

BEE552 Biometry Week 2

Maria Feiler

2/2/2022

My Learning Journey

Over the last week, I participated in Biometry in the following ways:

- I asked / answered **2** questions posed in class.
- I asked **2** questions in Slack.
- I answered **2** questions posed by other students on Slack.
- I came to Heather's office hours: **No**
- I came to Jose's office hours: **Yes**
- I met with Heather or Jose separately from office hours: **No**

Anything not falling into one of the above categories?

No

On a scale of 1 (no knowledge) to 10 (complete expert), how would I rate my comfort with R programming after this week?

5

Any topics from last week that you are still confused about?

- I tried to do a nested for loop to automate the bootstrap tests, and for the life in me I couldn't figure it out. How would you do it?

Problem Set

Question 1

Assume survival data (days from treatment to death) from two groups, a treatment group ($n_T = 20$) and a control/placebo group ($n_C = 15$) as follows:

```
treatment = c(90, 91, 68, 90, 167, 26, 38, 89, 88, 99, 41, 123, 76, 88, 79, 96, 79,
              122, 11, 23)

control = c(52, 104, 146, 27, 46, 120, 5, 15, 11, 48, 30, 40, 8, 42, 74)
```

a. Use a permutation method to test the null hypothesis that the treatment does not change survival time. Report the p -value associated with the two-tailed test of the null hypothesis. If the critical value $\alpha_c = 0.05$, would you reject the null hypothesis? (Would the observed difference be considered significant [i.e. reject the null hypothesis] if we set $\alpha_c = 0.01$? What about $\alpha_c = 0.10$?)

I will test the means of the two groups. Let \bar{y} represent the mean of the treatment group and \bar{z} represent the mean of the control group. Therefore $\hat{\theta} = \bar{y} - \bar{z}$. If the null is true, then $\hat{\theta} = 0$.

```
# This code was sourced in part from the original poster's updated code:
# https://stats.stackexchange.com/questions/176691/running-a-permutation-test-with-
# different-sample-sizes-in-r/176714

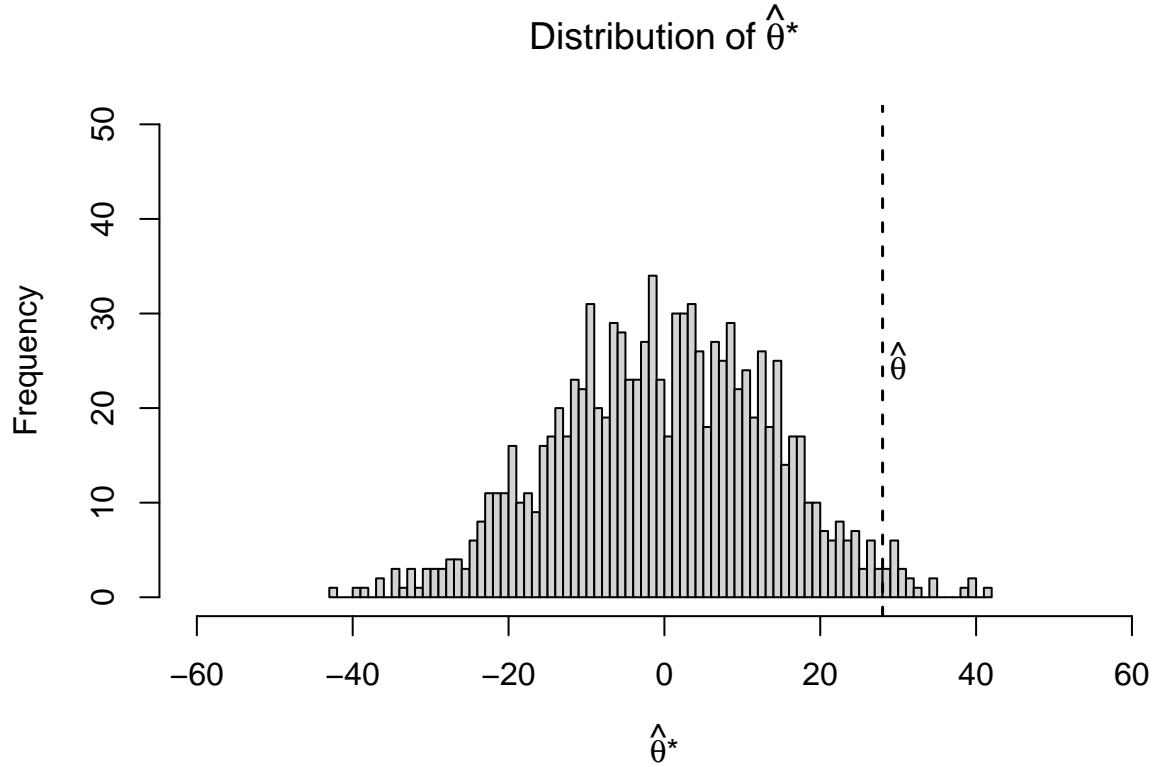
# Calculate the base statistic ybar - zbar
theta_hat <- mean(treatment) - mean(control)

# Create vector of zeros to compare the sample permutation test statistics
dif <- c()

# Create vector of the differences between the sample permutation's means and the
# null (0)
for (i in 1:1000) {
  samp <- sample(x = c(treatment, control),
                 replace = FALSE)

  samp_treat <- samp[1:20]
  samp_con <- samp[21:35]

  dif[i] <- mean(samp_treat) - mean(samp_con)
}
```



