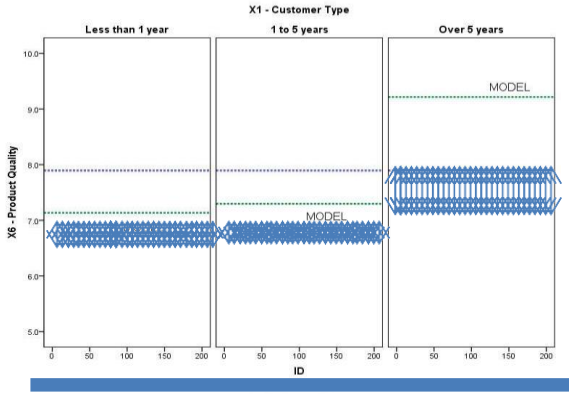
MANOVA

in R

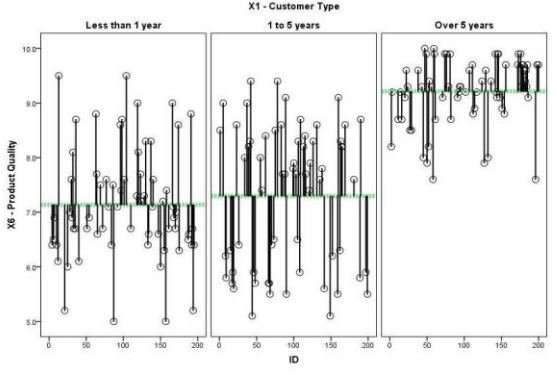
Lab Week 7 – Multivariate Analysis of variance

ANOVA Dataset

F Ratio =



Between Group Variance



WITHIN GROUP VARIANCE

x1	x6
1 to 5 years	8.5
Over 5 years	8.2
Over 5 years	9.2
Less than 1 year	6.4
1 to 5 years	9.0
Less than 1 year	6.5
Less than 1 year	6.9
1 to 5 years	6.2
1 to 5 years	5.8
Less than 1 year	6.4
Over 5 years	8.7
Less than 1 year	6.1
Less than 1 year	9.5
Over 5 years	9.2
1 to 5 years	6.3
Over 5 years	8.7
1 to 5 years	5.7
1 to 5 years	5.9
1 to 5 years	5.6
Over 5 years	9.1
Less than 1 year	5.2
Over 5 years	9.6
1 to 5 years	8.6

MANOVA Dataset

x1	x6	x7
1 to 5 years	8.5	3.9
Over 5 years	8.2	2.7
Over 5 years	9.2	3.4
Less than 1 year	6.4	3.3
1 to 5 years	9.0	3.4
Less than 1 year	6.5	2.8
Less than 1 year	6.9	3.7
1 to 5 years	6.2	3.3
1 to 5 years	5.8	3.6
Less than 1 year	6.4	4.5
Over 5 years	8.7	3.2
Less than 1 year	6.1	4.9
Less than 1 year	9.5	5.6
Over 5 years	9.2	3.9
1 to 5 years	6.3	4.5
Over 5 years	8.7	3.2
1 to 5 years	5.7	4.0
1 to 5 years	5.9	4.1
1 to 5 years	5.6	3.4
Over 5 years	9.1	4.5
Less than 1 year	5.2	3.8
Over 5 years	9.6	5.7
1 to 5 years	8.6	3.6

	X6	**X7**
X6	*Total Variance*	**Co-variance**
X7	**Co-variance**	*Total Variance*

ANOVA Test Statistics

Test of Homogeneity of Variances

X6 - Product Quality

Levene Statistic	df1	df2	Sig.
29.211	2	197	.000

MANOVA Test Statistics

Box's Test of Equality of Covariance Matrices[a]

Box's M	52.306
F	8.590
df1	6
df2	946341.218
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + x1

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
X6 - Product Quality	29.211	2	197	.000
X7 - E-Commerce	3.523	2	197	.031

