

Exercise 1

- (a) From the contingency table below, the proportions of the yes/no response are similar for the young age group and the old age group. So possibly there is no association between the age group and the decision to hire someone to do improvements they would have previously done themselves.

Table of agegrp by response			
agegrp	response		
Frequency Expected	no	yes	Total
older	440 444.86	634 629.14	1074
youngest	219 214.14	298 302.86	517
Total	659	932	1591

- (b) The cell counts are large enough, so the Chi-Square statistics will be fine here. The p-values of Pearson's Chi-Square and the Likelihood Ratio Chi-Square are 0.5977 and 0.5979 respectively (both larger than .05), indicating that we do not have enough evidence to reject the null hypothesis. Mantel-Haenszel is also appropriate because both variables can be seen as ordinal. It is also insignificant, indicating no evidence of a linear trend in association. We can conclude that there is no association between the age group and the answer to this question.

Statistic	DF	Value	Prob
Chi-Square	1	0.2785	0.5977
Likelihood Ratio Chi-Square	1	0.2782	0.5979
Continuity Adj. Chi-Square	1	0.2241	0.6359
Mantel-Haenszel Chi-Square	1	0.2783	0.5978
Phi Coefficient		-0.0132	
Contingency Coefficient		0.0132	
Cramer's V		-0.0132	

- (c) To test whether individuals under 30 have a higher probability of doing home improvements themselves that they would have previous hired someone to do, we can look at the Column 2 risk estimates (we can infer the same information from the column 1 estimates because those are the estimates for No response and will just be the negative of the Yes risks). From the column 2 risks, we see that the estimated risk for older age group is .59 with a 95% confidence interval of (.5609, .6197) and for the youngest age group, the estimated risk is .58 with a 95% confidence interval of (.5338, .6190). The difference is estimated to be .0139 with a confidence interval of (-.0378, .0657). The 95% confidence interval for the difference contains zero, so there is no significant difference in the probability of doing home improvement themselves for the older age group and younger age group.

Column 2 Risk Estimates						
	Risk	ASE	(Asymptotic) 95% Confidence Limits		(Exact) 95% Confidence Limits	
Row 1	0.5903	0.0150	0.5609	0.6197	0.5602	0.6199
Row 2	0.5764	0.0217	0.5338	0.6190	0.5325	0.6194
Total	0.5858	0.0123	0.5616	0.6100	0.5611	0.6101
Difference	0.0139	0.0264	-0.0378	0.0657		
Difference is (Row 1 - Row 2)						

Exercise 2

- (a) The following table gives us the yes and no counts by the type of work. Yes rates are highest for the skilled group and lowest for the unskilled group. The ratios of yes to no are roughly 5:3, 2:1 and a little less than 1:1 for the office, skilled and unskilled group, respectively. So we should suspect an association between the type of work and the response.

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

- (b) Again, Chi-Square tests will be sufficient given the cell counts, and we should not use Fisher's test because it is not needed and would be very computationally intensive. The Pearson's Chi-Square test and the Likelihood Ratio Chi-Square test are both significant at a .05 level. So we conclude there is in fact a significant association between the type of work and the response. Mantel-Haenszel's test would not be appropriate here because work type is not an ordinal variable.

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	

Exercise 3

- (a) The ANOVA table follows. We see that the overall model is significantly better than an error only model if we assume equal variances, as the F statistic of 1180.16 has a p-value less than .0001. Aside from normality, we should test the equal variance assumption using Levene's test. We see that Levene's test has a p-value <.0001, so there are significant differences between the sample variances and a Welch correction would be appropriate here. We should use Welch's ANOVA instead of the regular ANOVA table to test significance of the model. The result for Welch's ANOVA is given below.

Dependent Variable: PetalLength Petal Length
(mm)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	43710.28000	21855.14000	1180.16	<.0001
Error	147	2722.26000	18.51878		
Corrected Total	149	46432.54000			

R-Square	Coeff Var	Root MSE	PetalLength Mean
0.941372	11.45116	4.303345	37.58000

Levene's Test for Homogeneity of PetalLength Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Species	2	18996.5	9498.3	11.26	<.0001
Error	147	123986	843.4		

Welch's ANOVA for PetalLength			
Source	DF	F Value	Pr > F
Species	2.0000	1828.09	<.0001
Error	78.0730		

- (b) With the Welch's weighted correction, the ANOVA model F statistic has a p-value less than .0001. The model is highly significant. The R^2 value of .9414 tells us that about 94% of the variation in petal length can be described by species alone. Since a large amount of the variation can be explained by this model, it would be a very useful model for petal length.

(c) For all pairwise comparisons, we should use Tukey's test. The test results and confidence intervals will not take into account the unequal variance noted by Levene's test. If quoting the interval values, this should be noted, but that is not a big concern in this case since the differences are so significant (e.g. each confidence interval is far from 0).

Tukey's Studentized Range (HSD) Test for PetalLength

Comparisons significant at the 0.05 level are indicated by ***.				
Species Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
Virginica - Versicolor	12.9200	10.8822	14.9578	***
Virginica - Setosa	40.9000	38.8622	42.9378	***
Versicolor - Virginica	-12.9200	-14.9578	-10.8822	***
Versicolor - Setosa	27.9800	25.9422	30.0178	***
Setosa - Virginica	-40.9000	-42.9378	-38.8622	***
Setosa - Versicolor	-27.9800	-30.0178	-25.9422	***

We can conclude that Virginica petal lengths are significantly greater than Versicolor and Setosa petal lengths in general, and Versicolor petal lengths are significantly larger than Setosa petal lengths in general. Our point estimates for those differences of mean lengths would be 12.92mm, 40.9mm, and 27.98mm, respectively.