## CHI-SQUARE TEST OF INDEPENDENCE, ODDS AND RATIOS

**Question 1.**
We wish to investigate, in a population, if the stress is a risk factor for depression. A sample of 500 people has been observed and the results presented in the next table were obtained:

| | **Depression = yes** | **Depression = no** |
|---|---|---|
| **Stress = yes** | 100 | 120 |
| **Stress = no** | 70 | 210 |

Is the stress a risk factor for depression?

1. What is the $H_0$?
   $H_0$: The stress and depression are independent.

2. What is the $H_{1/a}$?
   $H_{1/a}$: The stress and depression are dependent.

4. Significance level: α = 0.05
5. Rejection region: $[3.84, \infty)$
6. Compute the expected contingency table.

| Observed table | Depression = yes | Depression = no | Total |
|---|---|---|---|
| Stress = yes | 100 | 120 | 220 |
| Stress = no | 70 | 210 | 280 |
| Total | 170 | 330 | 500 |

| Expected table | Depression = yes | Depression = no | Total |
|---|---|---|---|
| Stress = yes | =220*170/500 = 75 | =220*330/500 = 145 | 220 |
| Stress = no | =280*170/500 = 95 | =280*330/500 = 185 | 280 |
| Total | 170 | 330 | 500 |

7. Calculate the $\chi^2$ parameter

Hint: The test statistic is given by: $\chi^2 = \sum_{i=1}^{LC} \frac{(f_i^o - f_i^t)^2}{f_i^t}$, where $f_i^o$ and $f_i^t$ are observed and expected frequency respectively.

$\chi^2 = (100-75)^2/75 + (120-145)^2/145 + (70-95)^2/95 + (210-185)^2/185 = 22.60$

8. Interpret the results from statistical and clinical point of views.

Statistical interpretation: Since the value of $\chi^2$ (22.60) belong to the rejection region the null hypothesis is rejected at a significance level of 5%.
Clinical interpretation: The stress is a risk factor for depression.

**Question 2 (Source: http://www.brettscaife.net/statistics/introstat/06risk/exercise.html)**
The following data are taken from the paper Caries prevalence in northern Scotland before and 5 years after, water defluoridation (Stephen et al., 1987, BDJ 163: 324-326). They show the social composition of children recruited to two arms of the study one before and one after water defluoridation. What is the probability that we do not know the social class of a child in the fluoridated arm? What is the probability that a child in the defluoridated arm is from social class III?

| Social class | Fluoridated | Defluoridated |
|---|---|---|
| I & II | 16 | 32 |
| III | 45 | 53 |
| IV & V | 32 | 22 |
| Not known | 13 | 19 |
| Total | 106 | 126 |

What is the probability that we do not know the social class of a child in the fluoridated arm?

There are 106 children in total and we do not know the class of 13 of them. So, the probability we do not know the social class (p) is:

$$p = 13/106 = 0.12$$

Expressing this probability as a percentage chance is 12%.

What is the probability that a child in the defluoridated arm is from social class III?

The probability of a child in the defluoridated arm being in social class III is a bit more problematic. It is not really 53÷126, although this is the way I approached it in the lecture. In reality there is a problem caused by the not known category. These children will, in reality, belong to one of the other three categories (which we assume to be exhaustive). The best strategy to adopt here is probably to ignore the not known category and calculate the probability of being in social class III based on the adjusted total (126 - 19 = 107). So, the probability of being in social class III (p) is:

$$p = 53/107 = 0.50$$

This is, strictly, only correct if the 19 children in the not known category are split between the three other categories in the same proportions to everyone else. This is probably not the case but we don't know. We have here an illustration of one of the problems resulting from inaccurate data collection.

