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?

```
set.seed(5656)
alpha <- 0.05
n1 <- 20
(t.20 <- qt(0.95, df = n1-1))

## [1] 1.729133
```

Therefore $t_{20} \in [1.729133, \infty)$ provides statistically discernible support for the alternative hypothesis.

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

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

## [1] 1.699127
```

Therefore $t_{30} \in [1.699127, \infty)$ provides 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
it <- 1000
reject.count <- 0

for(i in 1:it){
  x <- rlaplace(n2, a, b)
  t20 <- mean(x[1:n1])/(sd(x[1:n1])/sqrt(n1))
  if(t20 >= t.20){
    reject.count <- reject.count + 1
  }
}
(error.rate <- reject.count / it)

## [1] 0.052
```

Therefore the Type I error rate is 0.052.

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

1

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
a1 <- 10
a2 <- 2
b1 <- 2
b2 <- 10
mean1 <- a1/(a1+b1)
mean2 <- a2/(a2+b2)
mean3 <- a1/(a1+b2)
tcrit.one <- qt(0.95, df = n-1)
tcrit.two <- qt(0.975, df = n-1)
rej.left.10.2 <- 0
rej.left.2.10 <- 0
rej.left.10.10 <- 0
rej.right.10.2 <- 0
rej.right.2.10 <- 0
rej.right.10.10 <- 0
rej.two.10.2 <- 0
rej.two.2.10 <- 0
rej.two.10.10 <- 0

for(i in 1:it){
  b.10.2 <- rbeta(n,a1,b1)
  t.10.2 <- (mean(b.10.2)-mean1) / (sd(b.10.2)/sqrt(n))
  b.2.10 <- rbeta(n,a2,b2)
  t.2.10 <- (mean(b.2.10)-mean2) / (sd(b.2.10)/sqrt(n))
  b.10.10 <- rbeta(n,a1,b2)
  t.10.10 <- (mean(b.10.10)-mean3) / (sd(b.10.10)/sqrt(n))
  if(t.10.2 <= -tcrit.one){
    rej.left.10.2 <- rej.left.10.2 + 1
  }
  if(t.2.10 <= -tcrit.one){
    rej.left.2.10 <- rej.left.2.10 + 1
  }
  if(t.10.10 <= -tcrit.one){
    rej.left.10.10 <- rej.left.10.10 + 1
  }
  if(t.10.2 >= tcrit.one){
    rej.right.10.2 <- rej.right.10.2 + 1
  }
  if(t.2.10 >= tcrit.one){
    rej.right.2.10 <- rej.right.2.10 + 1
  }
  if(t.10.10 >= tcrit.one){
    rej.right.10.10 <- rej.right.10.10 + 1
  }
  if(t.10.2 <= -tcrit.two | t.10.2 >= tcrit.two){
    rej.two.10.2 <- rej.two.10.2 + 1
  }
  if(t.2.10 <= -tcrit.two | t.2.10 >= tcrit.two){
    rej.two.2.10 <- rej.two.2.10 + 1
  }
  if(t.10.10 <= -tcrit.two | t.10.10 >= tcrit.two){
    rej.two.10.10 <- rej.two.10.10 + 1
  }
}
(error.left.10.2 <- rej.left.10.2 / it)

## [1] 0.023

(error.left.2.10 <- rej.left.2.10 / it)

## [1] 0.074

(error.left.10.10 <- rej.left.10.10 / it)

## [1] 0.056
```

Therefore the Type I error rate for a left-tailed test ranges from 0.023 to 0.074 in our set of distributions.

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

```
(error.right.10.2 <- rej.right.10.2 / it)
## [1] 0.078
(error.right.2.10 <- rej.right.2.10 / it)
## [1] 0.022
(error.right.10.10 <- rej.right.10.10 / it)
## [1] 0.055
```

Therefore the Type I error rate for a right-tailed test ranges from 0.022 to 0.078 in our set of distributions.

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

```
(error.two.10.2 <- rej.two.10.2 / it)
## [1] 0.059
(error.two.2.10 <- rej.two.2.10 / it)
## [1] 0.052
(error.two.10.10 <- rej.two.10.10 / it)
## [1] 0.054
```

Therefore the Type I error rate for a two-tailed test ranges from 0.052 to 0.059 in our set of distributions.

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

In a right-skewed distribution, e.g. Beta(10,2), the Type I error rate from a left-tailed test is substantially lower (0.023), while the error rate from a right-tailed test is substantially higher (0.078). In a left-skewed distribution, e.g. Beta(2,10), the reverse effect is present (0.074 vs. 0.022). In a symmetric distribution, e.g. Beta(10,10), error rate is not significantly impacted by the chosen t-test (ranges from 0.054 to 0.056).

References

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