Lab 6

key

This lab will focus on sample size decisions.

Part 1: Power

1. Explore the Shiny app created by Christian Stratton and Jenny Green.

Set the following characteristics in the app:

- Population Distribution: **Normal**
- Test Statistic: **Sample Mean**
- Null Value: **0**
- Population standard deviation: **1**
- Alternative Hypothesis: **Not Equal to**
- Alpha Level: **.05**
- Sample Size: **10**
- Theta: **0.5**

The app will generate two figures. Download and attach those figures. I have done the first one for you. Include a short description of the figures.

```
knitr::include_graphics("power.png")
```

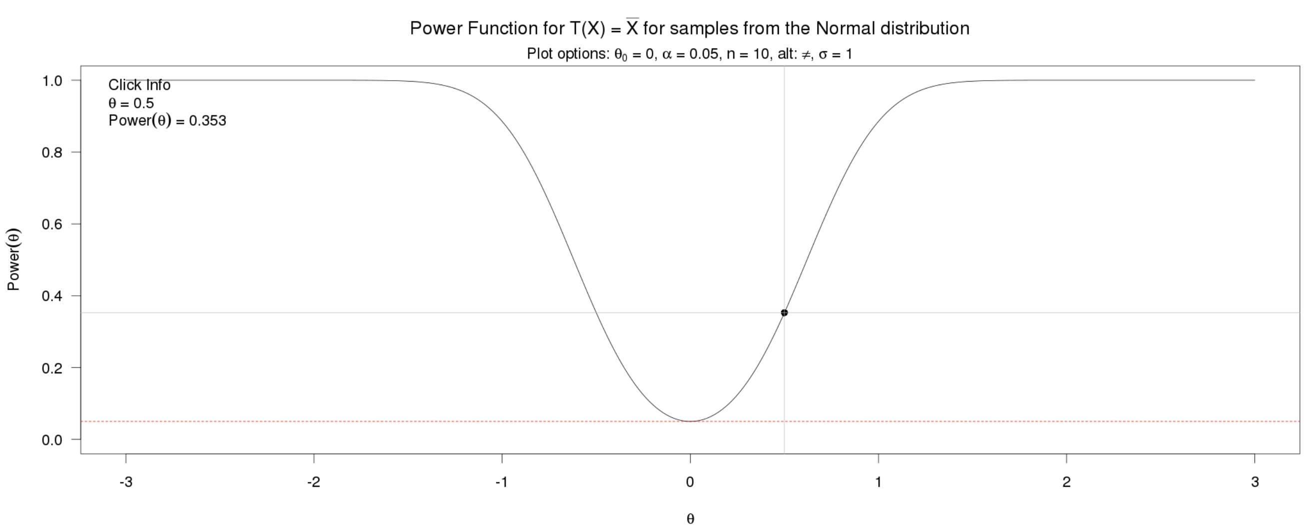


Figure 1: This figure shows the power for for detecting differences based on various alternative hypothesis (heta) values under the specified conditions. In particular, the power for theta = 1 (an actual difference of one unit is about .35).

```r
knitr::include_graphics("samp.png")
```

