

- 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

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

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?

```
# part a
zebrafinch.data <- read_csv("zebrafinches.csv")
mu0 <- 0
further.data <- zebrafinch.data$further
n <- length(further.data)

# t.test and t.stat
further.t.test <- t.test(x=further.data, mu = mu0, alternative = "less")
(t.further <- further.t.test$statistic[[1]])

## [1] -7.777991

# potential error calculation
error.num <- skewness(further.data) * (2*t.further^2 + 1) * dnorm(t.further)
error.denom <- 6 * sqrt(n)
(potential.error <- error.num/error.denom)

## [1] -1.226006e-13
```

Solution:The potential error in the computation of the p -value is $-1.2260063 \times 10^{-13}$ when testing $H_0 : \mu_X = 0; H_a : \mu_X < 0$ using the zebra finch further data.

- 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.

```
# part b
gg.errors <- rep(NA, length.out = 1000)
gg.tvals <- seq(-10,10,length.out=1000)

# create data for errors (further data)
for (i in 1:length(gg.tvals)){
  num <- skewness(further.data) * (2*gg.tvals[i]^2 + 1) * dnorm(gg.tvals[i])
  denom <- 6 * sqrt(n)
  gg.errors[i] <- num/denom
}

# plot
errors.plot <- ggplot()+
  geom_line(aes(x= gg.tvals, y = gg.errors))+
  theme_bw()+
```

```
ylab("Potential Error")+
xlab("t")+
ggtitle("Potential Error for t, from -10 to 10")+
geom_vline(aes(xintercept = t.further), color = "red")
```

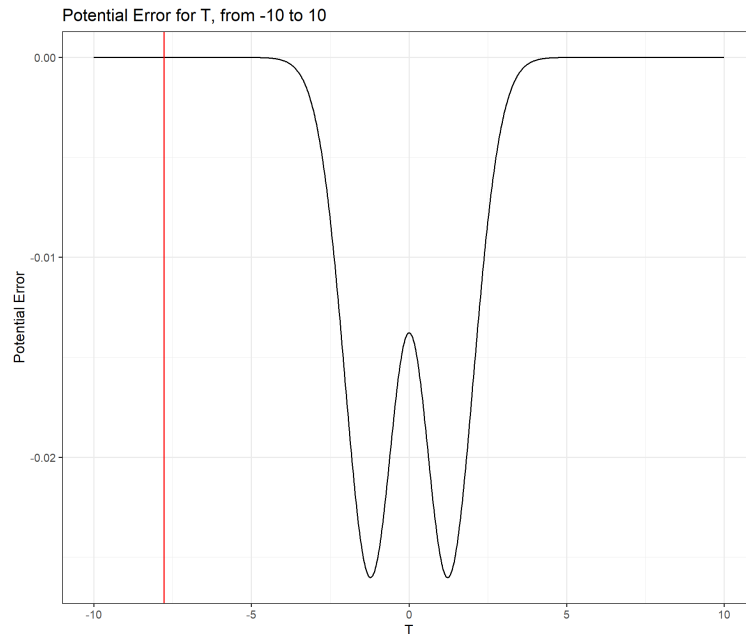


Figure 1: Potential error for t statistic ranging from -10 to 10, using the skewness of the zebra finch further data. The vertical line shows the t statistic for the zebra finch further data

Solution: Figure 1 plots the error for the t statistic, ranging from -10 to 10. The t statistic for the further data is -7.7779912, indicating an exceedingly small error (nearly 0).

- (c) 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:

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

which yields

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

```
alpha <- 0.05
t.alpha <- qnorm(alpha)

(min.nsize <- ((skewness(further.data) * (2*t.alpha^2 + 1) * dnorm(t.alpha))/
  (6 * 0.1 * alpha))^2)

## [1] 520.8876
```

Solution: The smallest sample size we would need to have a tail probability within 10% of the desired $\alpha = 0.05$ is 521. The experiment's $n = 25$ is significantly smaller than this.

