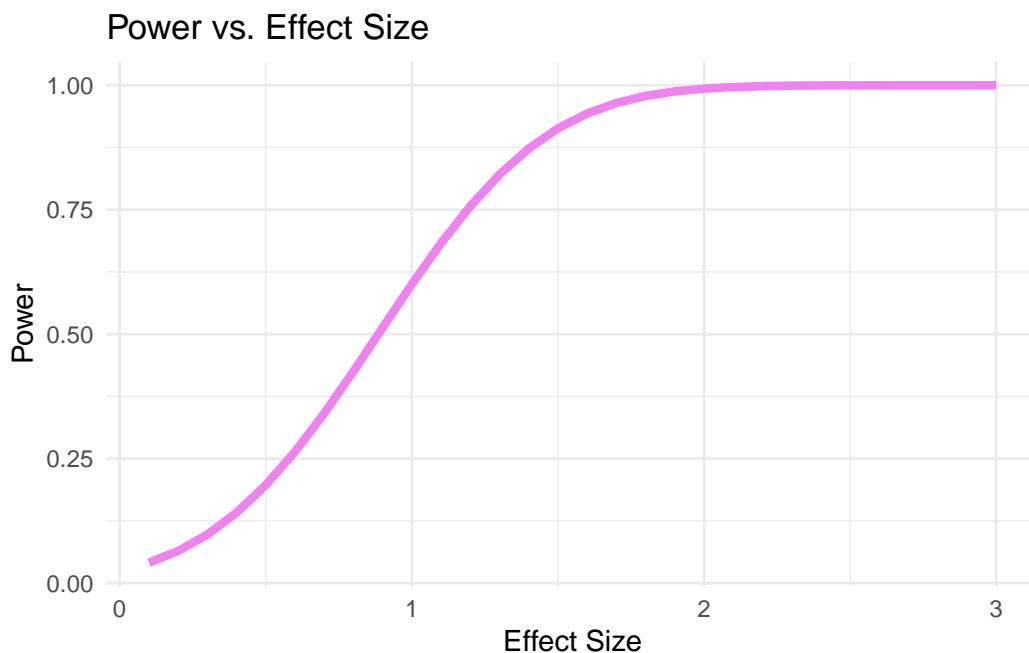
```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
v dplyr      1.1.4      v readr      2.1.5  
v forcats    1.0.0      v stringr    1.5.1  
v ggplot2     3.5.1      v tibble     3.2.1  
v lubridate  1.9.4      v tidyr      1.3.1  
v purrr       1.0.2  
-- Conflicts ----- tidyverse_conflicts() --  
x dplyr::filter() masks stats::filter()  
x dplyr::lag()     masks stats::lag()  
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(stats)
```

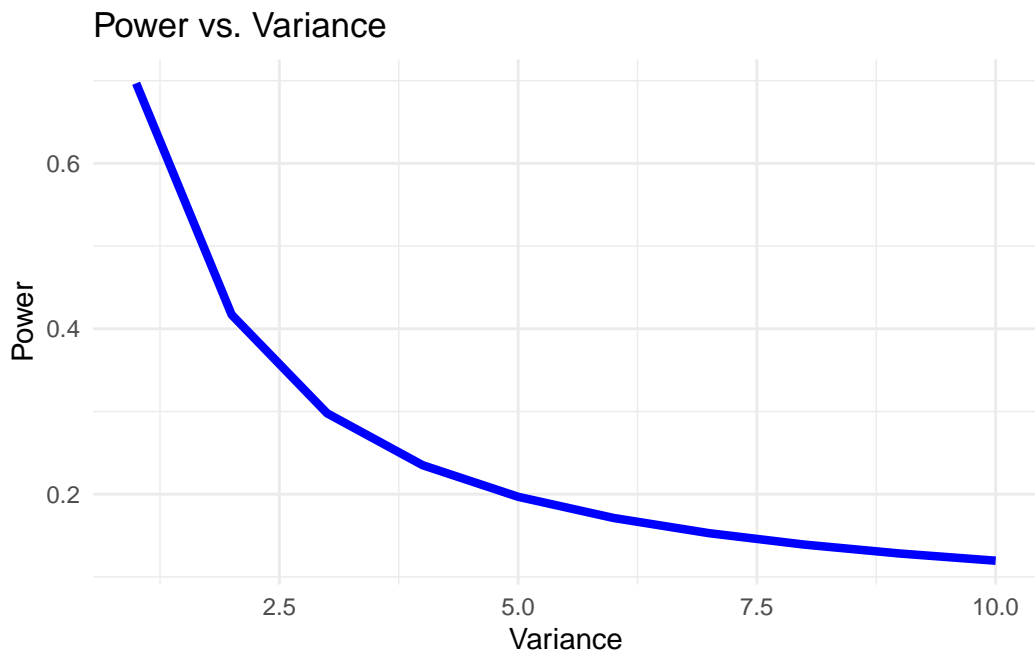
```
results <- data.frame(effect = numeric(0), var = numeric(0), sample = numeric(0), power = numeric(0))  
  
for (e in effect) {  
  for (v in variance) {  
    for (n in sample) {  
      power_result <- power.t.test(n = n, delta = e, sd = sqrt(v), sig.level = 0.05, type = "one.sided")  
      results <- rbind(results, data.frame(effect = e, variance = v, sample = n, power = power_result))  
    }  
  }  
}  
  
effect_plot <- ggplot(subset(results, variance == 5 & sample == 50), aes(x = effect, y = power)) +  
  geom_line(color = "violet", size = 1.5) +  
  labs(title = "Power vs. Effect Size", x = "Effect Size", y = "Power") +  
  theme_minimal()
```

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.