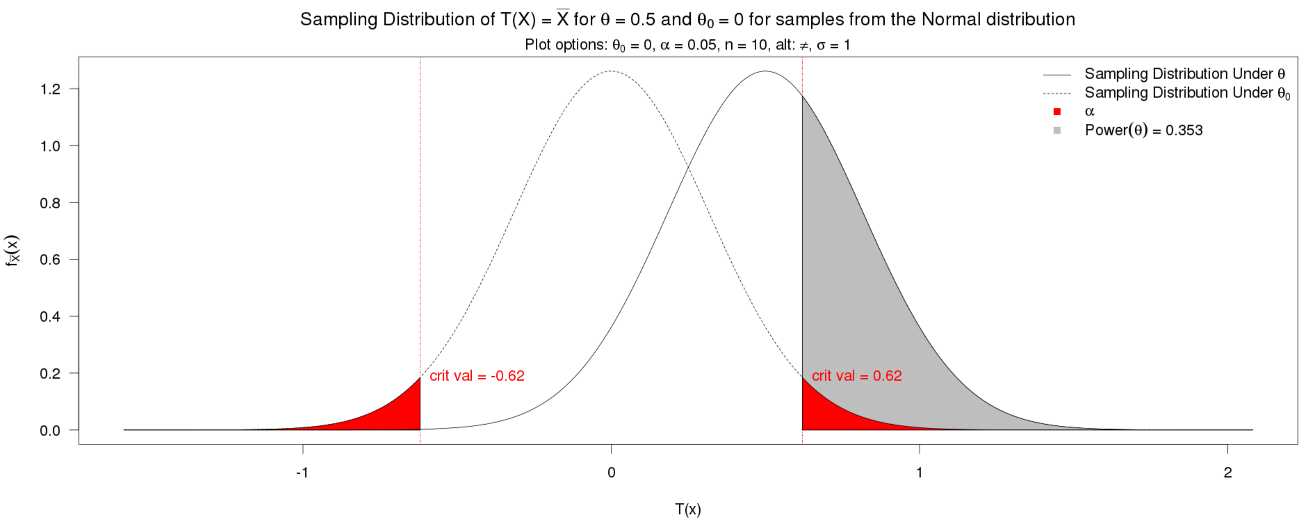


Figure 2: This figure shows the sampling distribution under the null and alternative hypotheses.

2. Write R code to calculate the power with the following. Note this can apply to our paired t-test type of experiments.

- Population Distribution: **Normal**
- Test Statistic: **Sample Mean**
- Null Value: **0**
- Population standard deviation: **1**
- Alternative Hypothesis: **Not Equal to**
- Alpha Level: **.05**
- Sample Size: **10**
- Theta: **0.5**

The sampling distribution for the null hypothesis is

$$N(\mu_0, \sigma^2/r) = N(0, 1/r)$$

so the critical points are at `qnorm(.025, mean = 0, sd = sqrt(1/10))` = -0.619795 and `qnorm(.975, mean = 0, sd = sqrt(1/10))` = 0.619795.

The sampling distribution for the alternative hypothesis is

$$N(\mu_1, \sigma^2/r) = N(0.5, 1/r)$$

so the probability of exceed the critical points (power) is `pnorm(qnorm(.025, mean = 0, sd = sqrt(1/10)), mean = .5, sd = sqrt(1/10)) + (1 - pnorm(qnorm(.025, mean = 0, sd = sqrt(1/10)), mean = .5, sd = sqrt(1/10)))` = 0.3526081

3. Write R code to create a simulation for the power with the following:

- Population Distribution: **Normal**
- Test Statistic: **Sample Mean**
- Null Value: **0**

- Population standard deviation: **1**
- Alternative Hypothesis: **Not Equal to**
- Alpha Level: **.05**
- Sample Size: **10**
- Theta: **0.5**

The sampling distribution for the null hypothesis is

$$N(\mu_0, \sigma^2/r) = N(0, 1/r)$$

so the critical points are at `qnorm(.025, mean = 0, sd = sqrt(1/10))` = -0.619795 and `qnorm(.975, mean = 0, sd = sqrt(1/10))` = 0.619795.

