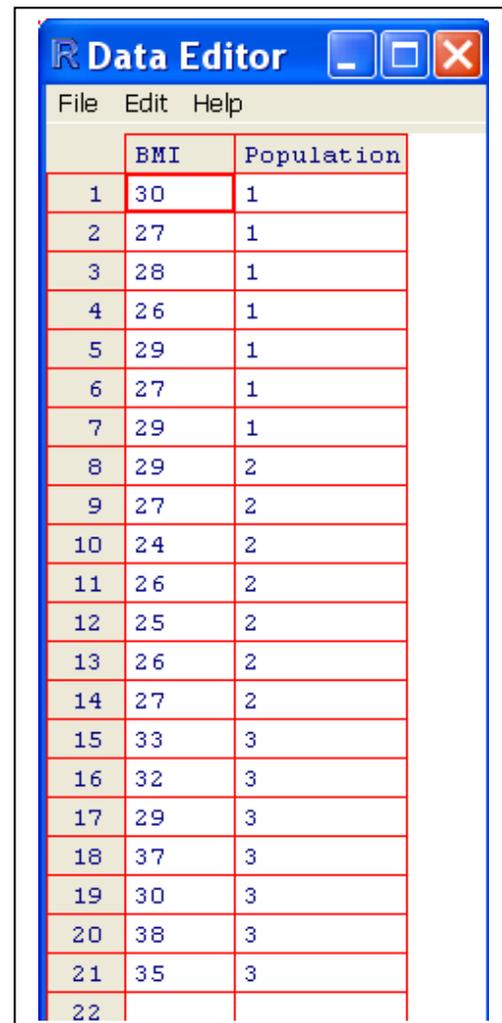


## One Way Analysis of Variance (ANOVA)

**Example:** Researchers wish to see if there is difference in average BMI among three different populations. Use the following data to test if there is significant **difference in average BMI among three different populations**, at 5% level of significance.

BMI		
Population 1	Population 2	Population 3
30	29	33
27	27	32
28	24	29
26	26	37
29	25	30
27	26	38
29	27	35

- 1) Create the data with two variables, BMI and Population. The arrangement of the data should be as in the Figure on the right for data collected from 21 subjects from the three independent samples.
- 2) If the **Population** variable data was entered as numeric, one should convert it into a **Factor variable**. The values of the Population variable, 1, 2, and 3 can be relabeled to Population 1, Population 2, and Population 3, in the process of converting to factor. To convert to factor, click on **Data** in the R Commander menu bar, and then select **Manage variable in active data set ...**, and select **Convert numeric variables to factors ...**. To convert re-label the Population variable, in the **Convert Numeric Variables to Factors dialog box**, select the Population variable and checked on **Use Numbers** option (or click on **Supply level names** option if one wishes to supply names for relabeling the numbers) and click on **OK** button, and also click on **OK** to overwrite variable.



The screenshot shows the R Data Editor window with a data table. The table has three columns: an unlabeled index column, BMI, and Population. The data is as follows:

	BMI	Population
1	30	1
2	27	1
3	28	1
4	26	1
5	29	1
6	27	1
7	29	1
8	29	2
9	27	2
10	24	2
11	26	2
12	25	2
13	26	2
14	27	2
15	33	3
16	32	3
17	29	3
18	37	3
19	30	3
20	38	3
21	35	3
22		

### Check normality assumption in One Way ANOVA

- 3) To check the normality assumption for ANOVA F-test, one can use the following `by()` function in R Commander **Script Window** and click **Submit** button to run the `by` function for **normality test**. Be sure that the Population variable is a factor variable.

R Script for normality test:

```
by(BMI$BMI, BMI$Population, shapiro.test)
```

Data set name.

So, BMI\$Population means the **Population** variable in **BMI** data set.

