1. A group of researchers is running an experiment over the course of 30 months, with a single observation collected at the end of each month. Let $X_1, ..., X_{30}$ denote the observations for each month. From prior studies, the researchers know that

$$X_i \sim f_X(x),$$

but the mean μ_X is unknown, and they wish to conduct the following test

$$H_0 : \mu_X = 0$$
$$H_a : \mu_X > 0.$$

At month k, they have accumulated data $X_1, ..., X_k$ and they have the t-statistic

$$T_k = \frac{\bar{X} - 0}{S_k/\sqrt{n}}.$$

The initial plan was to test the hypotheses after all data was collected (at the end of month 30), at level $\alpha = 0.05$. However, conducting the experiment is expensive, so the researchers want to "peek" at the data at the end of month 20 to see if they can stop it early. That is, the researchers propose to check whether t_{20} provides statistically discernible support for the alternative. If it does, they will stop the experiment early and report support for the researcher's alternative hypothesis. If it does not, they will continue to month 30 and test whether t_{30} provides statistically discernible support for the alternative.

(a) What values of t_{20} provide statistically discernible support for the alternative hypothesis?

```
library(tidyverse)
n20 <- 20
(val <- qt(0.95, df = n20-1))

## [1] 1.729133
```

To get a p-value which is less than 0.05 which gives us a statistically discernible support for the alternative, we need a t_{20} value of greater than 1.7291.

(b) What values of t_{30} provide statistically discernible support for the alternative hypothesis?

```
n30 <- 30
(val.30 <- qt(0.95, df = n30-1))

## [1] 1.699127
```

To get a p-value which is less than 0.05 which gives us a statistically discernible support for the alternative, we need a t_{30} value of greater than 1.6991.

(c) Suppose $f_X(x)$ is a Laplace distribution with $a = 0$ and $b = 4.0$. Conduct a simulation study to assess the Type I error rate of this approach.
Note: You can use the `rlaplace()` function from the `VGAM` package for `R` (Yee, 2010).

(d) **Optional Challenge:** Can you find a value of $\alpha < 0.05$ that yields a Type I error rate of 0.05?

```
# part C
# simulation study
library(VGAM)
num.sims <- 10000
a <- 0
b <- 4.0
mu0 <- 0

# errors
```

```r
type1error.20.count <- 0
type1error.30.count <- 0
no.error.count <- 0

for(i in 1:num.sims){
  # sims for n = 20 peek, n = 30
  sim30.curr <- rlaplace(n=30, location = a, scale = b)
  sim20.curr <- sim30.curr[1:20]

  # t test for n = 20 peek
  n20.stat <- t.test(x = sim20.curr, mu = mu0, alternative = "greater")

  if (n20.stat$p.value < 0.05){
    type1error.20.count <- type1error.20.count + 1
  }else{

    n30.stat <- t.test(x = sim30.curr, mu = mu0, alternative = "greater")

    if (n30.stat$p.value < 0.05){
      type1error.30.count <- type1error.30.count + 1
    } else {
      no.error.count <- no.error.count + 1
    }
  }
}


(prop.20 <- type1error.20.count/num.sims)

## [1] 0.0494

(prop.30 <- type1error.30.count/num.sims)

## [1] 0.0269

(prop.no <- no.error.count/num.sims)

## [1] 0.9237

(tot.type1.error <- (type1error.20.count + type1error.30.count)/num.sims)

## [1] 0.0763
```

A type 1 error is described as when we accidentally find statistically discernible support for the alternative although the null is correct. This entails falsely rejecting the null hypothesis for the alternative, which is not true. As can be seen through the proportions, for our simulation we saw about 7.04% type 1 error which means that we successfully made the right descion 92.96% of the time. When checking for the type 1 error after time 20, we see a type 1 error about 4.63% of the time and a type 1 error 2.41% of the time for time 30. As can be seen in the percentages, the time 30 percentage is lower than the time 20 percentage meaning the decsion was more accurate. NOTE: The percentages are subject to change as each simulation is unique and the ones in the paragraph are from a particular trial.

2. Perform a simulation study to assess the robustness of the T test. Specifically, generate samples of size $n = 15$ from the Beta(10,2), Beta(2,10), and Beta(10,10) distributions and conduct the following hypothesis tests against the actual mean for each case (e.g., $\frac{10}{10+2}$, $\frac{2}{10+2}$, and $\frac{10}{10+10}$).

(a) What proportion of the time do we make an error of Type I for a left-tailed test?

(b) What proportion of the time do we make an error of Type I for a right-tailed test?

(c) What proportion of the time do we make an error of Type I for a two-tailed test?

(d) How does skewness of the underlying population distribution effect Type I error across the test types?

References

Yee, T. W. (2010). The VGAM package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.