(Elementary) Regression Methods & Computational Statistics (405.952)

Part IV: Hypothesis Testing and Confidence Intervals (cont.)

Ass.-Prof. Dr. Wolfgang Trutschnig
Arbeitsgruppe Stochastik/Statistik
Fachbereich Mathematik
Universität Salzburg
www.trutschnig.net

Salzburg, January 2018

- We again return to the two-sided t-test.
- Suppose that $X \sim \mathcal{N}(\mu_x, \sigma^2)$, we do not know μ_x and σ^2.
- Suppose that $Y \sim \mathcal{N}(\mu_y, \sigma^2)$, we do not know μ_y and σ^2.
- Notice that the variance of X and Y is the same (and unknown).
- Given a sample $X_1, \ldots, X_n$ from X and a sample $Y_1, \ldots, Y_m$ from Y with $m, n \geq 2$ we now want to calculate a confidence interval for the parameter $\mu_D := \mu_x - \mu_y$.
- Remember that when testing for $H_0 : \mu_D = 0$ R also returned a 95%-confidence interval

```
mux <- muy <- 0
sigmax <- 1; sigmay <- 2
n <- 1000
x <- rnorm(n,mean=muy,sd=sigmax)
y <- rnorm(n,mean=muy,sd=sigmay)

test <- t.test(x,y,paired=FALSE,alternative="two.sided")
test
```

- yields

```
Welch Two Sample t-test

data:  x and y
t = 0.32697, df = 1515.4, p-value = 0.7437
alternative hypothesis: true difference in means is not equal to
     0
95 percent confidence interval:
  -0.1125703  0.1576068
sample estimates:
  mean of x    mean of y
0.009047213 -0.013471049
```

- How is this 95%-confidence interval calculated?
- We know that $S_{n,m}$ given by

$$S_{n,m} = \frac{\overline{X}_n - \overline{Y}_m - (\mu_x - \mu_y)}{\sqrt{S_p^2}\sqrt{\frac{1}{n} + \frac{1}{m}}},$$

  follows a t_{n+m-2}-distribution.
- A a consequence

$$\mathbb{P}\left(S_{n,m} \in \left[t_{n+m-2;\frac{\alpha}{2}}, t_{n+m-2;1-\frac{\alpha}{2}}\right]\right) = 1 - \alpha.$$

- Based on this we can easily derive the following confidence interval $C_{n,m}^{1-\alpha}$with coverage probability $1 - \alpha$:

$$C_{n,m}^{1-\alpha}(X_1, \dots, X_n, Y_1, \dots, Y_m) = C_{n,m}^{1-\alpha} = \left[\overline{X}_n - \overline{Y}_m - \Delta, \overline{X}_n - \overline{Y}_m + \Delta\right]$$

  with $\Delta = t_{n+m-2;1-\frac{\alpha}{2}}\sqrt{S_p^2}\sqrt{\frac{1}{n} + \frac{1}{m}}$.

```r
# t-test for H0: muD=mux-muy=0
mux <- 0
muy <- 0.2
sigmax <- sigmay <- 1
n <- m <- 50
x <- rnorm(n, mean=mux, sd=sigmax)
y <- rnorm(m, mean=muy, sd=sigmay)
test <- t.test(x,y, paired=FALSE, alternative="two.sided", var.
    equal=TRUE)
test

#confidence interval for muD manually
alpha <- 0.05
sp <- ((n-1)*var(x)+(m-1)*var(y))/(n+m-2)
Delta <- qt(p=1-alpha/2,df=n+m-2)*sqrt(sp *(1/n+1/m))
conf.int <- c(mean(x)-mean(y) - Delta, mean(x)-mean(y) + Delta)
test$conf.int[1:2]
conf.int
```

