The FREQ Procedure

Table of agegrp by response			
agegrp	**response**		
Frequency Expected	**no**	**yes**	**Total**
older	225 197.99	253 280.01	478
younger	434 461.01	679 651.99	1113
Total	659	932	1591

Exercise 1a):

We constructed a contingency table for **agegrp** and **response** variables. We also presented expected frequency and observed frequencies for all categories. We see that we have higher than expected observations in the cell where older individual did not do home improvements and in the cell where younger individual did home improvements. We have smaller than expected observations in the cell where younger individual did not do home improvements and in the cell where older individual did home improvements.

Exercise 1b):

We performed the following tests (on the next page). We see that all p-values are smaller than significance level. So, we reject null hypothesis and we accept alternative hypothesis which states there is association between variables. And there is a statistically significant association. It also provides us with the measures of association, however we see that it is pretty weak. (approx. 0.075)

The FREQ Procedure

Table of agegrp by response			
agegrp	**response**		
Frequency Expected	**no**	**yes**	**Total**
older	225 197.99	253 280.01	478
younger	434 461.01	679 651.99	1113
Total	659	932	1591

Statistics for Table of agegrp by response

Statistic	DF	Value	Prob
Chi-Square	1	8.9916	0.0027
Likelihood Ratio Chi-Square	1	8.9394	0.0028
Continuity Adj. Chi-Square	1	8.6618	0.0032
Mantel-Haenszel Chi-Square	1	8.9860	0.0027
Phi Coefficient		0.0752	
Contingency Coefficient		0.0750	
Cramer's V		0.0752	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	225
Left-sided Pr <= F	0.9988
Right-sided Pr >= F	0.0017
Table Probability (P)	0.0005
Two-sided Pr <= P	0.0032

Sample Size = 1591

The FREQ Procedure

Table of agegrp by response			
agegrp	**response**		
Frequency Row Pct	**no**	**yes**	**Total**
older	225 47.07	253 52.93	478
younger	434 38.99	679 61.01	1113
Total	659	932	1591

Statistics for Table of agegrp by response

Column 1 Risk Estimates						
	Risk	**ASE**	**95% Confidence Limits**		**Exact 95% Confidence Limits**	
Row 1	0.4707	0.0228	0.4260	0.5155	0.4252	0.5166
Row 2	0.3899	0.0146	0.3613	0.4186	0.3612	0.4193
Total	0.4142	0.0123	0.3900	0.4384	0.3899	0.4389
Difference	0.0808	0.0271	0.0276	0.1339		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	**ASE**	**95% Confidence Limits**		**Exact 95% Confidence Limits**	
Row 1	0.5293	0.0228	0.4845	0.5740	0.4834	0.5748
Row 2	0.6101	0.0146	0.5814	0.6387	0.5807	0.6388
Total	0.5858	0.0123	0.5616	0.6100	0.5611	0.6101
Difference	-0.0808	0.0271	-0.1339	-0.0276		
Difference is (Row 1 - Row 2)						

Sample Size = 1591

Exercise 1c):

We now tested if younger individuals have a significantly higher probability of doing home improvements they would have previously hired someone to do. We look at the second table (Column 2 Risk Estimates), we see that it is equal to 0.6101. Also, by looking at the difference, we see that it is negative (and do not involve 0), so younger individuals (Row2) have a significantly higher probability of doing home improvements. So, the difference is significant. It is estimated to be 0.0808 higher than the risk for older individuals of doing home improvements.

The FREQ Procedure

Table of work by response			
work	**response**		
Frequency Expected	**no**	**yes**	**Total**
office	301 323.91	481 458.09	782
skilled	119 149.11	241 210.89	360
unskill	239 185.98	210 263.02	449
Total	659	932	1591

Exercise 2a):

We constructed a contingency table for **work** and **response** variables. We also presented expected frequency and observed frequencies for all categories. We see that we have smaller than expected observations in the cell where individual who works at the office did not do home improvements and in the cell where skilled individual did not do home improvements and in the cell where unskilled individual did home improvements. We have higher than expected observations in the cell where individual who works at the office did home improvements and in the cell where skilled individual did home improvements and in the cell where unskilled individual did not do home improvements.

