R workshop - Addon: Bootstrap

Bernhard Piskernik

2024-03-14

Bootstrapping

Bootstrapping is an excellent way to get reliable/robust error estimations without distributional assumptions.

This workshop included three analyses:

- simple linear regression
- mediation
- moderation

The *mediation* already utilizes bootstrapping (multiplying the two parameters results in an indirect path with non-normal distributed errors, so bootstrapping is required). For the two other analyses, I will briefly demonstrate how to rewrite the code so that it uses bootstrapping.

Preparation

First we install the `lmboot` package. There are many other ways/packages to do bootstrapping in R, but this is one of easiest for `lm` objects.

```
install_load("lmboot")
```

Next, we set a seed, so that our results stay the same from run to run.

```
seed <- 42
```

simple linear regression

The original looked like this:

```
lm(
  SA_s ~ ILWOTA_s
  , data = ws
  ) %>%
  summary()
```

Call: lm(formula = SA_s ~ ILWOTA_s, data = ws)

Residuals: Min 1Q Median 3Q Max -2.61711 -0.46147 -0.04482 0.55426 2.29732

Coefficients: Estimate Std. Error t value Pr($>$|t|)
(Intercept) 1.375e-16 1.023e-01 0.00 1.000000
ILWOTA_s 3.776e-01 1.029e-01 3.67 0.000433 *** — Signif. codes: 0 '*** 0.001 ** 0.01 * 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9317 on 81 degrees of freedom (11 observations deleted due to missingness) Multiple R-squared: 0.1426, Adjusted R-squared: 0.132 F-statistic: 13.47 on 1 and 81 DF, p-value: 0.0004326

The bootstrapped version looks like that:

```
boot_simple_reg <- residual.boot(
    SA_s ~ ILWOTA_s
  , data = ws %>%
    # this part is a little awkward - scale() does not return a vector but a one-dimensional matrix
    # this is mostly irrelevant, because the look and feel alike,
    # but lmboot cannot deal with it, therefore we need to transform the columns into vectors
    select(SA_s, ILWOTA_s) %>%
    mutate_all(as.vector) %>%
    # and it cannot deal with missings
    # so either remove or impute them
    drop_na()
  , seed = seed
)


#plot the sampling distribution of the slope coefficient
hist(boot_simple_reg$bootEstParam[,2], main="Residual Bootstrap Sampling Distn.",
xlab="Slope Estimate")
```

Residual Bootstrap Sampling Distn.
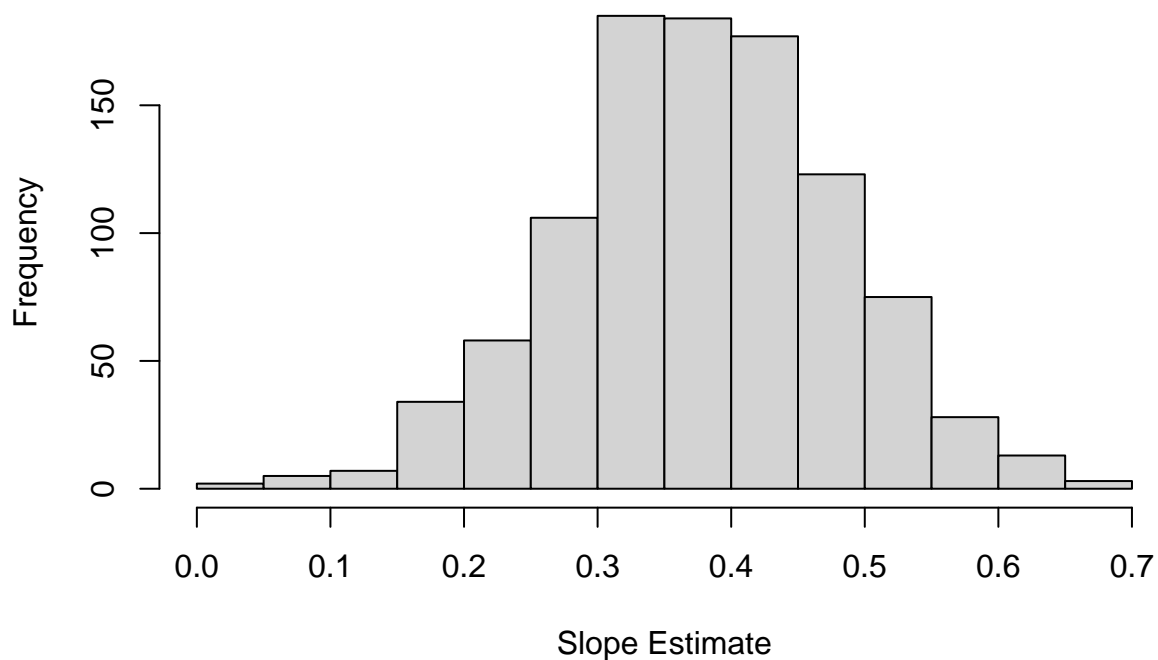

```
#bootstrap 95% CI for slope parameter (percentile method)
quantile(boot_simple_reg$bootEstParam[,2], probs=c(.025, .975))
```