Output shown in R Commander Output Window:

```
BMI$Population: 1
```

```
Shapiro-Wilk normality test
```

```
data: dd[x, ]  
W = 0.9524, p-value = 0.7518
```

```
-----  
BMI$Population: 2
```

```
Shapiro-Wilk normality test
```

```
data: dd[x, ]  
W = 0.9671, p-value = 0.8766
```

```
-----  
BMI$Population: 3
```

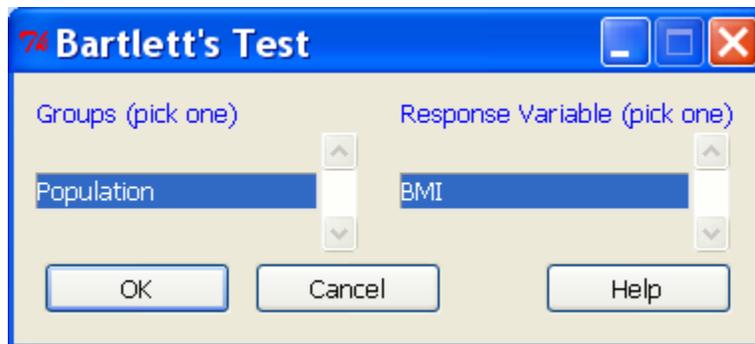
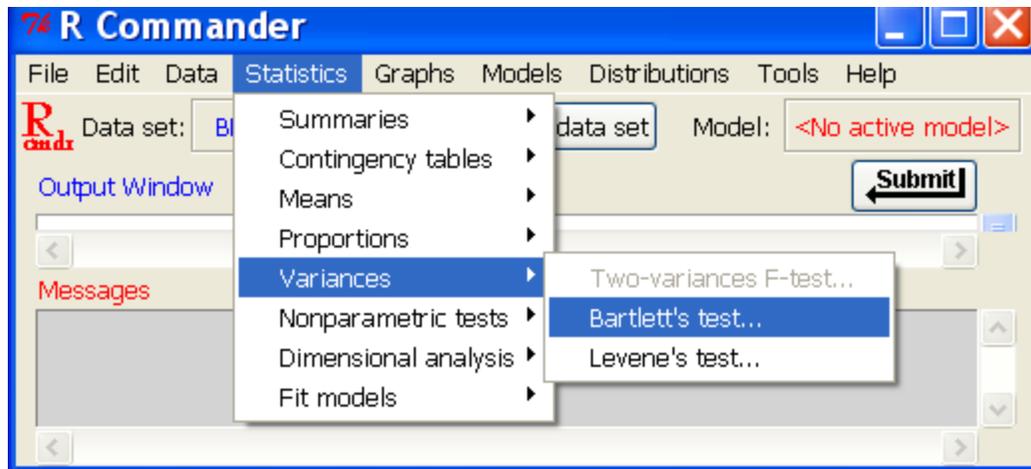
```
Shapiro-Wilk normality test
```

```
data: dd[x, ]  
W = 0.9534, p-value = 0.7602
```

The p-values from normality tests for all three samples are all greater 0.05. The normality assumption is acceptable.

### Check Homogeneity of variances assumption for One Way ANOVA

- 4) For checking the homogeneity of variances assumption, in R Commander, click on **Statistics** in menu bar, and select **Variances**, and then select **Bartlett's test ...** option, and then select the Population as the **group variable** and BMI as the **Response variable** and click OK. See the following figure:



Output from R Commander on Barlett's Test of Homogeneity of Variances:

```
> tapply(BMI$BMI, BMI$Population, var, na.rm=TRUE)
```

```
      1          2          3
2.000000  2.571429 11.619048
```

```
> bartlett.test(BMI ~ Population, data=BMI)
```

```
      Bartlett test of homogeneity of variances
```