The quartiles of the permutations were -42.47(0%), -9.68(25%), 0.23(50%), 9.71(75%), 41.42(100%). The difference of the permutation means had an average -0.03 ± 13.89 difference from the null hypothesis.

The p-value of the original sample is 0.042. At $\alpha_c = 0.05$, the null hypothesis is accepted, and the patients who recieved the treatment survived significantly longer than their control counterparts. At $\alpha_c = 0.01$, the null hypothesis is rejected, and the patients who recieved the treatment did not survive significantly longer than their control counterparts. At $\alpha_c = 0.10$, the null hypothesis is accepted, and the patients who recieved the treatment survived significantly longer than their control counterparts.

b. Given the context for the analysis, is 0.05 the best cut-off for significance? Why or why not? What might be an argument for using a different cut-off for significance?

I argue that 0.05 not a sufficient cutoff for this experiment this presents the issue of circulating a drug/treatment that is falsely proclaimed to extend life, or falsely rejecting the null hypothesis. In my opinion, a 5% chance that the treatment is falsely assumed to be effective is too high for a treatment that extends life, whereas a 1% chance is more acceptable for a matter of this significance. In addition, I argue that the sample size of the experiment testing the treatment is too small, which contributes to my opinion that a significance value of 0.05 is not sufficient to ensure a difference between the treatment and the control groups.

Question 2

Answer Question 1 using nonparametric bootstrap rather than permutation. Calculate confidence intervals for the difference in means between the two groups using bootstrap sampling. The distribution of the difference of means is $\overline{Treatment} - \overline{Control} \sim N\left(E[X_T] - E[X_C], \frac{Var[X_T]}{n_T} + \frac{Var[X_C]}{n_C}\right)$, where the standard deviation of $\overline{Treatment} - \overline{Control}$ should be $\sqrt{\frac{Var[X_T]}{n_T} + \frac{Var[X_C]}{n_C}}$.

I tested my bootstrap method for 1000, 10000, 100000, and 1000000 iterations, which produced the following differences between the the standard deviation of $\overline{Treatment} - \overline{Control}$ and $\sqrt{\frac{Var[X_T]}{n_T} + \frac{Var[X_C]}{n_C}}$:

Table 1: Number of Bootstrap Iterations and the Resulting Difference Between the Standard Deviation of the Difference in Means and the Covariance of the Original Samples