The FREQ Procedure

Table of work by response			
work	**response**		
Frequency Expected	**no**	**yes**	**Total**
office	301 323.91	481 458.09	782
skilled	119 149.11	241 210.89	360
unskill	239 185.98	210 263.02	449
Total	659	932	1591

Statistics for Table of work by response

Statistic	DF	Value	Prob
Chi-Square	2	38.9525	<.0001
Likelihood Ratio Chi-Square	2	38.7783	<.0001
Mantel-Haenszel Chi-Square	1	20.4480	<.0001
Phi Coefficient		0.1565	
Contingency Coefficient		0.1546	
Cramer's V		0.1565	

Sample Size = 1591

Exercise 2b):

We performed the following tests; we see it on the above table. We see that all p-values are smaller than significance level. So, we reject null hypothesis and we accept alternative hypothesis which states there is association between variables. And there is a statistically significant association. It also provides us with the measures of association; however we see that it is only approx. 0.16.

The FREQ Procedure

Table of Species by swgroup			
Species(Iris Species)	**swgroup**		
Frequency Expected	**Longer**	**Shorter**	**Total**
Setosa	42 22.333	8 27.667	50
Versicolor	8 22.333	42 27.667	50
Virginica	17 22.333	33 27.667	50
Total	67	83	150

Exercise 3a):

We constructed a contingency table for **Species** and **swgroup** variables. We see that we have higher than the expected observations in the cell where setosa has longer (>30mm) sepal width, and in the cell where Versicolor has shorter sepal width, and in the cell where virginica has shorter sepal width. We have lower than the expected observations in the cell where setosa has shorter sepal width, and in the cell where versicolor has longer sepal width, and in the cell where virginical has longer sepal width. So setosa has the highest probability of having longer longer sepal width and versicolor has the lowest probability of having shorter sepalwidth.

The FREQ Procedure

Table of Species by swgroup			
Species(Iris Species)	**swgroup**		
Frequency Expected	**Longer**	**Shorter**	**Total**
Setosa	42 22.333	8 27.667	50
Versicolor	8 22.333	42 27.667	50
Virginica	17 22.333	33 27.667	50
Total	67	83	150

Statistics for Table of Species by swgroup

Statistic	DF	Value	Prob
Chi-Square	2	50.2248	<.0001
Likelihood Ratio Chi-Square	2	54.1967	<.0001
Mantel-Haenszel Chi-Square	1	25.1191	<.0001
Phi Coefficient		0.5786	
Contingency Coefficient		0.5008	
Cramer's V		0.5786	

Sample Size = 150

Exercise 3b):

We performed the following tests; we see it on the above table. We see that all p-values are smaller than significance level. So, we reject null hypothesis and we accept alternative hypothesis which states there is association between variables. And there is a statistically significant association. It also provides us with the measures of association, all of them are slightly above 0.5, which is considered as a strong relationship. A look at the data shows that **setosa** species often have longer sepal width than expected due to chance, **versicolor** species less likely to have longer sepal width than expected. These can be observed by looking at the differences in sepalwidth measurements as well.

The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
Species	3	Setosa Versicolor Virginica

Number of Observations Read	150
Number of Observations Used	150

Exercise 4a):

We performed a one-way ANOVA for **sepalwidth** with **species** as the categorical predictor. The analysis of variance and diagnostics are on the next page. We see that we don't have any missing observations: number of observations read = number of observations used. The first ANOVA table was generated for the overall model. Further we state our null hypothesis; under the null hypothesis the model explains no more variation per degree of freedom than expected. So, based on the table we got F-value of 49.16 and p-value very small (<0.0001), so we reject null hypothesis. Our model is **statistically significant**, as the variation described by the model gets larger. The variation described by differences across species is significantly greater (from a statistical perspective) than expected due to chance.

Looking at the R-square value in the next table, of the next page we see that it is equal to 40%. We can say that is not a huge percentage, but it is rather weak percentage of **practical significance**. About 40% of variation in sepal width could be described by type of species alone.

The data in the third table would be identical to the values in the first table. So the conclusion species term describes more variation than expected and it is good to keep it in the model.

Lastly, it also generated a box plot that visualizes the distributions of the sepalwidth values for the different types of species. We may observe that **setosa species** generally have longer (>30mm) sepalwidth comparing to other types of species.

