

Sampling Intervals for Models

27-11-2024

Task 1

The subsection below contains the following code:

- Plotting data visualizing the comparison of means
- Sampling with replacement from each group
- Centering and resampling from the combined samples x1 and x2
- Bootstrap using both strategies
- Permutation version of the test
- Wilcoxon rank sum test statistics
- Comparison of results using t.test and wilcox.test

1. Plotting data visualizing the comparison of means

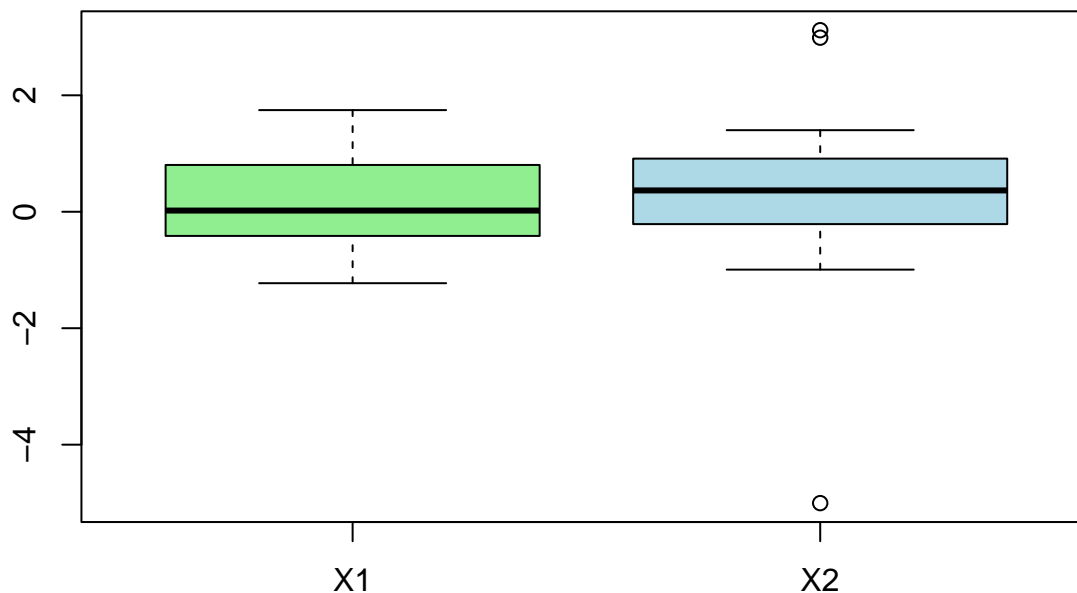
```
## Given data
```

```
x1 <- c(-0.673, -0.584, 0.572, -0.341, -0.218, 0.603, -0.415, -0.013, 0.763, 0.804, 0.054, 1.746, -0.47)
```

```
x2 <- c(0.913, -0.639, 2.99, -5.004, 3.118, 0.1, 1.128, 0.579, 0.32, -0.488, -0.994, -0.212, 0.413, 1.4
```

```
boxplot(x1, x2, names = c("X1", "X2"), col = c("lightgreen", "lightblue"), main = "Comparison of means")
```

Comparison of means



2. a) Sampling with replacement from each group
b) Centering and resampling from the combined samples x1 and x2

```

set.seed(1234)

n1 <- length(x1)
n2 <- length(x2)

## Bootstrap sampling

bootstrap_x1 <- replicate(10000, mean(sample(x1, n1, replace = TRUE)))
bootstrap_x2 <- replicate(10000, mean(sample(x2, n2, replace = TRUE)))

## Centering and resampling

x1_centre <- x1 - mean(x1)
x2_centre <- x2 - mean(x2)
combined_sample <- c(x1_centre, x2_centre)

bs_combined <- replicate(10000, {
  resampled <- sample(combined_sample, length(combined_sample), replace = TRUE)
  x1_resampled <- resampled[1:n1]
  x2_resampled <- resampled[(n1 + 1):length(resampled)]
  mean(x1_resampled) - mean(x2_resampled)
})

```

Centering and resampling is more natural when testing the null hypothesis since it directly aligns the resampling process with H_0 by creating a symmetric distribution around the mean. But to understand group-specific variability, it is more natural to resample independently within each group because it keeps the original group's structure intact.

Advantages: When centering the data, the mean difference between x_1 and x_2 is removed, making sure that the resampled distribution aligns with the null hypothesis. It also avoids group-specific bias.

Disadvantages: One of the biggest disadvantages is assuming that the two groups are exchangeable, which may not always be the case (if the variances or distributions have significant differences). Combining may also mean that the original group specific characteristic is lost, leading to an unrealistic bootstrap sample if x_1 and x_2 are very different.