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + x1

ANOVA Test Statistics

ANOVA

X6 - Product Quality

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	180.573	2	90.286	88.905	.000
Within Groups	200.060	197	1.016		
Total	380.633	199			

MANOVA Test Statistics

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.989	8899.900[b]	2.000	196.000	.000
	Wilks' Lambda	.011	8899.900[b]	2.000	196.000	.000
	Hotelling's Trace	90.815	8899.900[b]	2.000	196.000	.000
	Roy's Largest Root	90.815	8899.900[b]	2.000	196.000	.000
x1	Pillai's Trace	.478	30.942	4.000	394.000	.000
	Wilks' Lambda	.523	37.490[b]	4.000	392.000	.000
	Hotelling's Trace	.909	44.317	4.000	390.000	.000
	Roy's Largest Root	.906	89.284[c]	2.000	197.000	.000

a. Design: Intercept + x1

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

ANOVA Test Statistics

Tests of Between-Subjects Effects

Dependent Variable: X6 - Product Quality

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	180.573[a]	2	90.286	88.905	.000
Intercept	12415.961	1	12415.961	12226.056	.000
x1	180.573	2	90.286	88.905	.000
Error	200.060	197	1.016		
Total	12843.680	200			
Corrected Total	380.633	199			

a. R Squared = .474 (Adjusted R Squared = .469)

MANOVA Test Statistics

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	X6 - Product Quality	180.573[a]	2	90.286	88.905	.000
	X7 - E-Commerce	.319[b]	2	.159	.267	.766
Intercept	X6 - Product Quality	12415.961	1	12415.961	12226.056	.000
	X7 - E-Commerce	2834.132	1	2834.132	4758.317	.000
x1	X6 - Product Quality	180.573	2	90.286	88.905	.000
	X7 - E-Commerce	.319	2	.159	.267	.766
Error	X6 - Product Quality	200.060	197	1.016		
	X7 - E-Commerce	117.336	197	.596		
Total	X6 - Product Quality	12843.680	200			
	X7 - E-Commerce	2952.700	200			
Corrected Total	X6 - Product Quality	380.633	199			
	X7 - E-Commerce	117.655	199			

a. R Squared = .474 (Adjusted R Squared = .469)

b. R Squared = .003 (Adjusted R Squared = -.007)

ANOVA Test Statistics

Multiple Comparisons

Dependent Variable: X6 - Product Quality

LSD

(I) X1 - Customer Type	(J) X1 - Customer Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Less than 1 year	1 to 5 years	-.162	.1755	.358	-.508	.185
	Over 5 years	-2.079*	.1728	.000	-2.420	-1.739
1 to 5 years	Less than 1 year	.162	.1755	.358	-.185	.508
	Over 5 years	-1.918*	.1755	.000	-2.264	-1.572
Over 5 years	Less than 1 year	2.079*	.1728	.000	1.739	2.420
	1 to 5 years	1.918*	.1755	.000	1.572	2.264

Based on observed means.
The error term is Mean Square(Error) = 1.016.

*. The mean difference is significant at the 0.05 level.

MANOVA Test Statistics

Multiple Comparisons

LSD

Dependent Variable	(I) X1 - Customer Type	(J) X1 - Customer Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
X6 - Product Quality	Less than 1 year	1 to 5 years	-.162	.1755	.358	-.508	.185
		Over 5 years	-2.079*	.1728	.000	-2.420	-1.739
	1 to 5 years	Less than 1 year	.162	.1755	.358	-.185	.508
		Over 5 years	-1.918*	.1755	.000	-2.264	-1.572
	Over 5 years	Less than 1 year	2.079*	.1728	.000	1.739	2.420
		1 to 5 years	1.918*	.1755	.000	1.572	2.264
X7 - E-Commerce	Less than 1 year	1 to 5 years	-.098	.1344	.468	-.363	.167
		Over 5 years	-.056	.1324	.673	-.317	.205
	1 to 5 years	Less than 1 year	.098	.1344	.468	-.167	.363
		Over 5 years	.042	.1344	.756	-.223	.307
	Over 5 years	Less than 1 year	.056	.1324	.673	-.205	.317
		1 to 5 years	-.042	.1344	.756	-.307	.223

Based on observed means.
The error term is Mean Square(Error) = .596.

