Sofia Junqueiro | M2016106

Hugo Silveira da Cunha | M2016488

01-01-2017

1. INTRODUCTION

This work constitutes a practical approach to a set of statistical techniques commonly used in One-way ANOVA process and nonparametric counterparts.

The **main objective** was: to understand, to apply, to interpret and to compare the results of a limited set of parametric and non-parametric tests appropriate to solve a specific marketing problem.

> **Marketing Problem:**
>
> A specific company is interested in studying the overall purchases of its customers and understand if customers from different age groups have a distinct global behavior in relation to the amount spent during a certain period of time. The total sales were recorded over a period of two years, for each customer, belonging to different age groups, forming six groups independent of each other. To conduct this study, an independent random sample of customers was obtained from each of the six age groups.

2. METHODOLOGY

We have included output from two statistical software packages, R and SAS, and try to complement both data analysis and analysis of experiments with some code, mix some text with the code, and publish this work as a notebook.

Firstly, we will do an exploratory data analysis, describing the main statistics and plotting useful information to better understand the data under analysis.

Our objective is to test the equality of the six populations' means, i.e., test if for the different age groups the mean of purchases for the given period is the same in all its levels.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$
$$H_1 : \exists_{i,j(i \neq j)} : \mu_i \neq \mu_j$$

where $i, j = 1, 2, \ldots, 6$, are the number of levels.

If H_0 is true, the mean values are all equal, thus it was obtained a set of 6 samples withdrawn from the same population.

2.1. ONE-WAY ANOVA WITH FIXED EFFECTS

To test H_0, the analysis of the variances of different populations (age groups), popularly known as ANOVA, should be applied.

> **Experimental units**: customers.
>
> **Factors**: age group, with 6 levels: "18-25", "26-35", "36-45", "46-55", "56-65" and "66-90".
>
> **Groups**: 6 age groups (treatments)

STATISTICAL MODEL

Considering that the mean in each age group can be written as:

$$\mu_i = \mu + \alpha_i$$
(i^{th} population mean) = (overall mean) + (i^{th} population effect)

where $\alpha_i = \mu_i - \mu$ is the age effect in each group.

Then, the response $X_{i,j}$, distributed as $N(\mu + \alpha_i, \sigma^2)$, can be expressed in the terms of following **model**;

> $$X_{ij} = \mu_i + \epsilon_{ij} = \mu + \alpha_i + \epsilon_{ij}$$
> (sales) = (overall sales mean) + (age effect) + (random error)

where $\mu_i = \mu + \alpha_i$ is the mean value of the i^{th} age-group in the population, μ is the mean value of the population, α_i is the effect of the i^{th} age-group and $\epsilon_{ij} \sim N(0, \sigma^2)$ is a random residual.

In this formulation, testing the null hypothesis $H_0 : \mu_1 = \mu_2 = \ldots = \mu_6$, is equivalent to test: $H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_6$.

Assuming that H_0 is true, then this leads us to a restatement of the hypothesis of equality of means, satisfying the following constraint:

> $$\sum_{i=1}^{k} \alpha_i = \sum_{i=1}^{k} (\mu_i - \mu) = 0$$

We need this constraint in the model in order to be identifiable: $k + 1$ parameters for only k means, since there are an infinity of $(\mu; \alpha_1; \ldots; \alpha_6)$ that gives us the same means for each X_{ij}

Following this decomposition, the analysis of variance is based on the decompositions of the observations,

$$x_{i,j} = \bar{x} + (\bar{x}_i - \bar{x}) + (x_{i,j} - \bar{x}_l)$$
,
(sales) = (overall sample mean) + (estimated age effect) + (residual),

where $\bar{x}$ is an estimate of μ, $\hat{\alpha}_i = (\bar{x}_i - \bar{x})$ is an estimate of α_i, and $(x_{i,j} - \bar{x}_l)$ is an estimate of the error ϵ_{ij}.

To fit our model, we will use the sample mean:

$$\bar{X}_{i\bullet} = \frac{1}{n_i} \sum_{j=1}^{n_j} X_{ij} = \frac{X_{i\bullet}}{n_i}$$

corresponding to the i^{th} age-group.

However, to perform this analysis, we must validate the respective assumptions:

- Independence condition of the observations;
- Normal populations;
- Homoscedasticity.

2.1.1. Independence condition of the observations

To conduct this study, an independent random sample of customers was obtained from each of the six age groups and each customer can only belong to one of each age group.

2.1.2. Normal populations

$X_i \sim N(\mu_i, \sigma_i)$, unknown parameters, $i = 1 \ldots 6$

We will test the normality condition using the following test statistics:

- Lilliefors' test
- Shapiro-Wilk test

to test:

H_0 : all the six samples come from a population with a normal distribution
H_1 : exists one sample, at least, that doesn't come from a population with a normal distribution

Considering that we have a satisfactory transformation that gives us a good fit to normality for $\lambda = 0 : Y = ln(Sales_i)$, we will carry on both variables assuming, however, that they will follow different paths along the one-way ANOVA process and nonparametric counterparts. This approach will give us the possibility to go through multiple steps of the proposed methodology, the opportunity to apply multiple tests and, on the other hand, will also give us alternative results to be compared.

2.1.3. Populations with the same variance

If the normality condition is verified, the hypothesis of homoscedasticity will be tested:

$$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma^2$$
$$H_1: \exists_{i,j(i \neq j)}: \sigma_i^2 \neq \sigma_j^2$$

We will test the homoscedasticity condition using the following test statistics:

- Bartlett's test.
- Levene's test

2.1.4. ANOVA

To test H_0, the analysis of the variances of different age groups, the model errors are assumed to be normally and independently distributed random variables with mean zero and variance σ^2. The variance σ^2 is assumed to be constant for all Age Groups. In this experiment, we will use the transformed variable $Y = ln(X)$.

2.1.5. MULTIPLE COMPARISON TESTS

If we assume normality and homoscedasticity but we reject the null hypothesis, $H_0: \mu_1 = \ldots = \mu_6$, it only allows to conclude the non-equality between the mean values of the 6 groups. In this case, we will perform multiple comparisons tests to validate if there is evidence of differences between the mean behavior of the age groups regarding the amount spent, using the following test statistics:

- Tukey's HSD test;
- Hochberg (GF2);
- Scheffé'stest;

to test:

$$H_0: \mu_i = \mu_j$$
$$H_1: \mu_i \neq \mu_j, i \neq j$$

2.3. NONPARAMETRIC COUNTERPART OF ANALYSIS OF VARIANCE

For the original data, $X_i = Sales_i$, where the normality assumption is violated, we will use an alternative method to the analysis of variance that does not depend on normality assumptions, the nonparametric tests are applicable regardless of the distribution's form (distribution-free). However, in general, the parametric tests are more powerful than the nonparametric ones, that´s the reason we must validate the ANOVA assumptions before applying these tests.

2.3.1. Non Normal Populations

As we can't assume normality using original data, we will use the following test statistic:

- Kruskal-Wallis test

to test:

$H_0:$ The 6 samples come from the same population, or from identical populations
$H_1:$ Not all of the 6 samples come from the same population, or from identical populations in terms of location

2.3.2. Nonparametric Multiple Comparison Tests

If either normality or homocedasticity condition can't be assumed and we have rejected any of the hypothesis:

- H_0 : The 6 samples come from the same population, or from identical populations
- H_0 : The 6 normal populations have the same mean

then we will use Nonparametric Multiple Comparison Tests:

- Hodges-Lehmann test for independent samples;
- Dwass-Steel-Critchlow-Flignertest;
- Nemenyitest (or Nemenyi-Damico-Wolfe-Dunn test);
- Conover-Inman test;
- Wilcoxon-Mann-Whitney test (i.e., Mann-Whitney U test) with the Bonferroni correction.

to test if two populations have the same median:

$$H_0: \tilde{\mu}_i = \tilde{\mu}_j, (i \neq j)$$

3. RESULTS

In [2]:
```
# packages
library(gplots)
library(ggplot2)
library(dplyr)
library(MASS)
library(psych)
library(gridExtra)
library(nortest)
library(car)
library(graphics)
library(agricolae)
library(PMCMR)
library(conover.test)
library(NSM3)
options(jupyter.plot_mimetypes = c("text/plain", "image/png" ))
```

3.1. EXPLORATORY ANALYSIS

In our exploratory analysis we will compute some descriptive statistics, for the overall sample and for each group, and draw some graphs deemed adequate for the purpose of our study.

