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?

```
c20 <- qt(.95, df = 19)
print(c20)

## [1] 1.729133
```

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

```
c30 <- qt(.95, df=29)
print(c30)

## [1] 1.699127
```

(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).

```
a <- 0
b <- 4
n.sims <- 10000

#function
sim.experiment <- function() {
#30 samples
observations <- rlaplace(30, location=a, scale=b)

#t stat month 20
mean.20 <- mean(observations[1:20])
sd.20 <- sd(observations[1:20])
t.20 <- (mean.20 - 0) / (sd.20 / sqrt(20))

#rejection check at month 20
if (t.20 > c20) {
  return(1)  #reject H0, stop early
}

#t stat month 30
mean.30 <- mean(observations)
sd.30 <- sd(observations)
t.30 <- (mean.30 - 0) / (sd.30 / sqrt(30))

#reject check at month 30
if (t.30 > c30) {
  return(1)  #reject H0
} else {
```

```
  return(0)  #fail to reject H0
}
}

#run the simulation
results <- replicate(n.sims, sim.experiment())

#calc Type I error rate
type_I_error <- mean(results)
print(paste("Type I error rate:", round(type_I_error, 4)))

## [1] "Type I error rate: 0.0771"
```

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

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

   We saw Type I error rates of 0.03, 0.08 and 0.05 for the left tailed t-test under the Beta(10,2), Beta(2,10), and Beta(10,10) distributions, respectively, which we can see in Table 1.

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

   We saw Type I error rates of 0.08, 0.03 and 0.05 for the right tailed t-test under the Beta(10,2), Beta(2,10), and Beta(10,10) distributions, respectively, which we can see in Table 1.

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

   We saw Type I error rates of 0.06, 0.06 and 0.05 for the two tailed t-test under the Beta(10,2), Beta(2,10), and Beta(10,10) distributions, respectively, which we can see in Table 1.

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

Table 1: Type I error rate for Beta Distributions Depending on Test

   (d) How does skewness of the underlying population distribution effect Type I error across the test types? If we have a small sample size (n¡30) and our underlying population distribution is not normally distributed then these results indicate that the skewness distorts the sampling distribution of the t-statistic. It causes both inflated or deflated results depending on the direction of the test. However, if our underlying population is normally distributed then we would expect our results to be close to the expected level which we saw in the results.

References

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