*. The mean difference is significant at the .05 level.

MANOVA ASSUMPTIONS

Normal Distribution

The dependent variable should be normally distributed within groups. Overall, the F test is robust to non-normality, if the non-normality is caused by skewness rather than by outliers. Tests for **outliers** should be run before performing a MANOVA, and outliers should be transformed or removed.

Linearity

MANOVA assumes that there are linear relationships among all pairs of dependent variables, all pairs of covariates, and all dependent variable-covariate pairs in each cell. Therefore, when the relationship deviates from linearity, the power of the analysis will be compromised.

MANOVA ASSUMPTIONS

Homogeneity of Variances

Homogeneity of variances assumes that the dependent variables exhibit equal levels of variance across the range of predictor variables. The error variance is computed (SS error) by adding up the sums of squares within each group. If the variances in the two groups are different from each other, then adding the two together is not appropriate, and will not yield an estimate of the common within-group variance.

Homogeneity of Variances and Covariances

Since there are multiple dependent variables, it is also required that their intercorrelations (covariances) are homogeneous across the cells of the design. There are various specific tests of this assumption.

MANOVA in R

Loading data ‘MANOVA-1.CSV’ into data-frame in R

Sample data

- **Factor** X1-Customer Type (<1 year, 1-5years, >5years)
- **Responses**:

+ X6-Product Quality (cont variable)

+ X7-E-commerce (cont variable)

→ One way MANOVA

Normality test- outliers check

- **Shapiro-Wilk normality test or Kolmogorov-Smirnov**

SW

data: x$x6
W = 0.94985, p-value = 1.824e-06

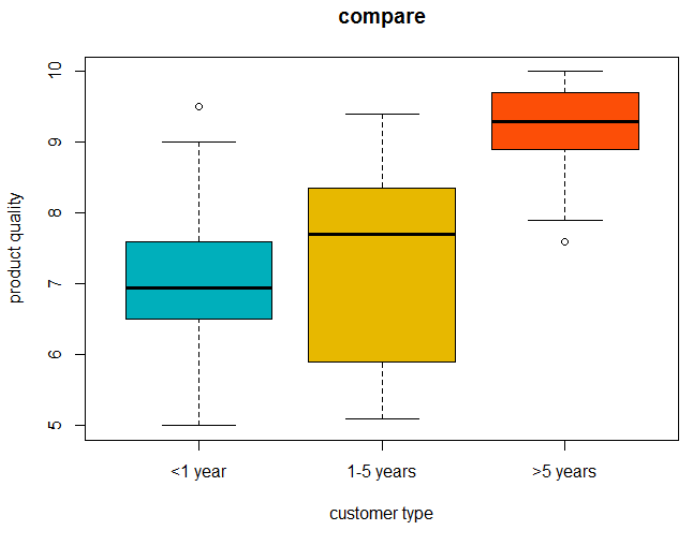
data: x$x7
W = 0.96247, p-value = 3.691e-05

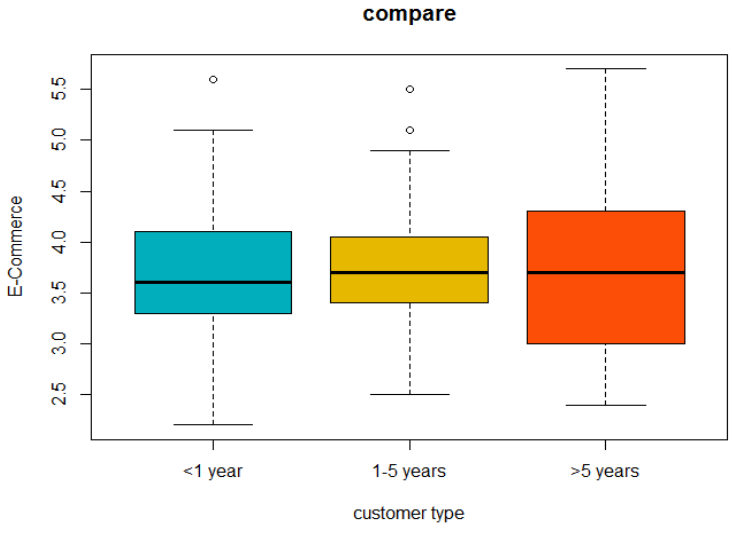
→ We want to see **non-significant** Result.