In [3]:
```
# import dataset
dataset = read.csv(file = "BD4R.csv", header=TRUE); dataset = dataset[,c(1,2)]
# define var Age as factor var
escEta = c("18-25", "26-35", "36-45", "46-55", "56-65", "66-90");
dataset$Age = factor(dataset$Age, labels = escEta, ordered=TRUE)

# setting sample size
obsGroups = c(50,50,50,50,50,50)

# subset by age group and number of observations
group.1 = subset(dataset,Age == escEta[1])[1:obsGroups[1],];group.2 = subset(dataset,Age == escEta[2])[1:obsGroups[2],]
group.3 = subset(dataset,Age == escEta[3])[1:obsGroups[3],];group.4 = subset(dataset,Age == escEta[4])[1:obsGroups[4],]
group.5 = subset(dataset,Age == escEta[5])[1:obsGroups[5],];group.6 = subset(dataset,Age == escEta[6])[1:obsGroups[6],]
groups = c(group.1,group.2,group.3,group.4,group.5,group.6)

# IMS 'data.frame': 300 obs. of  2 variables
IMS = rbind(group.1,group.2,group.3,group.4,group.5,group.6); str(IMS)
```

```
'data.frame':	300 obs. of  2 variables:
 $ Age  : Ord.factor w/ 6 levels "18-25"<"26-35"<..: 1 1 1 1 1 1 1 1 1 1 ...
 $ Sales: num  1500 35 432 497 885 ...
```

Our data frame IMS contains 300 observations of two variables:

- Age (customers age group, as a factor);
- $Sales$ (amount spent in purchases).

We have selected random samples from six different groups ($group_i$) corresponding to the following Age Groups (levels):

- $18 - 25$;
- $26 - 35$;
- $36 - 45$;
- $46 - 55$;
- $56 - 65$;
- $66 - 90$.

Inside each group we have computed some descriptive measures to better understand the distribution of $Sales_i$.

In [4]:
```
summary(IMS); describeBy(IMS$Sales, IMS$Age, mat=TRUE,type=3,digits=3)
```

```
   Age         Sales
 18-25:50   Min.   :  11.99
 26-35:50   1st Qu.: 138.41
 36-45:50   Median : 457.17
 46-55:50   Mean   : 799.34
 56-65:50   3rd Qu.: 989.37
 66-90:50   Max.   :9621.70
```

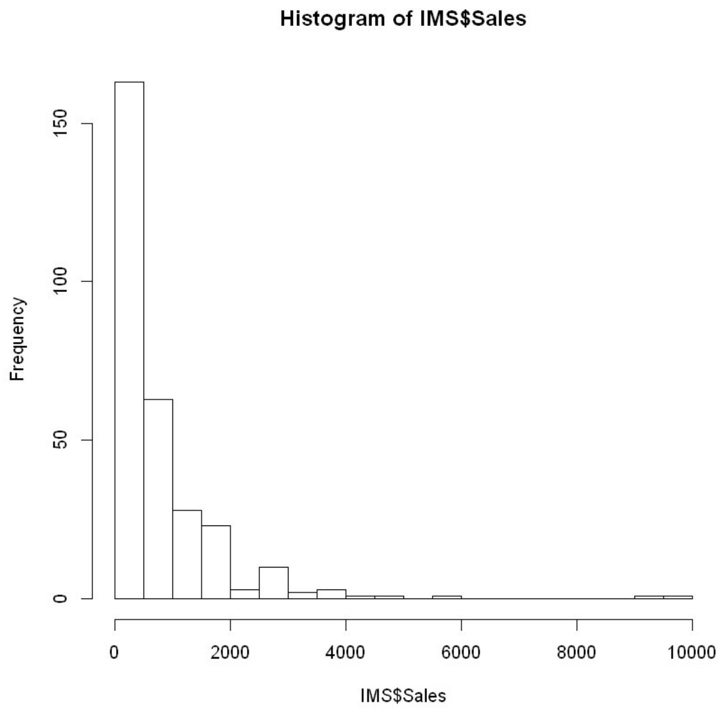
	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X11	1	18-25	1	50	499.761	539.446	390.465	406.759	423.312	11.99	2869.81	2857.82	2.033	5.564	76.289
X12	2	26-35	1	50	563.333	681.432	422.035	415.337	433.134	24.99	3823.46	3798.47	2.704	8.893	96.369
X13	3	36-45	1	50	1037.479	1593.022	469.905	697.744	609.964	13.99	9159.00	9145.01	3.253	12.449	225.287
X14	4	46-55	1	50	1191.420	1600.749	752.955	886.494	911.502	24.47	9621.70	9597.23	3.192	13.142	226.380
X15	5	56-65	1	50	928.595	971.622	586.445	743.451	669.446	20.79	4029.12	4008.33	1.596	1.978	137.408
X16	6	66-90	1	50	575.432	610.287	345.345	465.615	435.647	19.90	3023.65	3003.75	1.758	3.558	86.308

Table 1: Summary statistics from the original data

The values of the first four moments (mean, standard deviation, skewness and kurtosis) of the six samples show that probably the basic assumptions of our model, normality and homocesdasticity, can't be satisfied, so ANOVA cannot be applied appropriately. In fact, we have a positive skew in all groups and values for variance with substantial relative differences, i.e., a presumably presence of non normality and heteroscedasticity in residuals ϵ_{ij}.

We have drawn some graphs to visually explore and have an overview picture and capture some of the differences between the age groups.

In [5]:
```
hist(IMS$Sales, 30)
#plot(density(IMS$Sales, kernel = 'e'))
```



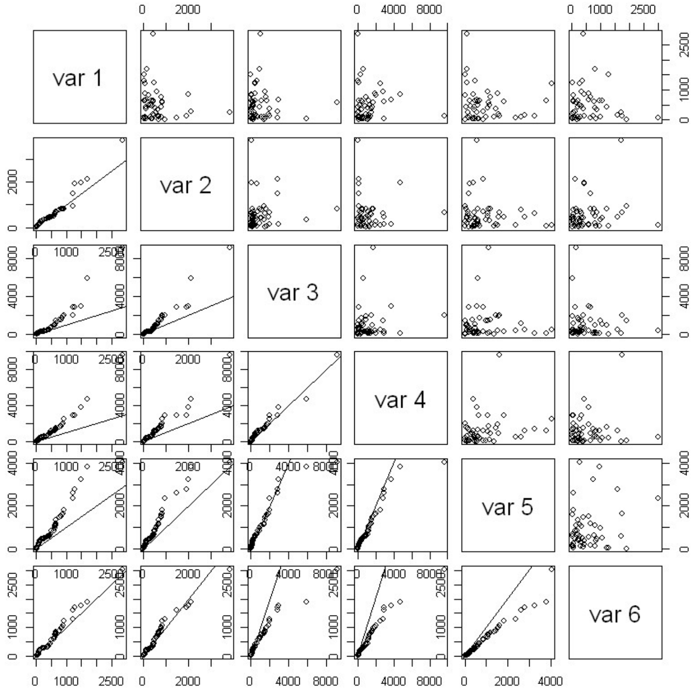
Graphic 1: Histogram of Sales

The histogram shows that the distribution of the customer purchases exhibits a heavy-tailed structure.

We can also see that the distributions are right-skewed, due to a concentration of customers with low amount of purchases. The difference between means and medians, within each group, is influenced by the presence of outliers, particularly in age groups 3 and 4, where the amount purchased exceeds the 9.000 monetary units.

The coefficient of variation is greater than 1 in all groups, which indicates a relatively high variation.

In [6]:
```
salesMatrix = cbind(group.1[,2], group.2[,2], group.3[,2],group.4[,2], group.5[,2], group.6[,2])
panel.qq = function(x, y, ...) {
  usr = par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1), new = TRUE)
  qqplot(x, y, xlab = deparse(substitute(x)), ylab = deparse(substitute(y)))
  abline(c(0,1), ...)
}
# pairs using Q-Q in the lower panel
pairs(salesMatrix, lower.panel = panel.qq)
```

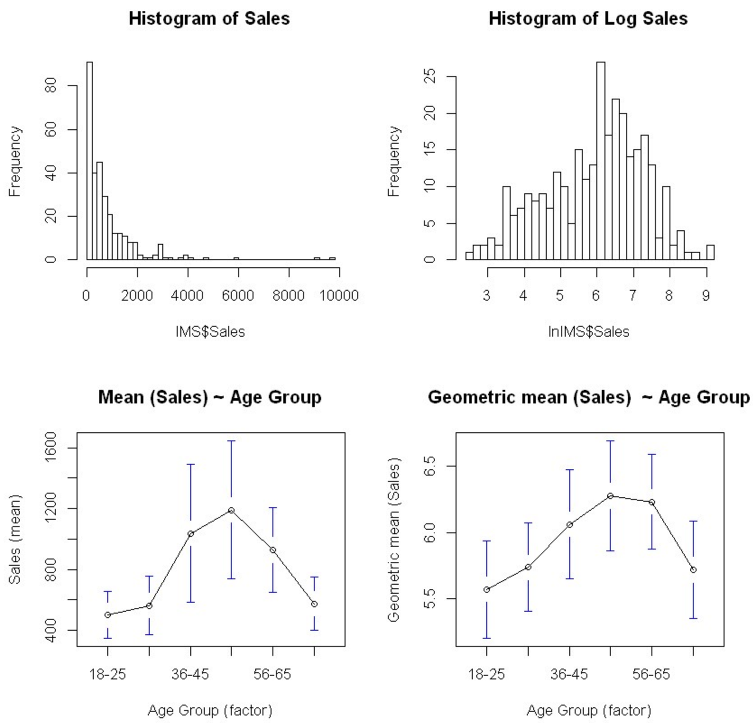


Graphic 2: Samples comparisons in upper diagonal and Q - Q Plots between samples in lower diagonal.

