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?

```
alpha = 0.05
n.20 = 20

(t.20 = qt(1 - alpha, n.20-1))

[1] 1.729133
```

The built in `qt()` function makes answering this question simple. It shows that a t value greater than approximately 1.729 with 19 degrees of freedom will get a p-value $< \alpha$ (0.05).

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

```
n.30 = 40

(t.30 = qt(1 - alpha, n.30-1))

[1] 1.684875
```

This code shows that a t value greater than approximately 1.685 with 29 degrees of freedom will get a p-value $< \alpha$ (0.05).

(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).

```
sims = 10000
err = 0
set.seed(7272)

for (i in 1:sims){
  lap = rlaplace(n = 30, location = 0, scale = 4)
  lap.20 = lap[1:20]
```

```r
  t.sim.20 = mean(lap.20)/(sd(lap.20)/sqrt(n.20))
  t.sim.30 = mean(lap)/(sd(lap)/sqrt(n.30))

  if((t.sim.20 > t.20) & (t.sim.30 < t.30)){
    err = err + 1
  }
}

(t1.err = err/sims)
```

```
[1] 0.0198
```

A type 1 error in this situation occurs if the researchers decide to peak at the results early and find a t value greater than 1.729, but if the trial continued for to 30 months they would have gotten a t value less than 1.685. In order to calculate the error rate from this approach I ran a simulation 10,000 times using `rlaplace()` from `VGAM`, and counted an error if the previously mentioned error condition was met. This error rate turned out to be 0.0198, which is low enough for researchers to make the decision to check early wihtout worrying about type 1 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?

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

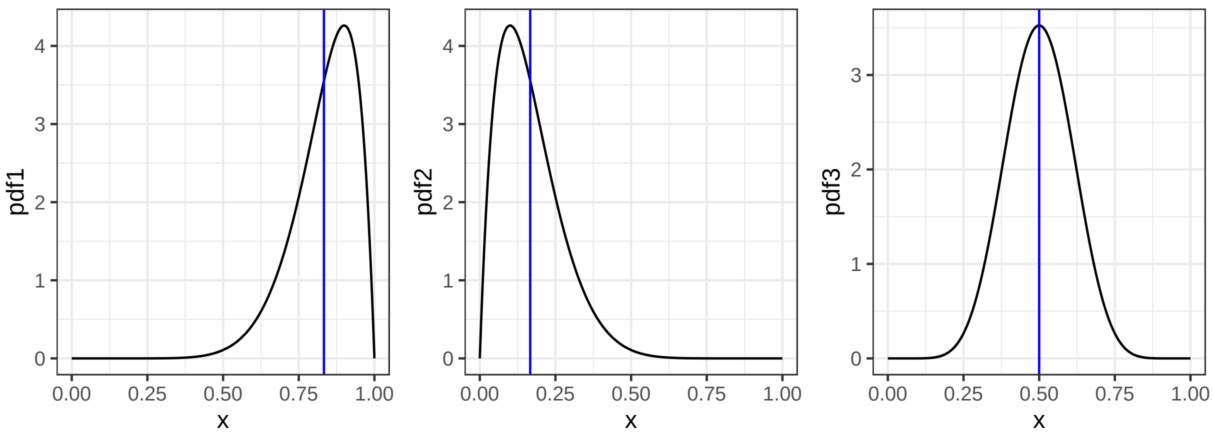
   (c) What proportion of the time do we make an error of Type I for a two-tailed test?
   In order to answer these questions efficiently I created the function below

```r
simulation = function(a, b, test){
  mu = (a/(a+b))
  err = 0

  t = c()
  mean = c()

  for (i in 1:sims){
    set.seed(7272+i)
    samp = rbeta(n = n, shape1 = a, shape2 = b)
    mean[i] = mean(samp)

    t[i] = (mean[i]-mu)/(sd(samp)/sqrt(n))
    if (t[i] < t.low & test == 'left.sided'){
      err = err + 1
    }else if (t[i] > t.high & test == 'right.sided'){
      err = err + 1
    }else if ((t[i] < t.2.side[1] | t[i] > t.2.side[2]) & test == 'two.sided'){
      err = err + 1
    }
  }
  mu.xbar = mean(mean)
  return(list(err/sims, mu.xbar))
}
```

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

α	β	μ	$\mu_{\bar{x}}$	left t error	right t error	two t error
10	2	0.8333	0.8330	0.0303	0.0796	0.0619
2	10	0.1667	0.1670	0.0796	0.0303	0.0619
10	10	0.5000	0.5001	0.0479	0.0525	0.0512

Table 1: Answers parts 1-b



The above table and graphs show that tests of the same direction as the model's skewness result in less type 1 errors while tests of the opposite direction result in more type 1 errors. In a symmetrical distribution type 1 error rate is approximately the same as alpha (confidence level not shape parameter) with a little variation due to randomness.

References

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