One-way ANOVA with two levels

Structure of the model

The simplest possible ANOVA has two levels for the factor, and can be written with two β parameters:

$$Y = \beta_1 X_1 + \beta_2 X_2 + e$$

where:

- Y is a vector of observed values
- β_1 is a parameter representing the mean at level 1, treated as unknown but constant.
- X_1 is a binary predictor that is 1 if this y has the factor at level 1, zero otherwise
- β_2 is a parameter representing the mean at level 2, treated as unknown but constant.
- X_2 is a binary predictor that is 1 if this y has the factor at level 2, zero otherwise
- e is a vector of independent, identically distributed $N(0, \sigma_e)$ random variables
- σ_e is the standard deviation of the error or residual terms e, which are assumed to have mean zero.

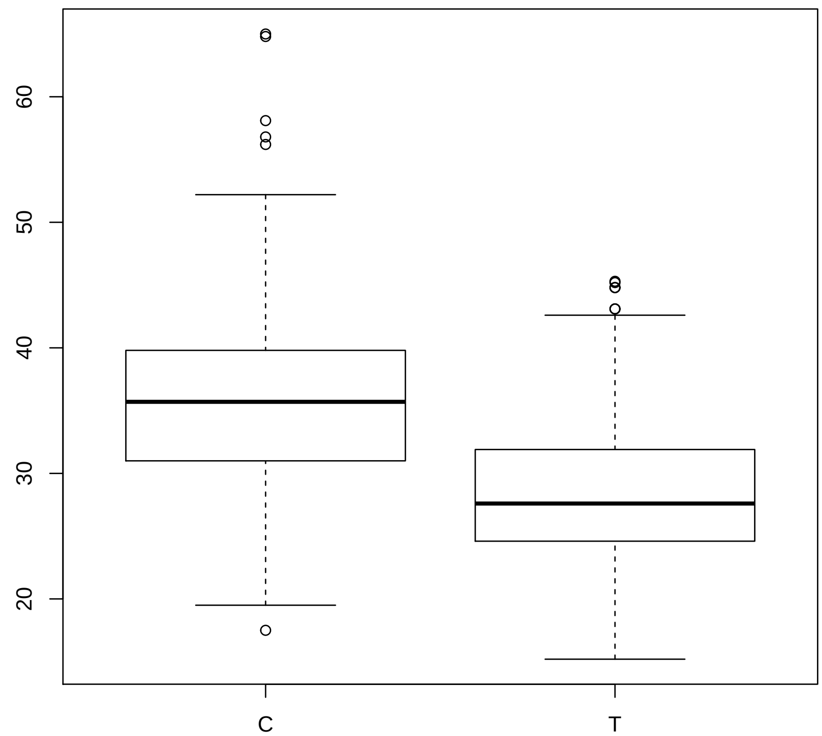
Example: Highway mileage by Car/Truck

Read the data:

```
df  = read.csv("epa.csv")
dfh = df[df$C.H=='H',]       #select just highway.  Don't forget the comma before the closing

CT = dfh$car.truck           #pick out car.truck and display its structure
str(CT)

##  Factor w/ 2 levels "C","T": 2 2 2 2 2 2 2 1 1 1 ...

mpg = dfh$mpg                #get miles per gallon
```

Check out the data with a boxplot

```
boxplot(mpg~CT)
```



Now run the one-way ANOVA.

As long as CT is a factor, R will determine the number of levels.

```
aov1 = aov(mpg ~ CT)

summary(aov1)

##                Df Sum Sq Mean Sq F value Pr(>F)
## CT              1  19056   19056   470.7 <2e-16 ***
## Residuals    1436  58133      40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This is the standard ANOVA table.

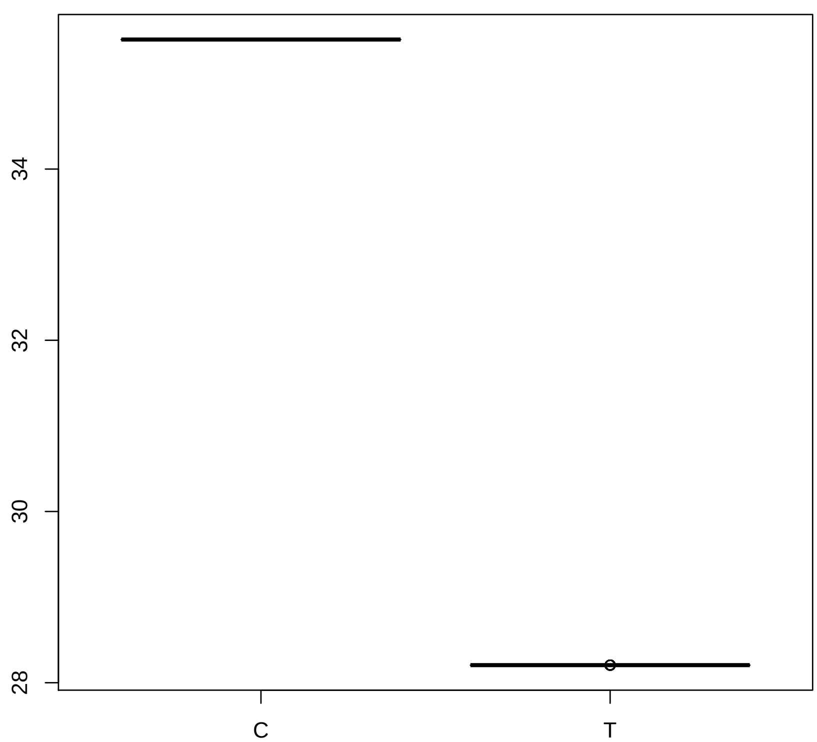
It has one line for each factor, and one line for `Residuals`.

Since we have only one factor, it has only two lines.

The `CT` line is for the car/truck factor. The F-statistic has `Pr(>F)` much less than .05, so it is very unlikely that this data was generated from a dataset where the mean mpg values of cars and trucks were nearly the same.

Like any linear model, this one has fitted values. We can examine them with a boxplot:

```
boxplot(aov1$fitted.values~CT)
```



The fitted values in any one-way ANOVA are the means for the levels:

```r
tapply(mpg,CT,mean)

##        C        T
## 35.51319 28.20533
```

These can be obtained from the coefficients, though the presentation can be confusing:

```r
aov1$coefficients

## (Intercept)         CTT
##   35.513188   -7.307861
```

You would interpret this to say that the expected mpg for cars is the `(Intercept)` value. To obtain the expected mpg for trucks, add the `(Intercept)` and `CTT` values. This will produce the same numbers as we got by computing the mean mpg values by level.