Visually, the direct comparison between groups quantiles, also confirms the behavior of residuals ϵ_{ij} (non normal and heteroscedastic) that we have identified before.

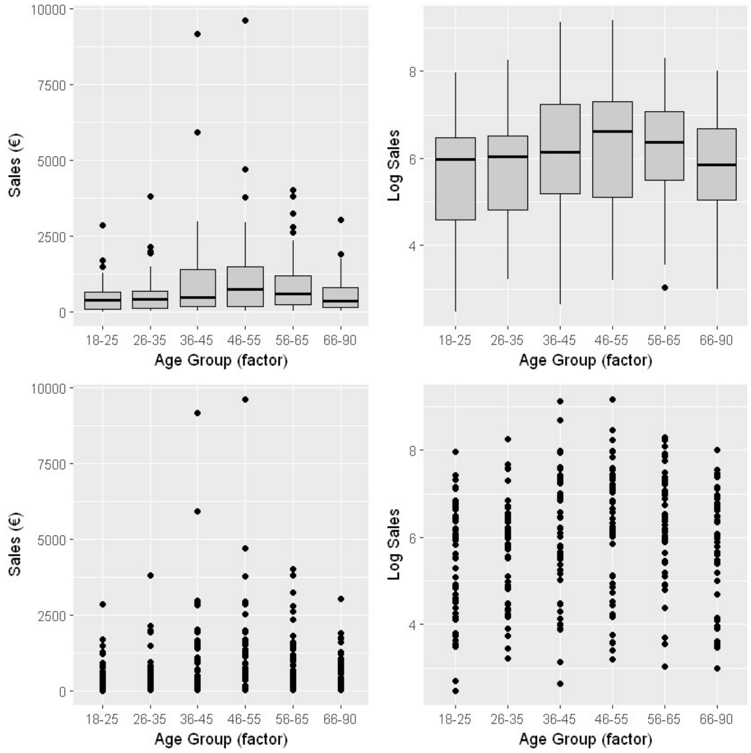
Since we have $Sales_{ij} > 0$, we have tested a power transformation for different values of λ's over the variable $Sales, Y = (Sales)^{\lambda}$, in order to explore the existence of some properties like symmetry or central tendency. We have high variability and a proportionality between the variance and the means of each group, so, a useful transformation for correcting skewness and, at the same time, to handle the outliers is the logarithmic function.

In [7]:
```
lnIMS = IMS; lnIMS$Sales = log(IMS$Sales)
par(mfrow=c(2,2));
hist(IMS$Sales, 50,  main = "Histogram of Sales");
hist(lnIMS$Sales, 30, main = "Histogram of Log Sales")
plotM = plotmeans(formula = Sales ~ Age,
 data = IMS,
 xlab = "Age Group (factor)",
 ylab = "Sales (mean)",
 n.label = FALSE,
 main="Mean (Sales) ~ Age Group")
plotLgM = plotmeans(formula = Sales ~ Age,
 data = lnIMS,
 xlab = "Age Group (factor)",
 ylab = "Geometric mean (Sales)",
 n.label = FALSE,
 main="Geometric mean (Sales)  ~ Age Group")
```



Graphic 3: Histograms and Plots of Group Means and Confidence Intervals from the original and from the transformed data

In [8]:
```
plot1 = ggplot(IMS, aes(x = Age, y = Sales)) +
    geom_boxplot(fill = "grey80", colour = "black") +
    scale_x_discrete() +
    xlab("Age Group (factor)") + ylab("Sales (€)")
plot2 = ggplot(lnIMS, aes(x = Age, y = Sales)) +
    geom_boxplot(fill = "grey80", colour = "black") +
    scale_x_discrete() +
    xlab("Age Group (factor)") + ylab("Log Sales")
plot3 = ggplot(IMS, aes(x = Age, y = Sales)) +
      xlab("Age Group (factor)") + ylab("Sales (€)")+ geom_point()
plot4 = ggplot(lnIMS, aes(x = Age, y = Sales)) +
      xlab("Age Group (factor)") + ylab("Log Sales")+ geom_point()
grid.arrange(plot1, plot2, plot3, plot4, ncol=2)
```



Graphic 4: Box Plots from the original and transformed data

In the box plot we can see the six age groups of customers through their quartiles of `Sales` and `Log Sales`. The `whiskers` gives us the variability, outside the upper and lower quartiles, and a unique `outlier`, in `56-65` group, plotted as an individual point.

The applied transformation has normalized the variable and removed almost all outliers from the six samples.

In [9]:
```
describeBy( lnIMS$Sales, lnIMS$Age, mat=TRUE, type=3, digits=3 )
```

	item	group1	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X11	1	18-25	1	50	5.566	1.307	5.967	5.636	1.301	2.484	7.962	5.478	-0.455	-0.728	0.185
X12	2	26-35	1	50	5.735	1.181	6.045	5.748	0.982	3.218	8.249	5.030	-0.228	-0.737	0.167
X13	3	36-45	1	50	6.059	1.457	6.151	6.091	1.537	2.638	9.122	6.484	-0.216	-0.602	0.206
X14	4	46-55	1	50	6.278	1.470	6.623	6.364	1.140	3.197	9.172	5.974	-0.469	-0.719	0.208
X15	5	56-65	1	50	6.233	1.262	6.374	6.325	1.159	3.034	8.301	5.267	-0.606	-0.165	0.178
X16	6	66-90	1	50	5.716	1.300	5.844	5.786	1.260	2.991	8.014	5.024	-0.478	-0.735	0.184

Table 2: Summary statistics from the transformed data

In [10]:
```
#par(mfrow=c(2,1))
IMS.model.1 = lm(Sales ~ Age, data = IMS); #summary(residuals(IMS.model.1)); hist(residuals(IMS.model.1),30,main="residuals")
IMS.model.lognormal = lm(Sales ~ Age, data = lnIMS); #summary(residuals(IMS.model.lognormal));
#hist(residuals(IMS.model.lognormal),30,main=" ")
```

3.2 DISTRIBUTION FITTING TESTS

Testing Normality

We have tested the normality condition for the response variable:
$X_i \sim N(\mu_i, \sigma_i)$, unknown parameters, $i = 1 \ldots 6$

using the following test statistics:

- Lilliefors' test
- Shapiro-Wilk test

The difference between our observed measures and the predictions of the statistical model we use, i.e., the residuals, was also inspected .

For the transformed variable, $Y_i = ln(X_i)$, we have followed the same process .

In [11]:
```
# normality tests
## response variable
### lillie.test
lillie.test.groups.response = tapply(IMS$Sales,IMS$Age,lillie.test); # lillie.test.groups.response
### shapiro.test
shapiro.test.groups.response = tapply(IMS$Sales,IMS$Age,shapiro.test); # shapiro.test.groups.response

## transformed variable Y
### lillie.test
lillie.test.groups.response.ln = tapply(lnIMS$Sales,lnIMS$Age,lillie.test); # lillie.test.groups.response.ln
### shapiro.test
shapiro.test.groups.response.ln = tapply(lnIMS$Sales,lnIMS$Age,shapiro.test); # shapiro.test.groups.response.ln
```

Age Group	Lilliefors: D	Lilliefors: p-value	Shapiro-Wilk: W	Shapiro-Wilk: p-value
G1	0.18294	0.0002349	0.79052	5.338e-07
ln G1	**0.13272**	**0.02783**	**0.95776**	**0.07162**
G2	0.22435	1.126e-06	0.68808	5.089e-09
ln G2	**0.12083**	**0.06556**	**0.96587**	**0.1564**
G3	0.26028	3.909e-09	0.60694	2.472e-10
ln G3	**0.083945**	**0.5094**	**0.98399**	**0.7273**
G4	0.233	3.142e-07	0.65773	1.554e-09
ln G4	**0.12935**	**0.03585**	**0.94985**	**0.03363**
G5	0.19654	4.675e-05	0.80191	9.703e-07
ln G5	**0.11178**	**0.1228**	**0.95889**	**0.07983**
G6	0.18851	0.0001233	0.80864	1.394e-06
ln G6	**0.12201**	**0.06051**	**0.95164**	**0.03986**

Lilliefors Critical Value	Shapiro-Wilk Critical Value α = (0.05; 0.01)
$D = \frac{0.886}{\sqrt{50}} = 0.1253$	W = (0.947; 0.930)

</pre>

Table 3: Output from the Lilliefors and Shapiro-Wilk tests

Lilliefors Test for Normality of original data

The hypothesis regarding normal distribution is rejected if the test statistic, D, is greater than the critical value. From the table, for $\alpha = 0.05$ and $N > 40$, we have $\frac{0.886}{\sqrt{50}} = 0.1253$.