**Question 3 (Source: http://www.brettscaife.net/statistics/introstat/06risk/exercise.html)**
A recent MDentSci project was looking at a number of risk factors thought to be associated with the health of oral implants in a population of elderly patients. One factor considered was smoking. The table below shows the number of healthy and non-healthy implants for smokers a non-smokers. Calculate and interpret the risk ratio and the odds ratio. The patients were selected for the study on the basis of the health of their implants (a case-control study). Which of the two ratios you have calculated would you use to report your results? Why?
A $\chi 2$ test was performed on these data, the results were: $\chi 2 = 2.023$, df = 1, p = 0.16. Interpret these results.

| | Healthy implant | | |
|---|---|---|---|
| | Yes | No | Total |
| Smoker | 32 | 48 | 80 |
| Non-smoker | 10 | 7 | 17 |
| Total | 42 | 55 | 97 |

I have calculated risks and odds on the basis of unhealthy implants as this made more sense to me. You may have done it the other way and calculated for healthy implants. In this case you answers will be different to mine.

**Risk ratio**

The risk of a smoker having an unhealthy implant is: 48/80 = 0.600
48 out of 80 smokers have unhealthy implants.

**The risk of a non-smoker having an unhealthy implant** is: 7/17 = 0.412
7 out of 17 non-smokers have unhealthy implants.

**The risk ratio for having a healthy implant for smokers compared to non-smokers** is 0.600/0.412 = 1.46
Smokers have about one and a half times the risk of having an unhealthy implant compared to non-smokers

**Odds ratio**

The odds of a smoker having an unhealthy implant are: 48/32 = 1.5
48 smokers have unhealthy implants, 32 do not.

**The odds of a non-smoker having an unhealthy implant** are: 7/10 = 0.7
7 non-smokers have unhealthy implants, 10 do not.

**The odds ratio for having a healthy implant for smokers compared to non-smokers** is 1.5/0.7= 2.14
Smokers have about twice the odds of having an unhealthy implant compared to non-smokers

**Which to use?**

As this is a retrospective study (the subjects were selected on the basis of the outcome - health of implant - rather than exposure - smoking status) we have to use the odds ratio.

**The Chi-squared test**

As p > 0.05 the $\chi 2$ test failed to show a significant association between smoking and health of implant. This leads us to believe that the odds ratio of 2.14 we calculated would be likely to have a confidence interval that includes 1 (indicating no difference in the odds of an unhealthy implant between the two groups).

**Question 4. (Source: http://www.brettscaife.net/statistics/introstat/06risk/exercise.html)**
A study (Erosion of dental enamel among competitive swimmers at a gas-chlorinated swimming pool, Centerwall et al., 1986, Am. J. Epid. 123: 641-647) was carried out to see if swimming in chlorinated water was linked to erosion of dental enamel. 49 swimmers with erosion of dental enamel (cases) were recruited along with 245 swimmers without erosion (controls). The data are summarized below.

| | Erosion of dental enamel | | |
|---|---|---|---|
| Hours of swimming per week | Yes | No | Total |
| 6 or more | 32 | 118 | 150 |
| less than 6 | 17 | 117 | 134 |
| Total | 49 | 235 | 284 |

Calculate the appropriate ratio to show the effect of excessive swimming on erosion of dental enamel. A test was performed on these data, the results were: $\chi2 = 4.802$, df = 1, p = 0.03. Interpret these results.

**Which ratio to use?**

As this is a retrospective study (the subjects were selected on the basis of the outcome - erosion of dental enamel - rather than exposure - hours of swimming) so we have to use the odds ratio.

**Odds ratio**

The odds of a frequent swimmer having erosion are: 32/118 = 0.27
32 frequent swimmers have erosion, 118 do not.

**The odds of an infrequent swimmer having erosion** are: 17/117 = 0.15
17 infrequent swimmers have erosion, 117 do not.

**The odds ratio for having erosion for frequent swimmers compared to infrequent swimmers** is 0.27/0.15= 1.87

Frequent swimmers have more than one and three-quarters the chance of having dental erosion compared to infrequent swimmers

**The Chi-squared test**

As $p < 0.05$ the $\chi2$ test leads us to reject the null hypothesis of no association. We would conclude that there is a significant association between frequency of swimming and erosion of dental enamel. We would expect that if we calculated a 95% confidence interval for the odds ratio above that it would not include 1.

