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?

```
  # part A
alpha <- 0.05
sample.sizeA <- 20
(hA.support20 = qt(p = 1-alpha, df = sample.sizeA-1))

## [1] 1.729133
```

Solution: Any value for $t_{20} \geq 1.729$ provides statistically discernible support for the alternative hypothesis.

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

```
# part B
alpha <- 0.05
sample.sizeB <- 30
(hA.support30 = qt(p = 1-alpha, df = sample.sizeB-1))

## [1] 1.699127
```

Solution: Any value for $t_{30} \geq 1.699$ 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).

```
# part C
# simulation study
mu0 <- 0
num.sims <- 10000
a <- 0
b <- 4.0

# errors
```

```r
type1error.n20.peek <- 0
type1error.n30.peek <- 0
type1error.n30.final <- 0

for(i in 1:num.sims){
# sims for n = 20 peek, n = 30
curr.sim30 <- rlaplace(n=sample.sizeB, location = a, scale = b)
curr.sim20 <- curr.sim30[1:20]

# t test for n = 20 peek
n20.test <- t.test(x = curr.sim20, mu = mu0, alternative = "greater")
n20.t <- n20.test$statistic

# t test for n = 30
n30.test <- t.test(x = curr.sim30, mu = mu0, alternative = "greater")
n30.t <- n30.test$statistic

# if hA is supported with n = 20
# if not continue on and see if hA is supported with n = 30
if (n20.t >= hA.support20){
  type1error.n20.peek <- type1error.n20.peek + 1
}

# situation where researchers do not peek the data
else if(n30.t >= hA.support30){
  type1error.n30.peek <- type1error.n30.peek + 1
}

if (n30.t >= hA.support30){
  type1error.n30.final <- type1error.n30.final + 1
}
}


(e.rate.final <- type1error.n30.final/num.sims)

## [1] 0.0505

(e.rate.peek <- (type1error.n20.peek + type1error.n30.peek)
              /num.sims)

## [1] 0.0754

# the final error rate may differ each time, due to the randomness of rlaplace()
```

Solution: When using the researchers' approach of "peeking" at the data after 20 months, the Type I error rate is 0.0754. However, when just continuing through the 30 months, the Type I error rate is 0.0505. As shown, the Type I error rate is discernibly higher when using the researchers' approach, above the accepted $\alpha = 0.05$, compared to the Type I error rate after waiting 30 months.

(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
# Problem 2
# set.seed() so results reproducible (can use xtable)
set.seed(123456)
betatype1.errors <- function(a, b){
num.sims <- 1000
alpha <- a
beta <- b
pop.mean <- alpha/(alpha + beta)
dist.skew <- (2*(beta-alpha)*sqrt(alpha+beta+1))/
             ((alpha + beta + 2)*sqrt(alpha*beta))
sample.size2 <- 15

# errors
righttail.errors <- 0
lefttail.errors <- 0
twotail.errors <- 0

for(i in 1:num.sims){
  # sims using rbeta, n = 15
  curr.sim <- rbeta(n=sample.size2, shape1 = alpha, shape2 = beta)

  # t test for right tail
  t.right <- t.test(x=curr.sim, mu = pop.mean, alternative = "greater")
  right.pval <- t.right$p.value
  # t test for left tail
  t.left <- t.test(x=curr.sim, mu = pop.mean, alternative = "less")
  left.pval <- t.left$p.value
  # t test for two tailed
  t.twotail <- t.test(x=curr.sim, mu = pop.mean, alternative = "two.sided")
  twotail.pval <- t.twotail$p.value

  # count errors
  if (right.pval < 0.05){
    righttail.errors <- righttail.errors + 1
  }
  if (left.pval < 0.05){
    lefttail.errors <- lefttail.errors + 1
  }
  if(twotail.pval < 0.05){
    twotail.errors <- twotail.errors + 1
  }
}

# find proportion
error.left <- lefttail.errors/num.sims
error.right <- righttail.errors/num.sims
error.twotail <- twotail.errors/num.sims

tibble(type = c("left tailed", "right tailed", "two tailed", "skewness"),
       errors = c(error.left, error.right, error.twotail, dist.skew)) |>
  pivot_wider(names_from = type, values_from = errors)
}
```

```r
errors.data <- tibble(bind_rows(betatype1.errors(a = 10, b = 2),
betatype1.errors(a = 2, b = 10),
betatype1.errors(a = 10, b = 10)))

# consolidate data
(errors.table <- tibble(distribution = c("Beta(10,2)", "Beta(2,10)",
                                          "Beta(10,10)"), errors.data))

## # A tibble: 3 x 5
##   distribution `left tailed` `right tailed` `two tailed` skewness
##   <chr>                <dbl>          <dbl>        <dbl>    <dbl>
## 1 Beta(10,2)           0.03           0.075        0.051   -0.921
## 2 Beta(2,10)           0.071          0.03         0.055    0.921
## 3 Beta(10,10)          0.051          0.048        0.044    0

xtable(errors.table)

## % latex table generated in R 4.4.2 by xtable 1.8-4 package
## % Wed Apr 23 15:52:08 2025
## \begin{table}[ht]
## \centering
## \begin{tabular}{rlrrrr}
##   \hline
##  & distribution & left tailed & right tailed & two tailed & skewness \\
##   \hline
## 1 & Beta(10,2) & 0.03 & 0.07 & 0.05 & -0.92 \\
##   2 & Beta(2,10) & 0.07 & 0.03 & 0.06 & 0.92 \\
##   3 & Beta(10,10) & 0.05 & 0.05 & 0.04 & 0.00 \\
##    \hline
## \end{tabular}
## \end{table}
```

distribution	left-tailed test	right-tailed test	two-tailed test	skewness
Beta(10,2)	0.030	0.075	0.051	-0.921
Beta(2,10)	0.071	0.030	0.055	0.921
Beta(10,10)	0.051	0.048	0.044	0.000

Table 1: Type I error proportions for different beta distributions, along with the skewness for each distribution.

(a) What proportion of the time do we make an error of Type I for a left-tailed test?
Solution: The Type I error proportions for a left-tailed test are 0.03 for the Beta(10, 2) distribution, 0.071 for the Beta(2, 10) distribution, and 0.051 for the Beta(10, 10) distribution.

(b) What proportion of the time do we make an error of Type I for a right-tailed test?
Solution: The Type I error proportions for a right-tailed test are 0.075 for the Beta(10, 2) distribution, 0.03 for the Beta(2, 10) distribution, and 0.048 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?
Solution: The Type I error proportions for a two-tailed test are 0.051 for the Beta(10, 2) distribution, 0.055 for the Beta(2, 10) distribution, and 0.044 for the Beta(10, 10) distribution.

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

Solution: The skewness and Type 1 error proportions of each distribution can be found in Table 1.
The Beta(10, 2) distribution is left-skewed, increasing the proportion of Type 1 errors for the right-tailed test and decreasing the proportion of Type 1 errors for the left-tailed test.
The Beta(2, 10) distribution is right-skewed, increasing the proportion of Type 1 errors for the left-tailed test and decreasing the proportion of Type 1 errors for the right-tailed test.
The Beta(10, 10) distribution is symmetric, so the proportion of Type 1 errors for all tests is just about the same.

References

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