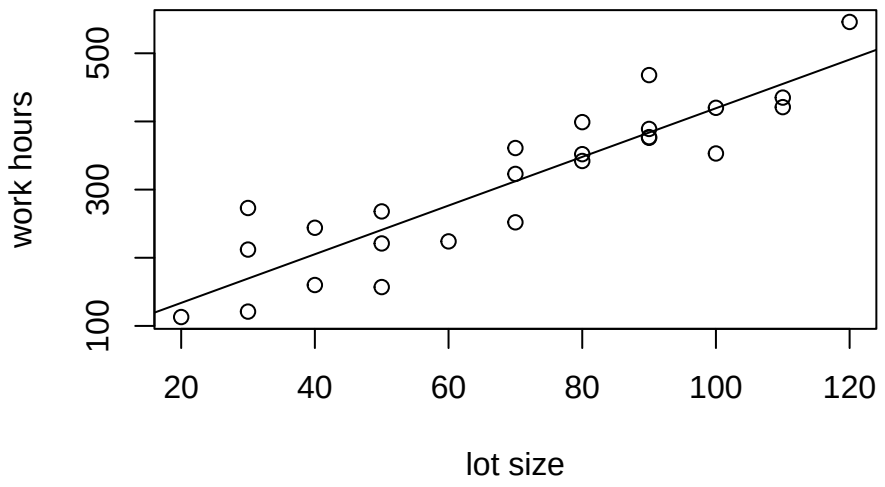
Rexample3

Lingxiao Zhou

Example for ANOVA (toluca)

Data on lot size (X) and work hours (Y) was obtained from 25 recent runs of a manufacturing process.

```
# find relationship between lot size (X) and work hours (Y)
toluca = read.table("http://www.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatasets/KutnerData/
    col.names = c("lotsize", "workhrs"))
plot(toluca$lotsize, toluca$workhrs, xlab = "lot size", ylab = "work hours")
toluca.reg = lm(workhrs ~ lotsize, data = toluca)
abline(toluca.reg)
```




```
anova(toluca.reg)
```

```
## Analysis of Variance Table
##
## Response: workhrs
##           Df Sum Sq Mean Sq F value    Pr(>F)
## lotsize    1 252378  252378  105.88 4.449e-10 ***
## Residuals 23  54825    2384
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(toluca.reg)
```

```
##
## Call:
## lm(formula = workhrs ~ lotsize, data = toluca)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -83.876 -34.088  -5.982  38.826 103.528
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   62.366     26.177   2.382   0.0259 *
## lotsize        3.570      0.347  10.290 4.45e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 48.82 on 23 degrees of freedom
## Multiple R-squared:  0.8215, Adjusted R-squared:  0.8138
## F-statistic: 105.9 on 1 and 23 DF,  p-value: 4.449e-10
```

- Note note that the F-test and t-test for β_1 are equivalent and the p-values are the same.

Perform the general linear test

Reduce Model: $Y_i = \beta_0$
Full Model: $Y_i = \beta_0 + \beta_1 x_i$
H_0 : Full model is not better than the reduced model (equivalently $\beta_1 = 0$)
H_a : Full model is better than the reduced model (equivalently $\beta_1 \neq 0$)

```r
Red <- lm(workhrs ~ 1, data = toluca)
Full <- toluca.reg

anova(Red, Full)
```

```
## Analysis of Variance Table
##
## Model 1: workhrs ~ 1
## Model 2: workhrs ~ lotsize
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     24 307203
## 2     23  54825  1    252378 105.88 4.449e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- P-value is $4.449 \times 10^{-10} < 0.05$. We reject null hypothesis and conclude that the full model is significantly better than the reduced model.