

- When conducting the work of Lab 11, we conducted the test that uses the Central Limit Theorem even though the sample size was “small” (i.e.,  $n < 30$ ). It turns out, that how “far off” the  $t$ -test is can be computed using a first-order Edgeworth approximation for the error. Below, we will do this for the the further observations.

(a) Boos and Hughes-Oliver (2000) note that

$$P(T \leq t) \approx F_Z(t) + \underbrace{\frac{\text{skew}}{\sqrt{n}} \frac{(2t^2 + 1)}{6}}_{\text{error}} f_Z(t),$$

where  $f_Z(\cdot)$  and  $F_Z(\cdot)$  are the Gaussian PDF and CDF and skew is the skewness of the data. What is the potential error in the computation of the  $p$ -value when testing  $H_0 : \mu_X = 0; H_a : \mu_X < 0$  using the zebra finch further data?

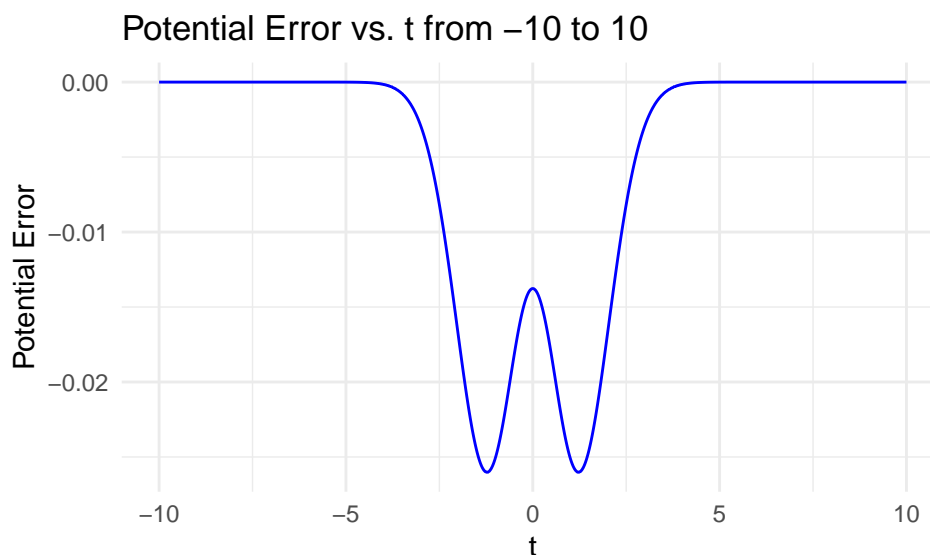
```
skew <- skewness(finches_data$further)
n <- 25
t <- t.test(finches_data$further, mu = 0,
             conf.level = 0.95, alternative = "less")$statistic
fz <- dnorm(t)
Fz <- pnorm(t)

(potential_error <- Fz+(skew/sqrt(n))*(((2*t^2)+1)/6)*fz)

##          t
## -1.189164e-13
```

The potential error is -1.189164e-13 which is approximately zero.

- Compute the error for  $t$  statistics from -10 to 10 and plot a line that shows the error across  $t$ . Continue to use the skewness and the sample size for the zebra finch further data.



- Suppose we wanted to have a tail probability within 10% of the desired  $\alpha = 0.05$ . Recall we did a left-tailed test using the further data. How large of a sample size would we need? That is, we need to solve the error formula equal to 10% of the desired left-tail probability:

$$0.10\alpha \stackrel{\text{set}}{=} \underbrace{\frac{\text{skew}}{\sqrt{n}} \frac{(2t^2 + 1)}{6}}_{\text{error}} f_Z(t),$$

which yields

$$n = \left( \frac{\text{skew}}{6(0.10\alpha)} (2t^2 + 1) f_Z(t) \right)^2.$$

```
skew <- skewness(finches_data$furthest)
a <- 0.05

t <- qnorm(0.05)
fz <- dnorm(t)

(n <- ((skew/(6*0.10*a))*((2*t^2)+1)*fz)^2)

## [1] 520.8876
```

A sample size of 520.8876 would be required. Since we cannot have a fraction of a sample, we would realistically need a sample of 521.

