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?

```
# (a) t-val for statistically discernible support for t20
# gives t val at 95th percentile
(val_t20 <- qt(0.95, df=19))

## [1] 1.729133
```

To determine whether the researchers should stop the experiment at month 20, we calculated the critical value of the one-sided t-test at the 5% significance level with 19 degrees of freedom (since $n = 20$). This gave a critical value of 1.729, so if $t_{20} > 1.729$, the researchers would reject H_0 and stop early.

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

```
  # (b) t-val for statistically discernible support for t30
  # gives t val at 95th percentile
  (val_t30 <- qt(0.95, 29))

## [1] 1.699127
```

If the experiment continues to month 30, the researchers use the full sample to compute a new test statistic and compare it to the critical value for a t-test with 29 degrees of freedom (since $n = 30$). This gave a critical value of 1.699, so if $t_{30} > 1.699$, they would reject H_0 at that point.

(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).

```
# (c) simulation to estimate type 1 error

library(VGAM)

## Loading required package:  stats4
## Loading required package:  splines

simulations <- 10000
alpha <- 0.05

type_1_error_count <- 0

for (i in 1:simulations) {
  # generate data from Laplace
```

```r
  data <- rlaplace(30, location=0, scale=4)

  # perform t test
  t20_result <- t.test(data[1:20], mu=0)
  t20 <- t20_result$statistic

  t30_result <- t.test(data, mu=0)
  t30 <- t30_result$statistic

  # check if we would reject the null at month 20
  # do the t-stats exceed the corresponding critical values?
    # if they do, reject the null
  if (t20 > val_t20) {
    type_1_error_count <- type_1_error_count + 1
    # if this doesnt reject, then we check t30
  } else if (t30 > val_t30) {
    type_1_error_count <- type_1_error_count + 1
  }
}

# estimate type 1 error rate
(type_1_error_rate <- type_1_error_count / simulations)

## [1] 0.0759
```

To examine the impact of this strategy, we simulated 10,000 experiments under the null hypothesis, where data were drawn from a Laplace distribution with mean zero and scale 4. In each simulation, we computed t_{20}. If it was significant, we stopped and counted it as rejection. Otherwise, we computed t_{30} and checked again. This procedure gave a type one error rate of approximately 0.073, or 7.3%, which is noticeably higher than the intended 5%. This confirms that checking twice without adjusting for it increases the overall chance of making a Type I error.

(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?

```r
  ################################################################################
# Question 2
################################################################################

simulations <- 10000
alpha <- 0.05
n <- 15

true_means <- c(10 / (10+2), 2 / (2+10), 10/(10+10))

# initialize counters
type_1_error_left <- c(0,0,0)
type_1_error_right<- c(0,0,0)
type_1_error_two_tailed <- c(0,0,0)

for (i in 1:simulations){
  # generate samples from Beta distributions
  data_beta1 <- rbeta(n, 10, 2)
  data_beta2 <- rbeta(n, 2, 10)
  data_beta3 <- rbeta(n, 10, 10)

  # conduct left tailed t test
  t_left_1 <- t.test(data_beta1, mu=true_means[1], alternative="less")
  t_left_2 <- t.test(data_beta2, mu=true_means[2], alternative="less")
  t_left_3 <- t.test(data_beta3, mu=true_means[3], alternative="less")

  # conduct right tailed t test
  t_right_1 <- t.test(data_beta1, mu=true_means[1], alternative="greater")
  t_right_2 <- t.test(data_beta2, mu=true_means[2], alternative="greater")
  t_right_3 <- t.test(data_beta3, mu=true_means[3], alternative="greater")

  # conduct two tailed t test
  t_two_1 <- t.test(data_beta1, mu=true_means[1], alternative="two.sided")
  t_two_2 <- t.test(data_beta2, mu=true_means[2], alternative="two.sided")
  t_two_3 <- t.test(data_beta3, mu=true_means[3], alternative="two.sided")
```

```r
    # count type one errors
    # Count Type I errors
    type_1_error_left[1] <- type_1_error_left[1] + (t_left_1$p.value < alpha)
    type_1_error_left[2] <- type_1_error_left[2] + (t_left_2$p.value < alpha)
    type_1_error_left[3] <- type_1_error_left[3] + (t_left_3$p.value < alpha)

    type_1_error_right[1] <- type_1_error_right[1] + (t_right_1$p.value < alpha)
    type_1_error_right[2] <- type_1_error_right[2] + (t_right_2$p.value < alpha)
    type_1_error_right[3] <- type_1_error_right[3] + (t_right_3$p.value < alpha)

    type_1_error_two_tailed[1] <- type_1_error_two_tailed[1] + (t_two_1$p.value < alpha)
    type_1_error_two_tailed[2] <- type_1_error_two_tailed[2] + (t_two_2$p.value < alpha)
    type_1_error_two_tailed[3] <- type_1_error_two_tailed[3] + (t_two_3$p.value < alpha)
  }


  type_1_error_rate_left <- type_1_error_left / simulations

  # (a) proportion of time we make a Type 1 error for left-tailed
  (type_1_error_rate_left <- type_1_error_left / simulations)

## [1] 0.0291 0.0820 0.0470
```

When using a left-tailed t-test, the Type 1 error rate was around 3% for Beta(10,2), 8% for Beta(2,10), and 5% for Beta(10,10). These results suggest that skewness can substantially influence error rates, with the right-skewed distribution (Beta(2,10)) producing an inflated rate of false positives when using a left-tailed test.

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

```r
  # (b) proportion of time we make a Type 1 error for right-tailed
(type_1_error_rate_right <- type_1_error_right / simulations)

## [1] 0.0800 0.0299 0.0533
```

For right-tailed tests, Type 1 error occurred about 8% of the time for Beta(10,2), 3% of the time for Beta(2,10), and 5% of the time for Beta(10,10). The left-skewed distribution Beta(10,2) distribution shows an elevated error rate, whereas the right-skewed Beta(2,10) suppresses it. Again, the symmetric distribution comes closest to the typical 5%.

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

```r
  # (c) proportion of time we make a Type 1 error for two-tailed
(type_1_error_rate_two_tailed <- type_1_error_two_tailed / simulations)

## [1] 0.0589 0.0629 0.0504
```

Two-tailed tests produced more stable results, with error rates of 6% for Beta(10,2), 6% for Beta(2,10), and 5% for Beta(10,10). The symmetric distribution aligns most closely with the expected Type 1 error, while skewed distributions only deviate slightly.

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

```r
 # (d) how does skewness effect Type 1 error
error_comparison <- data.frame(
  distribution = c("Beta(10,2)", "Beta(2,10)", "Beta(10,10)"),
  left_tailed = type_1_error_rate_left,
  right_tailed = type_1_error_rate_right,
  two_tailed = type_1_error_rate_two_tailed
)
```

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 1 Error Comparison

Skewness affects Type 1 error rates asymmetrically depending on the direction of the test. Right-skewed distributions like Beta(2,10) inflate the error rate for left-tailed tests and reduce it for right-tailed tests. Left-skewed distributions like Beta(10,2) show the opposite pattern. Two-tailed tests are less sensitive to skew, though they still deviate slightly from the nominal rate when the data are not symmetric. These patterns suggest that the t-test is less robust under skewed coniditions, especially for one-sided hypotheses with small samples.

References

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