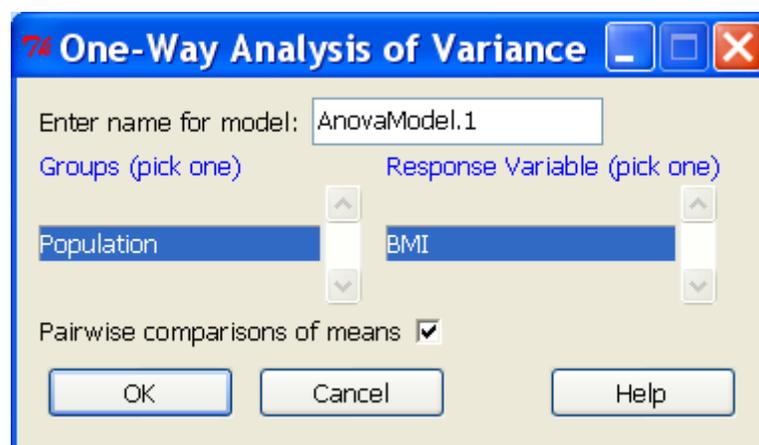
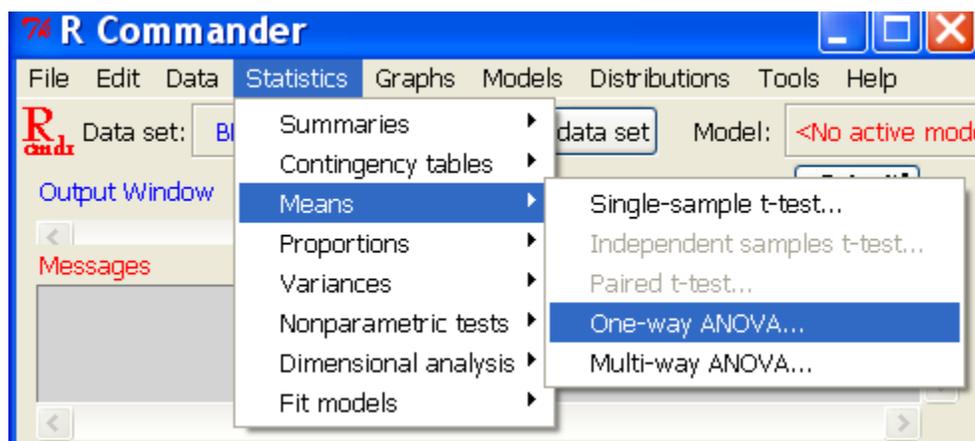
```
data: BMI by Population
```

```
Bartlett's K-squared = 5.4029, df = 2, p-value = 0.06711
```

The p-value of the test is **0.06711** which is greater than 0.05, therefore the homogeneity of variances assumption is acceptable. The difference in variances is not statistically significant.

Since both normality and homogeneity of variances assumptions are satisfied, we can run the One Way ANOVA F-test for equality of population means.

- 5) To run the One Way ANOVA for testing difference in average BMI's, click on **Statistics** and select **Means** and choose **One Way ANOVA ...** option, and in the One-way ANOVA dialog box, choose Population as the **Group** variable and BMI as the **Response** variable. If one wishes to also perform multiple comparisons for means, the **Pairwise comparisons of means** check box should be checked.



The output from R Commander Window:

```
> AnovaModel.1 <- aov(BMI ~ Population, data=BMI)

> summary(AnovaModel.1)
              Df  Sum Sq Mean Sq F value    Pr(>F)
Population    2  194.667   97.333  18.035 5.021e-05 ***
Residuals   18   97.143    5.397
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(The p-value of the F-test is **5.021e-05**, i.e. **0.00005021**, which is less than 0.05. One can conclude that the difference in means is significant.)

```

> numSummary(BMI$BMI , groups=BMI$Population,
+   statistics=c("mean", "sd"))
      mean      sd n
1 28.00000 1.414214 7
2 26.28571 1.603567 7
3 33.42857 3.408672 7

> .Pairs <- glht(AnovaModel.1, linfct =
+   mcp(Population = "Tukey"))

> confint(.Pairs) # confidence intervals

```

### Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts

Fit: aov(formula = BMI ~ Population, data = BMI)

Quantile = 2.5522

95% family-wise confidence level

Linear Hypotheses:

	Estimate	lwr	upr
2 - 1 == 0	-1.7143	-4.8847	1.4561
3 - 1 == 0	5.4286	2.2582	8.5990
3 - 2 == 0	7.1429	3.9725	10.3132

```

> cld(.Pairs) # compact letter display

```

```

2   3   1
"a" "b" "a"

```

(From Tukey's multiple comparisons, the difference between group 2 and group 1 is not significant with a confidence interval estimate for the difference of the two means covers 0. However, the differences between groups 3 and 1, also groups 3 and 2, are both statistically significant with confidence intervals cover a range of positive numbers. See Linear Hypotheses result above.)