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)
n <- 20
(val.20 <- qt(0.95,df = n-1))

## [1] 1.729133
```

To produce a one-sided p-value that is below 0.05 such that it provides statistically discernible support for the alternative hypothesis, the value of t_{20} would be 1.729 or greater. Since the researchers set up a right-tailed test, to reject the null, we need to find the t-statistic that is equal to 1.73 or larger, meaning beyond 0.05.

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

```
n2 <- 30
(val.30 <- qt(0.95,df = n2-1))

## [1] 1.699127
```

For statistically discernible support of the alternative hypothesis, we would need the t-test statistic to be 1.699 or larger.

(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).

```
## Part C: Laplace distribution a = 0, b = 4.0 ##
library(VGAM)
a <- 0
b <- 4.0
mu0 <- 0
n.simulations <- 1000
type1.count.20 <- 0
type1.count.30 <- 0
no.type1 <- 0
```

```r
for (i in 1:n.simulations){
  simulations <- rlaplace(n = 30, location = a, scale = b)

  # Up to 20 months
  first.20 <- simulations[1:20]
  t <- t.test(first.20,
              mu = mu0,
              alternative = "greater")

  if (t$p.value < 0.05){
    type1.count.20 <- type1.count.20 + 1
  }else{

    # 30 months
    t.2 <- t.test(simulations,
                  mu = mu0,
                  alternative = "greater")
    if (t.2$p.value < 0.05){
      type1.count.30 <- type1.count.30 + 1
    } else {
      no.type1 <- no.type1 + 1
    }
  }
}
proportion.20 <- type1.count.20/n.simulations
proportion.30 <- type1.count.30/n.simulations
no.type1.error <- no.type1/n.simulations
type1.error <- (type1.count.20 + type1.count.30)/n.simulations

type1.error.results <- tibble(
  Result = c("Type 1 Error Proportion (20)",
             "Fail to reject null at time 20, Discernible support Ha at time 30",
             "No Type 1 Error Proportion (30)",
             "Overall Type 1 Error"
             ),
  Proportion = c(round(proportion.20,3),
                 round(proportion.30,3),
                 round(no.type1.error,3),
                 round(type1.error,3))
)
library(xtable)
table1 <- xtable(type1.error.results,
                 caption = "Type 1 Error Proportions",
                 label = "tab:type1error")
```

Result	Proportion
Type 1 Error Proportion (20)	0.05
Fail to reject null at time 20, Discernible support Ha at time 30	0.03
No Type 1 Error Proportion (30)	0.92
Overall Type 1 Error	0.08

Table 1: Type 1 Error Proportions

A Type 1 error is when when we incorrectly find statistically discernible support for the alternative hypothesis. This means that the null hypothesis is true and we reject the null hypothesis in favor of the alternative. The researchers assume the null hypothesis to be $\mu_0 = 0$. For a Type 1 error to occur in this experiment, we want to reject the null when it is actually true, meaning the true mean is 0 but our simulated mean does not reflect that. The Type 1 error is larger when checked at time 20 with 4.7% of simulations incorrectly reject the null on the initial "peek". The Type 1 error for the end of the experiment at time 30 where there was not discernible support for the alternative at time 20 and discernible support for the alternative at time 30 (Type 1 Error) is smaller than initial "peek": 2.4%. We also found that in 92.9% of the simulations, there is not a Type 1 Error. The overall Type 1 error rate is high, at 7.1%, which is higher than 5%, indicating that Type 1 error is inflated when the researchers "peek" at the data.

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

```r
  # preform simulation study for assessment of robustness of T test
n <- 15
n.simulations <- 10000
# Beta (10,2)
true.mean1 <- 10/(10+2)
# Beta (2,10)
true.mean2 <- 2/(2+10)
# Beta (10,10)
true.mean3 <- 10/(10+10)

# Counters
leftcount <- c(0,0,0)
rightcount <- c(0,0,0)
dualcount <- c(0,0,0)

for (i in 1:n.simulations){
  beta1 <- rbeta(n, 10, 2)
  beta2 <- rbeta(n, 2, 10)
  beta3 <- rbeta(n, 10, 10)