> Critical Value (upper tail): $D_{crit} = 0.1253$
>
> Critical Region: Reject H_0 if $D_{obs} > 0.1253$

We should reject the null hypothesis at the 0.05 significance level since the value of the Lilliefors test statistic is greater than the critical value in all age groups. So, we assume that there is statistical evidence to claim that the six samples may not come from a population with a normal distribution.

> Reject H_0 for $Sales_i$, i.e., we reject the hypothesis that we have six samples that come from a population with a normal distribution against the alternative that exists one sample, at least, that doesn't come from a population with a normal distribution.

Shapiro-Wilk Test for Normality of original data

The hypothesis regarding normal distribution is rejected if the test statistic, W_{obs}, is lower than the critical valueW_{crit}. From the table, for $\alpha = 0.05$ and $N = 50$, we have $W_{crit} = 0.947$.

> Critical Value (upper tail): $W_{crit} = 0.947$
>
> Critical Region: Reject H_0 if $W_{obs} < 0.947$ ($p - value < 0.05$)

We should reject the null hypothesis at the 0.05 significance level since the value of the Shapiro-Wilk test statistic is lower than the critical value in all age groups. In fact, small values of W are evidence of departure from normality. So, we conclude that there is statistical evidence to claim that the six samples may not come from a population with a normal distribution.

> Reject H_0 for $Sales_i$, i.e., we reject the hypothesis that we have six samples that come from a population with a normal distribution against the alternative that exists one sample, at least, that doesn't come from a population with a normal distribution.

Lilliefors and Shapiro-Wilk tests for transformed data

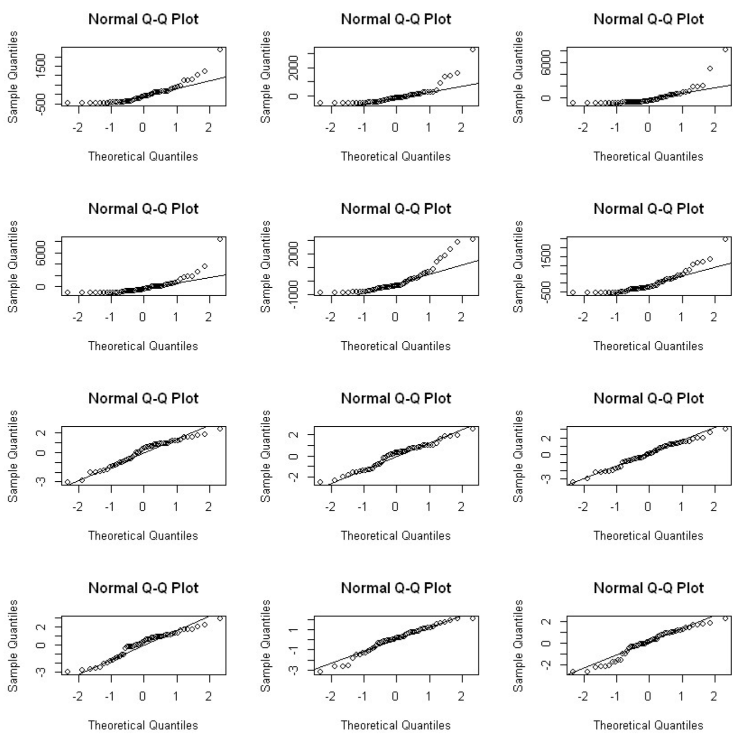
However, for the transformed variable $Y_i = ln(Sales_i)$, the result for the test statistics Lilliefors was almost all "positive" for the six samples, and for Shapiro-Wilk, was "positive" for all the six samples. So, we will assume that there is statistical evidence to claim that the six samples may come from a population with a **lognormal distribution** since the Shapiro-Wilk test usually presents a higher power when compared with other tests, including Lilliefors.

> Do Not Reject H_0 for the transformed variable $Y_i = ln(Sales_i)$, i.e., we accept the hypothesis that we have six samples that come from a population with a lognormal distribution against the alternative that exists one sample, at least, that doesn't come from a population with a lognormal distribution.

Residuals

A check of the normality assumption could be made by inspecting Q-Q Normal plots.

In [12]:
```
IMS.model.1 = lm(Sales ~ Age, data = IMS);
#summary(residuals(IMS.model.1)); hist(residuals(IMS.model.1),30,main="residuals")
IMS.model.lognormal = lm(Sales ~ Age, data = lnIMS); #summary(residuals(IMS.model.lognormal));
# Q-Q plots for groups residuals (6 original + 6 transformed)
functionQQNormLine = function(data){
  qqnorm(data)
  qqline(data)
}
par(mfrow=c(4,3))
qqNormPlots.residuals = tapply(residuals(IMS.model.1),IMS$Age,functionQQNormLine)
qqNormPlots.residuals.ln = tapply(residuals(IMS.model.lognormal),lnIMS$Age,functionQQNormLine)
```



Graphic 5: Q - Q Normal Plots: First 6 Q-Q, residuals from Sales ~ Age and, the last 6, residuals from Y ~ Age.

The Graphic 5, Q-Q Normal plots, illustrates the comparison between quantils of residuals in the six empirical samples and the quantils in a theoretical normal distribution (the line across the diagonal). In a first inspection on the Q-Q plots, in the first 6, we see some departure from normality, while for Y_i, the last 6, we can see some adherence to normality.

3.3 TESTS FOR EQUALITY OF VARIANCES

We will test the homoscedasticity condition for Y_i using the following test statistics:

- Bartlett's test;
- Levene's test (mean, median).

Bartlett

The critical value for Bartlett test of homogeneity of variances, is given by $\chi^2_{obs} > \chi^2_{1-\alpha,5}$, where $\chi^2_{1-\alpha,5}$ is the critical value of the chi-square distribution with 5 degrees of freedom and a significance level of α. From the table, for $\alpha = 0.05$, we have $P(\chi^2_{1-0.05,5} < 11.070) = 0.95$

In [13]:
```
bartlett.test(Sales ~ Age, lnIMS)
```

```
 Bartlett test of homogeneity of variances

data:  Sales by Age
Bartlett's K-squared = 3.5119, df = 5, p-value = 0.6216
```

Since we have $\chi^2_{obs} = 3.5119 < 11.070 = \chi^2_{crit}$, and a $p-value$ of 0.6216.

> Critical Value (upper tail): $\chi^2_{1-0.05,\,5} = 11.070$
>
> Critical Region: Reject H_0 if $\chi^2_{obs} = 3.5119 > \chi^2_{crit} = 11.070$

We fail to reject the null hypothesis at the 0.05 significance level since the value of the Bartlett test statistic is smaller than the critical value. We conclude that there is insufficient evidence to claim that the variances are not equal.

> Do Not Reject H_0 for $ln(Sales_i)$, i.e., we cannot reject the hypothesis that we have equality of variances across age groups against the alternative that variances are unequal for at least two age groups.

Levene Test

We have tested with the mean as the center of each group to Levene's test. The median is better for skewed data, however, as we have assumed the samples come from a lognormal distribution, we will use Levene test instead of Brown–Forsythe test.

In [14]:
```
leveneTest(Sales ~ Age, data=lnIMS, center="mean" )  # Levene Test
#leveneTest(Sales ~ Age, data=lnIMS, center="median") # Brown-Forsythe Test
```

	Df	F value	Pr(>F)
group	5	0.8619505	0.5070482
	294	NA	NA

> Critical Value (upper tail): $F_{(0.05;5;294)} = 2.21$
>
> Critical Region: Reject H_0 if $F_{obs} = 0.86195 > F_{(0.05;5;294)} = 2.21$

We fail to reject the null hypothesis at the 0.05 significance level since the value of the Levene test statistic is smaller than the critical value. We conclude that there is insufficient evidence to claim that the variances are not equal.

> Do Not Reject H_0 for $ln(Sales_i)$, i.e., we cannot reject the hypothesis that we have equality of variances across age groups against the alternative that variances are unequal for at least two age groups.

3.4. ANALYSIS OF VARIANCE

After the verification of ANOVA assumptions we were finally in conditions to apply the analysis of variance on Y_i.