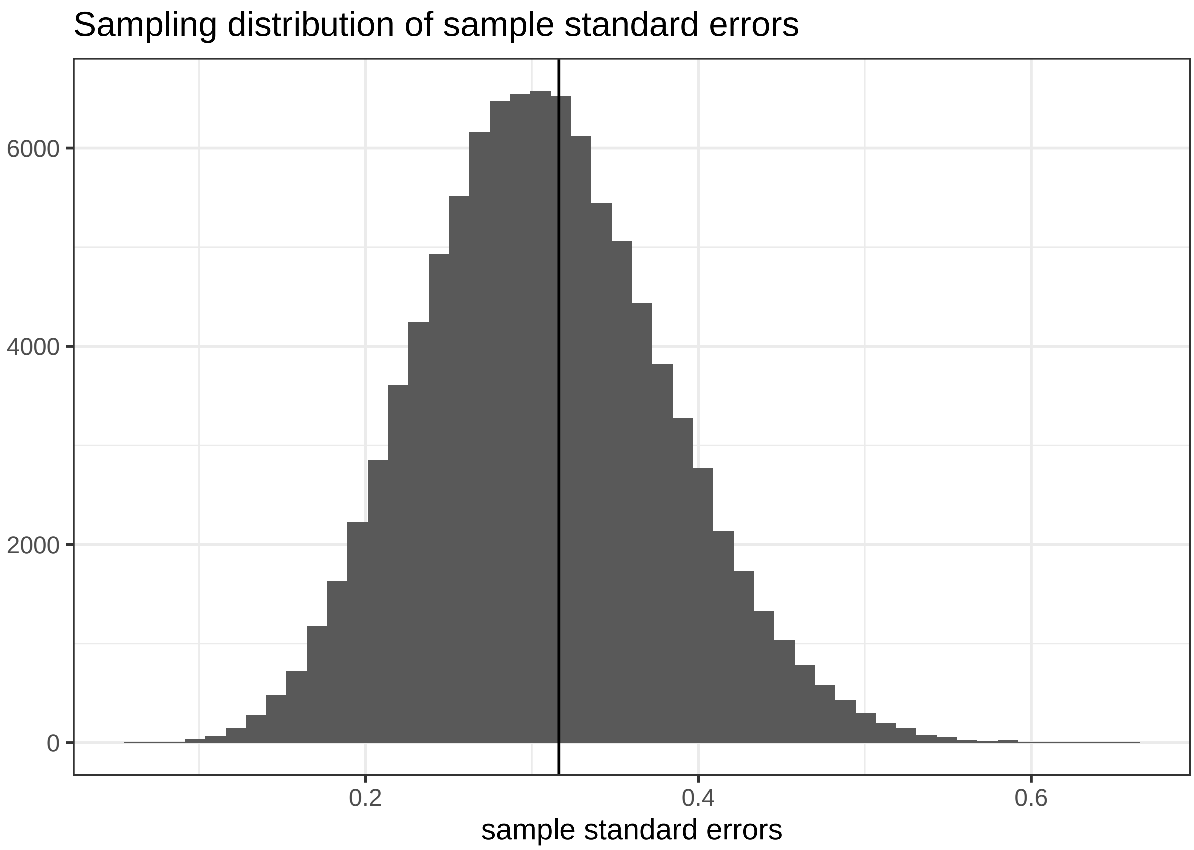
```
set.seed(03032022)
num_sims <- 100000
r <- 10
sample_means <- rowMeans(matrix(rnorm(num_sims * r, mean = .5, sd = 1 ), nrow = num_sims, ncol = r))
mean(sample_means < qnorm(.025, mean = 0, sd = sqrt(1/10)) |
       sample_means > qnorm(.975, mean = 0, sd = sqrt(1/10)) )
```

```
## [1] 0.35312
```