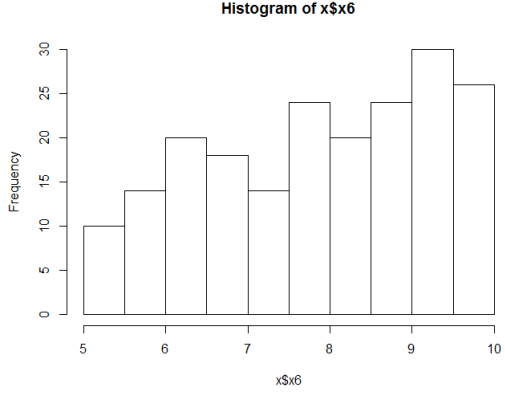
P_value for both x6 and x7 test of SW are equal to 0.000 which is less than 0.01 l.o.s, H0 is rejected and we conclude that the assumption of multivariate normality is not satisfied.

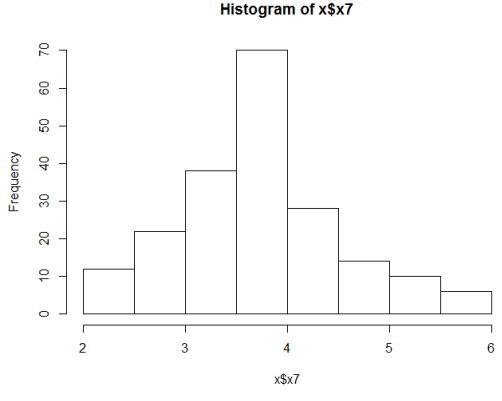
We will conduct the visualization to confirm what you have just concluded:

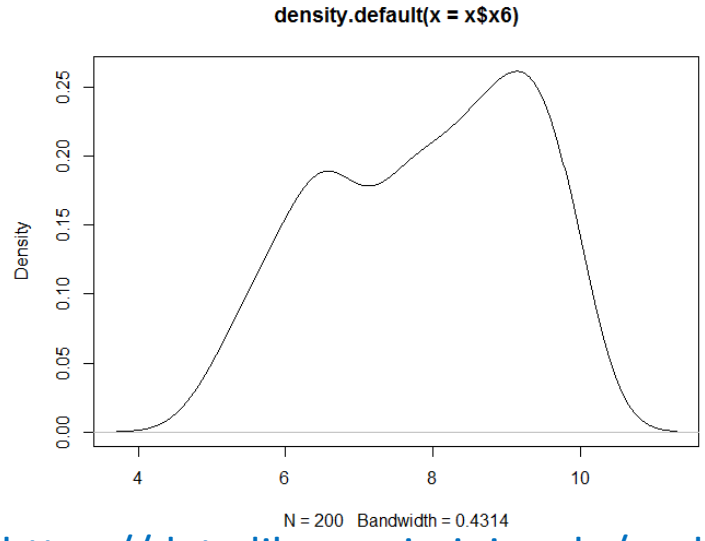
- **Boxplot/Histogram/QQPlot/Density Plot**