Iterations	Differences
1000	0.962
10000	0.483
100000	0.418
1000000	0.431

Since there are diminishing returns after 100000 iterations, I will use that number of repeats for the bootstrap.

```
# Define number of iterations
nit <- 100000

# Define the matrices into which the treatment bootstrap and control bootstrap samples
# will be stored, and the vector in which the differences between the means will be
# stored.

bootT <- matrix(data = NA,
                 nrow = nit,
                 ncol = length(treatment),
                 dimnames = list(c(1:nit),
                                c(1:length(treatment))
                                )
                 )

bootC <- matrix(data = NA,
                 nrow = nit,
                 ncol = length(control),
                 dimnames = list(c(1:nit),
                                c(1:length(control))
                                )
                 )

bootDiff <- c()

# Run bootstrap for the number of iterations as defined by nit
for (i in 1:nit) {
  # Collect bootstrap samples for the treatment
  bootT[i,] <- sample(x = treatment,
                      size = length(treatment),
```

```

        replace = TRUE
      )

      # Collect bootstrap samples for the control
      bootC[i,] <- sample(x = control,
                        size = length(control),
                        replace = TRUE
      )

      # Collect the difference between the means of the treatment and control bootstrap
      # samples
      bootDiff[i] <- mean(bootT[i,]) - mean(bootC[i,])
    }

# Test that the standard deviation of the difference of the bootstrapped means is
# close to the square root of the difference in variances
bootDiffSD <- sd(rowMeans(bootT)-rowMeans(bootC))

coVar <- sqrt(var(treatment)/length(treatment)+var(control)/length(control))

```

For this bootstrap run, the difference between the the standard deviation of $\overline{Treatment} - \overline{Control}$ and $\sqrt{\frac{Var[X_T]}{n_T} + \frac{Var[X_C]}{n_C}}$ is 0.45, which is acceptable.

a. Using your bootstrap samples, what is the standard error of the mean for the Treatment group ($\hat{se}_{boot,T}$). How does this compare to $\sqrt{\frac{Var[X_T]}{n_T}}$? Approximately how many bootstrap samples are required to get a decent estimate (you define what constitutes “decent”)?

```

# Find the standard error of the treatment bootstrap
bootTStErr <- sd(rowMeans(bootT))

# Find the standard deviation of the original treatment sample
treatmentVar <- sqrt(var(treatment)/length(treatment))

```

The standard error of the treatment bootstrap (8.14) is less than the standard deviation of the original treatment sample (8.38).

b. Using your bootstrap samples, what is the standard error of the mean for the Control group ($\hat{se}_{boot,C}$). How does this compare to $\sqrt{\frac{Var[X_C]}{n_C}}$? Approximately how many bootstrap samples are required to get a decent estimate (you define what constitutes “decent”)?

```

# Find the standard error of the control bootstrap
bootCStErr <- sd(rowMeans(bootC))

# Find the standard deviation of the original control sample
controlVar <- sqrt(var(control)/length(control))

```

The standard error of the control bootstrap (10.57) is less than the standard deviation of the original control sample (10.96).

c. Using ($\hat{se}_{boot,T}$) and ($\hat{se}_{boot,C}$), calculate the standard error of the difference in means ($\hat{se}_{boot,diff}$). Note that variances can be added and standard deviations are the square root of the variance.

```
# Calculate the standard error of the difference in means
bootDiffStErr <- sqrt(bootTStErr^2 + bootCStErr^2)
```

The standard error of the difference in means is 13.34.

d. Use $\hat{se}_{boot,diff}$ to estimate the 95th percentile confidence interval on the difference in means, assuming the interval is given by $(\text{mean} - 1.96\hat{se}_{boot,diff}, \text{mean} + 1.96\hat{se}_{boot,diff})$.

```
# Calculate the confidence interval using the standard error of the bootstrap
# differences and the mean of the bootstrap differences
upper2d <- mean(bootDiff)+(1.96*bootDiffStErr)
lower2d <- mean(bootDiff)-(1.96*bootDiffStErr)
```

The 95th percentile confidence interval for the difference in means is (1.87, 54.16). The since the null hypothesis (difference between the treatment and control samples equals zero) falls outside of the confidence interval, the null hypothesis is rejected.

