

Stat 539 - Biostatistics II – Spring 2006

Homework 8 due April 6 --- **Solutions**

(assignment for these solutions can be found on the last page)

The code for this analysis comes directly from the labs website
http://www.stat.unm.edu/~erike/courses/stat539/stat539_lab7.do
essentially just doing a find/replace of the three variables used.

a. *Use chi-squared (or binomial or Fisher's Exact) tests to explore the relationships of the various predictors to byssinosis.*

I run the following five commands to explore the relationship of byssinosis with the five other predictors. In summary, all of the predictors are all marginally related to byssinosis, however **dust** is extremely related (based on the very large Chi-squared statistic), followed by **sex**, **smoker**, **emplength**, **race**, respectively.

```
tabulate bys dust [fweight = count], cchi2 chi2 column exact expected lrchi2 row
tabulate bys race [fweight = count], cchi2 chi2 column exact expected lrchi2 row
tabulate bys sex [fweight = count], cchi2 chi2 column exact expected lrchi2 row
tabulate bys smoker [fweight = count], cchi2 chi2 column exact expected lrchi2 row
tabulate bys emplength [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

Dust is significantly related with byssinosis. The No's increase as dustiness decreases, and the Yes's decrease as dustiness decreases.

```
. tabulate bys dust [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

```
+-----+
| Key |
+-----+
| frequency |
| expected frequency |
| chi2 contribution |
| row percentage |
| column percentage |
+-----+
```

```
Enumerating sample-space combinations:
stage 3: enumerations = 1
stage 2: enumerations = 106
stage 1: enumerations = 0
```

Suffering from byssinosis		Dustiness of workplace, 1 = high, 2 = medium, 3 = low			Total
0 = no, 1 = yes	high	medium	low		
no	564	1,282	3,408	5,254	
	648.6	1,260.4	3,345.0	5,254.0	
	11.0	0.4	1.2	12.6	
	10.73	24.40	64.86	100.00	
	84.30	98.62	98.78	96.96	
yes	105	18	42	165	
	20.4	39.6	105.0	165.0	
	351.6	11.8	37.8	401.2	
	63.64	10.91	25.45	100.00	
	15.70	1.38	1.22	3.04	
Total	669	1,300	3,450	5,419	
	669.0	1,300.0	3,450.0	5,419.0	
	362.6	12.1	39.0	413.8	
	12.35	23.99	63.66	100.00	
	100.00	100.00	100.00	100.00	

```
Pearson chi2(2) = 413.8151 Pr = 0.000
likelihood-ratio chi2(2) = 252.1082 Pr = 0.000
Fisher's exact = 0.000
```

Race is significantly related with byssinosis. The No's nearly halve from white to other, while the Yes's decrease is not as dramatic. No surprise that whites are better off.

```
. tabulate bys race [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

```
+-----+
| Key |
+-----+
| frequency |
| expected frequency |
| chi2 contribution |
| row percentage |
| column percentage |
+-----+
```

Suffering from byssinosis			
. 0 = no, 1 = white, 2 = other			
1 = yes			
	white	other	Total
no	3,424	1,830	5,254
	3,408.9	1,845.1	5,254.0
	0.1	0.1	0.2
	65.17	34.83	100.00
	97.38	96.16	96.96
yes	92	73	165
	107.1	57.9	165.0
	2.1	3.9	6.0
	55.76	44.24	100.00
	2.62	3.84	3.04
Total	3,516	1,903	5,419
	3,516.0	1,903.0	5,419.0
	2.2	4.0	6.2
	64.88	35.12	100.00
	100.00	100.00	100.00

Pearson chi2(1) = 6.2195 Pr = 0.013
 likelihood-ratio chi2(1) = 6.0259 Pr = 0.014
Fisher's exact = 0.016
 1-sided Fisher's exact = 0.009

Sex is significantly related with byssinosis. The No's barely decrease from male to female, while the Yes's decrease to about a third.

```
. tabulate bys sex [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

```

+-----+
| Key |
+-----+
| frequency |
| expected frequency |
| chi2 contribution |
| row percentage |
| column percentage |
+-----+

Suffering |
from |
byssinosis |
. 0 = no, | 1 = male, 2 = female
1 = yes | male female | Total
+-----+
no | 2,788 2,466 | 5,254
| 2,827.2 2,426.8 | 5,254.0
| 0.5 0.6 | 1.2
| 53.06 46.94 | 100.00
| 95.61 98.52 | 96.96
+-----+
yes | 128 37 | 165
| 88.8 76.2 | 165.0
| 17.3 20.2 | 37.5
| 77.58 22.42 | 100.00
| 4.39 1.48 | 3.04
+-----+
Total | 2,916 2,503 | 5,419
| 2,916.0 2,503.0 | 5,419.0
| 17.9 20.8 | 38.7
| 53.81 46.19 | 100.00
| 100.00 100.00 | 100.00

Pearson chi2(1) = 38.6707 Pr = 0.000
likelihood-ratio chi2(1) = 41.3443 Pr = 0.000
Fisher's exact = 0.000
1-sided Fisher's exact = 0.000