3. Bootstrap using both strategies

```

n1 <- length(x1)
n2 <- length(x2)

## Bootstrap sampling

bootstrap_x1 <- replicate(10000, mean(sample(x1, n1, replace = TRUE)))
bootstrap_x2 <- replicate(10000, mean(sample(x2, n2, replace = TRUE)))

## Centering and resampling

x1_centre <- x1 - mean(x1)
x2_centre <- x2 - mean(x2)
combined_sample <- c(x1_centre, x2_centre)

bs_combined <- replicate(10000, {
  resampled <- sample(combined_sample, length(combined_sample), replace = TRUE)
  x1_resampled <- resampled[1:n1]
  x2_resampled <- resampled[(n1 + 1):length(resampled)]
})

```

```

    mean(x1_resampled) - mean(x2_resampled)
  })

## Observed mean difference
obs_diff <- mean(x1) - mean(x2)

t_test <- t.test(x1, x2)

p_value_x <- mean(abs(bootstrap_x1 - bootstrap_x2) >= abs(obs_diff))
p_value_combined <- mean(abs(bs_combined) >= abs(obs_diff))

ci_ind95 <- quantile(bootstrap_x1 - bootstrap_x2, c(0.025, 0.975))
ci_ind99 <- quantile(bootstrap_x1 - bootstrap_x2, c(0.005, 0.995))

ci_combined95 <- quantile(bs_combined, c(0.025, 0.975))
ci_combined99 <- quantile(bs_combined, c(0.005, 0.995))

cat("two_sample t.test")

## two_sample t.test
cat("\nt-statistic: ", t_test$statistic)

##
## t-statistic: -0.1188053
cat("\np-value: ", t_test$p.value)

##
## p-value: 0.90646
cat("\n95% CI: ", t_test$conf.int)

##
## 95% CI: -0.9556081 0.8518
cat("\n\nIndependent Sampling")

##
##
## Independent Sampling
cat("\np-value:", p_value_x)

##
## p-value: 0.9056
cat("\n95% CI:", ci_ind95)

##
## 95% CI: -0.8363058 0.8427785
cat("\n99% CI:", ci_ind99)

##
## 99% CI: -1.059151 1.129247

```

```

cat("\n\nBootstrap Centering and Combining:")

##
##
## Bootstrap Centering and Combining:
cat("\nP-value:", p_value_combined)

##
## P-value: 0.8877
cat("\n95% CI:", ci_combined95)

##
## 95% CI: -0.7707999 0.7864862
cat("\n99% CI:", ci_combined99)

##
## 99% CI: -1.037798 1.087438
## Decision making

if (p_value_x < 0.05) cat("Independent Sampling: Reject H0 at 0.05")
if (p_value_x < 0.01) cat("Independent Sampling: Reject H0 at 0.01")

if (p_value_combined < 0.05) cat("Combined Sampling: Reject H0 at 0.05")
if (p_value_combined < 0.01) cat("Combined Sampling: Reject H0 at 0.01")

```

Observation: Since the p-values for both cases are greater than the significant levels 0.05 and 0.01, there is no evidence to reject the null hypothesis.

4. Permutation version

```

combined_data <- c(x1, x2)

## Permutation test
permutation_diff <- replicate(10000, {
  p <- sample(combined_data)
  x1_p <- p[1:n1]
  x2_p <- p[(n1+1):length(p)]
  mean(x1_p) - mean(x2_p)
})

p_value_p <- mean(abs(permutation_diff) >= abs(obs_diff))

ci_perm95 <- quantile(permutation_diff, c(0.025, 0.975))
ci_perm99 <- quantile(permutation_diff, c(0.005, 0.995))

cat("Permutation Test:")

## Permutation Test:
cat("\nP-value:", p_value_p)

##
## P-value: 0.9069

```

```
cat("\n95% CI:", ci_perm95)
```

```
##
```

```
## 95% CI: -0.7448434 0.7793131
```

```
cat("\n99% CI:", ci_perm99)
```

```
##
```

```
## 99% CI: -0.9099949 0.9507404
```

Observation: The permutation test results show a p-value of 0.9004, which is greater than the significance levels of 0.05 and 0.01, meaning that there is no evidence to reject the null hypothesis H_0 .