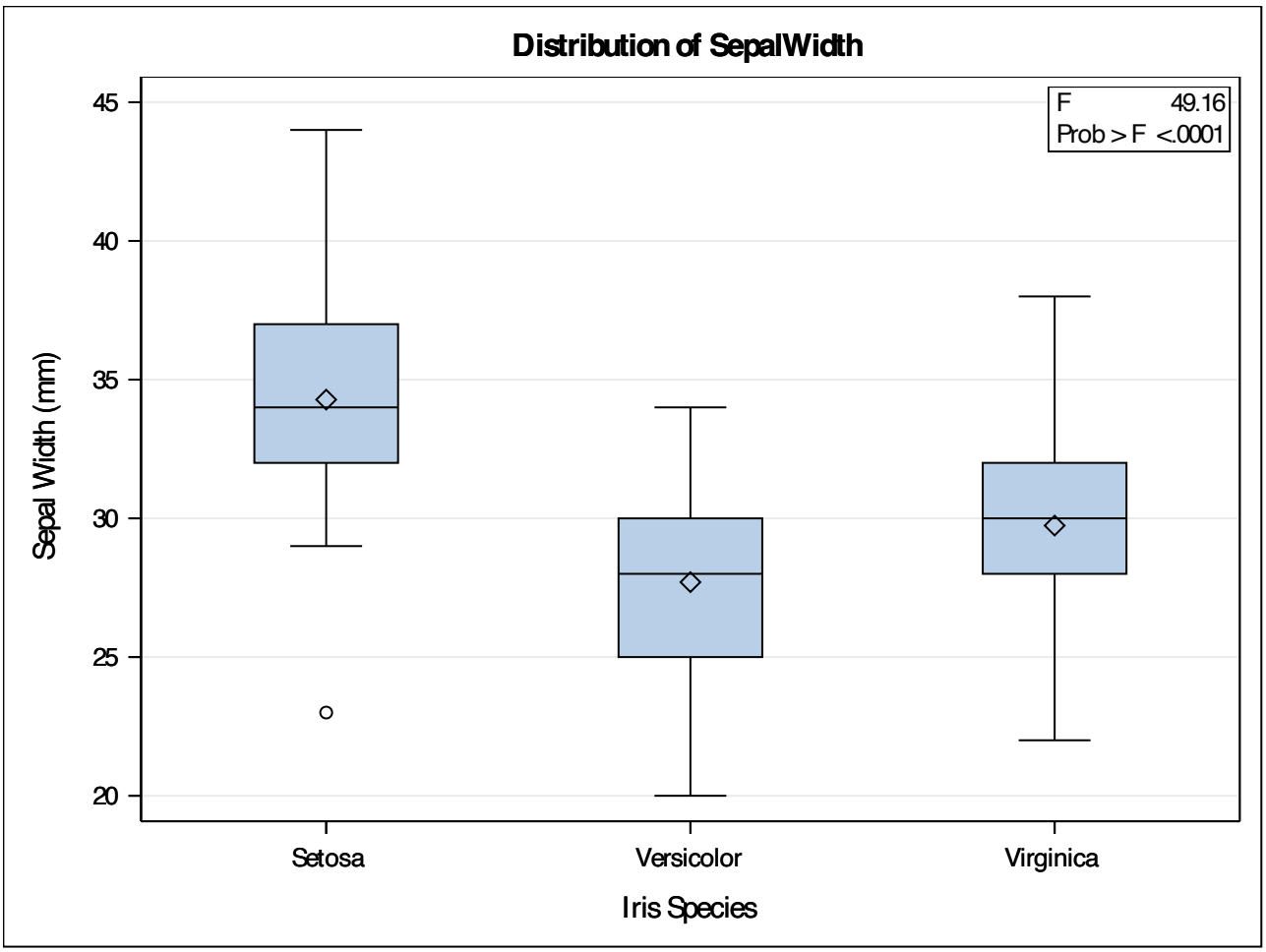
The ANOVA Procedure

Dependent Variable: SepalWidth Sepal Width (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1134.493333	567.246667	49.16	<.0001
Error	147	1696.200000	11.538776		
Corrected Total	149	2830.693333			

R-Square	Coeff Var	Root MSE	SepalWidth Mean
0.400783	11.11059	3.396877	30.57333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Species	2	1134.493333	567.246667	49.16	<.0001



The ANOVA Procedure

Class Level Information		
Class	**Levels**	**Values**
Species	3	Setosa Versicolor Virginica

Number of Observations Read	150
Number of Observations Used	150

Exercise 4b):

Next, we test equal variance assumption. First tables and box plot were the same tables that we previously got. Now we focus on the next table that includes Levene's test. Null hypothesis will be that the variances are the same for each type of species. We see that the p-value is larger than 0.05 (<0.4073), so we accept null hypothesis.

Moving now to Tukey's test for differences of sepalwidth means, a table containing alpha value, degrees of freedom and a few statistics needed for constructing confidence intervals for the differences of means is generated and followed by the comparison of means. The estimated differences of means are also provided along with 95% confidence intervals. From the table we can see that mean **sepalwidth** measurement for Setosa species is higher than for species Versicolor and Virginica. However, we may observe that the differences are not high (between 2.04-6.58). Moreover, all of the differences are statistically significant, since they don't involve 0 in the confidence intervals.

Generally, it matches the conclusion from Exercise 3b) about the difference measurements across species as stated above. However, there is some discrepancy in values.