  2.5%   97.5%

0.1708062 0.5785270

Since **0** is not included the confidence 95%-interval [2.5% - 97.5%] the result is significant on a alpha level of

5%.

moderation

The original moderation analysis looked like that:

```r
lm(
  SA ~ ILWOTA_c * PSYSAFE_c
  , data = ws
) %>%
  summary()
```

Call: lm(formula = SA ~ ILWOTA_c * PSYSAFE_c, data = ws)

Residuals: Min 1Q Median 3Q Max -29.5835 -5.1516 -0.6856 6.3027 26.2848

Coefficients: Estimate Std. Error t value Pr(>|t|)
(Intercept) 51.29651 1.19275 43.007 < 2e-16 * **ILWOTA__c 9.73199 3.07678 3.163 0.00222** PSYSAFE_c
-0.50720 1.41137 -0.359 0.72028
ILWOTA__c:PSYSAFE__c -0.09542 2.86678 -0.033 0.97353
— Signif. codes: 0 '*' *0.001 '*' 0.01* '' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.59 on 79 degrees of freedom (11 observations deleted due to missingness) Multiple R-squared: 0.144, Adjusted R-squared: 0.1115 F-statistic: 4.43 on 3 and 79 DF, p-value: 0.006245

The bootstrapped version looks like that:

We start by performing the analysis

```r
boot_moderation <- residual.boot(
    SA ~ ILWOTA_c * PSYSAFE_c
  , data = ws  %>%
    # this part is a little awkward - scale() does not return a vector but a one-dimensional matrix
    # this is mostly irrelevant, because the look and feel alike,
    # but lmboot cannot deal with it, therefore we need to transform the columns into vectors
    select(SA, ILWOTA_c, PSYSAFE_c) %>%
    mutate_all(as.vector) %>%
    # and it cannot deal with missings
    # so either remove or impute them
    drop_na()
  , seed = seed
)
```

Since there are more parameters we check there sequence.

```r
boot_moderation$bootEstParam  %>% colnames()
```

[1] "(Intercept)" "ILWOTA_c" "PSYSAFE_c"
[4] "ILWOTA__c:PSYSAFE__c"

Result for `ILWOTA_c`

```r
quantile(boot_moderation$bootEstParam[,2], probs=c(.025, .975))
```

 2.5%   97.5%

3.922308 15.292857

-> 0 not included -> significant

Result for `PSYSAFE_c`

```
quantile(boot_moderation$bootEstParam[,3], probs=c(.025, .975))
```

   2.5%   97.5%

-3.217050 2.015568

-> 0 included -> not significant

Result for `ILWOTA_c:PSYSAFE_c` (the interaction)

```
quantile(boot_moderation$bootEstParam[,4], probs=c(.025, .975))
```

   2.5%   97.5%

-5.420153 6.018665

-> 0 included -> not significant