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?

```
(t20.value <- qt(0.95, df = 20-1))

## [1] 1.729133
```

All t_{20} values greater than or equal to 1.729133 provide statistically discernible support for the alternative hypothesis.

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

```
(t30.value <- qt(0.95, df = 30-1))

## [1] 1.699127
```

All t_{30} values greater than or equal to 1.699127 provide statistically discernible support for the alternative hypothesis.

(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)
a = 0
b = 4.0
type1.20 <- 0
type1.30 <- 0
no.error <- 0

for (i in 1:10000){
  simulations <- rlaplace(n = 30, location = a, scale = b)

  t <- t.test(simulations[1:20],
              mu = 0,
              alternative = "greater")
```

```r
  # Outcome 1:
  if (t$p.value < 0.05){
    type1.20 <- type1.20 + 1
  } else {
    # Outcome 2:
    t2 <- t.test(simulations[1:30],
                 mu = 0,
                 alternative = "greater")

    if (t2$p.value < 0.05){
      type1.30 <- type1.30 + 1
    } else {
      # Outcome 3:
      no.error <- no.error + 1
    }
  }
}
(errors.20 <- type1.20/10000)

## [1] 0.0475

(errors.30 <- type1.30/10000)

## [1] 0.0258

(no.errors <- no.error/10000)

## [1] 0.9267

(type1.error <- (type1.20 + type1.30)/10000)

## [1] 0.0733
```

The type one error rate at the end of 20 months is about 0.05 Looking at the type one error between 20 and 30 months, it decreases to about 0.025 Overall, the type one error rate from 0 months to 30 months is about 0.075, leaving the rate of not getting a type one error to be about 0.925.

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
# sample size
n <- 15
# true mean for Beta(10,2)
mean.10.2 <- 10/(10+2)
# true mean for Beta(2,10)
mean.2.10 <- 2/(2+10)
# true mean for Beta(10,10)
mean.10.10 <- 10/(10+10)
# will loop over and add to
left_tail <- c(0,0,0)
right_tail <- c(0,0,0)
two_tail <- c(0,0,0)
# start loop
for (i in 1:10000){
  beta1 <- rbeta(n, 10, 2)
```

```r
  beta2 <- rbeta(n, 2, 10)
  beta3 <- rbeta(n, 10, 10)
# part (a): What proportion of the time do we make an error of Type I for a left-tailed test?
  # Left-Tailed Test
  left.10.2 <- t.test(beta1,
                mu = mean.10.2,
                alternative = "less")
  left.2.10 <- t.test(beta2,
                mu = mean.2.10,
                alternative = "less")
  left.10.10 <- t.test(beta3,
                mu = mean.10.10,
                alternative = "less")
  # Keeping track of type 1 errors for each distribution
  left_tail[1] <- left_tail[1] + (left.10.2$p.value < 0.05)
  left_tail[2] <- left_tail[2] + (left.2.10$p.value < 0.05)
  left_tail[3] <- left_tail[3] + (left.10.10$p.value < 0.05)
# part (b): What proportion of the time do we make an error of Type I for a right-tailed test?
  # Left-Tailed Test
  right.10.2 <- t.test(beta1,
                    mu = mean.10.2,
                    alternative = "greater")
  right.2.10 <- t.test(beta2,
                    mu = mean.2.10,
                    alternative = "greater")
  right.10.10 <- t.test(beta3,
                     mu = mean.10.10,
                     alternative = "greater")
  # Keeping track of type 1 errors for each distribution
  right_tail[1] <- right_tail[1] + (right.10.2$p.value < 0.05)
  right_tail[2] <- right_tail[2] + (right.2.10$p.value < 0.05)
  right_tail[3] <- right_tail[3] + (right.10.10$p.value < 0.05)
# part (c): What proportion of the time do we make an error of Type I for a two-tailed test?
  # Left-Tailed Test
  two_tail.10.2 <- t.test(beta1,
                     mu = mean.10.2,
                     alternative = "two.sided")
  two_tail.2.10 <- t.test(beta2,
                     mu = mean.2.10,
                     alternative = "two.sided")
  two_tail.10.10 <- t.test(beta3,
                      mu = mean.10.10,
                      alternative = "two.sided")
  # Keeping track of type 1 errors for each distribution
  two_tail[1] <- two_tail[1] + (two_tail.10.2$p.value < 0.05)
  two_tail[2] <- two_tail[2] + (two_tail.2.10$p.value < 0.05)
  two_tail[3] <- two_tail[3] + (two_tail.10.10$p.value < 0.05)
}
# part (a): What proportion of the time do we make an error of Type I for a left-tailed test?
(type1.error.left <- left_tail/10000)

## [1] 0.0283 0.0830 0.0496

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

```r
(type1.error.right <- right_tail/10000)

## [1] 0.0765 0.0301 0.0512

# part (c): What proportion of the time do we make an error of Type I for a two-tailed test?
(type1.error.two_sided <- two_tail/10000)

## [1] 0.0564 0.0655 0.0509
```

(a) What proportion of the time do we make an error of Type I for a left-tailed test? Using a left-tailed test, we make a type one error for the Beta(10,2) distribution at about a rate of 0.03, at about a rate of 0.08 for the Beta(2,10) distribution, and at about a rate of 0.05 for the beta(10,10) distribution.

```r
(type1.error.left <- left_tail/10000)

## [1] 0.0283 0.0830 0.0496
```

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

```r
(type1.error.right <- right_tail/10000)

## [1] 0.0765 0.0301 0.0512
```

Using a right-tailed test, we make a type one error for the Beta(10,2) distribution at a rate of about 0.08, at a rate of 0.03 for the Beta(2,10) distribution, and at a rate of about 0.05 for the beta(10,10) distribution.

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

```r
(type1.error.two_sided <- two_tail/10000)

## [1] 0.0564 0.0655 0.0509
```

Using a two-tailed test, we make a type one error for the Beta(10,2) distribution at a rate of about 0.06, at a rate of about 0.06 for the Beta(2,10) distribution, and at a rate of about 0.05 for the beta(10,10) distribution.

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

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

Table 1: Type 1 Error Rate Comparison

We can see from the table that the Beta(10,10) distribution, which is symmetric or not skewed, has similar rates of type one errors across all three tests. However, the skewed beta distributions have varying rates of type one errors across the tests. The Beta(10,2) distribution is left-tailed, so it has a lower rate of type one errors with a left-tailed test as opposed to with a right-tailed test. Since the Beta(2,10) distribution is right-tailed, it has a lower rate of type one errors with a right-tailed test as opposed to with a left-tailed test. When a two-tailed test is used, both distributions report similar rates of type 1 error.

References

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