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)
t.val.20.cutoff <- qt((1-.05),df= (20-1))
(t.val.20.cutoff)

## [1] 1.729133
```

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

```
t.val.30.cutoff <- qt((1-.05), df = (30-1))
(t.val.30.cutoff)

## [1] 1.699127
```

(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).

```
library(VGAM)

## Loading required package:  stats4
## Loading required package:  splines

set.seed(7272)
num.of.sims <- 1000
alpha <- 0.05
num.of.rejections <- 0

for (i in 1:num.of.sims){
  x <- rlaplace(30,location = 0,scale = 4.0)
  x.for.20.samples <- x[1:20]
  x.for.30.samples <- x[1:30]

  t.stat.for.20 <- mean(x.for.20.samples)/ (sd(x.for.20.samples)/ sqrt(20))
```

```
  t.stat.for.30 <- mean(x.for.30.samples)/ (sd(x.for.30.samples)/ sqrt(30))

  if (t.stat.for.20 > t.val.20.cutoff){
    num.of.rejections <- num.of.rejections + 1
  }
  else{
    if (t.stat.for.30 > t.val.30.cutoff){
      num.of.rejections <- num.of.rejections + 1
    }
  }
}

rate.for.type.1.errors <- num.of.rejections / num.of.sims
(rate.for.type.1.errors)

## [1] 0.098
```

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

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 $\frac{10}{10+10}$).

```
simulation.of.type.1.error.func <- function(num.sims, n, parameter.1, parameter.2, test.type = "two
  true.mean <- parameter.1 / (parameter.1 + parameter.2)
  errors <- 0

  for (i in 1:num.sims){
    dat <- rbeta(n,parameter.1,parameter.2)

    test <- t.test(dat, mu = true.mean, alternative = test.type)

    if (test$p.value < alpha){
      errors <- errors + 1
    }
  }

  return (errors/ num.sims)
}

num.sims <- 1000
n <- 15
alpha <- .05
```

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

```
left.error.10.2 <- simulation.of.type.1.error.func(num.sims, n, 10, 2, "less")
  (left.error.10.2)

## [1] 0.035

  left.error.2.10 <- simulation.of.type.1.error.func(num.sims,n, 2,10, "less")
  (left.error.2.10)

## [1] 0.079
```

```r
  left.error.10.10 <- simulation.of.type.1.error.func(num.sims,n,10,10,"less")
  (left.error.10.10)

## [1] 0.035
```

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

```r
right.error.10.2 <- simulation.of.type.1.error.func(num.sims,n, 10,2,"greater")
  (right.error.10.2)

## [1] 0.091

  right.error.2.10 <- simulation.of.type.1.error.func(num.sims,n, 2, 10, "greater")
  (right.error.2.10)

## [1] 0.032

  right.error.10.10 <- simulation.of.type.1.error.func(num.sims,n, 10,10,"greater")
  (right.error.10.10)

## [1] 0.057
```

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

```r
two.sided.error.10.2 <- simulation.of.type.1.error.func(num.sims,n, 10,2,"two.sided")
  (two.sided.error.10.2)

## [1] 0.065

  two.sided.error.2.10 <- simulation.of.type.1.error.func(num.sims,n, 2,10, "two.sided")
  (two.sided.error.2.10)

## [1] 0.062

  two.sided.error.10.10 <- simulation.of.type.1.error.func(num.sims,n, 10,10, "two.sided")
  (two.sided.error.10.10)

## [1] 0.052
```

(d) How does skewness of the underlying population distribution effect Type I error across the test types? Highly skewed distributions like Beta(10,2) and Beta(2,10) have type I error rates that are different (either above or below) from the original $\alpha = 0.05$. The symmetric Beta(10,10) has type I errors that are closer to .05, which means that the t-test is more robust with symmetric parameters for the beta distribution.

References

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