Exercise 1

Table of eyecolor by haircolor				
eyecolor	**haircolor**			
Frequency Expected Row Pct	**fair**	**medium**	**dark**	**Total**
light	688 382.8 47.12	584 642.86 40.00	188 434.34 12.88	1460
medium	343 436.29 20.61	909 732.69 54.63	412 495.03 24.76	1664
dark	98 309.91 8.29	403 520.45 34.09	681 351.64 57.61	1182
Total	1129	1896	1281	4306

Statistic	DF	Value	Prob
Chi-Square	4	944.6434	<.0001
Likelihood Ratio Chi-Square	4	923.8350	<.0001
Mantel-Haenszel Chi-Square	1	814.7860	<.0001
Phi Coefficient		0.4684	
Contingency Coefficient		0.4242	
Cramer's V		0.3312	

1a) Comparing observed and expected counts or looking for difference in row percentages within columns can provide evidence association. There is a noticeably larger than expected number of light eye color with fair hair, medium eye color with medium hair color, and dark eyes with dark hair. There are slightly lower than expected counts for medium values paired with non-medium values. Finally, there are far fewer than expected observations with dark hair and light eyes and fair hair with dark eyes. This suggests an association. In particular, observations with similar levels of hair darkness tend to occur more often and observations with greater difference in darkness of hair and eye color tend to happen less often. This suggests a positive association since both variables are ordinal.

1b) The sample size is large enough to look at asymptotic chi-square tests and the variables are ordinal, so we should look at Pearson's chi-square, the likelihood ratio chi-square, and the Mantel-Haenszel Test. Pearson and Likelihood Ratio are both highly significant, indicating there is a statistically significant association. Mantel-Haenszel is also statistically significant, indicating a significant linear trend. Based on these results and the trend noticed in part, we conclude there is a statistically significant positive association between eye color and hair color. Those with darker eyes are more likely to have dark hair; those with lighter eyes are more likely to have fairer hair; and those with medium eyes are more likely to have medium hair. Those with one characteristic darker and one lighter are much less likely.

Exercise 2

Table of eyecolor by haircolor			
eyecolor	**haircolor**		
Frequency Expected Row Pct	**fair**	**dark**	**Total**
light	688 416.03 78.54	188 459.97 21.46	876
dark	98 369.97 12.58	681 409.03 87.42	779
Total	786	869	1655

Statistic	DF	Value	Prob
Chi-Square	1	719.3494	<.0001
Likelihood Ratio Chi-Square	1	789.6714	<.0001
Continuity Adj. Chi-Square	1	716.7068	<.0001
Mantel-Haenszel Chi-Square	1	718.9147	<.0001
Phi Coefficient		0.6593	
Contingency Coefficient		0.5504	
Cramer's V		0.6593	

Column 1 Risk Estimates						
	Risk	**ASE**	**(Asymptotic) 95% Confidence Limits**		**(Exact) 95% Confidence Limits**	
Row 1	0.7854	0.0139	0.7582	0.8126	0.7567	0.8121
Row 2	0.1258	0.0119	0.1025	0.1491	0.1033	0.1512
Total	0.4749	0.0123	0.4509	0.4990	0.4506	0.4993
Difference	0.6596	0.0183	0.6238	0.6954		
Difference is (Row 1 - Row 2)						

2a) Just as before, light-fair and dark-dark occur mush more often than expected, suggesting a positive association.

2b) The sample is still large enough for asymptotic tests and the variables are ordinal, so Pearson's chi-square, Likelihood Ratio chi-square, and Mantel-Haenszel are all appropriate. Each is highly significant, indicating there is association and a linear trend in association. The phi coefficient (and Cramer's V) indicate that the association is positive with a moderate magnitude. Just as in exercise 1, similar darknesses of hair and eye color happen more frequently than expected due to chance, and those with different darkness values happen less likely.

2c) The question of interest is about fair hair, so column 1 can be used to directly answer the question. The risk for indiviuals with light eyes to have fair hair is estimated to be .7854, and for those with dark eyes to have fair hair is estimated to be .1258. This gives an estimated difference of risks of .6596. The confidence interval for this difference is completely positive and pretty far from 0, so we conclude that those with light eyes are much more likely to have fair hair than those with dark eyes.