- yields

```
[1] -0.5283843  0.2504880
[1] -0.5283843  0.2504880
```

- Check if the confidence interval does what it should.

```r
R <- 10000
error <- rep(0,R)
CI <- data.frame(lower=rep(0,R),upper=rep(0,R))
for(i in 1:R){
  mux <- 0
  muy <- 0.2
  sigmax <- sigmay <- 1
  n <- m <- 50
  x <- rnorm(n,mean=mux,sd=sigmax)
  y <- rnorm(m,mean=muy,sd=sigmay)
  test <- t.test(x,y,paired=FALSE,alternative="two.sided",var.equal=TRUE)
  CI[i,] <- test$conf.int[1:2]
}

CI$contained <- ifelse(CI$lower<= mux-muy & CI$upper>= mux-muy,1,0)
coverage <- mean(CI$contained)
coverage
[1] 0.9504
```

- What happens if we change the values of μ_x and μ_y?
- What happens if we change n and m?
- How is the hypothesis test for $H_0 : \mu_D = 0$ vs. the two-sided alternative related with the confidence interval?
- Answer: We reject H_0 if and only if $0 \notin C_{n,m}^{1-\alpha}$, i.e. if the confidence interval does not contain 0.

Exercise 39: Confirm the just-stated answer by simulations and proceed as follows:

- Choose some some values for μ_x and μ_y and simulate samples of X and Y.
- Apply the two-sided t-test and save the p-value as well as the confidence interval.
- Repeat the two steps $R = 1.000$ times and verify if in all R case we have that the p-value is less than 0.05 if and only if $0 \notin C_{n,m}^{1-\alpha}$.

- Suppose that $x_1, \dots, x_n$ is a sample from X and that $y_1, \dots, y_m$ is a sample from Y.
- We repeat the following steps R times:
  - Randomly draw n values from $x_1, \dots, x_n$ and m values from $y_1, \dots, y_m$ with (!) replacement
  - The resulting samples $x_1^*, \dots, x_n^*$, $y_1^*, \dots, y_m^*$ are called bootstrap samples or bootstrap replications.
  - Calculate $\overline{x_n^*} - \overline{y_m^*}$ and save this value.
- Let $d_1^*, \dots, d_R^*$ denote the resulting values (i.e. the differences of the means of the boostrap samples).
- The boostrap confidence interval $C_{n,m}^{*,1-\alpha}$ is then defined as the interval formed by the $\frac{\alpha}{2}$-quantile and the $(1-\frac{\alpha}{2})$-quantile of the sample $d_1^*, \dots, d_R^*$, i.e.

$$C_{n,m}^{*,1-\alpha} = \left[(F_d^*)^-\left(\frac{\alpha}{2}\right), (F_d^*)^-\left(1-\frac{\alpha}{2}\right)\right]$$

- Let's check the details in R.