```

Smoking is significantly related with byssinosis. The No's decrease by a third from smoking to nonsmoking, while the Yes's decrease to about a third.

```
. tabulate bys smoker [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

```
+-----+
| Key |
+-----+
| frequency |
| expected frequency |
| chi2 contribution |
| row percentage |
| column percentage |
+-----+
```

Suffering from byssinosis		1 = yes, 2 = no		Total
0 = no, 1 = yes	yes	no		
no	3,064	2,190	5,254	
	3,091.9	2,162.1	5,254.0	
	0.3	0.4	0.6	
	58.32	41.68	100.00	
	96.08	98.21	96.96	
yes	125	40	165	
	97.1	67.9	165.0	
	8.0	11.5	19.5	
	75.76	24.24	100.00	
	3.92	1.79	3.04	
Total	3,189	2,230	5,419	
	3,189.0	2,230.0	5,419.0	
	8.3	11.8	20.1	
	58.85	41.15	100.00	
	100.00	100.00	100.00	

```
Pearson chi2(1) = 20.0924 Pr = 0.000
likelihood-ratio chi2(1) = 21.4215 Pr = 0.000
Fisher's exact = 0.000
1-sided Fisher's exact = 0.000
```

Length of employment is significantly related with byssinosis. The No's are about what we'd expect, while the Yes's are low for <10 years, but high for >20 years.

```
. tabulate bys emplength [fweight = count], cchi2 chi2 column exact expected lrchi2 row
```

```
+-----+
| Key          |
+-----+
| frequency    |
| expected frequency |
| chi2 contribution |
| row percentage |
| column percentage |
+-----+
```

```
Enumerating sample-space combinations:
stage 3: enumerations = 1
stage 2: enumerations = 28
stage 1: enumerations = 0
```

Suffering		Length of employment, 1 = <10 years, 2 = 10-20 years, 3 = >20 years			Total
byssinosis		<10 years	10-20 yea	>20 years	
. 0 = no,		years			
1 = yes		<10 years	10-20 yea	>20 years	
no		2,666	686	1,902	5,254
		2,645.9	690.3	1,917.8	5,254.0
		0.2	0.0	0.1	0.3
		50.74	13.06	36.20	100.00
yes		97.69	96.35	96.16	96.96
		63	26	76	165
		83.1	21.7	60.2	165.0
		4.9	0.9	4.1	9.9
Total		38.18	15.76	46.06	100.00
		2.31	3.65	3.84	3.04
		2,729	712	1,978	5,419
		2,729.0	712.0	1,978.0	5,419.0
Total		5.0	0.9	4.3	10.2
		50.36	13.14	36.50	100.00
		100.00	100.00	100.00	100.00
		100.00	100.00	100.00	100.00

```
Pearson chi2(2) = 10.1604 Pr = 0.006
likelihood-ratio chi2(2) = 10.2359 Pr = 0.006
Fisher's exact = 0.005
```

b. *Compute observed odds ratios for probability of developing byssinosis from your analyses above. In all cases use the first category as the reference category.*

Each of the odds ratios below compares the odds of “success”, developing byssinosis, of the second or third category of a predictor variable to the first as the reference category.

Dust

Those with medium (2) dustiness in the workplace are $18 \cdot 564 / (1282 \cdot 105) = 0.0754$ times as likely to develop byssinosis as those with high (1) dustiness in the workplace.

Those with low (3) dustiness in the workplace are $42 \cdot 564 / (3408 \cdot 105) = 0.0662$ times as likely to develop byssinosis as those with high (1) dustiness in the workplace.

Race

Those who are not white (2) are $73 \cdot 3424 / (1830 \cdot 92) = 1.4846$ times as likely to develop byssinosis as those who are white (1).

Sex

Females (2) are $37 \cdot 2788 / (2466 \cdot 128) = 0.3268$ times as likely to develop byssinosis as Men (1).

Smoker

Nonsmokers (2) are $40 \cdot 3064 / (2190 \cdot 125) = 0.4477$ times as likely to develop byssinosis as smokers (1).

Emplength

Those working 10-20 years (2) are $26 \cdot 2666 / (686 \cdot 63) = 1.6039$ times as likely to develop byssinosis as those working less than 10 years (1).

Those working more than 20 years (3) are $76 \cdot 2666 / (1902 \cdot 63) = 1.6909$ times as likely to develop byssinosis as those working less than 10 years (1).

c. Do any of the predictor variables seem highly related? Which ones? How could that affect conclusions?

To see which predictor variables are related to eachother, I run the series of all pairs of predictors, just displaying the Pearson chi-squared value

```

tabulate dust race [fweight = count], chi2
tabulate dust sex [fweight = count], chi2
tabulate dust smoker [fweight = count], chi2
tabulate dust emplength [fweight = count], chi2
tabulate race sex [fweight = count], chi2
tabulate race smoker [fweight = count], chi2
tabulate race emplength [fweight = count], chi2
tabulate sex smoker [fweight = count], chi2
tabulate sex emplength [fweight = count], chi2
tabulate smoker emplength [fweight = count], chi2
  
```

Based on the Pearson chi-squared values below, the following pairs of predictors are highly significant (in decreasing order): (race, emplength), (dust, sex), (sex, smoker), (dust, race), followed by significant: (sex, emplength), (dust, smoker), (dust, emplength), (race,sex), and this pair was not significant (race, smoker).

Because some of these variables are highly related, the story is more complicated than concluding simply that race or emplength alone predicts the probability of developing byssinosis. A logistic regression, where many of these factors can be included together, will help explain the relationship between these factors and the probability of developing byssinosis.

