

Homework 7

Logistic Regression Exercises

Problem 1

The following data was taken from a five-year double blind randomized study testing whether regular intake of aspirin reduces mortality form cardiovascular disease. Participants in the study took either one aspirin tablet or a placebo every other day, depending on whether they were randomized into the treatment group or control group. The following counts were obtained.

Number of cases

Group	Myocardial Infarction (MI)		Total
	YES	No	
Placebo	189	10,845	11,034
Aspirin	104	10,933	11,037

This is summary data, so when you enter this data in SPSS be sure to weight the cases by count before you run any analyses. You can find the weight cases dialogue box under the data menu.

Perform the following analysis:

- a) Test whether a significant relationship exists between taking Aspirin and MI using traditional chi-square analysis. Calculate the odds ratio and relative risk. Interpret the results.
- b) Test whether a significant relationship exists between taking Aspirin and MI using logistic regression. Calculate the odds ratio and relative risk. Interpret the results.
- c) Explain the difference between odds ratio and relative risk.

Part A

From the below Chi-Square Test we can see that a significant relationship is shown between Myocardial Infarction (MI) and its drug treatment. The null hypothesis would have suggested that there was no relationship so we would reject it and accept the alternative.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	25.014 ^a	1	.000		
Continuity Correction ^b	24.429	1	.000		
Likelihood Ratio	25.372	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	25.013	1	.000		
N of Valid Cases	22071				

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	25.014 ^a	1	.000		
Continuity Correction ^b	24.429	1	.000		
Likelihood Ratio	25.372	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	25.013	1	.000		
N of Valid Cases	22071				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 146.48.

b. Computed only for a 2x2 table

Calculating an odds ratio would give us a better feeling for the strength of the effect of Asprin (drug treatment) on MI. To calculate an odds ratio we can consider the below cross table. When placebo was used for treatment of MI, the odds of MI occurring was 189 to 10,845 or 1.7% chance. When Asprin was used for treatment of MI, the odds were reduced to MI occurring was 104 to 10,933 or .9% chance. **The Odds ratio then**

would be $\frac{189 \div 10,845}{104 \div 10,933} = 1.9$.

Treatment * MI Crosstabulation

			MI		Total
			Yes	No	
Treatment	Placebo	Count	189	10845	11034
		Expected Count	146.5	10887.5	11034.0
		% within Treatment	1.7%	98.3%	100.0%
	Asprin	Count	104	10933	11037
		Expected Count	146.5	10890.5	11037.0
		% within Treatment	.9%	99.1%	100.0%
Total		Count	293	21778	22071
		Expected Count	293.0	21778.0	22071.0
		% within Treatment	1.3%	98.7%	100.0%

To examine the odds ratio and risk ratio further we can view the risk estimate table below to check our formula above to see that we are close. Our odds ratio highlighted in yellow and our risk ratio highlighted in green.

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Treatment (Placebo / Aspirin)	1.832	1.440	2.331
For cohort MI = Yes	1.818	1.433	2.306
For cohort MI = No	.992	.989	.995
N of Valid Cases	22071		

Part B

Part C

To understand the difference you have to understand the definition of odds ratio and relative risk.

The risk ratio gives you the percentage difference in classification between group one (Placebo) and group two (Aspirin) of our two possible drug treatments.

The odds ratio gives the ratio of the odds of suffering some fate (MI in our case).

Problem 2

In 1974 vehicles, seat belts use was considered in association with the ownership of the vehicle. Use the following data to test whether there is a relationship between car ownership type and seat belt use.

Number of Cases

Belt Use	Individuals	Rental	Lease	Other corporate
Lap and Shoulder	583	145	86	182
Lap Only	139	24	24	31
None	524	59	74	145

After doing a Chi-Square test, we can evaluate the Chi-Square results to determine if there is a significant relationship between car ownership type and seat belt use. From the highlighted Pearson Chi-Square value, we can see that we have a significance factor of .000 or 0% which is less than 5%. We would therefore reject the null hypothesis which would have assumed that there was no relationship and accept the alternative.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	26.722 ^a	6	.000
Likelihood Ratio	27.584	6	.000
Linear-by-Linear Association	1.469	1	.226
N of Valid Cases	2016		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 19.90.

Because this is not a 2x2 table, we cannot calculate an odds ratio or risk estimate.