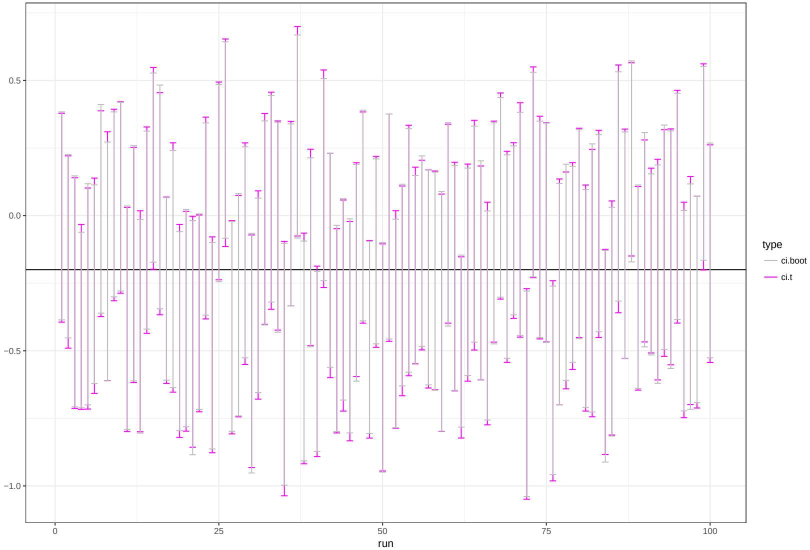
- ```r
  mux <- 0
  muy <- 0.2
  sigmax <- sigmay <- 1
  n <- m <- 50
  x <- rnorm(n,mean=mux,sd=sigmax)
  y <- rnorm(m,mean=muy,sd=sigmay)
  test <- t.test(x,y,paired=FALSE,alternative="two.sided",var.equal=TRUE)        #just
  test$conf.int[1:2]

  boot.diff <- rep(0,R)
  for(i in 1:R){
    x.boot <- sample(x,size = n,replace = TRUE)
    y.boot <- sample(y,size = m,replace = TRUE)
    boot.diff[i] <- mean(x.boot)-mean(y.boot)
  }
  ci.boot <- as.numeric(quantile(boot.diff,probs = c(alpha/2,1-alpha/2)))
  #compare the two intervals
  test$conf.int[1:2]
  ci.boot
  ```
- yields (lucky coincidence?)
- ```
  [1] -0.8296607  -0.0334093
  [1] -0.82211378  -0.03919495
  ```

```r
#systematical comparison of the two CIs
outer.R <- 100
Results <- data.frame(lower.t=rep(0,outer.R),lower.boot=rep(0,
      outer.R),upper.t=rep(0,outer.R),upper.boot=rep(0,outer.R))
for(k in 1:outer.R){
  mux <- 0; muy <- 0.2
  sigmax <- sigmay <- 1
  n <- m <- 50
  x <- rnorm(n,mean=mux,sd=sigmax)
  y <- rnorm(m,mean=muy,sd=sigmay)
  test <- t.test(x,y,paired=FALSE,alternative="two.sided",var.
      equal=TRUE)        #just
  Results[k,c(1,3)] <- test$conf.int[1:2]

  R <- 1000; boot.diff <- rep(0,R)
  for(i in 1:R){
    x.boot <- sample(x,size = n,replace = TRUE)
    y.boot <- sample(y,size = m,replace = TRUE)
    boot.diff[i] <- mean(x.boot)-mean(y.boot)
  }
  Results[k,c(2,4)] <- as.numeric(quantile(boot.diff,probs = c(
      alpha/2,1-alpha/2)))
}
```

The bootstrap confidence interval for $\mu_D = \mu_x - \mu_y$



Exercise 40:

- Fix $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- Generate a sample $X_1, \dots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Calculate a bootstrap confidence-interval $C_n^{1-\alpha}$ for the parameter μ based on the sample.
- Use the t-test to get an exact confidence interval and compare the interval with the bootstrap interval.
- Repeat the previous steps to get a more systematic picture of the performance of the bootstrap confidence interval.

Exercise 41:

- Fix $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- Generate a sample $X_1, \ldots, X_n$ from $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Calculate a bootstrap confidence-interval $C_n^{1-\alpha}$ for the parameter σ^2 based on the sample.
- Compare the exact confidence interval

$$I = \left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}} \right]$$

  and compare the interval with the bootstrap interval.
- Repeat the previous steps to get a more systematic picture of the performance of the bootstrap confidence interval.

Exercise 42:

- We have already mentioned the correspondence between two-sided hypothesis tests and confidence intervals.
- Return to the situation discussed in the slides (confidence interval for $\mu_D = \mu_x - \mu_y$) and use the boostrap confidence interval to derive a boostrap hypothesis test.
- Evaluate the performance of the test via simulations.

Exercise 43 @mathematicians:

- Fix $\lambda > 0$.
- Generate a sample $X_1, \ldots, X_n$ from $X \sim \mathcal{E}(\lambda)$ (exponential distribution).
- Calculate a bootstrap confidence-interval $C_n^{1-\alpha}$ for the parameter λ based on the sample.
- Evaluate the performance of the bootstrap confidence interval via simulations and compare the interval with an exact confidence interval (as derived in the UV 'Angewandte Statistik').