Assignment 4, Model Solutions

Q1. We begin by setting up the 2x2 table.

	Dependent on Injections	Not dependent on injections
Family history	24	4
No family history	42	30

a. Here is the relevant output from R:

<u>Pearson's Chi-squared test without Yate's continuity correction</u>

```
X-squared = 6.7354, df = 1, p-value = 0.009452
```

<u>Pearson's Chi-squared test with Yates' continuity correction</u>

```
X-squared = 5.5705, df = 1, p-value = 0.01827
```

<u>Fisher's Exact Test for Count Data</u>

```
p-value = 0.01003
```

All three tests provide evidence that the proportion of diabetic patients who are dependent on injections differs between those with and without family history. The P-value for the continuity test is bigger than that from the un-corrected test, as is the P-value for Fisher's test.

b. Because the family history groups constitute the rows in the above table, we can apply prop.test() to calculate estimates and obtain the confidence interval. I include only the relevant output.

```
2-sample test for equality of proportions with continuity correction
```

<u>95 percent confidence interval:</u>

```
0.0764760 0.4711431
```

```
2-sample test for equality of proportions without cont. corr
```

<u>95 percent confidence interval:</u>

```
0.1012776 0.4463415
```

<u>sample estimates:</u>

```
    prop 1    prop 2
0.8571429 0.5833333
```

The estimated risk difference is .857 – .583 = .274, with 95% confidence interval .076 to .471 (continuity corrected method) or .101 to .446 (un-corrected).

The relative risk is .857/.583 = 1.47. If one applies the methods appropriate under the Poisson assumption one obtains a confidence interval of .89 to 2.44. However the Poisson assumption doesn't really apply here – some students realized this and applied a different formula, one appropriate for estimating relative risks from binomial data. This approach (I'll cover this in class later) yields an interval of 1.15 to 1.88.

Q2. We start again by setting up the 2x2 table

	Crohn's patient	Control
Ate cornflakes	23	18
Didn't eat cornflakes	11	50

Applying fisher.test() in R we obtain a P–value <.0001 , and a confidence interval of 2.172 to 15.82. The estimate of 5.69 indicates an apparent increase in risk of Crohn's disease in those who eat cornflakes.

```
    Fisher's Exact Test for Count Data

p-value = 9.626e-05

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval:

  2.172565 15.824000
```

```
sample estimates:
odds ratio
  5.693577
```

The above results are based on exact theory. If we apply the simple asymptotic formula directly, we get an estimated odds ratio of

$\frac{23 \times 50}{11 \times 18} = 5.808$ with confidence interval

$$5.808 \times \exp^{\pm 1.96\sqrt{1/23 + 1/50 + 1/11 + 1/18}} = 2.37, 14.3 \ .$$

The statistical interpretation is that there appears to be a positive association between cornflake consumption and Crohn's disease, i.e. an apparent increased risk of Crohns's disease in those who eat cornflakes. However, because in this retrospective study we don't really know which came first, a prospective study would be more appropriate for establishing a true causal association.

b. The apparent relative risk is $23/41 \div 11/61 = 3.11$. We note that 102 is simply 1.5 times 68, so we simply multiply the frequencies for the controls by 1.5, yielding the following table:

	Crohn's patient	Control
Ate cornflakes	23	27
Didn't eat cornflakes	11	75

c. The odds ratios for both tables are 5.8, whereas the apparent relative risks are 3.11 and 3.60. This illustrates the fact that one should not calculate relative risks from retrospective data.