$SQT = SQTr + SQE$

Total variation in $ln(Sales)$ = Variation explained by Age + Variation due to error

In [15]:
```
model.anova = aov( Sales ~ Age, lnIMS ) # one-way anova model
summary(model.anova) # display Type I ANOVA table
```

```
             Df Sum Sq Mean Sq F value Pr(>F)  
Age           5   22.3   4.453   2.504 0.0306 *
Residuals   294  522.8   1.778                 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA

In our statistical model, the factor age of the customer is only responsible for 4% of the total variability.

$$\eta^2 = \frac{SQTr}{SQE} = \frac{22.3}{22.3+522.8} = 0.04$$

In other words, of all the variability that exists within the dependent variable $Y = ln(Sales)$, only 4% is associated with variability in the independent variable "Age Group".

The critical value for the F test, at a significance level of 0.05, is such that $P(F_{(5,294)} < 2.21) = 0.95$.

> Critical Value (upper tail): $F_{(5,294)} = 2.21$
>
> Critical Region: Reject H_0 if $F_{obs} = 2.5037 > F_{crit} = 2.21$

We should reject the null hypothesis at the 0.05 significance level since the value of $F_{obs} = 2.5037$ is greater than the critical value $F_{crit} = 2.21$, with a $p-value$ of 0.0306. In fact, large values of F are evidence against H_0. So, we assume that there is statistical evidence that the six age groups do not have an equal behavior, in mean value, in relation to $Y = ln(Sales)$.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$
$$H_1 : \exists_{i,j(i \neq j)} : \mu_i \neq \mu_j$$

> Reject H_0 for $Y_i = ln(Sales_i)$, i.e., we reject the hypothesis that we have six samples with equal means against the alternative that exists one sample, at least, that doesn't have the same mean.

3.5. MULTIPLE COMPARISON TESTS

3.5.1. Tukey HSD Test

We have applied the Tukey method for each of the 15 different comparisons between μ_i and μ_j for $i \neq j$ of Y.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mu_2 - \mu_1$	$\mu_3 - \mu_1$	$\mu_4 - \mu_1$	$\mu_5 - \mu_1$	$\mu_6 - \mu_1$	$\mu_3 - \mu_2$	$\mu_4 - \mu_2$	$\mu_5 - \mu_2$	$\mu_6 - \mu_2$	$\mu_4 - \mu_3$	$\mu_5 - \mu_3$	$\mu_6 - \mu_3$	$\mu_5 - \mu_4$	$\mu_6 - \mu_4$	$\mu_6 - \mu_5$

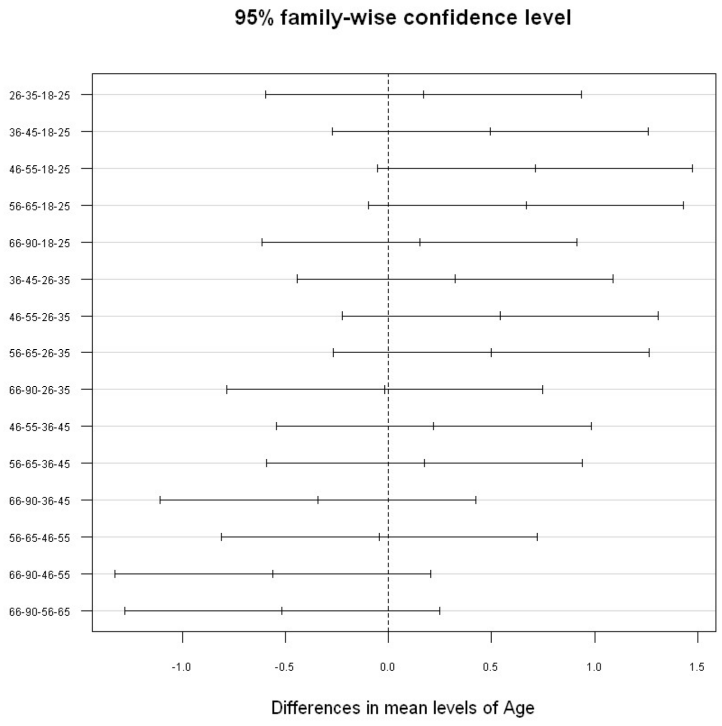
</pre>

In [16]:
```
# Tukey Honestly Significant Differences
TukeyHSD(model.anova, conf.level = 0.95) #  Tukey HSD
plot(TukeyHSD(model.anova, conf.level = 0.95), las=1 ,cex.axis=0.60) # plot geometric diff means and C.I.
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Sales ~ Age, data = lnIMS)

$Age
                    diff         lwr       upr     p adj
26-35-18-25  0.16954455 -0.59554950 0.9346386 0.9882440
36-45-18-25  0.49291241 -0.27218164 1.2580065 0.4365625
46-55-18-25  0.71187170 -0.05322235 1.4769657 0.0846618
56-65-18-25  0.66682736 -0.09826669 1.4319214 0.1273629
66-90-18-25  0.15040659 -0.61468746 0.9155006 0.9932278
36-45-26-35  0.32336786 -0.44172619 1.0884619 0.8306126
46-55-26-35  0.54232714 -0.22276691 1.3074212 0.3259705
56-65-26-35  0.49728280 -0.26781125 1.2623769 0.4262403
66-90-26-35 -0.01913796 -0.78423201 0.7459561 0.9999997
46-55-36-45  0.21895928 -0.54613477 0.9840533 0.9635170
56-65-36-45  0.17391494 -0.59117911 0.9390090 0.9868020
66-90-36-45 -0.34250582 -1.10759987 0.4225882 0.7935560
56-65-46-55 -0.04504434 -0.81013839 0.7200497 0.9999808
66-90-46-55 -0.56146510 -1.32655915 0.2036289 0.2874553
66-90-56-65 -0.51642076 -1.28151481 0.2486733 0.3821627
```



Graphic 6: Pairwise differences in means, 95% family-wise confidence level

The simultaneous pairwise comparisons indicate that the differences between all six means are not significantly different from 0 because the value 0 is inside all confidence intervals, with a confidence coefficient for the set equal to 0.95. In this post-ANOVA comparisons, the required percentile of the studentized range distribution is $q_{(0.05;6;294)} = 4.030$.

This comparison of sample means is applied simultaneously to the set of all pairwise comparisons. However, this test does not tell whether any particular sample mean significantly differs from any particular other but for the purpose of our study this conclusion is sufficient.

3.5.2. Hochberg (GF2)

This is similar to Tukey method but critical values are based on the studentized maximum modulus distribution and is always more conservative than the Tukey test for balanced designs.

SAS OUTPUT

Page Break

The ANOVA Procedure

Studentized Maximum Modulus (GT2) Test for ValorCompras

Note: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	294
Error Mean Square	1.778382
Critical Value of Studentized Maximum Modulus	2.95059
Minimum Significant Difference	0.787

Means with the same letter are not significantly different.			
SMM Grouping	**Mean**	**N**	**EscalaoEtario**
A	6.2777	50	4. [46,55]
A			
A	6.2326	50	5. [56,65]
A			
A	6.0587	50	3. [36,45]
A			
A	5.7353	50	2. [26,35]
A			
A	5.7162	50	6. [66,90]
A			
A	5.5658	50	1. [18,25]

Graphic 7: SAS output for Hochberg (GF2)

So, we may conclude that, at a significance level of 0.05, we have no statistical evidence to presume that exists one means that is significantly different from another.

3.5.3. Scheffé's test

We have rejected the null hypothesis in One-Way Analysis of Variance, it means that at least one of the means isn't the same as the other means, but we don't know wich of them are different.

The critical value for the Scheffe' test, at a significance level of 0.05, is $df_b * F_{crit}$, where F_{crit} is the same that the one from the one-way ANOVA.

> Critical Value (upper tail): $5 * F_{(5,294)} = 5 * 2.21 = 11.05$
>
> Critical Region: Reject H_0 if $F_{obs} > F_{crit} = 11.05$

In [17]:
```
lnIMS.means.comparisons = scheffe.test(model.anova,"Age", group=TRUE, console=TRUE,
main="Log Sales different Age Groups")
#attributes(scheffe.test)
```

```
Study: Log Sales different Age Groups

Scheffe Test for Sales

Mean Square Error  : 1.778382

Age,  means

         Sales      std  r      Min      Max
18-25 5.565784 1.306936 50 2.484073 7.962001
26-35 5.735328 1.181081 50 3.218476 8.248911
36-45 6.058696 1.456912 50 2.638343 9.122492
46-55 6.277655 1.470444 50 3.197448 9.171776
56-65 6.232611 1.261684 50 3.034472 8.301303
66-90 5.716190 1.300236 50 2.990720 8.014220

alpha: 0.05 ; Df Error: 294
Critical Value of F: 2.244703

Minimum Significant Difference: 0.8935256

Means with the same letter are not significantly different.