Exercise 3

Table of BP_Status by Chol_Status				
BP_Status(Blood Pressure Status)	**Chol_Status(Cholesterol Status)**			
Frequency Expected Row Pct	**Desirable**	**Borderline**	**High**	**Total**
Optimal	20 14.985 29.85	25 25.512 37.31	22 26.503 32.84	67
Normal	65 54.797 26.53	92 93.29 37.55	88 96.913 35.92	245
High	36 51.218 15.72	89 87.198 38.86	104 90.584 45.41	229
Total	121	206	214	541

Statistic	DF	Value	Prob
Chi-Square	4	11.7368	0.0194
Likelihood Ratio Chi-Square	4	11.9792	0.0175
Mantel-Haenszel Chi-Square	1	9.9553	0.0016
Phi Coefficient		0.1473	
Contingency Coefficient		0.1457	
Cramer's V		0.1042	

3a) We can look at either the expected counts or the row percentsges to see that Optimal and Normal blood pressures occur more often with Desirable cholesterol level, Borderline cholesterol level are consistent with expected values in all three groups, High blood pressure occurs less often with Desirable cholesterol, High cholesterol and High blood pressure happen more often together, and High cholesterol happens less often with Optimal or Normal blood pressure. The differences are noticeable, but not particularly large, with the exception of High blood pressure for the Desirable and High cholesterol levels. This suggests there is a positive association, but it may not be very strong in general.

3b) The variables are ordinal and data is large enough for asymptotic tests. The p-values for Pearson's chi-square and the Likelihood Ratio chi-square are both a little less than .02 so these are significant at a .05 level and indicate there is an association. The Mantel-Haenszel test is even more significant with a p-value of .0016, indicating a pretty noticeable linear trend. Higher cholesterol levels tend to be associated with higher blood pressures.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	25210.845	12605.422	6.67	0.0014
Error	538	1016631.488	1889.650		
Corrected Total	540	1041842.333			

R-Square	Coeff Var	Root MSE	Cholesterol Mean
0.024198	18.65388	43.47010	233.0351

Source	DF	Anova SS	Mean Square	F Value	Pr > F
BP_Status	2	25210.84472	12605.42236	6.67	0.0014

Levene's Test for Homogeneity of Cholesterol Variance ANOVA of Squared Deviations from Group Means					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
BP_Status	2	9964224	4982112	0.56	0.5719
Error	538	4.7912E9	8905491		

4a) Since this is a one-way analysis, Levene's test for homogeneity of variance should be performed. The p-value of 0.5719 is highly insignificant so an equal variance assumption is fine here. The analysis of variance table has a p-value of 0.0014, so the model is statistically significant. More of the variation in cholesterol can be described by blood pressure level than expected due to chance. The r-squared value of .0242, however, is very small. Only 2.42% of the variation in cholesterol can be described by the blood pressure levels.

As an added note, r-square will tend to go smaller for when a categorical predictor has been based on cutoffs for a continuous variable. If there is a linear relationship between the two continuous variables, we would expect the response just to the left of the cutoff to be pretty close to the expected response just to the right of the cutoff, but the observations would be in different groups. The reduced granularity in the predictors from binning tends to flatten out the predictions and reduce explanation of variation.

Comparisons significant at the 0.05 level are indicated by ***.				
BP_Status Comparison	**Difference Between Means**	**Simultaneous 95% Confidence Limits**		
High - Normal	11.543	2.153	20.934	***
High - Optimal	18.647	4.456	32.837	***
Normal - High	-11.543	-20.934	-2.153	***
Normal - Optimal	7.103	-6.982	21.188	
Optimal - High	-18.647	-32.837	-4.456	***
Optimal - Normal	-7.103	-21.188	6.982	

4b) The best test to use for testing all pairwise comparisons is Tukey's test. Expected cholesterol levels are significantly different for high and normal blood pressures and for high and optimal blood pressures. We expect individuals with high blood pressures to have cholesterol levels 11.54 higher than those with normal blood pressures on average and 18.65 higher than those with optimal blood pressure on average. The difference between normal and optimal blood pressure groups is not significant.