One-sample t-test as a model comparison

André Meichtry

Contents

Simulation of data with parameters μ and σ assumed known

Simulate some two-group-data from model with parameters μ_i and σ assumed known:

- Parameterization: $Y_i = \mu + \epsilon_i, i = 1, ..., n_i,$ $\boxed{Y_i \sim N(\mu, \sigma^2)}$

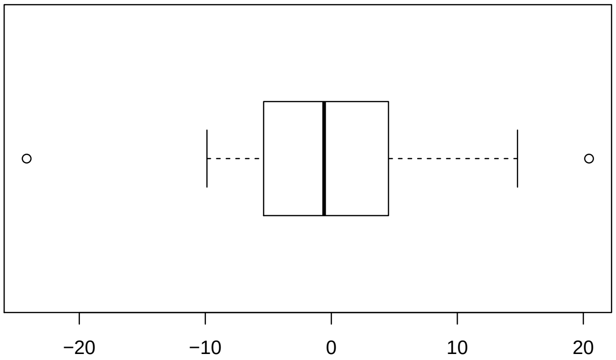
```
n <- 20
mu <- 2
sigma <-10
set.seed(9)
X <- rnorm(n,mu,sigma)
X
```

```
 [1]  -5.66796  -6.16458   0.58465  -0.77605   6.36307  -9.86873  13.91987
 [8]   1.81810  -0.48085  -1.62937  14.77571  -2.68897   2.71054  -0.66038
[15]  20.45257  -6.39450   1.22552 -24.17706  10.87884  -5.07491
```

```
summary(X)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-24.177  -5.223  -0.571   0.457   3.624  20.453
```

```
boxplot(X,horizontal=TRUE)
```



Models

Unconstrained, unrestricted or full model

μ is unknown and has to be estimated:

$$\boxed{X_i \sim N(\mu, \sigma^2)}$$

```
modU <- lm(X~1) ##regression of X on only intercept
```

Constrained or restricted model

μ is assumed known and has not to be estimated, i.e.

$$\boxed{H_0 : \mu = 0, \quad X_i \sim N(0, \sigma^2)}$$

```
modR <- lm(X~0) ##formula for excluding the intercept
```

in R with ANOVA for model comparison

```
anova(modR,modU)
```

```
Analysis of Variance Table

Model 1: X ~ 0
Model 2: X ~ 1
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1     20 1832
2     19 1827  1      4.18 0.04   0.84
```

By hand

Residual sum of squares and explained sum of squares

```
(RSSU <-sum(modU$residuals^2))
```

```
[1] 1827.4
```

```r
(RSSR <-sum(modR$residuals^2))
```

```
[1] 1831.6
```

```r
(RSSR-RSSU)
```

```
[1] 4.182
```

degrees of freedom

```r
(dfU <- n-1)
```

```
[1] 19
```

```r
(dfR <- n-(1-1))
```

```
[1] 20
```

F-test

The F-statistic is the amount of available fit that is actually achieved,

$$F = \frac{(RSS_R - RSS_U)/(df_R - df_U)}{RSS_U/df_U} = \frac{"Explained"}{"Not\,explained"}$$

```r
F <- (RSSR-RSSU)/(dfR-dfU)/(RSSU/(dfU))
p <- 1-pf(F,df1=dfR-dfU,df2=dfU)
sigmaU <- sqrt(RSSU/dfU)
sigmaR <- sqrt(RSSR/dfR)
print(data.frame(RSSR,RSSU,SSExplained=RSSR-RSSU,F,p,sigmaU,sigmaR),row.names=FALSE)
```

```
   RSSR   RSSU SSExplained        F       p sigmaU sigmaR
 1831.6 1827.4       4.182 0.043482 0.83704 9.8071 9.5697
```

log-Likelihood of both models

These are the log-Likelhoods of the model at MLE's (the maximum likelihood estimates).

```r
logLik(modU)
```

```
'log Lik.' -73.528 (df=2)
```

```r
logLik(modR)
```

```
'log Lik.' -73.551 (df=1)
```

AIC and BIC

Adding penalties for model complexity:

$$AIC = -2l + 2p$$

with l as the log-likelihood and p the number of parameters in the model.

$$BIC = -2l + 2\log(n)$$

```r
AIC(modR,modU)
```

```
     df    AIC
modR  1 149.10
modU  2 151.06
```

```r
BIC(modR,modU)
```

```
     df    BIC
modR  1 150.10
modU  2 153.05
```

Result

Smaller AIC and BIC (smaller negative penalized likelihoood) are better. We do NOT reject the constrained model in favor of the unconstrained model.