Groups, Treatments and means
a    46-55    6.278
a    56-65    6.233
a    36-45    6.059
a    26-35    5.735
a    66-90    5.716
a    18-25    5.566
```

So, we may conclude that, at a significance level of 0.05, we have no statistical evidence to presume that exists one means that is significantly different from another.

3.6. NONPARAMETRIC COUNTERPART OF ANALYSIS OF VARIANCE

As we have seen in 3.2., our response variable $Sales$ failed in the assumption of normality. To test if the samples of the age groups come from the same population, or from identical populations, we will use nonparametric tests.

3.6.1. Kruskal-Wallis Test

H_0 : All the 6 samples come from the same population, or from identical populations

H_1 : Not all of the 6 samples come from the same population, or from identical populations in terms of location

In [18]:
```
kruskal.test(Sales ~ Age, data = IMS)
```

```
 Kruskal-Wallis rank sum test

data:  Sales by Age
Kruskal-Wallis chi-squared = 13.22, df = 5, p-value = 0.0214
```

The critical value for the H_0 test, at a significance level of 0.05, is such that $P(\chi^2_{(0.05;5)} < 11.070) = 0.95$

> Critical Value (upper tail): $\chi^2_{(0.05;5)} = 11.070$
>
> Critical Region: Reject H_0 if $\chi^2_{obs} = 13.22 > \chi^2_{crit} = 11.070$

We should reject the null hypothesis at the 0.05 significance level since the value of $\chi^2_{obs} = 13.22$ is greater than the critical value $\chi^2_{crit} = 11.070$, with a $p-value$ of 0.0214. So, we assume that there is statistical evidence that all of the 6 samples may not come from the same population, or from identical populations in terms of location.

> Reject H_0 for $Sales_i$, i.e., we reject the hypothesis that all the 6 samples come from the same population, or from identical populations, against the alternative that exists one sample, at least, that come from a different population.

3.7. NONPARAMETRIC MULTIPLE COMPARISON TESTS

3.7.1. Hodges-Lehmann test for independent samples

The $Hodges-Lehmann$ estimator is the median of the combined data points and Walsh averages. It's the same as the pseudo-median returned as a by-product of the $Wilcoxon$ test.

In [19]:
```
# Hodges-Lehmann
hodges.Lehmann.estimate = function(data){
  wilcox.test(data,  conf.int = TRUE)$estimate
}
tapply(IMS$Sales,IMS$Age,hodges.Lehmann.estimate)
# pairwise difference in location
wilcox.test(group.1$Sales, group.2$Sales, conf.int = TRUE)$estimate
wilcox.test(group.1$Sales, group.3$Sales, conf.int = TRUE)$estimate
wilcox.test(group.1$Sales, group.4$Sales, conf.int = TRUE)$estimate
wilcox.test(group.1$Sales, group.5$Sales, conf.int = TRUE)$estimate
wilcox.test(group.1$Sales, group.6$Sales, conf.int = TRUE)$estimate
wilcox.test(group.2$Sales, group.3$Sales, conf.int = TRUE)$estimate
wilcox.test(group.2$Sales, group.4$Sales, conf.int = TRUE)$estimate
wilcox.test(group.2$Sales, group.5$Sales, conf.int = TRUE)$estimate
wilcox.test(group.2$Sales, group.6$Sales, conf.int = TRUE)$estimate
wilcox.test(group.3$Sales, group.4$Sales, conf.int = TRUE)$estimate
wilcox.test(group.3$Sales, group.5$Sales, conf.int = TRUE)$estimate
wilcox.test(group.3$Sales, group.6$Sales, conf.int = TRUE)$estimate
wilcox.test(group.4$Sales, group.5$Sales, conf.int = TRUE)$estimate
wilcox.test(group.4$Sales, group.6$Sales, conf.int = TRUE)$estimate
wilcox.test(group.5$Sales, group.6$Sales, conf.int = TRUE)$estimate
```

18-25	414.984978074357
26-35	433.579994400496
36-45	735.830018226812
46-55	871.115002417179
56-65	749.959998387523
66-90	479.355054545273

difference in location: -17.3956545286404

difference in location: -136.034168200745

difference in location: -338.280979444067

difference in location: -226.997166008208

difference in location: -27.9717760049179

difference in location: -104.046506950634

difference in location: -323.997532410119

difference in location: -196.269645744521

difference in location: 0.11780429789453

difference in location: -93.8243466185963

difference in location: -68.6769720610071

difference in location: 89.8882646169815

difference in location: 45.4573585384222

difference in location: 276.899733261655

difference in location: 187.913337808548

SAS OUTPUT

Hodges-Lehmann Estimation	
Location Shift (1. [18,25] - 2. [26,35])	-17.4000

Hodges-Lehmann Estimation	
Location Shift (2. [26,35] - 3. [36,45])	-104.1050

Hodges-Lehmann Estimation	
Location Shift (3. [36,45] - 5. [56,65])	-68.6050

95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-167.1200	96.6100	-35.2550	67.2793

Hodges-Lehmann Estimation			
Location Shift (1. [18,25] - 3. [36,45])	-136.0350		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-371.7700	16.0200	-177.8750	98.9278

Hodges-Lehmann Estimation			
Location Shift (1. [18,25] - 4. [46,55])	-338.6100		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-646.6800	-47.9800	-347.3300	152.7324

Hodges-Lehmann Estimation			
Location Shift (1. [18,25] - 5. [56,65])	-226.9350		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-445.3500	-33.8600	-239.6050	104.9739

Hodges-Lehmann Estimation			
Location Shift (1. [18,25] - 6. [66,90])	-28.4700		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-189.0000	88.1300	-50.4350	70.6977

95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-345.0900	57.0300	-144.0300	102.5835

Hodges-Lehmann Estimation			
Location Shift (2. [26,35] - 4. [46,55])	-323.9700		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-635.6600	-26.9900	-331.3250	155.2758

Hodges-Lehmann Estimation			
Location Shift (2. [26,35] - 5. [56,65])	-196.3200		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-421.5800	-15.0000	-218.2900	103.7213

Hodges-Lehmann Estimation			
Location Shift (2. [26,35] - 6. [66,90])	0.1150		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-170.0100	124.0900	-22.9600	75.0269

Hodges-Lehmann Estimation			
Location Shift (3. [36,45] - 4. [46,55])	-93.8100		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-418.9200	125.0000	-146.9600	138.7577

95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-308.9100	147.8800	-80.5150	116.5302

Hodges-Lehmann Estimation			
Location Shift (3. [36,45] - 6. [66,90])	89.7700		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-63.0100	303.1000	120.0450	93.3971

Hodges-Lehmann Estimation			
Location Shift (4. [46,55] - 5. [56,65])	46.1400		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
-185.6400	375.8700	95.1150	143.2450

Hodges-Lehmann Estimation			
Location Shift (4. [46,55] - 6. [66,90])	276.9900		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
20.0800	581.9200	301.0000	143.3292

Hodges-Lehmann Estimation			
Location Shift (5. [56,65] - 6. [66,90])	187.8200		
95% Confidence Limits		Interval Midpoint	Asymptotic Standard Error
0.8900	407.0700	203.9800	103.6193

Graphic 8: SAS output for Hodges-Lehmann test

The median for Sales on Age Groups are:

- 18-25: € 414.98
- 26-35: € 433.58
- 36-45: € 735.83
- 46-55: € 871.12
- 56-65: € 749.96
- 66-90: € 479.36

A non-parametric 0.95 confidence interval for HL_{Δ} provides an estimate of the probability that a randomly chosen customer from $AgeGroup_i$ has a higher value of sales than a randomly chosen customer from $AgeGroup_j$. Based on Hodges-Lehmann Estimation, at a confidence level of 95%, we can observe a significant separation between the following groups:

- 1 vs 4 and 5,
  the pseudo median of group 1 is almost half of the medians of groups 4 and 5, a difference in location of € 338.28 for group 4 and €227.00 for group 5.
- 2 vs 4 and 5, difference in location of € 324.00 for group 4 and € 196.26 for group 5;
- 4 vs 6, difference in location of € 277.90;
- 5 vs 6, difference in location of € 187.91.

3.7.2. Dwass-Steel-Critchlow-Flignertest

To compare the medians of all pairs of groups we use the Steel-Dwass-Critchlow-Fligner pairwise ranking method. This method controls the error rate simultaneously for all the 15 comparisons.

In [20]:
```
# critical value for the Dwass, Steel, Critchlow-Fligner W distribution and parwise comparisons
cSDCFlig(alpha=0.05,n=rep(50,6),method="Monte Carlo",n.mc=500) # iter > 500 take some time to exec.
pSDCFlig(tapply(IMS$Sales,IMS$Age,list),method="Monte Carlo",n.mc=500) # take some time to exec.
```

```
Monte Carlo Approximation (with  500  Iterations) used:

Group sizes: 50 50 50 50 50 50
For the given experimentwise alpha=0.05, the upper cutoff value is Dwass, Steel, Critchlow-Fligner W=4.0849707634956, with true experimentwise alpha level=0.05

Ties are present, so p-values are based on conditional null distribution.
Group sizes: 50 50 50 50 50 50
Using the Monte Carlo (with 500 Iterations) method:

For treatments 1 - 2, the Dwass, Steel, Critchlow-Fligner W Statistic is 0.546.
The smallest experimentwise error rate leading to rejection is 1 .

For treatments 1 - 3, the Dwass, Steel, Critchlow-Fligner W Statistic is 2.4276.
The smallest experimentwise error rate leading to rejection is 0.554 .

For treatments 1 - 4, the Dwass, Steel, Critchlow-Fligner W Statistic is 3.7827.
The smallest experimentwise error rate leading to rejection is 0.094 .

For treatments 1 - 5, the Dwass, Steel, Critchlow-Fligner W Statistic is 3.383.
The smallest experimentwise error rate leading to rejection is 0.2 .

For treatments 1 - 6, the Dwass, Steel, Critchlow-Fligner W Statistic is 0.7409.
The smallest experimentwise error rate leading to rejection is 0.996 .