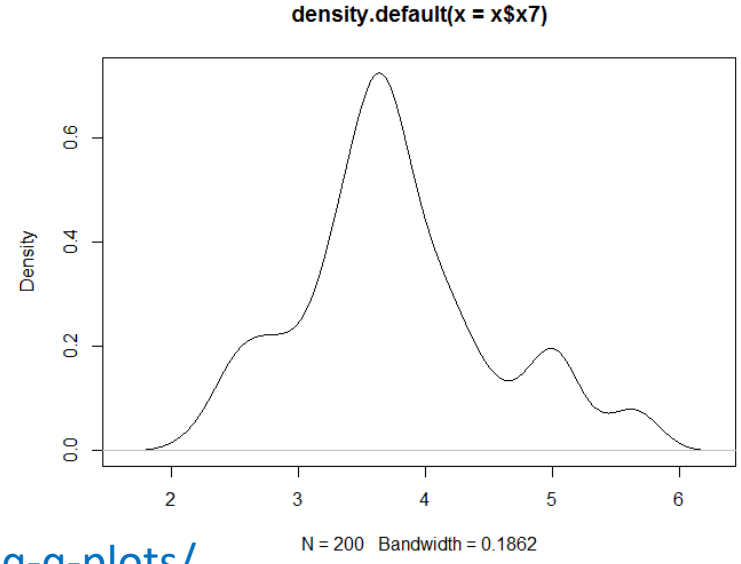




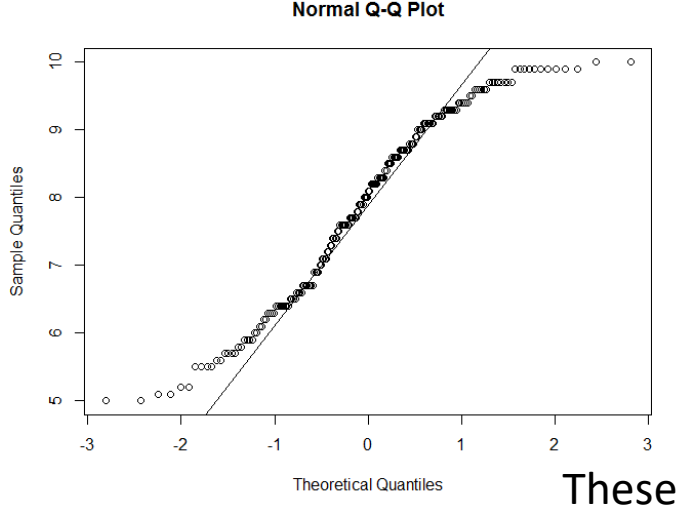


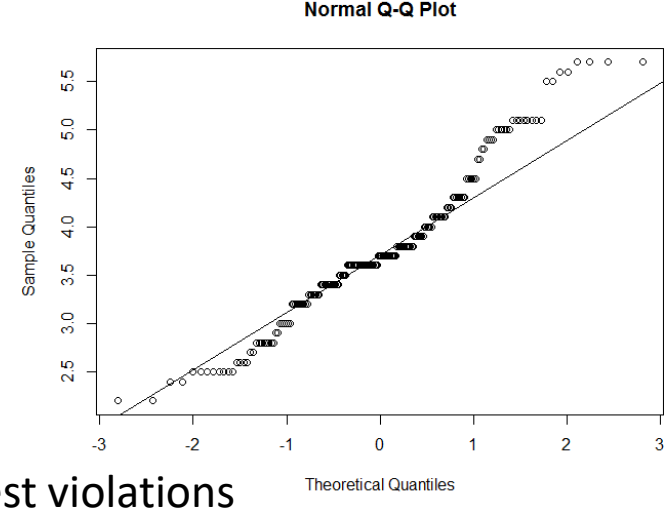




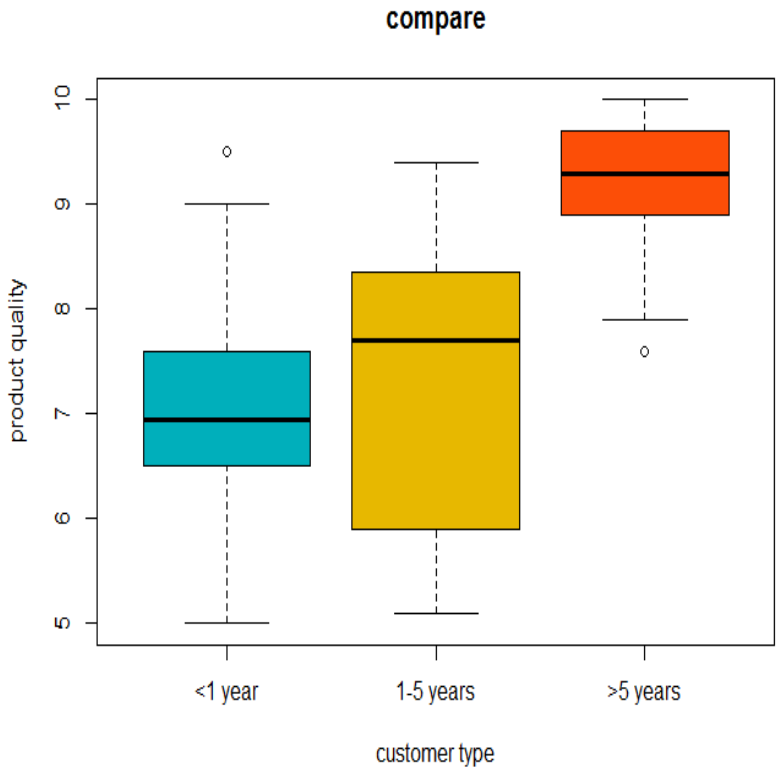


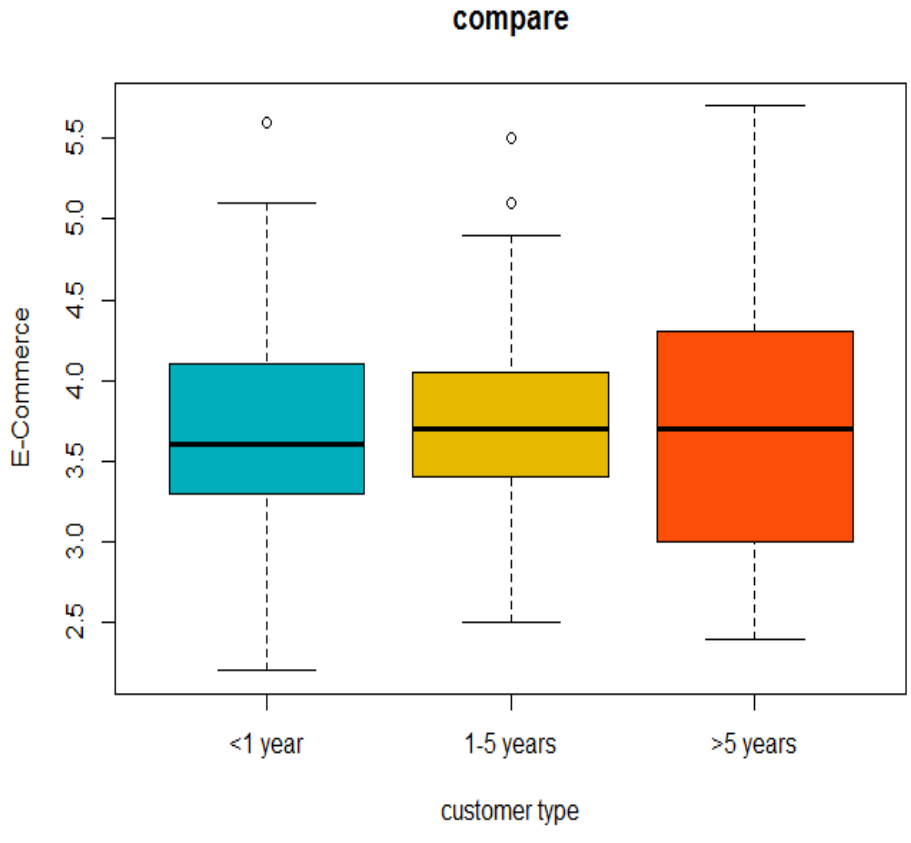
https://data.library.virginia.edu/understanding-q-q-plots/





These graphs suggest violations of normality for both x6 and x7


compare
product quality
5
6
7
8
9
10
<1 year
1-5 years
>5 years
customer type


compare
E-Commerce
2.5
3.0
3.5
4.0
4.5
5.0
5.5
<1 year
1-5 years
>5 years
customer type

OUTLIERS CHECK

We can detect all the outliers using boxplot visualization and remove them out of the dataset to make your data more robust.