2. Complete the following steps to revisit the analyses from lab 11 using the bootstrap procedure.

- (a) Now, consider the zebra finch data. We do not know the generating distributions for the closer, further, and difference data, so perform resampling to approximate the sampling distribution of the  $T$  statistic:

$$T = \frac{\bar{x}_r - 0}{s/\sqrt{n}},$$

where  $\bar{x}_r$  is the mean computed on the  $r^{\text{th}}$  resample and  $s$  is the sample standard deviation from the original samples. At the end, create an object called `resamples.null.closer`, for example, and store the resamples shifted to ensure they are consistent with the null hypotheses at the average (i.e., here ensure the shifted resamples are 0 on average, corresponding to  $t = 0$ , for each case).

```
# resamples.null.closer

n <- 25
R <- 10000
resamples.null.closer <- tibble(t = rep(NA, R))

for(i in 1:R){
  curr.resample <- sample(finches_data$closer,
                        size = nrow(finches_data),
                        replace = T)

  resamples.null.closer$t[i] <- (mean(curr.resample))/(sd(finches_data$closer)/sqrt(n))
}

# Center
resamples.null.closer <- (resamples.null.closer)$t
resamples.null.closer.shifted <- resamples.null.closer - mean(resamples.null.closer)
(mean(resamples.null.closer.shifted))

## [1] -7.001941e-16

# resamples.null.further

n <- 25
R <- 10000
resamples.null.further <- tibble(t = rep(NA, R))
```

```

for(i in 1:R){
  curr.resample <- sample(finches_data$furthest,
                        size = nrow(finches_data),
                        replace = T)

  resamples.null.furthest$t[i] <- (mean(curr.resample))/(sd(finches_data$furthest)/sqrt(n))
}
# Center
resamples.null.furthest <- (resamples.null.furthest)$t
resamples.null.furthest.shifted <- resamples.null.furthest - mean(resamples.null.furthest)
(mean(resamples.null.furthest.shifted))

## [1] -1.915357e-16

# resamples.null.diff

n <- 25
R <- 10000
resamples.null.diff <- tibble(t = rep(NA, R))

for(i in 1:R){
  curr.resample <- sample(finches_data$diff,
                        size = nrow(finches_data),
                        replace = T)

  resamples.null.diff$t[i] <- (mean(curr.resample))/(sd(finches_data$diff)/sqrt(n))
}
# Center
resamples.null.diff <- (resamples.null.diff)$t
resamples.null.diff.shifted <- resamples.null.diff - mean(resamples.null.diff)
(mean(resamples.null.diff.shifted))

## [1] 6.38134e-16

```

- (b) Compute the bootstrap  $p$ -value for each test using the shifted resamples. How do these compare to the  $t$ -test  $p$ -values?

```

# Closer
t_closer <- t.test(finches_data$closer, mu = 0,
                  conf.level = 0.95, alternative = "two.sided")$statistic
(p_closer <- mean(resamples.null.closer.shifted >= t_closer))

## [1] 0

# Further
t_further <- t.test(finches_data$furthest, mu = 0,
                  conf.level = 0.95, alternative = "two.sided")$statistic
(p_further <- mean(resamples.null.furthest.shifted <= t_further))

## [1] 0

# Diff
t_diff <- t.test(finches_data$diff, mu = 0,
                conf.level = 0.95, alternative = "two.sided")$statistic
low <- -(t_diff)
high <- (t_diff)

```

```
p.low = mean(resamples.null.diff.shifted <= low)
p.high = mean(resamples.null.diff.shifted >= high)
(p_diff = p.low + p.high)

## [1] 0
```

All three of the p-values are 0 which is what we expected. In lab 11, all three of the t-test p-values were also virtually zero.

