

Homework 2

Jiangyan Feng (NetID: jf8)

Exercise 1

a)

Table of safety by acceptable					
safety	acceptable				
Frequency Expected	unacc	acc	good	vgood	Total
low	576 403.33	0 128	0 23	0 21.667	576
med	357 403.33	180 128	39 23	0 21.667	576
high	277 403.33	204 128	30 23	65 21.667	576
Total	1210	384	69	65	1728

The frequency table for safety and acceptable is shown as above. Comparing the real data with the expected value, there is clearly large deviation from the expected value and therefore there are apparent associations. Since the real value is much larger than the expected value along the diagonal, it appears to be positive association.

b)

Statistics for Table of safety by acceptable

Statistic	DF	Value	Prob
Chi-Square	6	479.3224	<.0001
Likelihood Ratio Chi-Square	6	628.0670	<.0001
Mantel-Haenszel Chi-Square	1	333.3407	<.0001
Phi Coefficient		0.5267	
Contingency Coefficient		0.4660	
Cramer's V		0.3724	

Sample Size = 1728

According to the frequency table in part a, the expected value in each cell is much larger than 5, therefore the chi-square test is used (results are shown as above).

Since both variables (safety and acceptable) are ordinal, the first three Chi-Square in the above table can all be used to test the association. Since the Value is large (~479, ~628, ~333) and the Prob (<0.001) is much smaller than 0.05 in all three Chi-Square, we conclude that there is strong association between safety and acceptability.

Next, since the Phi Coefficient is positive and large (0.5267), the magnitude of contingency coefficient is large (0.4660), and the Cramer's V is positive and relatively large (0.3724), we conclude that there is strong positive association between safety and acceptability.

c)

Table of safety by acceptyesno			
safety	acceptyesno		
Frequency Expected	no	yes	Total
low	576 403.33	0 172.67	576
med	357 403.33	219 172.67	576
high	277 403.33	299 172.67	576
Total	1210	518	1728

Statistics for Table of safety by acceptyesno

Statistic	DF	Value	Prob
Chi-Square	2	396.3444	<.0001
Likelihood Ratio Chi-Square	2	547.6844	<.0001
Mantel-Haenszel Chi-Square	1	369.4969	<.0001
Phi Coefficient		0.4789	
Contingency Coefficient		0.4319	
Cramer's V		0.4789	

Sample Size = 1728

The frequency table for safety and acceptyesno and the chi-square test results are shown as above. According to the frequency table, the real value deviate from the expected value a lot, therefore suggesting an association. Since acceptyesno is not ordinal, we should not suggest a positive nor a negative association.

Since the expected value in each cell is much larger than 5, Chi-Square test is used. Since the acceptyesno is not ordinal variable, Mantel-Haenszel Chi-Square is not suitable here. We will focus on the first two rows: Chi-Square and the Likelihood Ratio Chi-Square. Since Value is large and Prob (<0.0001) is much smaller than 0.05 in both rows, we conclude there is statistically significant association between safety and acceptyesno.

According to the magnitude of Phi Coefficient, contingency coefficient, and Crmaer's V, we conclude that there is strong association between safety and acceptyesno.

Comparing findings for the four-valued and binary acceptability ratings:

1. Both cases suggest strong association.
2. For the binary case, the Value in both Chi-Square and Likelihood Ratio Chi-Square, and the value of Phi Coefficient decrease. This suggests that the association from the binary rating is weaker compared with the four-valued one.
3. Unlike the four-valued rating, the binary case is not ordinal and therefore is not indicative of positive or negative association.

Exercise 2

a)

Table of safety by acceptable					
safety	acceptable				
Frequency Expected	unacc	acc	good	vgood	Total
med	357 317	180 192	39 34.5	0 32.5	576
high	277 317	204 192	30 34.5	65 32.5	576
Total	634	384	69	65	1152

The safety=low data are removed. The frequency table for safety and acceptable is shown as above. Comparing the real data with the expected value, there is clearly large deviation from the expected value and therefore there are apparent associations. Since the real value is much larger than the expected value along the diagonal, it appears to be positive association.

b)

Statistics for Table of safety by acceptable

Statistic	DF	Value	Prob
Chi-Square	3	77.7686	<.0001
Likelihood Ratio Chi-Square	3	102.9090	<.0001
Mantel-Haenszel Chi-Square	1	50.5369	<.0001
Phi Coefficient		0.2598	
Contingency Coefficient		0.2515	
Cramer's V		0.2598	

According to the frequency table in part a, the expected value in each cell is much larger than 5, therefore the chi-square test is used (results are shown as above).

Since both variables (safety and acceptable) are ordinal, the first three Chi-Square in the above table can all be used to test the association. Since the Value is large (~77, ~102, ~50) and the Prob (<0.001) is much smaller than 0.05 in all three Chi-Square, we conclude that there is strong association between safety and acceptability.

