Stat 448, homework 2
Shuhui Guo

Exercise 1

a)

Table of eyecolor by haircolor				
eyecolor	**haircolor**			
Frequency Expected	**fair**	**medium**	**dark**	**Total**
light	688 382.8	584 642.86	188 434.34	1460
medium	343 436.29	909 732.69	412 495.03	1664
dark	98 309.91	403 520.45	681 351.64	1182
Total	1129	1896	1281	4306

According to the above contingency table, for the light eye color group, fair hair color appears more frequent than other kinds of hair colors. For the medium eye color group, medium hair color appears more frequent than other kinds of hair colors. For the dark eye color group, dark hair color appears more frequent than other kinds of hair colors. Therefore, there might be some linear associations between eye color and hair color.

b)

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	

Since the expected counts for each cell are all greater than 5 and the variables are ordered, Mantel-Haenszel Chi-Square test is appropriate for the table. The p-value for Mantel-Haenszel Chi-Square test is less than 0.05. Therefore under the significant level of 5%, the null hypothesis of no association is rejected. We can conclude that there is significant linear dependency between eye color and hair color.

Exercise 2

a)

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

According to the above contingency table, for the light eye color group, fair hair color appears more frequent than dark hair color. For the dark eye color group, dark hair color appears more frequent than fair hair color. Therefore, there might be some linear associations between these eye color and hair color groups.

b)

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	

Fisher's Exact Test	
Cell (1,1) Frequency (F)	688
Left-sided Pr <= F	1.0000
Right-sided Pr >= F	<.0001
Table Probability (P)	<.0001
Two-sided Pr <= P	<.0001

Since the expected counts for each cell are all greater than 5 and the variables are ordered, Mantel-Haenszel Chi-Square test is appropriate for the table. The p-value for Mantel-Haenszel Chi-Square test is less than 0.05. Therefore under the significant level of 5%, the null hypothesis of no

association is rejected. We can conclude that there is significant linear dependency between these eye color and hair color groups.

This 2×2 table and the full 3×3 table in Exercise 1 all conclude that there is significant linear dependency between eye and hair color groups. The associations concluded from 2×2 table are between the light and dark eye color and fair and dark hair color, while the associations concluded from 3×3 table include medium eye color and medium hair color. Also, the results of Fisher's Exact Test are included in 2×2 table while not in 3×3 table. Smaller sample size in 2×2 table might be the reason.

c)

Column 1 Risk Estimates						
	Risk	ASE	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)						

The last row of the above table shows that the confidence interval of risk differences does not contain 0, which indicates that there is significant difference between light and dark eye color having fair hair. Since the difference is positive, it is reasonable to say that those with light eye color are significantly more likely to have fair hair color than those with dark eye color.

Exercise 3

a)

Table of BP_Status by Chol_Status				
BP_Status(Blood Pressure Status)	Chol_Status(Cholesterol Status)			
Frequency Expected	Desirable	Borderline	High	Total
Optimal	20 14.985	25 25.512	22 26.503	67
Normal	65 54.797	92 93.29	88 96.913	245
High	36 51.218	89 87.198	104 90.584	229
Total	121	206	214	541

According to the above contingency table, for both the optimal and normal blood pressure status, borderline cholesterol status appears more frequent than other cholesterol statuses. For the high blood pressure status, high cholesterol status appears more frequent than other cholesterol statuses. Therefore, there might be some relationships between blood pressure and cholesterol statuses.

b)

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	

Since the expected counts for each cell are all greater than 5 and the variables are ordered, Mantel-Haenszel Chi-Square test is appropriate for the table. The p-value for Mantel-Haenszel Chi-Square test is less than 0.05. Therefore under the significant level of 5%, the null hypothesis of independence is rejected. We can conclude that there is significant linear dependency between blood pressure and cholesterol statuses.

Base on the above results, it is reasonable to say that higher blood pressure status associates with higher cholesterol level.

Exercise 4

a)

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			

The above table is the results of F test, which tests whether all parameters are zero in the one-way anova model. The test statistic follows an F distribution with degree of freedom 2 and 538 under the null hypothesis. The p-value is less than 0.05. Therefore under the significant level of 5%, the null hypothesis that all parameters are zero is rejected. We can conclude that not all parameters are zero and the model is significant.

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

The R-square is 0.024198, which indicates that about 2.4198% of variation of Cholesterol can be described by the model.

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		

The above table is the results of Levene's test, which tests the equal-variance assumption. The test statistic follows an F distribution with degree of freedom 2 and 538 under the null hypothesis. The p-value is greater than 0.05. Therefore under the significant level of 5%, the null hypothesis which is homogeneity of Cholesterol variance cannot be rejected. We can conclude that the Cholesterol variable has equal variance.

b)

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	

According to the above table, the group with high blood pressure status has significantly greater cholesterol than the group with normal blood pressure status. The group with high blood pressure status has significantly greater cholesterol than the group with optimal blood pressure status. Based on the results, higher blood pressure status associates with higher cholesterol level.

The above results and the results from part b of exercise 3 give the same conclusion of higher blood pressure status associates with higher cholesterol level.