Periodontal conditions and coronary artery disease

Coronary artery disease (CAD) is the most common type of heart disease. It is the leading cause of death in the United States in both men and women. For this reason it is important to identify potential risk factors for CAD. A study in Belgium discussed in the 2004 *Journal of Periodontology* explored the possible connection between periodontal disease and CAD. It is reasonable to think that such a connection could exist, as there is a growing body of evidence supporting the idea that chronic infections can play a role in the initiation and progression of CAD. In the study 108 older CAD patients were examined, with 98 found to have periodontitis, while of 62 healthy older controls 41 were found to have periodontitis. Is there evidence that periodontal disease is related to CAD?

This is a two–sample problem, referring to Binomial proportions. We want to examine if the observed proportions in the two samples can be considered equal. Let $\overline{p}$ be the sample frequency estimate of p, the probability of periodontitis. We know that

$$var(\overline{p}) \approx \overline{p}\,(1 - \overline{p}\,)/n,$$

where "≈" is representing "can be estimated by." This is all that is needed to construct a **two–sample z–test** for a comparison of Binomial proportions. It is similar to the two–sample t–test for the comparison of means that assumes unequal variances, except that we assume that the degrees of freedom are large enough so that we can use a normal approximation. An exact test would be desirable here, and that is also provided by `Minitab` (it is called *Fisher's exact test*; see the output below). Suppose we have two Binomial samples, with the observed proportions of periodontitis being $\overline{p}_1$ and $\overline{p}_2$, respectively, based on samples of size n_1 and n_2, respectively. These sample proportions are estimates of the true probabilities of periodontitis, p_1 and p_2, respectively. We wish to test the hypotheses

$$H_0 : p_1 = p_2$$

versus

$$H_a : p_1 \neq p_2.$$

The z–statistic then has the form

$$z = \frac{\overline{p}_1 - \overline{p}_2}{\sqrt{\overline{p}_1(1 - \overline{p}_1)/n_1 + \overline{p}_2(1 - \overline{p}_2)/n_2}}.$$

How does this testing procedure apply to this study? We have $\overline{p}_{\text{CAD}} = 98/108 = .9074$ and $\overline{p}_{\text{control}} = 41/62 = .6613$, so the test statistic is

$$z = \frac{.9074 - .6613}{\sqrt{.000778 + .0036126}} = 3.71,$$

which has a tail probability of .0001. Thus, the study provides highly statistically significant evidence that periodontitis is associated with having coronary artery disease.

An equivalent way to approach this question is by using confidence intervals. Since the standard error of $\overline{p}_1 - \overline{p}_2$ can be approximated by

$$\text{SE}(\overline{p}_1 - \overline{p}_2) \approx \sqrt{\overline{p}_1(1-\overline{p}_1)/n_1 + \overline{p}_2(1-\overline{p}_2)/n_2},$$

a 95% confidence interval for the true difference in probabilities $p_1 - p_2$ has the form

$$\overline{p}_1 - \overline{p}_2 \pm (1.96)\sqrt{\overline{p}_1(1-\overline{p}_1)/n_1 + \overline{p}_2(1-\overline{p}_2)/n_2}.$$

If the 95% confidence interval does not contain zero, the hypothesis of equality of the two proportions can be rejected at the 5% level of significance.

Substituting the observed values into the equation above gives the 95% confidence interval, which is $(0.1162, 0.376)$. This result confirms that the difference in proportions is deemed statistically significantly different from zero, and is consistent with the true difference in periodontitis probability being roughly between 11.5% and 37.5%.

Does this mean that 91% of people with periodontitis have CAD? Certainly not, since the probabilities given here are conditional probabilities of periodontitis given CAD status, not CAD given periodontitis status (since they refer to the two groups of people with CAD and people without CAD). We know how to get from this probability to the one we want, of course:

$$P(\text{CAD}|\text{Periodontitis}) = \frac{P(\text{Periodontitis}|\text{CAD})P(\text{CAD})}{P(\text{Periodontitis})},$$

by the definition of conditional probability. According to published research, 35% of seniors have CAD, and 55% of seniors have periodontitis, so that would imply

$$\hat{P}(\text{CAD}|\text{Periodontal disease}) = \frac{(.9074)(.35)}{.55} = .5774.$$

Similarly,

$$\begin{aligned}\hat{P}(\text{CAD}|\text{No periodontitis}) &= \frac{P(\text{No periodontitis}|\text{CAD})P(\text{CAD})}{P(\text{No periodontitis})} \\ &= \frac{(.0926)(.35)}{.45} \\ &= .0720.\end{aligned}$$

This illustrates that using periodontitis as a diagnostic tool for CAD is a reasonable strategy, since the estimated probability of CAD is 8 times higher for those with periodontitis versus those without it.

This sort of experimental design, where sampling is done on the basis of the outcome of interest (CAD) rather than on the basis of a potential cause of the outcome (periodontitis) is called a **case-control** design. Such designs suffer from the difficulty noted here (that the probability that is estimable from the data is the "wrong" one), and a regression method that addresses this issue in an effective way is **logistic regression**.

Below is `Minitab` output corresponding to the results given above.

```
Test and CI for Two Proportions

Sample   X    N  Sample p
1       98  108  0.907407
2       41   62  0.661290


Difference = p (1) - p (2)
Estimate for difference:  0.246117
95% CI for difference:  (0.116246, 0.375988)
Test for difference = 0 (vs not = 0):  Z = 3.71  P-Value = 0.000

Fisher's exact test: P-Value = 0.000
```

Minitab commands

A confidence interval and hypothesis test for the equality of two Binomial proportions can be obtained by clicking `Stat` → `Basic Statistics` → `2 Proportions`. Data can be given in stacked form (zeroes and ones in one column, with a second column identifying groups), unstacked form (two columns of zeroes and ones), or as summarized values of the number of trials and number of successes in each group.