Problem 3

The following data come from a study examining prevalence of lung cancer among smokers and nonsmokers for 8 studies conducted in 8 different cities in china.

Number of People with Lung Cancer for Smokers and Nonsmokers

Study City	Smokers	Lung Cancer	No Lung Cancer
Beijing	Smokers	126	100
	Nonsmokers	35	61
Shanghai	Smokers	908	688
	Nonsmokers	497	807
Shenyang	Smokers	913	747
	Nonsmokers	336	598
Nanjing	Smokers	235	172
	Nonsmokers	58	121
Harbin	Smokers	402	308
	Nonsmokers	121	215
Zhengzhou	Smokers	182	156
	Nonsmokers	72	98
Taiyuan	Smokers	60	99
	Nonsmokers	11	43
Nanchang	Smokers	104	89
	Nonsmokers	21	36

What does the data suggest about the relationship between smoking and lung cancer? Is the risk Consistent across the 8 cities? Be sure to weight the cases by count.

Part A

The data once analyzed shows that there is a significant relationship between smoking and lung cancer. Our significance factor from a Chi-Square Test is .000 or 0%. It is lower than 5% so we would reject the null hypothesis which would have assumed no relationship and accept the alternative.

Cancer Count * Smoking Status

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	273.091 ^a	1	.000		
Continuity Correction ^b	272.346	1	.000		
Likelihood Ratio	275.584	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	273.058	1	.000		
N of Valid Cases	8419				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 1517.23.

b. Computed only for a 2x2 table

We can further evaluate the odds ratio of getting lung cancer from smoking to gain a better confidence on the relationship. The below Crosstab shows that those who smoke have a 71.8% chance of having lung cancer as opposed to the 28.2% that don't – that is they are 2.1 times more likely to have lung cancer than those that don't (highlighted in yellow in our risk estimation results).

The risk ration would be 1.320 (highlighted in green in the Risk Estimation results).

Crosstab

			Smoking Status		Total
			Smokers	Nonsmokers	
Cancer Count	Cancer	Count	2930	1151	4081
		Expected Count	2563.8	1517.2	4081.0
		% within Cancer Count	71.8%	28.2%	100.0%
No Cancer	No Cancer	Count	2359	1979	4338
		Expected Count	2725.2	1612.8	4338.0
		% within Cancer Count	54.4%	45.6%	100.0%
Total		Count	5289	3130	8419
		Expected Count	5289.0	3130.0	8419.0
		% within Cancer Count	62.8%	37.2%	100.0%

Risk Estimate

	Value	95% Confidence Interval	
		Lower	Upper
Odds Ratio for Cancer Count (Cancer / No Cancer)	2.136	1.950	2.338
For cohort Smoking Status = Smokers	1.320	1.277	1.365
For cohort Smoking Status = Nonsmokers	.618	.583	.656
N of Valid Cases	8419		

Part B

We can evaluate the crosstab results comparing cancer count to city to see if the risk is consistent across the 8 cities evaluated. We can see in the highlighted examples below (Shanghai and Shenyang) that the risk of lung cancer is higher than the other six cities. In the other six, it is roughly the same.

			Crosstab								Total
			City								
			Beijing	Shanghai	Shenyang	Nanjing	Harbin	Zhengzhou	Taiyuan	Nanchang	
Cancer Count	Cancer	Count	161	1405	1249	293	523	254	71	125	4081
		Expected Count	156.1	1405.7	1257.4	284.1	507.0	246.2	103.2	121.2	4081.0
		% within Cancer Count	3.9%	34.4%	30.6%	7.2%	12.8%	6.2%	1.7%	3.1%	100.0%
No Cancer	No Cancer	Count	161	1495	1345	293	523	254	142	125	4338
		Expected Count	165.9	1494.3	1336.6	301.9	539.0	261.8	109.8	128.8	4338.0
		% within Cancer Count	3.7%	34.5%	31.0%	6.8%	12.1%	5.9%	3.3%	2.9%	100.0%
Total	Total	Count	322	2900	2594	586	1046	508	213	250	8419
		Expected Count	322.0	2900.0	2594.0	586.0	1046.0	508.0	213.0	250.0	8419.0
		% within Cancer Count	3.8%	34.4%	30.8%	7.0%	12.4%	6.0%	2.5%	3.0%	100.0%