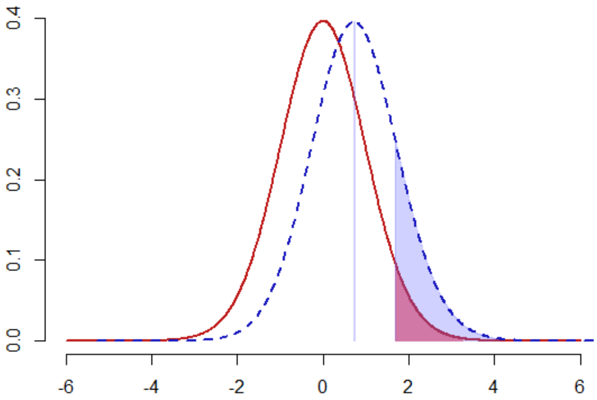
BACS HW5
Student ID: 110065508

Due: 2022/03/20
Special thanks to 108071001

Question 1. Confirming from the following plot, we cannot reject the H_{null}. The power is equal to the shaded area under the alternative distribution.



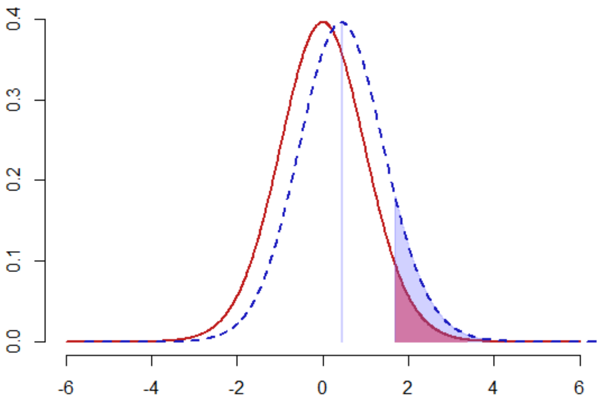
Using the R command

```
> power.t.test(n=50,delta=0.3,sd=2.9,alternative="one.sided")
```

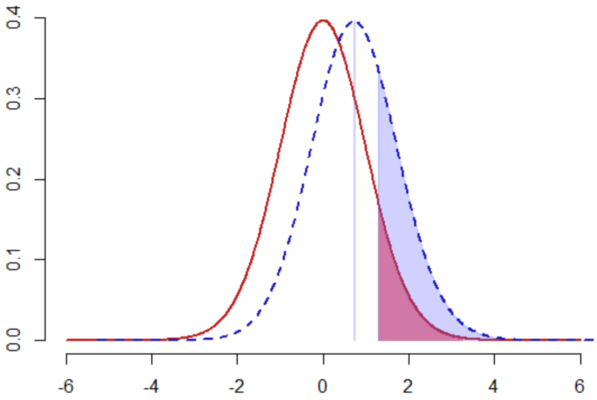
We obtain the following information

```
Two-sample t test power calculation
n = 50
delta = 0.3
sd = 2.9
sig.level = 0.05
power = 0.1289906
alternative = one.sided
```

(a) In this scenario, we should adopt stratified sampling to analysis the data. Otherwise there will be a bias (or a systematic error). So `diff` and sd will be affected. Given that the older customers use the product much less every day, `diff` increases and `sd` decreases. Hence the power increase. This leads to a decrements of β, which makes type II error becomes more unlikely to happen. As for type I error, since α is not change, it has the same likelihood to happens.

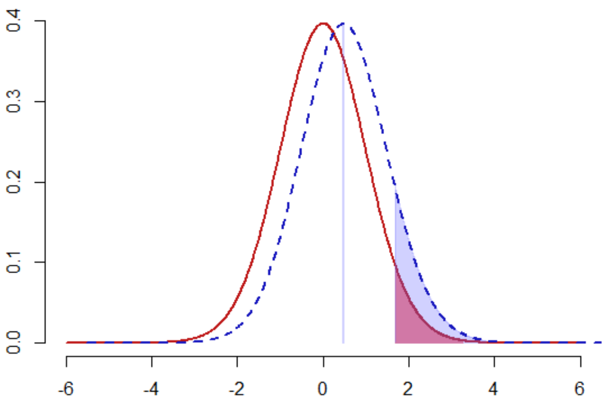


(b) According to Wikipedia [1], this is a system error. Now n decreases to 30 and sd increases. `diff` is not change accroding to the given statement. Hence the power decreases. Type II error becomes more likely to happen. This does not affect that of type I error.



(c) This does not create any error, but α is affected and drop to 0.9. The power increase and hence Both type II error becomes more unlikely to happen. However type I becomes more likely to happen.

(d) This scenario creates systematic error that cause **`diff`** to decrease. So the power decrease. Type II error becomes more unlikely to happen. As for type I error, since α is not change, it has the same likelihood to happens.



■

Question 2. First, read the data with the following codes:

```r
# Question 2
# Read the data and get some basic statistics
library(tidyverse)
ver_time <- read_csv('verizon.csv')$Time # read file
ver_mean <- mean(ver_time) # mean of sample
ver_sd <- sd(ver_time)
ver_size <- length(ver_time)
hyp <- 7.6 # Null hypothesis
```

(a) The complete code of (a) is given by:

```r
# 2(a) Recreate the traditional hypothesis test
t.test(ver_time, mu = hyp, alternative="greater", conf.level=0.99) # (i)
power.t.test(n=ver_size,
             delta=ver_mean-hyp,
             type="one.sample",
             sd=ver_sd,
```

```
                sig.level=0.01,
                alternative="one.sided") # (ii)
```

(i) The 99% confidence interval of the mean is $(7.683604, \infty)$, the t-value is $t = 2.5608$, and the p-value is 0.005265. Since the mean is $\bar{X} = 8.522009 \in (7.683604, \infty)$, we cannot reject the null hypothesis.

(ii) The power is $\beta = 0.5918705$.

(b) The complete code of (b) is given by:

```
# 2(b) bootstrap
bootstrap_null_alt <- function(sample0, hyp_mean) {
  resample <- sample(sample0, length(sample0), replace=TRUE)
  resample_se <- sd(resample) / sqrt(length(resample))
  t_stat_alt <- (mean(resample) - hyp_mean) / resample_se
  t_stat_null <- (mean(resample) - mean(sample0)) / resample_se
  c(t_stat_alt, t_stat_null)
}

# (i) original t_value
t_value <- 2.5608

# (ii) Bootstrap the null and alternative t-distributions
boot_t_stats <- replicate(2000, bootstrap_null_alt(ver_time, hyp))

# (iii) Find the 99% cutoff value
t_alt <- boot_t_stats[1,]
t_null <- boot_t_stats[2,]
ci_99 <- quantile(t_null, probs=c(0, 0.99)) # one tailed

# (iv) Compute the p-value and power of our bootstrapped test
null_probs <- ecdf(t_null)
one_tailed_pvalue <- 1 - null_probs(t_value)
alt_probs <- ecdf(t_alt)
one_tailed_power <- 1 - alt_probs(ci_99[2]) # one tailed
```

(iii) Since the one-sided 99% CI of t is $(-3.345003, 2.067544)$, the 99% cutoff value for critical null values of t is 2.067544. Also, since $t = 2.5608 \notin (-3.345003, 2.067544)$, we can reject the null hypothesis.

(iv) The power is 0.7125, the p-value is 0.001.

■

References. [1] https://en.wikipedia.org/wiki/Errors_and_residuals