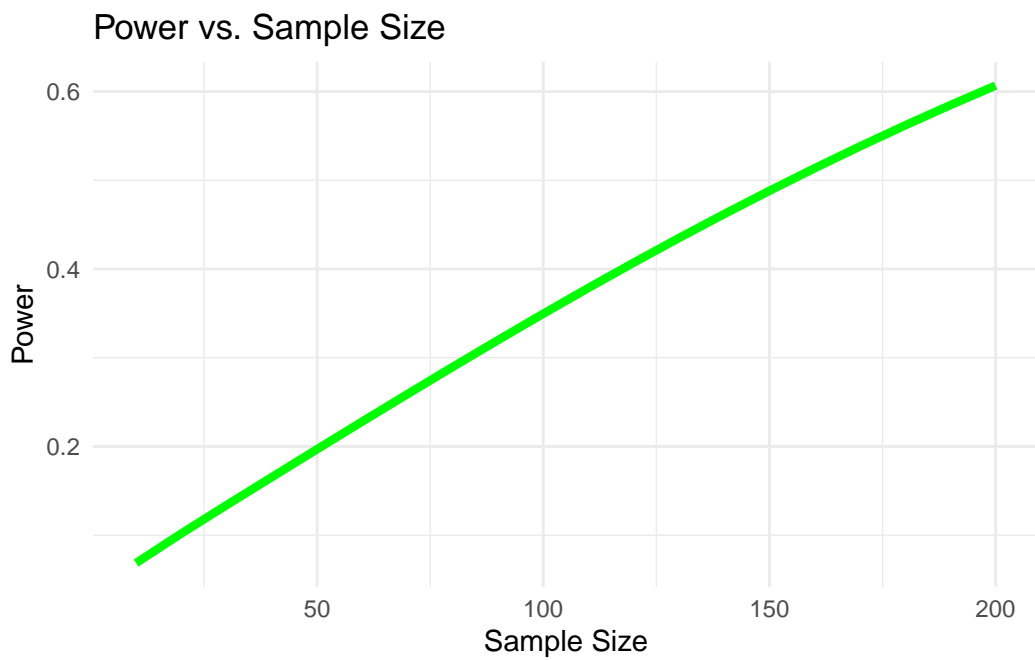
```
variance_plot <- ggplot(subset(results, effect == 0.5 & sample == 50), aes(x = variance, y =  
  geom_line( color = "blue", size = 1.5) +  
  labs(title = "Power vs. Variance", x = "Variance", y = "Power") +  
  theme_minimal()  
  
sample_plot <- ggplot(subset(results, effect == 0.5 & variance == 5), aes(x = sample, y = pow  
  geom_line(color = "green", size = 1.5) +  
  labs(title = "Power vs. Sample Size", x = "Sample Size", y = "Power") +  
  theme_minimal()  
  
print(effect_plot)
```



```
print(variance_plot)
```



```
print(sample_plot)
```



iii

As the sample size increases, the power increases. This is a linear increase.

As the effect size increases, power also increases. Something with a big effect would be easier to detect. Our graph shows that once we get to a certain power as well, the effect flat lines.

As variance increases, power decreases. This is because are data, or results become less interpretable with the more variance we have.

c

i

```
set.seed(1)

reject <- numeric(2000)

for (i in 1:2000) {
  X <- rnorm(n = 20, mean = 5, sd = sqrt(0.1))
  u <- rnorm(n = 20, mean = 0, sd = sqrt(1))
  Y <- 10 + 5*X + u
  mod <- lm(Y~X)
  p_value <- summary(mod)$coefficients[2,4]
  reject[i] <- p_value < 0.05
}

mean(reject)
```

```
[1] 0.9995
```

ii

The increase in our sample now gives our model the ability to reject 100% of the null hypothesis, as opposed to .995. The reason for this is the bigger the sample within the model, the increased likelihood of us getting the true effect. In general, tests become more precise with more data.

```
set.seed(123)

for (i in 1:2000) {
  x <- rnorm(n = 100, mean = 5, sd = sqrt(0.1))
  u <- rnorm(n = 100, mean = 0, sd = sqrt(1))
  y <- 10 + 5 * x + u
  mod <- lm(y~x)
  p_value <- summary(mod)$coefficients[2,4]
  reject[i] <- p_value < 0.05
}

mean(reject)
```

```
[1] 1
```