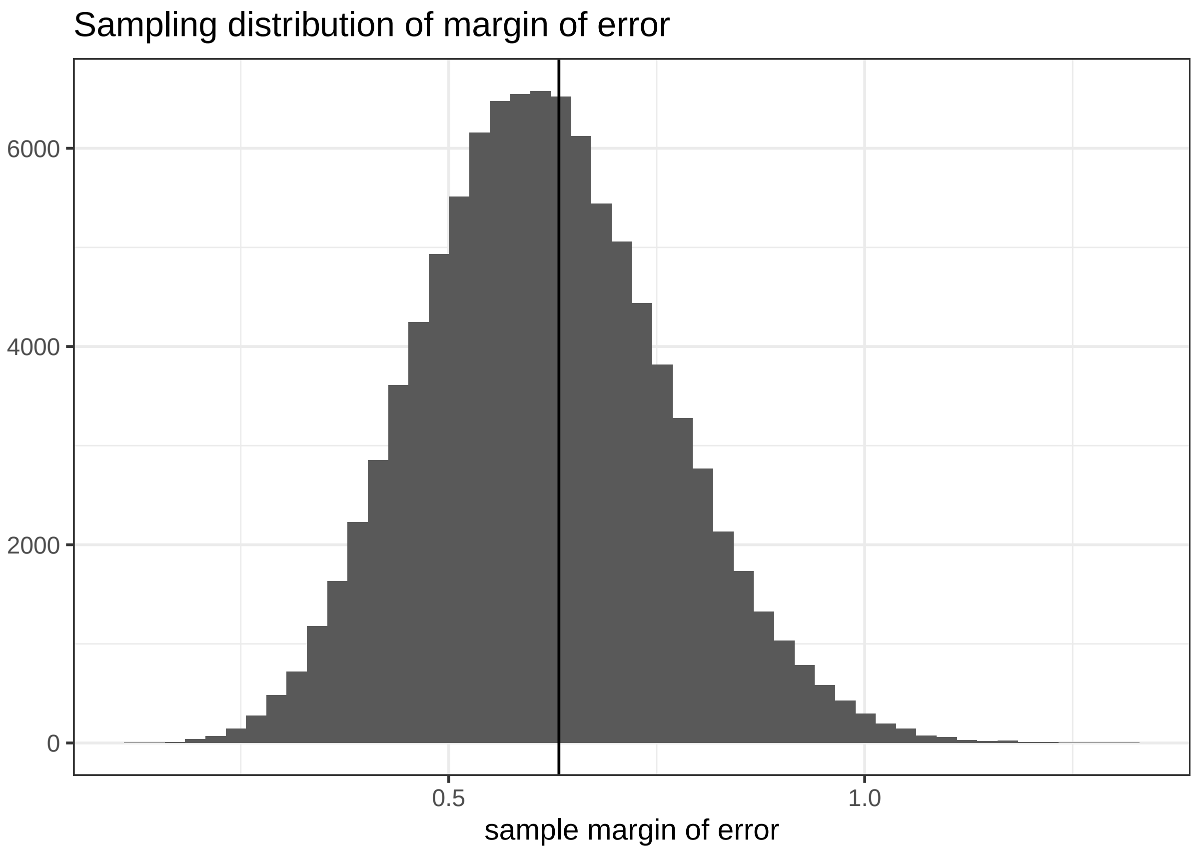
Part 2: Uncertainty Bounds

4. Under the same conditions as above, create a distribution of the estimated standard error ($\frac{\hat{\sigma}}{r}$) and uncertainty interval widths.

