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)
library(VGAM)

## Warning:  package 'VGAM' was built under R version 4.4.3
## Loading required package:  stats4
## Loading required package:  splines

#Question 1

#Part a
t_val20 = qt(0.95, df=19)
t_val20

## [1] 1.729133

#Part b
t_val30 = qt(0.95, df=29)
t_val30

## [1] 1.699127

#Part c
R=10000
err20 = 0
err30 = 0

for(i in 1:R){
set.seed(7272+i)
x20 = rlaplace(20,0,4) #Calculating sample
t20 = mean(x20) / (sd(x20)/sqrt(20)) #Calculating t20
if (t20>t_val20) #If type1 error, add 1
  err20=err20+1

x30 = rlaplace(30,0,4) #Calculating sample
t30 = mean(x30) / (sd(x30)/sqrt(30)) #Calculating t30
if (t30>t_val30)
  err30=err30+1
}
#Calculating percentage of t error
perc_err20 = err20/R
perc_err30= err30/R

perc_err20

## [1] 0.0507

perc_err30

## [1] 0.05
```

t values that satisfy $|t| > 1.729$ provide statistically discernible support.

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

t values that satisfy $|t| > 1.699$ provide statistically discernible support.

(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).

When simulating type I error, I calculating a 0.0507 rate of error when peeking at the data, and a 0.05 error rate when examining the data after 30 months. This suggests that looking at the data at 20 months increases the chance of a type I error.

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

```
############################
#####    QUESTION 2    #####
############################
n=15 #Initializing
R=10000
mean1 = 10/12
mean2 = 2/12
mean3 = 10/20
err1_left=0
err2_left=0
err3_left=0

err1_right=0
err2_right=0
err3_right=0

err1_two=0
err2_two=0
err3_two=0
for(i in 1:R){
  set.seed(7272+i)
  sample1 = rbeta(n, 10, 2) #Calculating samples
  sample2 = rbeta(n, 2, 10)
  sample3 = rbeta(n, 10, 10)

  #Calculating t values
  t1 = (mean(sample1)-mean1)/(sd(sample1)/sqrt(n))
  t2 = (mean(sample2)-mean2)/(sd(sample2)/sqrt(n))
  t3 = (mean(sample3)-mean3)/(sd(sample3)/sqrt(n))
  #Checking if type 1 error occurred for left test
  if (t1<qt(0.05, df=14))
    err1_left=err1_left+1
  if (t2<qt(0.05, df=14))
    err2_left=err2_left+1
  if (t3<qt(0.05, df=14))
    err3_left=err3_left+1


  #Checking if type 1 error occurred for right test
  if (t1>qt(0.95, df=14))
    err1_right=err1_right+1
  if (t2>qt(0.95, df=14))
    err2_right=err2_right+1
  if (t3>qt(0.95, df=14))
    err3_right=err3_right+1


  #Checking if type 1 error occurred for two tailed test
  if (abs(t1)>qt(0.975, df=14))
    err1_two=err1_two+1
  if (abs(t2)>qt(0.975, df=14))
    err2_two=err2_two+1
  if (abs(t3)>qt(0.975, df=14))
    err3_two=err3_two+1

}

err1_left/R
```

```
## [1] 0.0303

err2_left/R

## [1] 0.0769

err3_left/R

## [1] 0.0477
```

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

For the left-tailed test, we make a type I error at a rate of 0.0303, 0.0769, and 0.0477 for respective parameter values of Beta(10,2), Beta(2,10), and Beta(10,10).

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

```
err1_right/R

## [1] 0.0796

err2_right/R

## [1] 0.029

err3_right/R

## [1] 0.049
```

For the right-tailed test, we make a type I error at a rate of 0.0796, 0.0290, and 0.0490 for respective parameter values of Beta(10,2), Beta(2,10), and Beta(10,10).

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

```
err1_two/R

## [1] 0.0619

err2_two/R

## [1] 0.0582

err3_two/R

## [1] 0.0503
```

For the two-tailed test, we make a type I error at a rate of 0.0619, 0.0582, and 0.0503 for respective parameter values of Beta(10,2), Beta(2,10), and Beta(10,10).

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

We can see that skewness can cause left-tailed versus right-tailed to have a noticeably different type I error rates. For the left skewed beta(10,2) distribution, there are more type I errors when conducting a right-tailed test compared to a left-tailed test. The opposite is true for the right skewed beta(2,10) distribution. For the beta(10,10) distribution, where the distribution is not skewed, there is not a large difference between type I error across differing test types. Further, we can notice that the type I error rate for the differently skewed distributions is similar for the two-tailed test because their increased likelihoods for opposing tests average out.

References

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