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?

```
qt(0.95,19,0) #df = 19 because df = n-1

## [1] 1.729133
```

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

```
qt(0.95,29,0) #df = 29 because df = n-1

## [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).

```
n = 30
a = 0
b = 4

early = c()
normal = c()
counter = 0


for(i in 1:10000){
curr.sample = rlaplace(n,a,b)
early.sample = curr.sample[1:20]

normal = t.test(curr.sample, alternative = "greater")[["statistic"]][["t"]]
early = t.test(early.sample, alternative = "greater")[["statistic"]][["t"]]

if (early > qt(0.95, df = 19)) {
  counter = counter + 1
} else {
  if (normal > qt(0.95, df = 29)) {
    counter = counter + 1
  }
}


}

(typeierror = counter/10000)

## [1] 0.0737
```

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

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

```
n=15
t210 = c()
t102 = c()
t1010 = c()

for(i in 1:10000){
  curr.sample210 = rbeta(n,2,10)
  curr.sample102 = rbeta(n,10,2)
  curr.sample1010 = rbeta(n,10,10)

  t210[i] = t.test(curr.sample210, mu = 1/6, alternative = "less")[["statistic"]][["t"]]
  t102[i] = t.test(curr.sample102, mu = 5/6, alternative = "less")[["statistic"]][["t"]]
  t1010[i] = t.test(curr.sample1010, mu = 0.5, alternative = "less")[["statistic"]][["t"]]

}


length(which(t210 < qt(0.05,14)))/10000 #Type 1 for Beta(2,10)

## [1] 0.0766

length(which(t102 < qt(0.05,14)))/10000 #Type 1 for Beta(10,2)

## [1] 0.0282

length(which(t1010 < qt(0.05,14)))/10000 #Type 1 for Beta(10,10)

## [1] 0.0524
```

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

```
n=15
t210 = c()
t102 = c()
t1010 = c()

for(i in 1:10000){
  curr.sample210 = rbeta(n,2,10)
  curr.sample102 = rbeta(n,10,2)
  curr.sample1010 = rbeta(n,10,10)

  t210[i] = t.test(curr.sample210, mu = 1/6, alternative = "greater")[["statistic"]][["t"]]
  t102[i] = t.test(curr.sample102, mu = 5/6, alternative = "greater")[["statistic"]][["t"]]
  t1010[i] = t.test(curr.sample1010, mu = 0.5, alternative = "greater")[["statistic"]][["t"]]

}


length(which(t210 > qt(0.95,14)))/10000 #Type 1 for Beta(2,10)

## [1] 0.0315

length(which(t102 > qt(0.95,14)))/10000 #Type 1 for Beta(10,2)

## [1] 0.0817

length(which(t1010 > qt(0.95,14)))/10000 #Type 1 for Beta(10,10)

## [1] 0.0461
```

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

```
n=15
t210 = c()
t102 = c()
t1010 = c()

for(i in 1:10000){
  curr.sample210 = rbeta(n,2,10)
  curr.sample102 = rbeta(n,10,2)
```

```r
  curr.sample1010 = rbeta(n,10,10)

  t210[i] = t.test(curr.sample210, mu = 1/6)[["statistic"]][["t"]]
  t102[i] = t.test(curr.sample102, mu = 5/6)[["statistic"]][["t"]]
  t1010[i] = t.test(curr.sample1010, mu = 0.5)[["statistic"]][["t"]]

}


length(which(t210 > qt(0.975,14) | t210 < qt(0.025,14)))/10000 #Type 1 for Beta(2,10)

## [1] 0.0601

length(which(t102 > qt(0.975,14) | t102 < qt(0.025,14)))/10000 #Type 1 for Beta(10,2)

## [1] 0.0592

length(which(t1010 > qt(0.975,14) | t1010 < qt(0.025,14)))/10000 #Type 1 for Beta(10,10)

## [1] 0.0525
```

(d) How does skewness of the underlying population distribution effect Type I error across the test types?
Explanation: If the distribution is skewed towards the tail of the test than that increases the p-value thereby increasing the probability of a type 1 error. This is why we see a higher type I error for the beta(2,10) distribution in the left-tailed test and the higher type I error in the beta(2,10) distribution in the right-tailed test. This equalizes a little bit in the two-tailed test however it still increases it because there is more data in the tails for skewed distributions proving why the beta(10,10) distribution has the lowest Type I error in this test.

References

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