  # Left-Tailed Test
  left1 <- t.test(beta1,
                  mu = true.mean1,
                  alternative = "less")
  left2 <- t.test(beta2,
                  mu = true.mean2,
                  alternative = "less")
  left3 <- t.test(beta3,
                  mu = true.mean3,
                  alternative = "less")

  # Type 1 Error Count
  leftcount[1] <- leftcount[1] + (left1$p.value < 0.05)
  leftcount[2] <- leftcount[2] + (left2$p.value < 0.05)
  leftcount[3] <- leftcount[3] + (left3$p.value < 0.05)
```

```r
  # Right-Tailed Test
  right1 <- t.test(beta1,
                   mu = true.mean1,
                   alternative = "greater")
  right2 <- t.test(beta2,
                   mu = true.mean2,
                   alternative = "greater")
  right3 <- t.test(beta3,
                   mu = true.mean3,
                   alternative = "greater")

  # Type 1 Error Count
  rightcount[1] <- rightcount[1] + (right1$p.value < 0.05)
  rightcount[2] <- rightcount[2] + (right2$p.value < 0.05)
  rightcount[3] <- rightcount[3] + (right3$p.value < 0.05)


  # Two-Tailed Test
  dual1 <- t.test(beta1,
                  mu = true.mean1,
                  alternative = "two.sided")
  dual2 <- t.test(beta2,
                  mu = true.mean2,
                  alternative = "two.sided")
  dual3 <- t.test(beta3,
                  mu = true.mean3,
                  alternative = "two.sided")

  # Type 1 Error Count
  dualcount[1] <- dualcount[1] + (dual1$p.value < 0.05)
  dualcount[2] <- dualcount[2] + (dual2$p.value < 0.05)
  dualcount[3] <- dualcount[3] + (dual3$p.value < 0.05)

}
```

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

```r
## Part A: Proportion of Time we make a Type 1 error for left-tailed test ##
type1.proportion.left <- leftcount/n.simulations
(round(type1.proportion.left, 4))

## [1] 0.0316 0.0851 0.0491
```

We make a Type 1 Error for the left-tailed test 3.07% (Beta(10,2)), 7.91% (Beta(2,10)), and 4.8% (Beta(10,10)) of the time.

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

```r
## Part B: Proportion of Time we make a Type 1 error for right-tailed test ##
type1.proportion.right <- rightcount/n.simulations
(round(type1.proportion.right, 4))

## [1] 0.0789 0.0311 0.0495
```

We make a Type 1 Error for the right-tailed test 8.92% (Beta(10,2)), 2.75% (Beta(2,10)), and 4.8% (Beta(10,10)) of the time.

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

```
## Part C: Proportion of Time we make a Type 1 error for two-tailed test ##
type1.proportion.dual <- dualcount/n.simulations
(round(type1.proportion.dual, 4))

## [1] 0.0625 0.0666 0.0514
```

We make a Type 1 Error for the two-tailed test 6.63% (Beta(10,2)), 6.03% (Beta(2,10)), and 5.03% (Beta(10,10)) of the time.

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

```
## Part D: Skewness of underlying population distribution effect Type 1 across test types? ##
distribution.results <- tibble(
  Distribution = c("Beta(10,2)", "Beta(2,10)", "Beta(10,10)"),
  `Left-Tailed Test` = type1.proportion.left,
  `Right-Tailed Test` = type1.proportion.right,
  `Two-Tailed Test` = type1.proportion.dual
)
library(xtable)
table <- xtable(distribution.results,
                caption = "Type 1 Error Comparison Across Tests",
                label = "tab:distresults")
```

Distribution	Left-Tailed Test	Right-Tailed Test	Two-Tailed Test
Beta(10,2)	0.03	0.08	0.06
Beta(2,10)	0.09	0.03	0.07
Beta(10,10)	0.05	0.05	0.05

Table 2: Type 1 Error Comparison Across Tests

Table 2 shows the three distributions and the differences in the tests cases for the Type 1 proportions. Beta(10,2) is left-skewed which results in the left-tailed test rejecting the null less often than 5%; the right-tailed test rejects more often, and the two-tailed test is slightly inflated. Beta(2,10) is right-skewed which results in the left-tailed test rejecting the null more often than 5%; the right-tailed test rejects less often, and the two-tailed test is slightly inflated. Beta(10,10) is symmetric which results in the all of the tests being close to 5%.

References

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