**Question 5 (Source: http://www.brettscaife.net/statistics/introstat/06risk/exercise.html)**
The following data from a prospective study are taken from the paper Dental caries in pre-school children: associations with social class, tooth brushing habit and consumption of sugars and sugar-containing foods (Gibson & Williams 1999, Caries Research 33: 101-113). They show the number of children with caries according to three different risk factors: social class; tooth brushing frequency; and frequency of consumption of sugary foods. Which of these three factors has most impact on the likelihood of a child developing caries?

| | Caries | | |
|---|---|---|---|
| Social class | Yes | No | Total |
| Manual | 162 | 574 | 736 |
| Non-manual | 64 | 574 | 638 |
| Total | 226 | 1148 | 1374 |

| | Caries | | |
|---|---|---|---|
| Brushing frequency | Yes | No | Total |
| 0 or 1 per day | 114 | 477 | 591 |
| >1 per day | 112 | 671 | 783 |
| Total | 226 | 1148 | 1374 |

| | Caries | | |
|---|---|---|---|
| Sugary foods | Yes | No | Total |
| <3 times a day | 61 | 347 | 408 |
| 3 or more times a day | 165 | 801 | 966 |
| Total | 226 | 1148 | 1374 |

As this is a prospective study it makes sense to calculate the risk ratios. I've expressed all my risk ratios in terms of the most 'risky' categories. If you've done it the other way round it doesn't matter, as long as your interpretation ends up the same.

**Social class**

**The risk of a 'manual' child having caries** is: 162/736 = 0.220
162 out of 736 'manual' children have caries.

**The risk of a 'non-manual' child having caries** is: 64/638 = 0.100
64 out of 638 'non-manual' children have caries.

**The risk ratio for having caries for 'manual' children compared to 'non-manual' children** is 0.220/0.100= 2.19
'Manual' children have about twice the risk of caries compared to 'non-manual' children
Brushing frequency

**The risk of a child brushing less than twice a day having caries** is: 114/591 = 0.193
114 out of 591 children brushing less than twice a day have caries.

**The risk of a child brushing at least twice a day having caries** is: 112/783 = 0.143

112 out of 783 children brushing at least twice a day have caries.

**The risk ratio for having caries for children brushing less than twice a day compared to children brushing at least twice a day** is 0.193/0.143= 1.35

Children brushing less than twice a day have about one and a third the risk of caries compared to children who brush at least twice a day

**Sugary foods**

**The risk of a child who eats sugary foods 3 or more times a day having caries** is: 165/966 = 0.171

165 out of 966 children eating sugary foods 3 or more times a day have caries.

**The risk of a child who eats sugary foods less than 3 times a day having caries** is: 61/408 = 0.150

61 out of 408 children eating sugary foods less than 3 times a day have caries.

**The risk ratio for having caries for children eating sugary foods 3 or more times a day compared to children eating sugary foods less than 3 times a day** is 0.171/0.150= 1.14

Children brushing less than twice a day have about 1·14 the risk of caries compared to children who brush at least twice a day

**Which risk factor is worse?**

The risk ratio for manual compared to non-manual is by far the biggest of the three we have calculated. We may be justified on this evidence in claiming that the social class of a child has a greater impact on the risk of developing caries than either frequency of tooth brushing or frequency of eating sugary foods.

Of course, the three factors we have looked at are unlikely to be totally independent. Frequency of tooth brushing and frequency of eating sugary foods are likely to be influenced by social class and possibly by each other. We would probably do better to analyze these data with more sophisticated modeling techniques that could take these possible interactions into account. But these are too much for you!