For treatments 2 - 3, the Dwass, Steel, Critchlow-Fligner W Statistic is 1.6769.
The smallest experimentwise error rate leading to rejection is 0.85 .

For treatments 2 - 4, the Dwass, Steel, Critchlow-Fligner W Statistic is 3.3343.
The smallest experimentwise error rate leading to rejection is 0.21 .

For treatments 2 - 5, the Dwass, Steel, Critchlow-Fligner W Statistic is 3.0418.
The smallest experimentwise error rate leading to rejection is 0.316 .

For treatments 2 - 6, the Dwass, Steel, Critchlow-Fligner W Statistic is -0.0146.
The smallest experimentwise error rate leading to rejection is 1 .

For treatments 3 - 4, the Dwass, Steel, Critchlow-Fligner W Statistic is 1.1504.
The smallest experimentwise error rate leading to rejection is 0.98 .

For treatments 3 - 5, the Dwass, Steel, Critchlow-Fligner W Statistic is 0.8384.
The smallest experimentwise error rate leading to rejection is 0.996 .

For treatments 3 - 6, the Dwass, Steel, Critchlow-Fligner W Statistic is -1.6379.
The smallest experimentwise error rate leading to rejection is 0.868 .

For treatments 4 - 5, the Dwass, Steel, Critchlow-Fligner W Statistic is -0.6922.
The smallest experimentwise error rate leading to rejection is 0.996 .

For treatments 4 - 6, the Dwass, Steel, Critchlow-Fligner W Statistic is -3.2124.
The smallest experimentwise error rate leading to rejection is 0.254 .

For treatments 5 - 6, the Dwass, Steel, Critchlow-Fligner W Statistic is -2.8078.
The smallest experimentwise error rate leading to rejection is 0.366 .
```

SAS OUTPUT

Page Break

The NPAR1WAY Procedure

Pairwise Two-Sided Multiple Comparison Analysis Dwass, Steel, Critchlow-Fligner Method Variable: ValorCompras			
EscalaoEtario	Wilcoxon Z	DSCF Value	Pr > DSCF
1. [18,25] vs. 2. [26,35]	-0.3861	0.5460	0.9989
1. [18,25] vs. 3. [36,45]	-1.7166	2.4276	0.5207
1. [18,25] vs. 4. [46,55]	-2.6748	3.7827	0.0803
1. [18,25] vs. 5. [56,65]	-2.3922	3.3830	0.1588
1. [18,25] vs. 6. [66,90]	-0.5239	0.7409	0.9952
2. [26,35] vs. 3. [36,45]	-1.1857	1.6769	0.8439
2. [26,35] vs. 4. [46,55]	-2.3577	3.3343	0.1714
2. [26,35] vs. 5. [56,65]	-2.1509	3.0418	0.2612
2. [26,35] vs. 6. [66,90]	0.0103	0.0146	1.0000
3. [36,45] vs. 4. [46,55]	-0.8135	1.1504	0.9652
3. [36,45] vs. 5. [56,65]	-0.5929	0.8384	0.9915
3. [36,45] vs. 6. [66,90]	1.1582	1.6379	0.8566
4. [46,55] vs. 5. [56,65]	0.4895	0.6922	0.9966
4. [46,55] vs. 6. [66,90]	2.2715	3.2124	0.2057

5. [56,65] vs. 6. [66,90]	1.9854	2.8078	0.3506

Graphic 9: SAS output for Dwass-Steel-Critchlow-Flignertest test

The critical value for the $Dwass, Steel, Critchlow-Fligner$ test, at a significance level of 0.05, is $W = 3.9387$.

> Critical Value (upper tail): $W = 3.9387$
>
> Critical Region: Reject H_0 if $W > W_{crit} = 3.9387$

From all tests under Dwass-Steel-Critchlow-Flignertest method, $max|W_{obs}| = 3.78 < W_{crit} = 3.9387$, DSCF value from SAS output, so we conclude that there are no significant differences between groups.

3.7.3. Nemenyi-Damico-Wolfe-Dunn test

Pairwise multiple comparisons between group levels and (mean) rank sums of independent samples.

$H_0:$ for each pairwise comparison, the probability of observing a randomly selected value from the first group that is larger than a randomly selected value from the second group equals one half.

This null hypothesis corresponds to that of the Wilcoxon-Mann-Whitney rank-sum test (see 3.7.5).

In [21]:
```
# Pairwise Test for Multiple Comparisons of Mean Rank Sums
output.nemenyi = posthoc.kruskal.nemenyi.test(Sales ~ Age, data = IMS, dist="Chisquare")
print(output.nemenyi$statistic)
```

```
Warning message in posthoc.kruskal.nemenyi.test.default(c(1500.45, 35, 431.82, 496.98, :
"Ties are present. Chi-sq was corrected for ties."
```

```
          18-25       26-35     36-45     46-55    56-65
26-35 0.1604737          NA        NA        NA       NA
36-45 2.9940273 1.768192573        NA        NA       NA
46-55 7.4297295 5.406374932 0.9908754        NA       NA
56-65 5.7383291 3.979583042 0.4424314 0.1090796       NA
66-90 0.2179739 0.004393696 1.5963035 5.1025222 3.719514
```

The critical value for the $Nemenyi$ test, at a significance level of 0.05, is such that $P(\chi^2_{(0.05;5)} < 11.070) = 0.95$

> Critical Value (upper tail): $\chi^2_{(0.05;5)} = 11.070$
>
> Critical Region: Reject H_0 if $\chi^2_{obs} > \chi^2_{crit} = 11.070$

We should not reject the null hypothesis at the 0.05 significance level since all values of χ^2_{obs} are smaller than the critical value $\chi^2_{crit} = 11.070$. So, we assume that there are no statistical evidence that the (mean) rank sums comparisons between our samples are diferent.

> Do not Reject H_0 for $Sales_i$, i.e., we do not reject the hypothesis that the samples have the same mean rank sums against the alternative that exists differences between mean rank sums.

3.7.4. Conover-Inman test

Since we have rejected the $Kruskal-Wallis$, we have applied the $Conover-Inman$ test to verify if the CDF of one age group does not cross the CDF of other, among multiple pairwise comparisons, i.e., we tested the stochastic dominance among six age groups. However, the pair-wise rank sum tests that we are comparing aren't the same that Kruskal-Wallis have tested. We must have in mind that, in this case, if we want to compare the results, we should use Dunn's test, since it preserves a pooled variance for the tests implied by the Kruskal-Wallis null hypothesis.

In [22]:
```
#conover.test(IMS$Sales,IMS$Age, label=TRUE,  table=TRUE, alpha=0.05)
#conover.test(IMS$Sales,IMS$Age,  method="bonferroni", label=TRUE, table=TRUE, alpha=0.05)
```

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1. [18,25]	50	161279.0	169175.0	15121.7793	3225.580
2. [26,35]	50	177071.0	169175.0	15121.7793	3541.420

Conover Two-Sample Test	
Statistic	161279.0000
Z	-0.5222
One-Sided Pr < Z	0.3008
Two-Sided Pr > |Z|	0.6016

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1. [18,25]	50	89587.0	169175.0	15121.7793	1791.740
3. [36,45]	50	248763.0	169175.0	15121.7793	4975.260

Conover Two-Sample Test	
Statistic	89587.0000
Z	-5.2631
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
3. [36,45]	50	201976.00	169174.750	15121.6685	4039.520
5. [56,65]	50	136373.50	169174.750	15121.6685	2727.470

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	201976.0000
Z	2.1692
One-Sided Pr > Z	0.0150
Two-Sided Pr > |Z|	0.0301

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1. [18,25]	50	98069.0	169175.0	15121.7793	1961.380
4. [46,55]	50	240281.0	169175.0	15121.7793	4805.620

Conover Two-Sample Test	
Statistic	98069.0000
Z	-4.7022
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1. [18,25]	50	109133.00	169174.750	15121.6069	2182.660
5. [56,65]	50	229216.50	169174.750	15121.6069	4584.330

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	109133.0000
Z	-3.9706
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
4. [46,55]	50	197701.00	169174.750	15121.6795	3954.020
5. [56,65]	50	140648.50	169174.750	15121.6795	2812.970

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	197701.0000
Z	1.8864
One-Sided Pr > Z	0.0296
Two-Sided Pr > |Z|	0.0592

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
1. [18,25]	50	149060.0	169175.0	15121.7793	2981.20
6. [66,90]	50	189290.0	169175.0	15121.7793	3785.80

Conover Two-Sample Test	
Statistic	149060.0000
Z	-1.3302
One-Sided Pr < Z	0.0917
Two-Sided Pr > |Z|	0.1835

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
2. [26,35]	50	94183.0	169175.0	15121.7793	1883.660
3. [36,45]	50	244167.0	169175.0	15121.7793	4883.340

Conover Two-Sample Test	
Statistic	94183.0000
Z	-4.9592
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
5. [56,65]	50	214998.50	169174.750	15121.6157	4299.970
6. [66,90]	50	123351.00	169174.750	15121.6157	2467.020

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	214998.5000
Z	3.0303
One-Sided Pr > Z	0.0012
Two-Sided Pr > |Z|	0.0024

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
2. [26,35]	50	104678.0	169175.0	15121.7793	2093.560
4. [46,55]	50	233672.0	169175.0	15121.7793	4673.440

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	104678.0000
Z	-4.2652
One-Sided Pr < Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
2. [26,35]	50	114234.00	169174.750	15121.6157	2284.680
5. [56,65]	50	224115.50	169174.750	15121.6157	4482.310

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	114234.0000
Z	-3.6333
One-Sided Pr < Z	0.0001
Two-Sided Pr > |Z|	0.0003

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
3. [36,45]	50	236754.0	169175.0	15121.7793	4735.080
6. [66,90]	50	101596.0	169175.0	15121.7793	2031.920

Conover Two-Sample Test	
Statistic	236754.0000
Z	4.4690
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
2. [26,35]	50	149725.0	169175.0	15121.7793	2994.50
6. [66,90]	50	188625.0	169175.0	15121.7793	3772.50

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	149725.0000
Z	-1.2862
One-Sided Pr < Z	0.0992
Two-Sided Pr > |Z|	0.1984

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
3. [36,45]	50	154602.0	169175.0	15121.7793	3092.040
4. [46,55]	50	183748.0	169175.0	15121.7793	3674.960

Conover Two-Sample Test	
Statistic	154602.0000
Z	-0.9637
One-Sided Pr < Z	0.1676
Two-Sided Pr > |Z|	0.3352

Conover Scores for Variable ValorCompras
Classified by Variable EscalaoEtario

EscalaoEtario	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
4. [46,55]	50	230888.0	169175.0	15121.7793	4617.760
6. [66,90]	50	107462.0	169175.0	15121.7793	2149.240

Average scores were used for ties.

Conover Two-Sample Test	
Statistic	230888.0000
Z	4.0811
One-Sided Pr > Z	<.0001
Two-Sided Pr > |Z|	<.0001

Graphic 10: SAS output for Conover-Inman test

From SAS output, Conover tests show a significant separation between the groups:

- 1 vs 3, 4 and 6;
- 2 vs 3, 4 and 5;
- 3 vs 5 and 6;
- 4 vs 6;
- 5 vs 6.

3.7.5. Wilcoxon-Mann-Whitney test with the Bonferroni correction

To test if they come from distinct populations and the samples do not affect each other we used the Wilcoxon-Mann-Whitney test. To adjust p-values for multiple comparisons and to control the family-wise error rate we used the Bonferroni correction.

$H_0:$ for each pairwise comparison, the probability of observing a randomly selected value from the first group that is larger than a randomly selected value from the second group equals one half.

In [23]:
```
# Wilcoxon-Mann-Whitney test with the Bonferroni correction
pairwise.wilcox.test(IMS$Sales,IMS$Age, p.adj = "bonf")
```