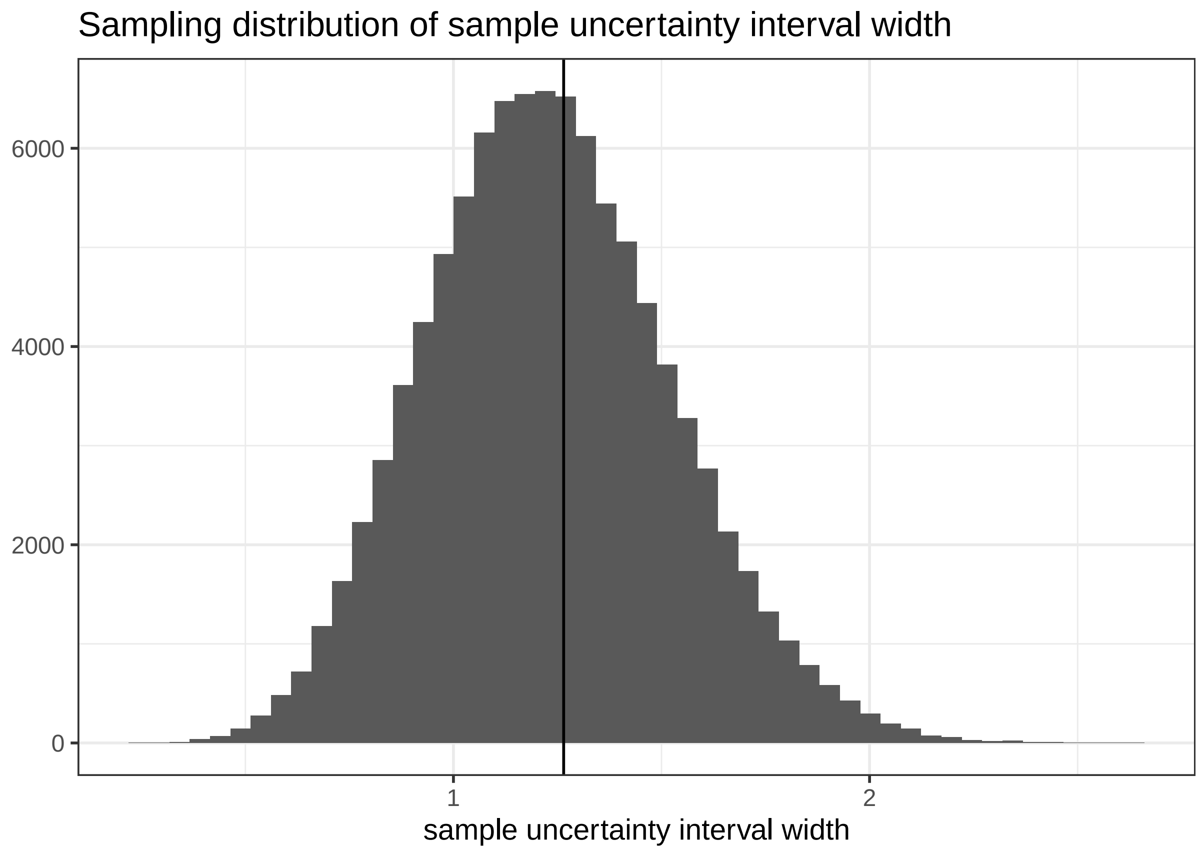
```
num_sims <- 100000
r <- 10
sample_se <- tibble(se = apply(matrix(rnorm(num_sims * r, mean = .5, sd = 1 ),
                                      nrow = num_sims, ncol = r),1, sd) / sqrt(r))

sample_se %>% ggplot(aes(x = se)) + geom_histogram(bins = 50) + theme_bw() +
  geom_vline(xintercept = sqrt(1/10)) + xlab('sample standard errors') + ylab('') +
  ggtitle("Sampling distribution of sample standard errors")
```




```
sample_se %>% ggplot(aes(x = se * 2)) + geom_histogram(bins = 50) + theme_bw() +
  geom_vline(xintercept = sqrt(1/10) * 2) + xlab('sample margin of error') + ylab('') +
  ggtitle("Sampling distribution of margin of error")
```