There are finally 4 (x6) and 8 (x7) observations should be omitted.

```
> print(outliersx6) [1] 9.5 9.5 7.6 7.6
```

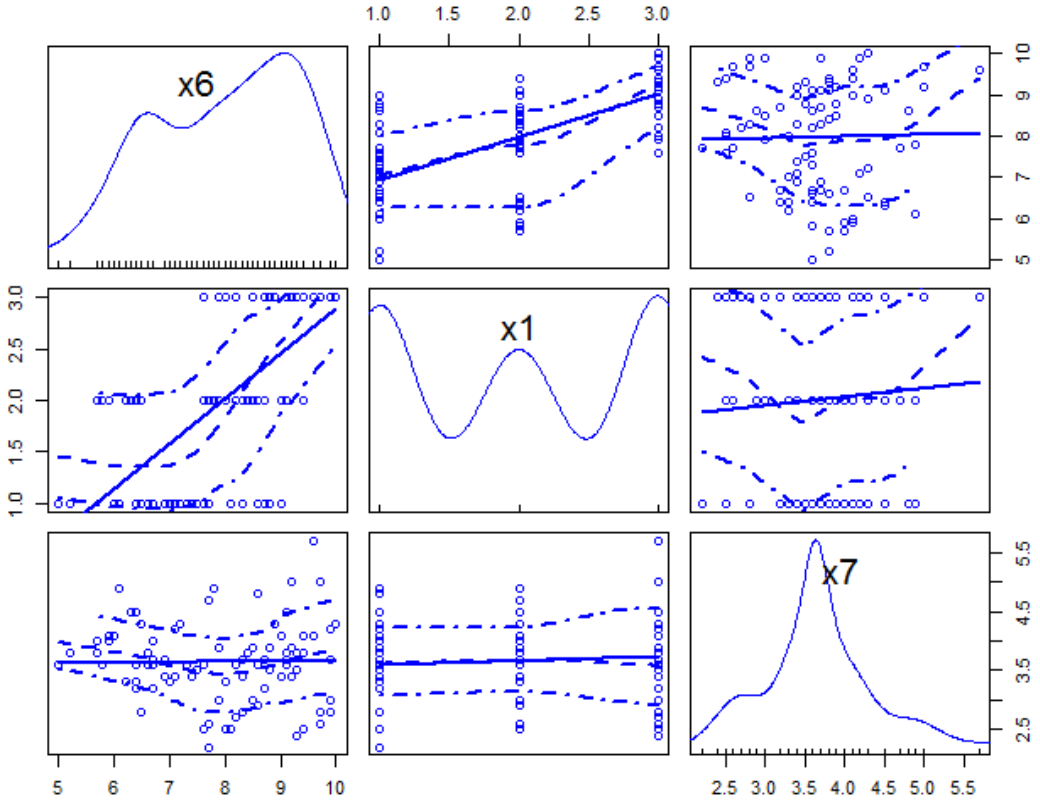
```
> print(outliersx7) [1] 5.6 5.6 5.1 5.1 5.5 5.1 5.1 5.5
```

X6-> observations 13,104, 58,196

X7 -> observations 44,57,90,121,149,199, 13,104

Therefore, 10 observations will be deleted as outliers

Linearity between dependent variables and covariates



Looking at each pair of variables, the visualization shows no violation in linearity assumption

Testing homogeneity of variance & covariance

We are using:
Levene's Test (variance) and Box's Test(cov)

We want to see non-significant results (p>l.o.s). This means the assumption of homogeneity is met

Levene's Test

X6 Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)	
group	2	18.322	**5.039e-08**	***
	197			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

X7 Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)	
group	2	3.6137	**0.02875**	*
	197			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

P_value of Levene's for X7 is 0.02<0.05 → assumption is not satisfied
P_value of Levene's for X6 is 0.000<0.05 → assumption is not satisfied

Box's Test of Equality of Covariance Matrices[a]

Box's M	52.306
F	8.590
df1	6
df2	946341.218
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + x1

P_value of Box test is equal .000 < level of sig 0.05 → homogeneity of cov matrices assumption not met. We should use Pillai's Trace for interpretation of MANOVA results.

