

# R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
# dataset
aq <- airquality
head(aq)
```

```
##   Ozone Solar.R Wind Temp Month Day
## 1    41     190  7.4   67     5   1
## 2    36     118  8.0   72     5   2
## 3    12     149 12.6   74     5   3
## 4    18     313 11.5   62     5   4
## 5    NA      NA 14.3   56     5   5
## 6    28      NA 14.9   66     5   6
```

```
str(aq)
```

```
## 'data.frame':   153 obs. of  6 variables:
## $ Ozone   : int  41 36 12 18 NA 28 23 19 8 NA ...
## $ Solar.R: int  190 118 149 313 NA NA 299 99 19 194 ...
## $ Wind    : num  7.4 8 12.6 11.5 14.3 14.9 8.6 13.8 20.1 8.6 ...
## $ Temp    : int  67 72 74 62 56 66 65 59 61 69 ...
## $ Month   : int  5 5 5 5 5 5 5 5 5 5 ...
## $ Day     : int  1 2 3 4 5 6 7 8 9 10 ...
```

```
1
```

```
1.
```

```
## [1] 1
```

```
# gof test for temp variable
# test normality
# using confirmative test : kolmogorov-smirov test for sample size 153
ks.test(aq$Temp, 'pnorm')
```

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  aq$Temp
## D = 1, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

Interpretation: we can see that the p-value is less than 0.05 .So we can conclude that the sample is not normal

2.

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.3
```

```
## Loading required package: carData
```

```
## Warning: package 'carData' was built under R version 4.3.3
```

```
leveneTest(Temp~factor(Month),data = aq)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group    4  2.5849 0.03941 *
##          148
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation: Since the categories of months are greater than 2 , we use leveneTest. Since , p-value <0.05, we can conclude there is not equal variance between these two variables

3. Since the p-value is less than 0.05 , we need to use post-hoc test.

4. The best independent sample statistical test for the data is TukeyHSD test

```
TukeyHSD(aov(Temp~factor(Month),data = aq))
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Temp ~ factor(Month), data = aq)
##
## $'factor(Month)'
```

	diff	lwr	upr	p adj
## 6-5	13.55161290	8.84386422	18.259362	0.0000000
## 7-5	18.35483871	13.68583759	23.023840	0.0000000
## 8-5	18.41935484	13.75035372	23.088356	0.0000000
## 9-5	11.35161290	6.64386422	16.059362	0.0000000
## 7-6	4.80322581	0.09547713	9.510974	0.0430674
## 8-6	4.86774194	0.15999325	9.575491	0.0388654
## 9-6	-2.20000000	-6.94617992	2.546180	0.7038121
## 8-7	0.06451613	-4.60448499	4.733517	0.9999995
## 9-7	-7.00322581	-11.71097449	-2.295477	0.0006215
## 9-8	-7.06774194	-11.77549062	-2.359993	0.0005376

Here we can see that month 8 -7 has highest p adj. Also months 18 7-5 ans 8-5 has difference in mean 18.