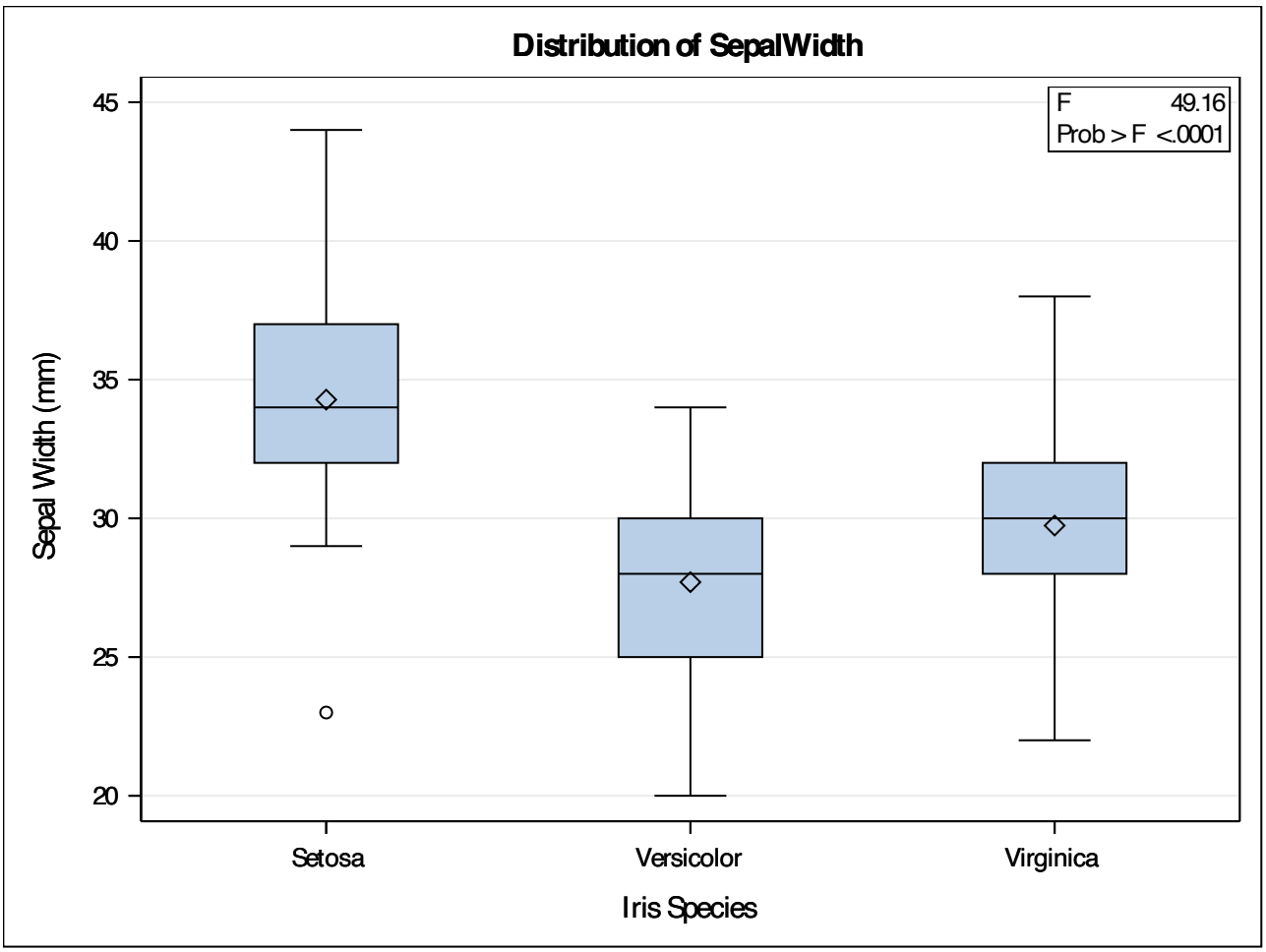
The ANOVA Procedure

Dependent Variable: SepalWidth Sepal Width (mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	1134.493333	567.246667	49.16	<.0001
Error	147	1696.200000	11.538776		
Corrected Total	149	2830.693333			

