MSDS 6372: UNIT 11 HW

non-smokers		
	exposure	
mortality	yes	no
deaths	17	6
referents	98	103

smokers		
	exposure	
mortality	yes	no
deaths	131	69
referents	274	240

The table above references a case-referent (case-control) study of lung cancer deaths (Deaths) and non-deaths (Referents) and their relationship to asbestos exposure (yes,no) and the worker's smoking status. The data for analysis with SAS is in hw11Data.xlsx. The study was conducted by Liddell in 2001 from previous study Liddell had performed in 1984. A link to the 2001 paper is:

https://pdfs.semanticscholar.org/db24/28cdad85c64e3ec9643aa658805875af679e.pdf

Let's assume the data were collected by sampling 223 workers who died from lung cancer and 715 that did not die from lung cancer. These workers were then classified as either exposed to asbestos dust or not as well as classified as a smoker or non-smoker. Note that the response is listed in the rows in the above table and the explanatory variable is listed in the columns. When you run proc freq in SAS make sure the explanatory variable is listed in the row and the response in the columns. Also, sort the input data with the code below so the analysis by all students is consistent:

```
proc sort data=datIn;
   by descending exposure descending response;
run;
```

A. Combine the smoking and non-smoking data to create one table of data with the responses (Death / Referents (non-deaths) on the columns and the explanatory variable (Exposure) on the rows. Be sure and clearly label your rows and columns. Simply cut and paste your table below.

Obs	exposure	response	n	smoke Status
1	noExposure	referent	240	Smoker
2	noExposure	death	69	Smoker
3	exposure	referent	274	Smoker
4	exposure	death	131	Smoker
5	noExposure	referent	103	NonSmoker
6	noExposure	death	6	NonSmoker
7	exposure	referent	98	NonSmoker
8	exposure	death	17	NonSmoker

B. Calculate a chi-square statistic to evaluate if exposure affects deaths. Do your calculations manually and confirm them with SAS. What do you conclude from the results?

Tables Shown in Homework MS WORD					
All-smokers			Expected Frequencies		
	Mortality			Mortality	
Exposure	referents	deaths	Exposure	referents	deaths

no exposure	343	75	**no exposure**	318.6247	99.37527
Exposure	372	148	**exposure**	396.3753	123.6247
	938	715	223		

Tables To Use for Calculations

All-smokers				**Chi-square**	
	Mortality				
Exposure	**referents**	**deaths**		**referents**	**deaths**
no exposure	48	34	**no exposure**	1.864744	5.978888
exposure	52	66	**exposure**	1.498967	4.806106
	100%	100%	**Total**		**14.14871**

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of exposure by response		
	response(response)		
exposure(exposure)	referent	death	Total
noExposure	343 36.57 82.06 47.97	75 8.00 17.94 33.63	418 44.56
exposure	372 39.66 71.54 52.03	148 15.78 28.46 66.37	520 55.44
Total	715 76.23	223 23.77	938 100.00

Statistics for Table of exposure by response

Statistic	DF	Value	Prob
Chi-Square	1	14.1487	0.0002
Likelihood Ratio Chi-Square	1	14.4060	0.0001
Continuity Adj. Chi-Square	1	13.5742	0.0002
Mantel-Haenszel Chi-Square	1	14.1336	0.0002
Phi Coefficient		0.1228	
Contingency Coefficient		0.1219	

Reject H_0 if $X^2 > 3.84$ ($\alpha = 0.05$, $df = 1$)
Exposure and deaths are not statistically independent since chi-square of 14.1487 manually and from SAS

C. Use the table from part A to test if the proportion (odds, if proportions are not appropriate) of dying from lung cancer is greater for those that are exposed to asbestos than those that are not. Provide the null and alternative hypothesis of your test as well as any p-values and conclusion. Please include a 95% confidence interval in your conclusion. Do you calculations manually and check them with SAS.

Odds Ratio	1.819498208
LN	0.598560753

SE	0.160292589
CI Lower	1.328947505
CI Upper	2.49112453

Odds Ratio and Relative Risks			
Statistic	Value	95% Confidence Limits	
Odds Ratio	1.8195	1.3290	2.4911
Relative Risk (Column 1)	1.1470	1.0691	1.2306
Relative Risk (Column 2)	0.6304	0.4929	0.8064

Sample Size = 938

Risk Difference Test	
H0: P1 - P2 = 0	Wald Method
Risk Difference	0.1052
ASE (H0)	0.0280
Z	3.7615
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	0.0002
Column 1 (response = referent)	

There is strong evidence that lung cancer for those exposed to asbestos is greater than non-asbestos exposure with p-value <.0001 and 95% confidence interval of (1.3290, 2.4911)

D. Using SAS conduct a Mantel-Haenszel test to test if the proportion of dying of lung cancer is greater for those exposed to asbestos than those that are not after accounting for smoking. Be sure and first check the assumption of the test with the Breslow-Day test. Provide an analysis of your results.

Contents

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)				
Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	12.7527	0.0004
2	Row Mean Scores Differ	1	12.7527	0.0004
3	General Association	1	12.7527	0.0004

Common Odds Ratio and Relative Risks				
Statistic	Method	Value	95% Confidence Limits	
Odds Ratio	Mantel-Haenszel	1.7816	1.2949	2.4511
	Logit	1.7717	1.2863	2.4403
Relative Risk (Column 1)	Mantel-Haenszel	1.1368	1.0608	1.2183
	Logit	1.1280	1.0589	1.2015
Relative Risk (Column 2)	Mantel-Haenszel	0.6499	0.5103	0.8276
	Logit	0.6598	0.5181	0.8403

Breslow-Day Test for Homogeneity of the Odds Ratios	
Chi-Square	1.2527
DF	1
Pr > ChiSq	0.2630

Total Sample Size = 938

The large p-value for the Breslow-Day test(0.2630) indicates no difference between exposed and non-exposed after accounting for smoking.

The Mantel-Haenszel test significant P-value of 0.0004 indicates that the association between asbestos and lung cancer remains strong after adjusting for smoking.