Next, since the Phi Coefficient is positive and relatively large (0.2598), the magnitude of contingency coefficient is relatively large (0.2515), and the Cramer's V is positive and relatively large (0.2598), we conclude that there is positive association between safety and acceptability, and the strength of the association is not medium level.

c)

Table of safety by acceptyesno			
safety	acceptyesno		
Frequency Expected	no	yes	Total
med	357 317	219 259	576
high	277 317	299 259	576
Total	634	518	1152

Statistics for Table of safety by acceptyesno

Statistic	DF	Value	Prob
Chi-Square	1	22.4498	<.0001
Likelihood Ratio Chi-Square	1	22.5264	<.0001
Continuity Adj. Chi-Square	1	21.8921	<.0001
Mantel-Haenszel Chi-Square	1	22.4304	<.0001
Phi Coefficient		0.1396	
Contingency Coefficient		0.1383	
Cramer's V		0.1396	

Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.6198	0.0202	0.5801	0.6594	0.5787	0.6596
Row 2	0.4809	0.0208	0.4401	0.5217	0.4394	0.5226
Total	0.5503	0.0147	0.5216	0.5791	0.5211	0.5793
Difference	0.1389	0.0290	0.0820	0.1958		
Difference is (Row 1 - Row 2)						

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.3802	0.0202	0.3406	0.4199	0.3404	0.4213
Row 2	0.5191	0.0208	0.4783	0.5599	0.4774	0.5606
Total	0.4497	0.0147	0.4209	0.4784	0.4207	0.4789
Difference	-0.1389	0.0290	-0.1958	-0.0820		
Difference is (Row 1 - Row 2)						

The frequency table for safety and acceptyesno, the chi-square test results, and the risk test results are shown as above. According to the frequency table, the real value deviate from the expected value, therefore suggesting an association. Since acceptyesno is not ordinal, we should not suggest a positive nor a negative association.

Since the expected value in each cell is much larger than 5, Chi-Square test is used. Since the acceptyesno is not ordinal variable, Mantel-Haenszel Chi-Square is not suitable here. We will focus on the first two rows: Chi-Square and the Likelihood Ratio Chi-Square. Since Value is large and Prob (<0.0001) is much smaller than 0.05 in both rows, we conclude there is statistically significant association between safety and acceptyesno.

According to the magnitude of Phi Coefficient, contingency coefficient, and Crmaer's V, we conclude that there is association between safety and acceptyesno and the strength of this association is much weaker compared Exercise 1 in which all data are used.

Comparing findings for the four-valued and binary acceptability ratings:

1. Both causes suggest weak association.
2. For the binary case, the Value in both Chi-Square and Likelihood Ratio Chi-Square, and the value of Phi Coefficient decrease. This suggests that the association from the binary rating is weaker compared with the four-valued one.
3. Unlike the four-valued rating, the binary case is not ordinal and therefore is not indicative of positive or negative association.

To understand whether or not high safety vehicles are more likely to be rated at least acceptable than medium safety, we performed risk tests. We focus on Row 2 (high safety) and Column 2 (acceptyesno=yes) Risk. According to the table, Risk is large (0.5191) for high safety and the Risk Difference is large negative value (-0.1389) between medium and high safety.

Therefore, we suggest that high safety vehicles are more likely to be rated at least accepted than medium safety vehicles are.

Exercise 3

a)

Dependent Variable: sgpt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	5520.9511	1380.2378	3.74	0.0054
Error	340	125450.2373	368.9713		
Corrected Total	344	130971.1884			

R-Square	Coeff Var	Root MSE	sgpt Mean
0.042154	63.17422	19.20863	30.40580

Source	DF	Anova SS	Mean Square	F Value	Pr > F
drinkgroup	4	5520.951147	1380.237787	3.74	0.0054

Levene's Test for Homogeneity of sgpt Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
drinkgroup	4	4022355	1005589	0.43	0.7899
Error	340	8.0263E8	2360662		

We first performed Levene's test to check the equal variance assumption. Since the P value (0.7899) is much larger than 0.05, there is not significant evidence to reject the null hypothesis and therefore the equal variance assumption can be trusted.

Since the equal variance assumption is reasonable, we performed the normal one-way ANOVA without adjustment. Since the F value in the first table is 3.74 (larger than 1) and P value = 0.0054 (smaller than 0.05), we conclude that there is statistical significance of the model. Since the R-Square = 0.042154 in the second table, we suggest that ~4.2% of variation is described by the model. Since the R-Square is not large and the F Value = 3.74 in table 1 is not large, we suggest there may not be practical significance of this model.

b)

Tukey's Studentized Range (HSD) Test for sgpt

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

Alpha	0.05
Error Degrees of Freedom	340
Error Mean Square	368.9713
Critical Value of Studentized Range	3.87844