Use Levene's test and Box's M test

- If box's M test is not significant (wish to) we look at the **Wilks' Lambda statistics**
- If significant we look at the **Pillai's Trace**

- By Levene's test we check the assumptions of homogeneity of variance →If Levene's test is significant, we look at **Tamhane**, otherwise we will use **Tukey**

Main results of MANOVA

Multivariate tests for the main effects

1. Hotelling's T Square
2. **Pillai-Bartlett trace**
3. **Wilks's lambda**
4. Roy's largest root

Main effects

These four numbers give you the p-values for the four different multivariate tests. These results tell you if there is a significant effect of the IVs on all of the DVs, **considered as a group**. If you are asked "overall, is there a significant effect of something on some set of variables, as a group," then you would run a MANOVA and look at these multivariate tests for your conclusion.

EX: look at Pillai's Trace result. X1 has a statistically significant effect on X6 and X7 as a group (F= 30.9, p=0.000<0.01)

	Df	Pillai approx	F num	Df den Df	Pr(>F)
as.factor(x1)	2	0.47809	**30.942**	4 394	**< 2.2e-16 *****
Residuals	197				

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Tests of between-subjects effects

Response 1 (x6) :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(x1)	2	180.57	90.286	**88.906**	**< 2.2e-16 *****
Residuals	197	200.06	1.016		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response 2 (x7):

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(x1)	2	0.319	0.15926	**0.2674**	**0.7656**
Residuals	197	117.336	0.59562		

→ There is not a significant relationship between x1 and x7 (F_value=0.26, x=0.00<0.01 l.o.s) and there is a significant relationship between x1 and x6 (F_value = 88.9, p=0.7>0.01 l.o.s). OR The p-values of X1 tell you that X1-Customer type has a significant effect on the results of the product quality (p = .000), and has a non-significant effect on the results of the E-commerce (p = .766)

More specific, there is a mean difference in x6 (product quality) according type of customers. Or different level of customer (x1) will result in different product quality.

Multiple comparison tables

Table 1

data: x$x6 and as.factor(x$x1)
alternative hypothesis: two.sided
P value adjustment method: T2 (Sidak)
H0

	t value	Pr(>|t|)	
2 - 1 == 0	0.794	0.8134	
3 - 1 == 0	14.446	<2e-16	***
3 - 2 == 0	10.762	<2e-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Table 2

data: x$x7 and as.factor(x$x1)
alternative hypothesis: two.sided
P value adjustment method: T2 (Sidak)
H0

	t value	Pr(>|t|)	
2 - 1 == 0	0.804	0.80755	
3 - 1 == 0	0.404	0.96935	
3 - 2 == 0	-0.302	0.98674	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- As main effects in previous table shown, no significant results for each level of x1 on x7 found.

A Tamhane post hoc test revealed that the quality was statistically significantly increased after taking the customer type over 5 years compared to the 1 to 5 years and less than 1 year.

In the relationship with X6 there is not a significant difference between less than 1 year and 1 to 5 years predictors (p_value = 0.813) while there are significant differences between less than 1 year and over 5 years (p_value = 0.000); and 1 to 5 years and over 5 years (p_value = 0.000).

In the relationship with X7 all *p*-value are **GREATER THAN .05**, then there are non-significant differences between the independent groups or levels of the categorical predictor variables.

The mean difference between over 5 years and less than 1 year is the highest (difference value is equal to approx. 14.4)