Public Health 241 - Spring 2018
Assignment 5

1. (a)

$$\widehat{OR} = \frac{16 \times 32}{8 \times 24} = 2.667$$

(b)

$$\widehat{var}(\widehat{OR}) = \frac{1}{16} + \frac{1}{8} + \frac{1}{24} + \frac{1}{32} = 0.2604$$

(c) For a 90% confidence interval we a need a z-value such that $(100\% - 90\%)/2 = 5\%$ of the standard normal density are greater than z, i.e. we need the 95th percentile of the standard normal density:

```
. display invnormal(0.95)
1.6448536
```

$$\log(2.667) \pm 1.645\sqrt{0.2604} = (0.141, 1.820)$$

(d)

$$(e^{0.141}, e^{1.820}) = (1.15, 6.17)$$

(e) Since the confidence interval does not include the null value 1, the data provide evidence (at a 10% significance level) for an association between employment history and the risk for oral cancer.

(f)

$$\chi^2 = \frac{80(16 \times 32 - 8 \times 24)^2}{40 \times 40 \times 24 \times 56} = 3.81 \qquad (p = 0.051)$$

The χ^2 test rejects the null hypothesis of independence at a 10% significance level (but would accept the null hypothesis at the 5% significance level). As is to be expected, this agrees with the confidence interval we calculated.

(g)

$$OR_{ss} = \frac{16 \times 32}{9 \times 25} = 2.28$$

The regular point estimate differs from the small-sample adjusted point estimate by about 17%. This is a considerable discrepancy so that the approximations used in the original confidence interval calculation may be somewhat suspect and exact confidence interval may be preferable.

(h)

$$\begin{aligned}
(\log \widehat{OR})_{SS} &= \log \frac{16.5 \times 32.5}{8.5 \times 24.5} = 0.946 \\
\widehat{var}(\log \widehat{OR})_{SS} &= \frac{1}{16.5} + \frac{1}{8.5} + \frac{1}{24.5} + \frac{1}{32.5} = 0.250 \\
90\% CI \text{ for } \log(OR) &= 0.946 \pm 1.645 \times \sqrt{0.250} = (0.124, 1.768) \\
90\% CI \text{ for } OR &= (e^{0.124}, e^{1.768}) = (1.13, 5.86)
\end{aligned}$$

The upper limit of the approximate confidence interval differs from the upper limit of the small-sample confidence interval by about 5%. This is not a huge difference, but again we may be worried that the approximations used in the original confidence interval calculation may be somewhat suspect so that exact confidence interval may be preferable.

(i) . cci 16 24 8 32, level(90)

```
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        16          24  |         40      0.4000
        Controls |         8          32  |         40      0.2000
-----------------+------------------------+------------------------
           Total |        24          56  |         80      0.3000
                 |                        |
                 |      Point estimate    |    [90% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         2.666667       |    1.033422     7.03807 (exact)
 Attr. frac. ex. |             .625       |    .0323408    .8579156 (exact)
 Attr. frac. pop |              .25       |
                 +-------------------------------------------------
                               chi2(1) =     3.81  Pr>chi2 = 0.0510
```

The exact confidence interval $(1.03, 7.04)$ is substantially wider than the approximate confidence intervals calculated above, even if small-sample adjustments are used. The normal approximation that forms the basis for those approximate confidence intervals may not be appropriate.

2. (a)

```
. generate alc2 = 0

. replace alc2 = 1 if alcgp >=2

. tab alc2 alcgp

           |                     alcgp
      alc2 |         0          1          2          3 |     Total
-----------+--------------------------------------------+----------
         0 |        35         38          0          0 |        73
         1 |         0          0         31         31 |        62
-----------+--------------------------------------------+----------
     Total |        35         38         31         31 |       135

. tab alc2 casestatus [freq=freq]

           |      casestatus
      alc2 |         0          1 |     Total
-----------+----------------------+----------
         0 |       666        104 |       770
         1 |       109         96 |       205
-----------+----------------------+----------
     Total |       775        200 |       975
```

(b)
```
. cc casestatus alc2 [freq=freq], wo

                 | alc2                   |           Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        96         104  |        200       0.4800
        Controls |       109         666  |        775       0.1406
-----------------+------------------------+------------------------
           Total |       205         770  |        975       0.2103
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         5.640085       |    4.000589    7.951467 (Woolf)
 Attr. frac. ex. |         .8226977       |    .7500368    .8742371 (Woolf)
 Attr. frac. pop |         .3948949       |
                 +-------------------------------------------------
                               chi2(1) =   110.26  Pr>chi2 = 0.0000
```

(c)
```
. cc casestatus alc2 [freq=freq]
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        96         104  |        200       0.4800
        Controls |       109         666  |        775       0.1406
-----------------+------------------------+------------------------
           Total |       205         770  |        975       0.2103
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         5.640085       |    3.937435    8.061794 (exact)
 Attr. frac. ex. |         .8226977       |    .7460276    .8759581 (exact)
 Attr. frac. pop |         .3948949       |
                 +-------------------------------------------------
                               chi2(1) =   110.26  Pr>chi2 = 0.0000
```

The end points of the approximate confidence interval are less than 2% off from those of the exact confidence interval. The sample sizes of cases and controls appear large enough to justify the normal approximation.

(d) Either of the two last tables gives a χ^2-statistic of 110.26. Stata rounds the p-value down to zero, but we can obtain an actual p-value using the `chi2tail()` command:

```
. display chi2tail(1, 110.26)
8.595e-26
```

Like the confidence interval, the χ^2-test provides strong evidence against the null hypothesis that alcohol consumption is independent of the risk of esophageal cancer.