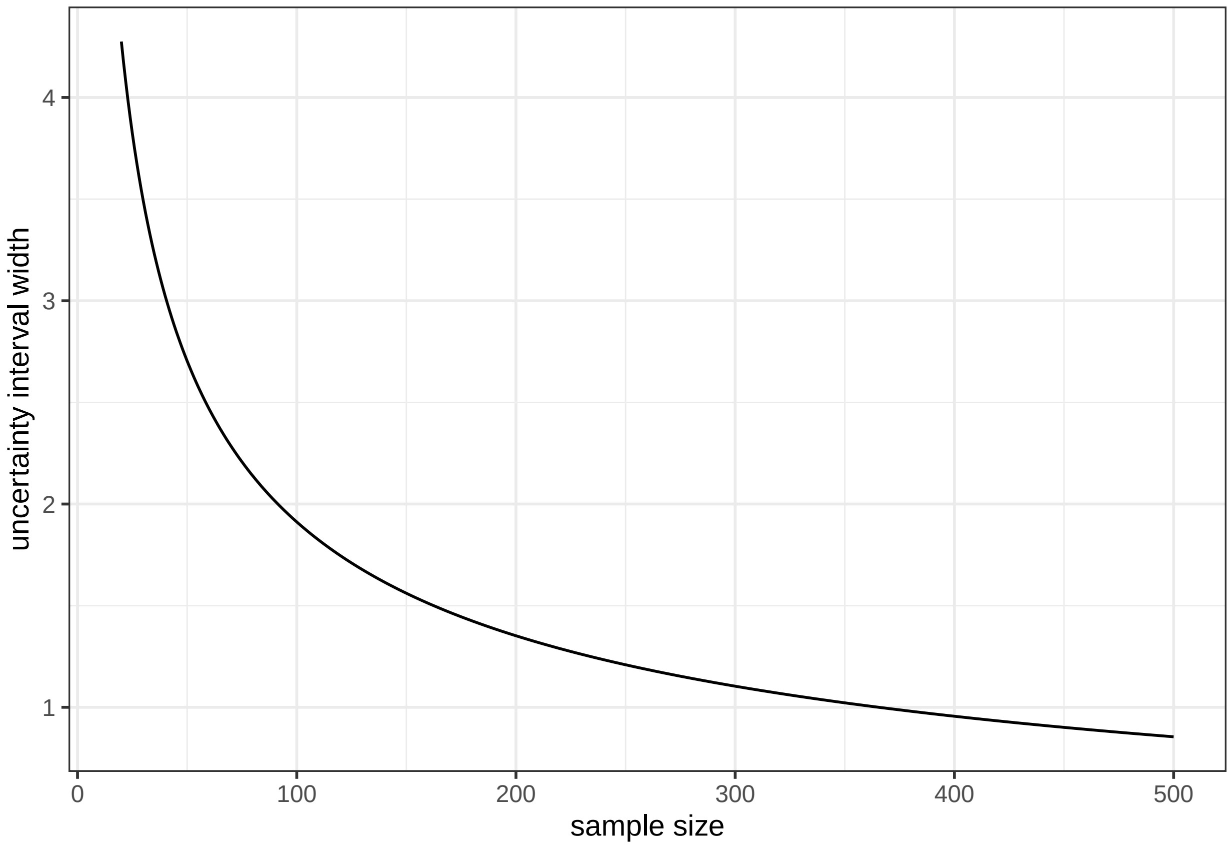



```
sample_se %>% ggplot(aes(x = se * 4)) + geom_histogram(bins = 50) + theme_bw() +
  geom_vline(xintercept = sqrt(1/10) * 4) + xlab('sample uncertainty interval width') +
  ylab('') + ggtitle("Sampling distribution of sample uncertainty interval width")
```



5. Assume you'd like your estimated width of your uncertainty interval to be less than 1. You can also assume that the standard deviation of the data from the pilot study is 4.78 (this is the value from our airplane example). Estimate how many samples will be needed to achieve this precision (possibly with probability .95).

```
sigma <- 4.78
r_seq <- 20:500
estimated_se <- 4.78 / sqrt(r_seq)
interval_width <- estimated_se * 4

combined <- tibble(r = r_seq, width = interval_width)

combined %>% ggplot(aes(x = r_seq, y = width )) +
  geom_line() + theme_bw() +
  xlab('sample size') + ylab('uncertainty interval width')
```



You'd need at least 366 samples to have an estimated uncertainty interval width less than 1