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^{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).

```
# part a
R <- 1000
# data
further.data <- zebrafinch.data$further
closer.data <- zebrafinch.data$closer
diff.data <- zebrafinch.data$diff
# set.seed() to make results reproducible
set.seed(13345)
# resampling
resamples <- tibble(resamps.further = rep(NA, R),
                    resamps.closer = rep(NA, R),
                    resamps.diff = rep(NA, R))
for (i in 1:R){
  # further
  further.resample <- sample(x = further.data,
                           size = length(further.data),
                           replace = T)
  resamples$resamps.further[i] <- (mean(further.resample - mu0))/
    (sd(further.data)/sqrt(n))

  # closer
  closer.resample <- sample(x = closer.data,
                          size = length(closer.data),
                          replace = T)
  resamples$resamps.closer[i] <- (mean(closer.resample - mu0))/
    (sd(closer.data)/sqrt(n))

  # diff
  diff.resample <- sample(x = diff.data,
                        size = length(diff.data),
                        replace = T)
  resamples$resamps.diff[i] <- (mean(diff.resample - mu0))/
    (sd(further.data)/sqrt(n))
}

# shifting data
resamples <- resamples |>
  mutate(resamps.further.null = resamps.further - mean(resamps.further),
         resamps.closer.null = resamps.closer - mean(resamps.closer),
         resamps.diff.null = resamps.diff - mean(resamps.diff))
```

Solution: Above is the code using to conducting resampling for the t statistic for the zebra finch data. Since we did resampling on the t statistic, which we want to be centered at 0, we shift the resamples by simply subtracting the mean of all the resamples to center. `set.seed()` was used to create reproducible results.

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

```
# part b
boot.pvals <- resamples |>
summarize(p.further = mean(resamps.further.null <= mean(resamps.further)),
          p.closer = mean(resamps.closer.null >= mean(resamps.closer)),
          p.low.diff = mean(resamps.diff.null <= -mean(resamps.diff)),
          p.high.diff = mean(resamps.diff.null >= mean(resamps.diff)),
          pdiff = p.low.diff + p.high.diff)

# p vals from t test
# further
t.pval.further <- t.test(x=further.data, mu = mu0, alternative = "less")$p.value
# closer
t.pval.closer <- t.test(x=closer.data, mu = mu0, alternative = "greater")$p.value
# diff
t.pval.further <- t.test(x=diff.data, mu = mu0, alternative = "two.sided")$p.value

# creating table
all.pvals <- tibble(test = c("T-test", "Bootstrapping"),
                    p.further = c(t.pval.further, boot.pvals$p.further),
                    p.closer = c(t.pval.closer, boot.pvals$p.closer),
                    p.diff = c(t.pval.closer, boot.pvals$pdiff))

xtable(all.pvals)

## % latex table generated in R 4.4.2 by xtable 1.8-4 package
## % Fri May 2 02:11:53 2025
## \begin{table}[ht]
## \centering
## \begin{tabular}{rlrrr}
## \hline
## & test & p.further & p.closer & p.diff \\
## \hline
## 1 & T-test & 0.00 & 0.00 & 0.00 \\
## 2 & Bootstrapping & 0.00 & 0.00 & 0.00 \\
## \hline
## \end{tabular}
## \end{table}
```

	Type	p.further	p.closer	p.diff
1	T-test	0.00	0.00	0.00
2	Bootstrapping	0.00	0.00	0.00

Table 1: Table comparing p -values for further, closer, and differences in the zebra finch data using the T-test and Bootstrapping

Solution: The p -values calculated using bootstrapping is so small that it is effectively 0 and treated as such, for the further, closer, and difference zebra finch data. In order to calculate the

two-sided p -value for the difference data, you need to find the p -value on both tails, which can be found by adding or subtracting the mean of the difference data by δ . However, since $\mu_0 = 0$, we can just using the negative of the mean of the difference data to get the “mirror” observation on the other tail. p -values are compared in Table 1.

- (c) What is the 5th percentile of the shifted resamples under the null hypothesis? Note this value approximates $t_{0.05, n-1}$. Compare these values in each case.
 - (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
 - (b) Compute the randomization test p -value for each test.
 - (c) Compute the randomization confidence interval by iterating over values of μ_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.