e. We're now going to answer for 2d in a slightly different way using the percentiles of the bootstrapped samples. Calculate a 95th percentile confidence interval for the difference in mean survival days $\overline{\text{Treatment}} - \overline{\text{Control}}$ based on the quantiles of the bootstrapped samples.

```
# Calculate the confidence interval using the quantiles of the bootstrap mean
# differences
lower2e <- unname(quantile(bootDiff, probs = 0.025))
upper2e <- unname(quantile(bootDiff, probs = 0.975))
```

The 95th percentile confidence interval for the difference in means is (1.15, 53.43). The since the null hypothesis (difference between the treatment and control samples equals zero) falls outside of the confidence interval, the null hypothesis is rejected.

Question 3

Briefly (less than a paragraph, with drawings as needed) explain the relationship between testing a hypothesis and estimating an effect size with a confidence interval. Can you use an effect size and its confidence interval to test a hypothesis? If so, how? Why might providing an effect size and its confidence interval be preferable to providing just a p-value?

Testing a hypothesis and estimating the effect size with a confidence interval are both ways to determine if your data are significant. In the case of the treatment and control groups presented in Question 1, we are testing if the differences between the treatment and control groups are significant determining if the means of the bootstrapped samples of treatment and control survival days differ significantly from the null hypothesis that $\overline{Treatment} - \overline{Control} = 0$. With a hypothesis test, you would receive a p-value that tells you if the null is accepted or rejected. Though that is helpful, it only tells you that your treatment is different from your control without describing the relationship. By including a confidence interval, you can also understand what your underlying population looks like. The difference between the original treatment sample and the control sample is not significant ($p = 0.0520243$), but we again are learning very little about the actual relationship between these two conditions and survival time. For the bootstrap done in Question 2, our null hypothesis is significantly different than the bootstrap sample mean difference distribution ($p = 0$), which still tells us very little. However, the confidence interval of our bootstrap sample difference distribution (1.15, 53.43) tells us that those who received the treatment survived for a longer time than the control. This gets us closer to the actual relationship between the treatment and survival time.

Bonus Question

Using the 'fitdistr' function available from the MASS package, find the distribution that reasonably well fits the treatment and control datasets. Using that distribution, answer Question 2e again using a parametric bootstrap approach.

A selection of distributions were tested for the treatment and control datasets. The log likelihoods of each are as follows:

Table 2: Log Likelihood for Test Distributions Representing the Treatment Group

Distribution Type	Log Likelihood
exponential	-100.28
log-normal	-100.34
logistic	-100.37
normal	-103.84
Poisson	-107.44
weibull	-249.21

Table 3: Log Likelihood for Test Distributions Representing the Control Group

Distribution Type	Log Likelihood
exponential	-73.49
log-normal	-73.95
logistic	-74.04
normal	-76.97
Poisson	-76.99
weibull	-276.1

Since the exponential distribution fits both the treatment and control datasets the best, those parameters will be used.

```
# Worked with Isabel Martins on this code
# Get distributions for treatment and control datasets
treatDist <- fitdistr(x = treatment,
                     densfun = "exponential"
                     )

conDist <- fitdistr(x = control,
                   densfun = "exponential"
                   )

# Sample the new distribution and calculate the skew
treatPara <- c()
for (i in 1:100000) {
  temp <- rexp(n = treatDist$n,
              rate = treatDist$estimate)
  treatPara <- c(treatPara, skewness(temp))
}
```



```

conPara <- c()
for (i in 1:100000) {
  temp <- rexp(n = conDist$n,
              rate = conDist$estimate)
  conPara <- c(conPara, skewness(temp))
}

# Get the difference between the parametric tests
paraDiff <- c(treatPara - conPara)

# Calculate the 95th percentile of the difference in between the control and
# treatment parametric bootstraps using an exponential distribution
upper3 <- quantile(paraDiff, probs = 0.975)
lower3 <- quantile(paraDiff, probs = 0.025)

```

The 95th percentile confidence interval for the difference in means is (-1.62, 1.95). Since the null hypothesis (difference between the treatment and control samples equals zero) falls within the confidence interval, the null hypothesis is accepted.