One-Way Anova

Divit Vasu

14 August 2018

One Way ANOVA

One-way ANOVA (Analysis of Variance) is used to examine the differences between means of more than two independent samples. It is used when we have a **categorical independent variable** (treatment) (with more than two categories) and a normally distributed **interval or ratio dependent variable** (Response).

The objective is to test the differences in means of the dependent variable (response) across various levels of independent variables (treatment).

Example

Ages of 3 particlular groups as reported by a survey of a population are given below:

```
Group1 <- c(53, 48, 32, 49, 69, 72, 84)
Group2 <- c(22, 30, 19, 47, 49, 55, 58)
Group3 <- c(61, 71, 77, 81, 59, 68, 53)
```

Perform one-way ANOVA to test whether the difference between the ages of the three groups is significant or not.

The hypothesis testing problem is:

H_0 : there is no difference between ages of all three groups
against
H_1 : difference among means is significant.

Checking normality of data

```
res <- shapiro.test(Group1)
res
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Group1
## W = 0.96325, p-value = 0.8461
```

```
res <- shapiro.test(Group2)
res
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Group2
## W = 0.88581, p-value = 0.2535
```

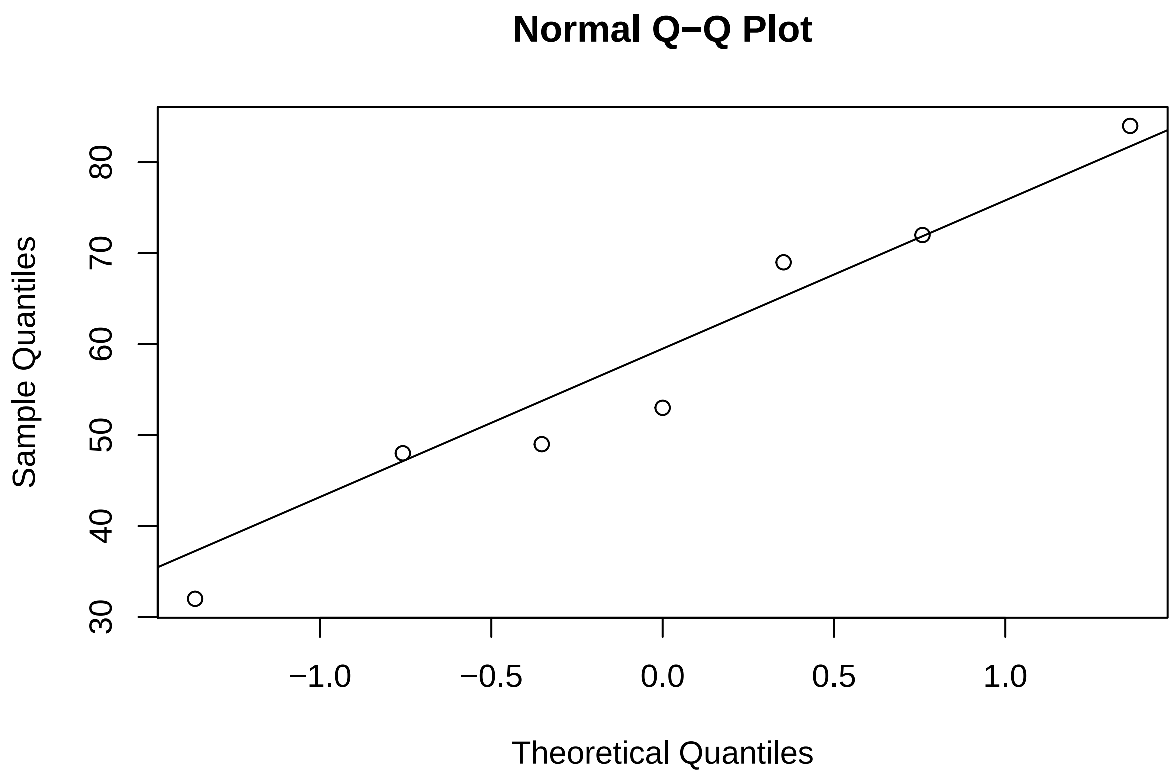
```
res <- shapiro.test(Group3)
res
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Group3
## W = 0.97084, p-value = 0.9044
```

As the p-values are 0.8461, 0.2535 and 0.9044 respectively for Group1, Group2 and Group3 which are all more than 0.05, we accept the null hypothesis that the data is normalized, and further one way ANOVA can be performed on it.

Testing Normality Graphically

```r
qqnorm(Group1) # Quantile-Quantile Plot
qqline(Group1, distribution=qnorm)
```



Similarly we can check normal line for Group2 and Group3.

Testing Homogeneity of Variances

Bartlett's test is used to test if k samples have equal variances. Equal variances across samples is called homogeneity of variances. Before applying ANOVA we have to check the assumption about the homogeneity of variances.

The hypothesis testing problem is

$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$
$H_1 : \sigma_i^2 = \sigma_j^2$ for at least one pair (i, j)

```
result <- data.frame(Group1, Group2, Group3)
res.var<-bartlett.test(list(result$Group1, result$Group2, result$Group3))
res.var
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  list(result$Group1, result$Group2, result$Group3)
## Bartlett's K-squared = 1.771, df = 2, p-value = 0.4125
```

As the p-value 0.4125 is not less than 0.05, we fail to reject the null hypothesis H_0 at 5% level of significance. We conclude that the variances are equal across samples.

One Way ANOVA

```
result
```

```
##   Group1 Group2 Group3
## 1     53     22     61
## 2     48     30     71
## 3     32     19     77
## 4     49     47     81
## 5     69     49     59
## 6     72     55     68
## 7     84     58     53
```

```
st_result <- stack(result)
res <- oneway.test(values~ind, data = st_result)
res
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  values and ind
## F = 6.7983, num df = 2.000, denom df = 11.215, p-value = 0.01165
```

Since p-value which 0.01165 is very less than 0.05 we reject the null hypothesis.

Pair-wise Comparison

```
res.anova <- aov(values~ind, data = st_result)
TK <- TukeyHSD(res.anova)
TK
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = values ~ ind, data = st_result)
##
## $ind
##                    diff        lwr       upr     p adj
```

```
## Group2-Group1 -18.14286 -38.549308  2.263594 0.0864170
## Group3-Group1   9.00000 -11.406451 29.406451 0.5112080
## Group3-Group2  27.14286   6.736406 47.549308 0.0086204
```

```r
plot(TK)
```