

# Sta442/1008f05 Overheads 4: The Berkeley Data

---

The three variables are sex, department and admit.

```
proc freq;
  tables sex*admit / nopercnt nocol chisq;
  tables dept*sex / nopercnt nocol chisq;
  tables dept*admit / nopercnt nocol chisq;
  tables dept*sex*admit / nopercnt nocol chisq;
  weight count;
```

---

Table of sex by admit

sex	admit		
Frequency			
Row Pct	No	Yes	Total
Male	1493 55.48	1198 44.52	2691
Female	1278 69.65	557 30.35	1835
Total	2771	1755	4526

The FREQ Procedure

Statistics for Table of sex by admit

Statistic	DF	Value	Prob
Chi-Square	1	92.2053	<.0001

Now we look at the output from dept\*sex\*admit: Sex by admit, controlling for dept.

Table 1 of sex by admit  
Controlling for dept=A

sex	admit		
Frequency	No	Yes	Total
Male	313 37.94	512 62.06	825
Female	19 17.59	89 82.41	108
Total	332	601	933

Statistics for Table 1 of sex by admit  
Controlling for dept=A

Statistic	DF	Value	Prob
Chi-Square	1	17.2480	<.0001

Table 2 of sex by admit  
Controlling for dept=B

sex	admit		
Frequency	No	Yes	Total
Male	207 36.96	353 63.04	560
Female	8 32.00	17 68.00	25
Total	215	370	585

Statistics for Table 2 of sex by admit  
Controlling for dept=B

Statistic	DF	Value	Prob
Chi-Square	1	0.2537	0.6145

Table 3 of sex by admit  
Controlling for dept=C

sex	admit		
Frequency			
Row Pct	No	Yes	Total
Male	205 63.08	120 36.92	325
Female	391 65.94	202 34.06	593
Total	596	322	918

Statistics for Table 3 of sex by admit  
Controlling for dept=C

Statistic	DF	Value	Prob
Chi-Square	1	0.7535	0.3854

Table 4 of sex by admit  
Controlling for dept=D

sex	admit		
Frequency			
Row Pct	No	Yes	Total
Male	279 66.91	138 33.09	417
Female	244 65.07	131 34.93	375
Total	523	269	792

Statistics for Table 4 of sex by admit  
Controlling for dept=D

Statistic	DF	Value	Prob
Chi-Square	1	0.2980	0.5852

Table 5 of sex by admit  
Controlling for dept=E

sex	admit		
Frequency	No	Yes	Total
Male	138 72.25	53 27.75	191
Female	299 76.08	94 23.92	393
Total	437	147	584

Statistics for Table 5 of sex by admit  
Controlling for dept=E

Statistic	DF	Value	Prob
Chi-Square	1	1.0011	0.3171

Table 6 of sex by admit  
Controlling for dept=F

sex	admit		
Frequency	No	Yes	Total
Male	351 94.10	22 5.90	373
Female	317 92.96	24 7.04	341
Total	668	46	714

Statistic	DF	Value	Prob
Chi-Square	1	0.3841	0.5354

Department	Percent Applicants Female	Percentage Applicants Accepted
A	11.58	64.42
B	4.27	63.25
C	64.60	35.08
D	47.35	33.96
E	67.29	25.17
F	47.76	6.44

**Pooling the chisquare tests from subtables:** Just add chisquare values and degrees of freedom

$$17.2480+0.2537+0.7535+0.2980+1.0011+0.3841 = 19.9384$$

with 6 degrees of freedom. Using any statistics text (except this one), we can look up the critical value at the 0.05 significance level. It's 12.59; since  $19.9 > 12.59$ , the pooled test is significant at the 0.05 level. To get a p-value for our pooled chisquare test, we can use SAS. Here's the output from the program on the next page:

```

Chisquare =          X          PVAL
                19.9384 df=6, p = 0.0028402

```

```

/***** berkeley.sas *****/
options linesize=79 pagesize=35 noovp formdlim='_';
title 'Berkeley Graduate Admissions Data: ';

proc format;
  value sexfmt 1 = 'Female' 0 = 'Male';
  value ynfmt 1 = 'Yes' 0 = 'No';
data berkley;
  input line sex dept $ admit count;
  format sex sexfmt.; format admit ynfmt.;
  datalines;
1      0      A      1      512
2      0      B      1      353
3      0      C      1      120
4      0      D      1      138
5      0      E      1      53
6      0      F      1      22
7      1      A      1      89
8      1      B      1      17
9      1      C      1      202
10     1      D      1      131
11     1      E      1      94
12     1      F      1      24
13     0      A      0      313
14     0      B      0      207
15     0      C      0      205
16     0      D      0      279
17     0      E      0      138
18     0      F      0      351
19     1      A      0      19
20     1      B      0      8
21     1      C      0      391
22     1      D      0      244
23     1      E      0      299
24     1      F      0      317
;
proc freq;
  tables sex*admit / nopercnt nocol chisq;
  tables dept*sex / nopercnt nocol chisq;
  tables dept*admit / nopercnt nocol chisq;
  tables dept*sex*admit / nopercnt nocol chisq;
  weight count;

/* Get p-value */
proc iml;
  x = 19.9384;
  pval = 1-probchi(x,6);
  print "Chisquare = " x "df=6, p = " pval;

```