- (c) What is the 5<sup>th</sup> percentile of the shifted resamples under the null hypothesis? Note this value approximates  $t_{0.05, n-1}$ . Compare these values in each case.

```
# Closer
(firth_percentile_closer <- quantile(resamples.null.closer.shifted, .05))

##          5%
## -1.599281

(firth_percentile_closer_t <- qt(0.05, df=length(finches_data$closer-1)))

## [1] -1.708141

# Further
(firth_percentile_further <- quantile(resamples.null.further.shifted, .05))

##          5%
## -1.674602

(firth_percentile_further_t <- qt(0.05, df=length(finches_data$further-1)))

## [1] -1.708141

# Diff
(firth_percentile_diff <- quantile(resamples.null.diff.shifted, .05))

##          5%
## -1.563979

(firth_percentile_diff_t <- qt(0.05, df=length(finches_data$diff-1)))

## [1] -1.708141
```

In all three cases the values are very similar to one another which is what we expected.

- (d) Compute the bootstrap confidence intervals using the resamples. How do these compare to the t-test confidence intervals?
3. Complete the following steps to revisit the analyses from lab 11 using the randomization procedure.
- (a) Now, consider the zebra finch data. We do not know the generating distributions for the closer, further, and difference data, so perform the randomization procedure

```
# Randomization Test Closer

rand.closer <- tibble(t = rep(NA, R))

for(i in 1:R){
  curr.rand <- resamples.null.closer.shifted *
    sample(x = c(-1, 1),
           size = length(resamples.null.closer.shifted),
           replace = T)
```

```

    rand.closer$t[i] <- mean(curr.rand)
  }
  # Thinking is hard
  rand.closer.shifted <- rand.closer |>
    mutate(t = t + mean(resamples.null.closer)) # shifting back

  # Randomization Test Further

  rand.further <- tibble(t = rep(NA, R))

  for(i in 1:R){
    curr.rand <- resamples.null.further.shifted *
      sample(x = c(-1, 1),
            size = length(resamples.null.further.shifted),
            replace = T)

    rand.further$t[i] <- mean(curr.rand)
  }
  # Thinking is hard
  rand.further.shifted <- rand.further |>
    mutate(t = t + mean(resamples.null.further)) # shifting back

  # Randomization Test Diff

  rand.diff <- tibble(t = rep(NA, R))

  for(i in 1:R){
    curr.rand <- resamples.null.diff.shifted *
      sample(x = c(-1, 1),
            size = length(resamples.null.diff.shifted),
            replace = T)

    rand.diff$t[i] <- mean(curr.rand)
  }
  # Thinking is hard
  rand.diff.shifted <- rand.diff |>
    mutate(t = t + mean(resamples.null.diff)) # shifting back

```

- (b) Compute the randomization test  $p$ -value for each test.

```

# closer randomization p-value
(p_closer_randomization <- mean(rand.closer >= t_closer))

## [1] 0

# further randomization p-value
(p_further_randomization <- mean(rand.further <= t_further))

## [1] 0

# diff randomization p-value
low <- -(t_diff)
high <- (t_diff)
p.low.rand = mean(rand.diff <= low)
p.high.rand = mean(rand.diff >= high)
(p_diff_rand = p.low.rand + p.high.rand)

```

```
## [1] 0
```

- (c) Compute the randomization confidence interval by iterating over values of  $\mu_0$ .

**Hint:** You can “search” for the lower bound from  $Q_1$  and subtracting by 0.0001, and the upper bound using  $Q_3$  and increasing by 0.0001. You will continue until you find the first value for which the two-sided  $p$ -value is greater than or equal to 0.05.

4. **Optional Challenge:** In this lab, you performed resampling to approximate the sampling distribution of the  $T$  statistic using

$$T = \frac{\bar{x}_r - 0}{s/\sqrt{n}}.$$

I’m curious whether it is better/worse/similar if we computed the statistics using the sample standard deviation of the resamples ( $s_r$ ), instead of the original sample ( $s$ )

$$T = \frac{\bar{x}_r - 0}{s_r/\sqrt{n}}.$$

- (a) Perform a simulation study to evaluate the Type I error for conducting this hypothesis test both ways.
- (b) Using the same test case(s) as part (a), compute bootstrap confidence intervals and assess their coverage – how often do we ‘capture’ the parameter of interest?

## References

Boos, D. D. and Hughes-Oliver, J. M. (2000). How large does  $n$  have to be for  $z$  and  $t$  intervals? *The American Statistician*, 54(2):121–128.