APA Statement using Resampling Techniques for GLM using R

This book is a reference on how to perform resampling techniques (e.g., bootstrapping and permutation testing) to write a more informed APA statement.

Ekarin E. Pongpipat, Phillip F. Agres, Paulina Skolasinska

2019-11-16

Contents

Chapter 1

Introduction

Currently, the APA statement includes:

- Estimate
- Name of Statistic
- Degrees of Freedom (df)
- Statistical Value
- *p*-value

However, this is not always informative. The APA statement should include:

- Estimate and its Confidence Interval (CI)
- Name of Statistic
- Degrees of Freedom (df)
- Statistical Value
- Mean Squared Error (MSE)
- Adjusted R^2 and its CI
- Permutated P-Value

Chapter 2

Bootstrap

```
library(tidyverse)
df <- carData::Salaries
```

2.1 Simple Linear Regression

```
# set number of bootstraps
n_bootstrap <- 1000

# create empty dataframes for coefficients and r-squared
bootstrap_coef <- tibble(n_iter = NA,
                         .rows = n_bootstrap)
bootstrap_rsq <- tibble(n_iter = NA,
                         .rows = n_bootstrap)

# for loop for bootstrap
for (i in 1:n_bootstrap) {

  # randomly sample with replacement from the rows
  idx <- sample(1:nrow(df), nrow(df), replace = T)
  df_boot <- df[idx,]

  # run linear model
  model <- lm(salary ~ yrs.since.phd + yrs.service, df_boot)

  # extract estimates and r^2 value
  summary_model <- summary(model)
  t_stat <- summary_model$coefficients[, "t value"]
```

```
  df_denom <- summary_model$df[[2]]
  r_sq <- t_stat^2 / (t_stat^2 + df_denom)

  # write bootstrap iteration
  bootstrap_coef[i, 1] <- i
  bootstrap_rsq[i, 1] <- i

  # determine number of coefficients
  n_coef <- length(model$coefficients)

  # write estimate and r^2 to table for looping across variables
  for (j in 1:n_coef) {

    # bootstrap estimate confidence interval
    bootstrap_coef[i, names(model$coefficients[j])] <- model$coefficients[[j]]

    # bootstrap R^2 CI
    bootstrap_rsq[i, names(model$coefficients[j])] <- r_sq[j]
  }
}

# print estimate CI and R^2 CI for each variable
for (k in 2:n_coef+1) {

  # estimate CI
  cat("\n", colnames(bootstrap_coef[,k]), "\n")
  cat("\n estimate CI\n")
  bootstrap_ci <- quantile(as.matrix(bootstrap_coef[,k]), probs = c(0.025, .975))
  print(bootstrap_ci)

  # R^2 CI
  cat("\n R^2 CI\n")
  bootstrap_rsq_ci <- quantile(as.matrix(bootstrap_rsq[,k]), probs = c(0.025, .975))
  print(bootstrap_rsq_ci)
}
```

```
##
##  yrs.since.phd
##
##  estimate CI
##      2.5%     97.5%
##  979.3215 2072.2106
##
##  R^2 CI
##       2.5%      97.5%
```

```
## 0.03122567 0.15621742
##
##  yrs.service
##
##  estimate CI
##         2.5%        97.5%
## -1162.692563     4.480129
##
##  R^2 CI
##         2.5%        97.5%
## 0.0002034577 0.0535608331
```

Chapter 3

Permutation

Chapter 4

APA Statement

We have finished a nice book.

Chapter 5

Assumptions

Chapter 6

Multiple Comparison Correction