5. Wilcoxon rank sum test statistics

```
## Wilcoxon test observed
```

```
wil_obs <- wilcox.test(x1, x2, exact = FALSE)$statistic
```

```
## Bootstrap Wilcoxon
```

```
bootstrap_wilcoxon <- replicate(10000, {  
  x1_sample <- sample(x1, n1, replace = TRUE)  
  x2_sample <- sample(x2, n2, replace = TRUE)  
  wilcox.test(x1_sample, x2_sample, exact = FALSE)$statistic  
})
```

```
wil_p_value <- mean(bootstrap_wilcoxon >= wil_obs)
```

```
ci_wilcox95 <- quantile(bootstrap_wilcoxon, c(0.025, 0.975))
```

```
ci_wilcox99 <- quantile(bootstrap_wilcoxon, c(0.005, 0.995))
```

```
cat("Wilcoxon Rank Sum Test:")
```

```
## Wilcoxon Rank Sum Test:
```

```
cat("\nObserved:", wil_obs)
```

```
##
```

```
## Observed: 181
```

```
cat("\nP-value:", wil_p_value)
```

```
##
```

```
## P-value: 0.5068
```

```
cat("\n95% CI:", ci_wilcox95)
```

```
##
```

```
## 95% CI: 109 254
```

```
cat("\n99% CI:", ci_wilcox99)
```

```
##
```

```
## 99% CI: 87 279
```

Observation: The Wilcoxon rank sum test shows a p-value of 0.5038, which is greater than both significance levels, showing no evidence to reject the null hypothesis. The confidence levels for the test statistic supports that there is no significant difference between the samples.

6. Comparison of results using t.test and wilcox.test

```
t_test_stat = t_test$statistic
t_test_pval = t_test$p.value
t_test_ci = t_test$conf.int

## Table to compare results

results <- data.frame(
  Method = c(
    "Two-Sample t-Test",
    "Bootstrap Independent",
    "Bootstrap Combining",
    "Permutation Test",
    "Wilcoxon Test"
  ),
  `p-value` = c(
    round(t_test_pval, 3),
    round(p_value_x, 3),
    round(p_value_combined, 3),
    round(p_value_p, 3),
    round(wil_p_value, 3)
  ),
  `Statistic / Observed` = c(
    round(t_test_stat, 3),
    "N/A",
    "N/A",
    "N/A",
    round(wil_obs, 3)
  ),
  `95% CI` = c(
    paste(round(t_test_ci[1], 3), round(t_test_ci[2], 3), sep = " to "),
    paste(round(ci_ind95[1], 3), round(ci_ind95[2], 3), sep = " to "),
    paste(round(ci_combined95[1], 3), round(ci_combined95[2], 3), sep = " to "),
    paste(round(ci_perm95[1], 3), round(ci_perm95[2], 3), sep = " to "),
    paste(round(ci_wilcox95[1], 3), round(ci_wilcox95[2], 3), sep = " to ")
  ),
  `99% CI` = c(
    "N/A",
    paste(round(ci_ind99[1], 3), round(ci_ind99[2], 3), sep = " to "),
    paste(round(ci_combined99[1], 3), round(ci_combined99[2], 3), sep = " to "),
    paste(round(ci_perm99[1], 3), round(ci_perm99[2], 3), sep = " to "),
    paste(round(ci_wilcox99[1], 3), round(ci_wilcox99[2], 3), sep = " to ")
  )
)

print(results)
```

##	Method	p.value	Statistic...Observed	X95..CI
## 1	Two-Sample t-Test	0.906	-0.119	-0.956 to 0.852
## 2	Bootstrap Independent	0.906	N/A	-0.836 to 0.843
## 3	Bootstrap Combining	0.888	N/A	-0.771 to 0.786
## 4	Permutation Test	0.907	N/A	-0.745 to 0.779
## 5	Wilcoxon Test	0.507	181	109 to 254
##	X99..CI			

```
## 1          N/A
## 2 -1.059 to 1.129
## 3 -1.038 to 1.087
## 4  -0.91 to 0.951
## 5      87 to 279
```

Observation: The results show that p-values are much higher than both significance levels, showing no evidence to reject the null hypothesis. While the Wilcoxon test reports an observed statistic of 181, the p-value aligns with the other methods, showing that the results are consistent. Moreover, the confidence intervals for all methods include 0, reinforcing the conclusion that there is no significant difference between the two samples.

Task 2

The subsection below contains the following code: - Creating the sample of size 200

- Residual bootstrap for linear regression
- Pairs bootstrap for linear regression
- Comparison of two approaches

1. Creating the sample

```
set.seed(1234)

n <- 200
x1 <- rnorm(n, mean = 2, sd = sqrt(3))
x2 <- runif(n, min=2, max = 4)
x3 <- runif(n, min=-2, max = 2)
e <- rt(n, df = 5)

y <- 3 + (2 * x1) + x2 + e
```

2. Residual bootstrap for linear regression

```
model <- lm(y ~ x1 + x2 + x3)

n_boot <- 1000
boot_coeff <- replicate(n_boot, {
  residuals <- residuals(model)
  resampled_residuals <- sample(residuals, replace = TRUE)

  y_boot <- fitted(model) + resampled_residuals

  coef(lm(y_boot ~ x1 + x2 + x3))
})

ci_residual <- apply(boot_coeff, 1, function(coefs) {
  quantile(coefs, probs = c(0.025, 0.975))
})

cat("Residual bootstrap\n")

## Residual bootstrap
print(ci_residual)
```

```
##      (Intercept)      x1      x2      x3
```

```
## 2.5%      1.462101 1.881458 0.885742 -0.1286506
## 97.5%     3.373326 2.076352 1.477529  0.1483873
```

Observation: The residual bootstrap confidence interval for x3 includes a zero. This shows that x3 is not statistically significant at the confidence level 95%. But we cannot exclude x3 from the model since the data does not provide enough evidence to confirm that it doesn't have an effect.