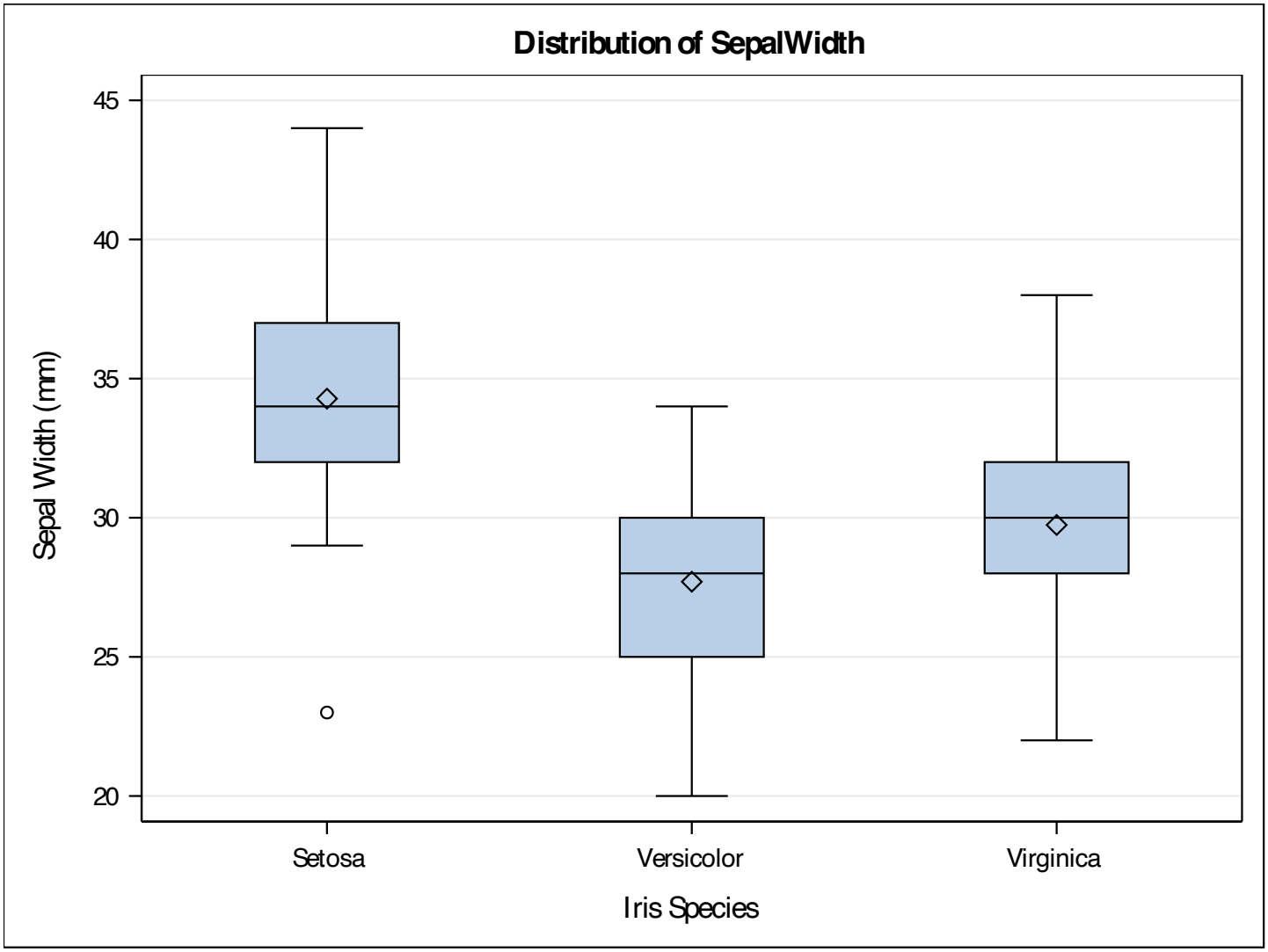
R-Square	Coeff Var	Root MSE	SepalWidth Mean
0.400783	11.11059	3.396877	30.57333

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Species	2	1134.493333	567.246667	49.16	<.0001



The ANOVA Procedure

Levene's Test for Homogeneity of SepalWidth Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Species	2	584.3	292.2	0.90	0.4073
Error	147	47521.1	323.3		



The ANOVA Procedure

Tukey's Studentized Range (HSD) Test for SepalWidth

Note: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	147
Error Mean Square	11.53878
Critical Value of Studentized Range	3.34842
Minimum Significant Difference	1.6086

Comparisons significant at the 0.05 level are indicated by *.**				
Species Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
Setosa - Virginica	4.5400	2.9314	6.1486	***
Setosa - Versicolor	6.5800	4.9714	8.1886	***
Virginica - Setosa	-4.5400	-6.1486	-2.9314	***
Virginica - Versicolor	2.0400	0.4314	3.6486	***
Versicolor - Setosa	-6.5800	-8.1886	-4.9714	***
Versicolor - Virginica	-2.0400	-3.6486	-0.4314	***