```
 Pairwise comparisons using Wilcoxon rank sum test

data:  IMS$Sales and IMS$Age

      18-25 26-35 36-45 46-55 56-65
26-35 1.00  -     -     -     -
36-45 1.00  1.00  -     -     -
46-55 0.11  0.28  1.00  -     -
56-65 0.25  0.48  1.00  1.00  -
66-90 1.00  1.00  1.00  0.35  0.71

P value adjustment method: bonferroni
```

The Sales of the six Age Groups were compared using the Wilcoxon–Mann–Whitney two-sample rank-sum test. The difference between Age Groups was quantified using the Hodges–Lehmann (HL) estimator, which is consistent with the Wilcoxon test for the median of all possible differences in value of purchases between a customer in $AgeGroup_i$ and the purchases of a customer in $AgeGroup_j$.

A non-parametric 0.95 confidence interval for HL_Δ provides an estimate of the probability that a randomly chosen customer from $AgeGroup_i$ has a higher value of sales than a randomly chosen customer from $AgeGroup_j$. The median for Sales on Age Groups are given in $3.7.1$.

However, in this case, we have inconclusive results due to the Bonferoni correction, which adjusts p-values for the number of tests we have done. We may adjust the false discovery rate in order to have a less conservative correction but this isn't in the scope of our study.

4. CONCLUSION

According to our marketing problem, either by rejecting the ANOVA null hypothesis that all the treatment groups have identical mean values, or by rejecting the Kruskal-Wallis null hypothesis that all the 6 samples come from the same population, or from identical populations, we can conclude that customers from different age groups have a distinct global behavior in relation to the amount spent during the given period of time. In fact, either by testing the original or transformed data, we are led to the same overall conclusion. However, when we perform the multiple-comparisons tests, the same conclusions cannot be reached. Based on parametric multiple-comparison tests, we have no statistical evidence to assume that exists one mean that is significantly different from another.

On the other hand, based on non-parametric counterparts, there are some tests that show a significant separation between some pairs of age groups, in particular, Conover-Inman test and Hodges-Lehmann Estimate. It is important to note that nonparametric tests are subject to the same errors as parametric tests, such as Type I and Type II errors. More, nonparametric tests can be subject to low power mainly due to small sample size. Moreover, although pairwise comparisons are a useful way to fully describe the pattern of mean differences, we might at the same time be increasing the chances of making a Type I error. That´s why the Bonferroni correction is used, to reduce the chances of obtaining false-positive results (type I errors). We could consider to lower the significance level α, but as we lower α we increase β, the chance of a Type II error, that represents the probability of a false negative, failing to find a difference in age group means or medians when there actually is a difference. Another point we wish to remind is that nonparametric multiple-comparisons tests are not testing a hypothesis equivalent to ANOVA Age Group means comparison. These comparisons methods are only useful to check if the probability of a random customer of one age group will score higher than a customer of another age group. The hypotheses they are testing are different and they are not comparable.

Taking into account the presented results, it´s debatable whether we should choose the parametric or nonparametric approach to solve this specific marketing problem. According to original data, the median better represents the center of our distribution, so we might consider the nonparametric tests for marketing purposes. However, our recommendation would be to increase the sample sizes so we can be able to provide more consistent conclusions.

REFERENCES

Bartlett, M. S. (1937). "Properties of sufficiency and statistical tests".
Proceedings of the Royal Statistical Society, Series A 160, 268–282

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing.
Journal of the Royal Statistical Society. Series B (Methodological) 57, 289–300.

Box, G. E. P. and Cox, D. R. (1964), "An analysis of transformations"
Journal of the Royal Statistical Society, Series B, 26, 211-252.

Brown, Morton B.; Forsythe, Alan B. (1974). "Robust tests for the equality of variances".
Journal of the American Statistical Association. 69: 364–367.

Conover, W. J. and Iman, R. L. (1979). On multiple-comparisons procedures.
Technical Report LA-7677-MS, Los Alamos Scientific Laboratory.

Crichtlow, D. E. and Fligner, M. A. (1991). On distribution-free multiple comparisons in the one-way analysis of variance.
Communications in Statistics—Theory and Methods, 20(1):127.

Dunn, O. J. 1964. "Multiple comparisons using rank sums".
Technometrics 6: 241–252.

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. and Altman, D.G. (2016): "Statistical Tests, P-values, Confidence Intervals, and Power: A Guide to Misinterpretations."
European Journal of Epidemiology, Volume 31, Issue 4, pp 337–350

Fisher, R. A. (1922a). On the interpretation of Chi^2 from contingency tables, and the calculation of p.
Journal of the Royal Statistical Society, 84, pp. 87–94.

Hodges, J.L., and Lehmann, E.L. (1963), Estimates of location based on rank tests.
The Annals of Mathematical Statistics, 34, 598–611.

Kruskal, W. H., & Wallis, W. A. (1952), "Use of ranks in one-criterion variance analysis."
Journal of the American Statistical Association, Vol 47, pp. 583–621.

Levene, H. (1960). In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin et al. eds.
Stanford University Press, pp. 278-292.

Lilliefors, H. (June 1967), "On the Kolmogorov–Smirnov test for normality with mean and variance unknown"
Journal of the American Statistical Association, Vol. 62, pp. 399–402.

Pearson, K. (1900), "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.
Philosophical Magazine, 50, pp. 157–175.

Mann, Henry B.; Whitney, Donald R. (1947). "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other".
Annals of Mathematical Statistics. 18 (1): 50–60.

Shapiro, S. S. and Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)".
Biometrika, 52, 3 and 4, pages 591-611.

Student, A. (1908). "The probable error of a mean"
Biometrika, 6, pp. 1–2.

Tukey, John (1949). "Comparing Individual Means in the Analysis of Variance"
Biometrics. 5 (2): 99–114. JSTOR 3001913

W. Tukey, (1957), "On the comparative anatomy of transformations"
Annals of Mathematical Statistics, vol. 23, pp. 604.

Welch, B. L. (1951). On the comparison of several mean values: An alternative approach.
Biometrika, Vol. 38, pp. 330–336.

Wilcoxon, Frank (1945). "Individual comparisons by ranking methods"
Biometrics Bulletin. 1 (6): 80–83.

In [ ]:
```
#sessionInfo()
```