3. Pairs bootstrap for linear regression

```
boot_coeff_pairs <- replicate(n_boot, {
  i <- sample(1:n, size = n, replace = TRUE)

  x1_boot <- x1[i]
  x2_boot <- x2[i]
  x3_boot <- x3[i]
  y_boot <- y[i]

  coef(lm(y_boot ~ x1_boot + x2_boot + x3_boot))
})

ci_pairs <- apply(boot_coeff_pairs, 1, function(coefs) {
  quantile(coefs, probs = c(0.025, 0.975))
})

cat("Pairs bootstrap for linear regression\n")
```

```
## Pairs bootstrap for linear regression
print(ci_pairs)
```

```
##      (Intercept)  x1_boot  x2_boot  x3_boot
## 2.5%      1.482433 1.866536 0.8928894 -0.1270905
## 97.5%      3.214083 2.078609 1.4800340  0.1601752
```

Observation: Taking the pairs bootstrap for linear regression, the confidence interval for x3 also includes 0. This shows that x3 is not statistically significant at the confidence level 95%. But we cannot exclude x3 from the model since the data does not provide enough evidence to confirm that it doesn't have an effect.

4. Comparison of two approaches

Residual Bootstrap:

- This approach assumes that the model structure $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e$ is correct (where b is beta and e is epsilon).
- It resamples the residuals from the fitted model and creates new y-values by adding these resamples residuals to the fitted values (\hat{y}).
- It keeps x_1 , x_2 , and x_3 fixed, and the variability in the predictors are ignored.

Observation: The confidence level for x3 is narrower, showing a smaller range of uncertainty. This is because the bootstrap assumed the predictors are fixed.

Pairs Bootstrap:

- It resamples pairs (entire rows) of the dataset (x_1 , x_2 , x_3 , y) using replacement. .
- New bootstrap datasets are created by resampling both the predictors and the response simultaneously.
- The method allows for variability in predictors and ensures the relationship between x_1 , x_2 , x_3 , and y are retained.

Observation: The confidence level for x3 is wider, showing a greater range of uncertainty. This is because the pairs bootstrap incorporates the variability in the predictors.

Differences in sampling approach

The residual bootstrap is clearly more optimistic because it assumed that the model is correct and ignores the uncertainty, focusing only on the resampling residuals. This can result smaller confidence intervals, underestimating the true uncertainty.

The pairs bootstrap gives a more conservative estimate because since it allows variability in both the response and the predictors, providing a wider confidence interval. This captures the real-world uncertainty better.

Task 3

Bootstrapping Methodology

Bootstrapping is a resampling method that is used to estimate the distribution of a statistic by repeatedly resampling the observed data. It also assumes that the observed data represents the population. The method can be used for testing hypothesis, creating confidence intervals, or estimating the variability of model parameters.

Residual Bootstrapping

Advantages

1. Residual bootstrapping was easy to apply (example, to assess the significance of x_3).
2. It provides robust results in exercises where traditional assumption couldn't hold (such as e being t -distributed and x_2 being uniformly distributed).
3. The method successfully constructed the confidence intervals. Residual bootstrapping provided a narrow intervals, due to the assumption of fixed indicators.

Disadvantages 1. The method assumed the model is correctly specified, underestimating uncertainty and leading to narrower confidence intervals.

2. In Task 2, the residual bootstrap ignores individual variability, that potentially cause the overconfidence in intervals (causing narrower intervals).
3. In Task 2, the confidence interval for x_3 included 0, making it harder to conclude if it actually has an effect.

Pairs Bootstrapping

Advantages

1. Pairs bootstrapping was easy to apply (example, to assess the significance of x_3).
2. It is useful to capture the real-world variability in responses and predictors.
3. The method was able to construct the confidence intervals for coefficients successfully. The pairs bootstrapping were more conservative with the results, using predictor variability.

Disadvantages

1. The pairs bootstrap provides a wider interval, leading to more conservative estimates (this is a context-specific disadvantage).
2. The pairs bootstrap also included 0 in the x_3 confidence level, making it harder to conclude the effect of x_3 .