Comparisons significant at the 0.05 level are indicated by ***.				
drinkgroup Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
9 or more - 6 to 8	7.050	-6.124	20.225	
9 or more - 3 to 5	13.652	0.858	26.446	***
9 or more - 1 or 2	13.941	0.321	27.562	***
9 or more - less than 1	13.967	1.482	26.452	***
6 to 8 - 9 or more	-7.050	-20.225	6.124	
6 to 8 - 3 to 5	6.602	-1.940	15.143	
6 to 8 - 1 or 2	6.891	-2.845	16.627	
6 to 8 - less than 1	6.917	-1.154	14.987	
3 to 5 - 9 or more	-13.652	-26.446	-0.858	***
3 to 5 - 6 to 8	-6.602	-15.143	1.940	
3 to 5 - 1 or 2	0.289	-8.925	9.504	
3 to 5 - less than 1	0.315	-7.118	7.748	
1 or 2 - 9 or more	-13.941	-27.562	-0.321	***
1 or 2 - 6 to 8	-6.891	-16.627	2.845	
1 or 2 - 3 to 5	-0.289	-9.504	8.925	
1 or 2 - less than 1	0.026	-8.754	8.806	
less than 1 - 9 or more	-13.967	-26.452	-1.482	***
less than 1 - 6 to 8	-6.917	-14.987	1.154	
less than 1 - 3 to 5	-0.315	-7.748	7.118	
less than 1 - 1 or 2	-0.026	-8.806	8.754	

Tukey's method is used to check the differences of means among five drinking groups. According to the table as shown above, the following pairwise groups are different: 9 or more vs. 3 to 5, 9 or more vs. 1 or 2, 9 or more vs. less than 1. According to the positive sign of the difference between means for 9 or more, means of alanine aminotransferase in 9 or more group is significantly larger than these 3 other groups (3 to 5, 1 or 2, less than 1).

Exercise 4

a)

Dependent Variable: gammagt

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	54286.2218	13571.5555	9.70	<.0001
Error	340	475791.9405	1399.3881		
Corrected Total	344	530078.1623			

R-Square	Coeff Var	Root MSE	gammagt Mean
0.102412	97.71272	37.40840	38.28406

Source	DF	Anova SS	Mean Square	F Value	Pr > F
drinkgroup	4	54286.22180	13571.55545	9.70	<.0001

Levene's Test for Homogeneity of gammagt Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
drinkgroup	4	2.059E8	51473760	2.11	0.0791
Error	340	8.2919E9	24387981		

We first performed Levene's test to check the equal variance assumption. Since the P value (0.0791) is larger than 0.05, there is not significant evidence to reject the null hypothesis and therefore the equal variance assumption can be trusted.

Since the equal variance assumption is reasonable, we performed the normal one-way ANOVA without adjustment. Since the F value in the first table is 9.70 (larger than 1) and P value = 0.0001 (smaller than 0.05), we conclude that there is statistical significance of the model. Since the R-Square = 0.102412 in the second table, we suggest that ~10.2% of variation is described by the model. Since the R-Square is large and the F Value = 9.70 in table 1 is also large, we suggest there is practical significance of this model.

b)

Tukey's Studentized Range (HSD) Test for gammagt

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

Alpha	0.05
Error Degrees of Freedom	340
Error Mean Square	1399.388
Critical Value of Studentized Range	3.87844

Comparisons significant at the 0.05 level are indicated by ***.				
drinkgroup Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
9 or more - 6 to 8	19.290	-6.367	44.947	
9 or more - 1 or 2	34.876	8.351	61.402	***
9 or more - 3 to 5	35.971	11.055	60.887	***
9 or more - less than 1	45.396	21.082	69.709	***
6 to 8 - 9 or more	-19.290	-44.947	6.367	
6 to 8 - 1 or 2	15.586	-3.374	34.547	
6 to 8 - 3 to 5	16.681	0.047	33.315	***
6 to 8 - less than 1	26.106	10.388	41.823	***
1 or 2 - 9 or more	-34.876	-61.402	-8.351	***
1 or 2 - 6 to 8	-15.586	-34.547	3.374	
1 or 2 - 3 to 5	1.094	-16.850	19.039	
1 or 2 - less than 1	10.519	-6.579	27.618	
3 to 5 - 9 or more	-35.971	-60.887	-11.055	***
3 to 5 - 6 to 8	-16.681	-33.315	-0.047	***
3 to 5 - 1 or 2	-1.094	-19.039	16.850	
3 to 5 - less than 1	9.425	-5.051	23.901	
less than 1 - 9 or more	-45.396	-69.709	-21.082	***
less than 1 - 6 to 8	-26.106	-41.823	-10.388	***
less than 1 - 1 or 2	-10.519	-27.618	6.579	
less than 1 - 3 to 5	-9.425	-23.901	5.051	

Tukey's method is used to check the differences of means among five drinking groups. According to the table as shown above, the following pairwise groups are different: 9 or more vs. 3 to 5, 9 or more vs. 1 or 2, 9 or more vs. less than 1; 6 to 8 vs. 3 to 5, 6 to 8 vs. less than 1. Considering no significant difference between 9 or more and 6 to 8, we suggest these two can be grouped together and different the three groups.

According to the sign of the difference between means, means of gamma-glutamyl transpeptidase in 9 or more group and 6 and 8 is significantly larger than 3 other groups (3 to 5, 1 or 2, less than 1).

Since the magnitude of the difference between means is much larger in gamma-glutamyl transpeptidase (~40 for 9 or more) than the alanine aminotransferase (~13 for 9 or more), we suggest there is more significant difference in gamma-glutamyl transpeptidase across the drinking groups.