```

. tabulate dust race [fweight = count], chi2

Dustiness |
  of |
workplace, |
  1 = high, |
    2 = |
  medium, 3 | 1 = white, 2 = other
    = low |   white   other |   Total
-----+-----+-----
    high |      267      402 |      669
    medium |     855      445 |     1,300
    low |    2,394    1,056 |    3,450
-----+-----+-----
    Total |    3,516    1,903 |    5,419

      Pearson chi2(2) = 214.3310   Pr = 0.000
  
```

```

. tabulate dust sex [fweight = count], chi2

Dustiness |
  of |
workplace, |
  1 = high, |
    2 = |
  medium, 3 | 1 = male, 2 = female
    = low |   male   female |   Total
-----+-----+-----
    high |      607       62 |      669
    medium |     501      799 |     1,300
    low |    1,808    1,642 |    3,450
-----+-----+-----
    Total |    2,916    2,503 |    5,419
  
```

Pearson chi2(2) = 491.6613 Pr = 0.000

. tabulate dust smoker [fweight = count], chi2

Dustiness of workplace,		1 = yes, 2 = no		Total
1 = high, 2 = medium, 3 = low	yes	no		
high	451	218		669
medium	734	566		1,300
low	2,004	1,446		3,450
Total	3,189	2,230		5,419

Pearson chi2(2) = 24.1530 Pr = 0.000

. tabulate dust emplength [fweight = count], chi2

Dustiness of workplace,		Length of employment, 1 = <10 years, 2 = 10-20 years, 3 = >20 years			Total
1 = high, 2 = medium, 3 = low	<10 years	10-20 yea	>20 years		
high	359	87	223		669
medium	686	145	469		1,300
low	1,684	480	1,286		3,450
Total	2,729	712	1,978		5,419

Pearson chi2(4) = 12.5409 Pr = 0.014

. tabulate race sex [fweight = count], chi2

1 = white, 2 = other		1 = male, 2 = female		Total
	male	female		
white	1,945	1,571		3,516
other	971	932		1,903
Total	2,916	2,503		5,419

Pearson chi2(1) = 9.1591 Pr = 0.002

. tabulate race smoker [fweight = count], chi2

1 = white, 2 = other		1 = yes, 2 = no		Total
	yes	no		
white	2,085	1,431		3,516
other	1,104	799		1,903
Total	3,189	2,230		5,419

Pearson chi2(1) = 0.8441 Pr = 0.358

. tabulate race emplength [fweight = count], chi2

| Length of employment, 1 = <10

Stat 539 – Biostatistics II
Homework 8 solutions

```

      | years, 2 = 10-20 years, 3 = >20
1 = white, |
2 = other | <10 years  10-20 yea  >20 years |      Total
-----+-----+-----+-----+-----
white |      1,071      604      1,841 |      3,516
other |      1,658      108      137 |      1,903
-----+-----+-----+-----+-----
Total |      2,729      712      1,978 |      5,419

```

Pearson chi2(2) = 1.6e+03 Pr = 0.000

```
. tabulate sex smoker [fweight = count], chi2
```

```

1 = male, |      1 = yes, 2 = no
2 = female |      yes      no |      Total
-----+-----+-----+-----
male |      2,059      857 |      2,916
female |      1,130      1,373 |      2,503
-----+-----+-----+-----
Total |      3,189      2,230 |      5,419

```

Pearson chi2(1) = 360.6466 Pr = 0.000

```
. tabulate sex emplength [fweight = count], chi2
```

```

      | Length of employment, 1 = <10
      | years, 2 = 10-20 years, 3 = >20
1 = male, |
2 = female | <10 years  10-20 yea  >20 years |      Total
-----+-----+-----+-----+-----
male |      1,308      446      1,162 |      2,916
female |      1,421      266      816 |      2,503
-----+-----+-----+-----+-----
Total |      2,729      712      1,978 |      5,419

```

Pearson chi2(2) = 79.6952 Pr = 0.000

```
. tabulate smoker emplength [fweight = count], chi2
```

```

      | Length of employment, 1 = <10
      | years, 2 = 10-20 years, 3 = >20
1 = yes, 2 |
= no | <10 years  10-20 yea  >20 years |      Total
-----+-----+-----+-----+-----
yes |      1,587      481      1,121 |      3,189
no |      1,142      231      857 |      2,230
-----+-----+-----+-----+-----
Total |      2,729      712      1,978 |      5,419

```

Pearson chi2(2) = 26.7018 Pr = 0.000

d. *Write a summary of your analyses using these “classical” methods.*

