Breslow-Day Test for Homogeneity of the Odds Ratio

To test the hypothesis that the OR between X and Y is the same at each level of Z.

$$H_0 : \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(k)}$$

We calculate the Breslow-Day test statistic as follows:

$$\chi^2 = \sum_i \sum_j \sum_k \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

- The $\hat{\mu}_{ijk}$ are calculated assuming the null is true, that there is a common odds ratio across the level of the third variable.
- The test statistic has an approximate chi-square distribution with df $= k - 1$.
- Rejection Region: $\chi^2 > \chi^2_{\alpha,k-1}$
- p-value: $\Pr(\chi^2 > \chi^2_{obs})$
- Similar to our previous analysis of 2×2 tables:

(1) n_{ijk} = observed frequency in row i, column j, of table k

(2) $\hat{\mu}_{ijk}$ = expected frequency in row i, column j, of table k

In particular,

$$\hat{\mu}_{ijk} = \frac{n_{i+k}\, n_{+jk}}{n_k}.$$

- We need at least 80% of the expected cell counts to be > 5 for the chi-square approximation to be valid.
- If we fail to reject H_0 using the Breslow-Day statistic, then we can assume homogeneous association holds, and we can form a common value for the odds ratio.

The Mantel-Haenszel estimate of the common odds ratio, θ

Let each partial table (indexed by k) have the form:

	$Y = 1$	$Y = 0$
$X = 1$	a_k	b_k
$X = 0$	c_k	d_k

Then

$$\hat{\theta}_{MH} = \frac{\sum_{k=1}^{K} \frac{a_k d_k}{n_k}}{\sum_{k=1}^{K} \frac{b_k c_k}{n_k}}, \quad n_k = a_k + b_k + c_k + d_k$$

- Add the numerators and denominators of the individual odds ratios separately, then divide
- The conditional independence is stronger than homogeneous association. Not only are all of the OR equal, but they all equal 1!!

Cochran-Mantel-Haenszel Test for Conditional Independence

To test the hypothesis that X and Y are conditionally independent given a third variable Z:

$$H_0 : \theta_{XY(1)} = \theta_{XY(2)} = \cdots = \theta_{XY(k)} = 1$$

$$H_a : \text{At least one } \neq 1$$

Form the chi-square statistic:

$$\chi^2_{CMH} = \frac{\sum_{k=1}^{K}\left(n^*_{ijk} - \frac{n_{i+k}\, n_{+jk}}{n_k}\right)^2}{\sum_{k=1}^{K}\frac{n_{1+k}\, n_{+1k}\, n_{2+k}\, n_{+2k}}{n_k^2(n_k-1)}}$$

- Which follows an approximate chi-square distribution with $v = 1$ df.

- Note: If all $\theta_{XY(k)} = 1$, then the CMH statistic is close to zero. However if some or all $\theta_{XY(k)} > 1$ (or < 1, but not both), then the CMH statistic is large.

Let's revisit the Berkley Admissions Data problem

Let G = Gender, A = Admissions Status

$$H_0 : \theta_{GA(A)} = \theta_{GA(B)} = \cdots = \theta_{GA(F)} = 1$$

$$H_a : \text{At least one } \neq 1$$

```
Mantel-Haenszel X-squared = 1.4269, df = 1, p-value = 0.2323

alternative hypothesis:  true common odds ratio is not equal to 1

95 percent confidence interval:

0.7719074 1.0603298

sample estimates:

common odds ratio

0.9046968
```

- Fail to reject the null, so we can assume conditional independence!

- Which means that there was not bias in the admissions process against women, once the departments they applied to were taken into account

- Conditional independence also implies homogeneous association!

- If you compute the Mantel-Haenszel estimate of the common odds ratio using data given in Activity 2, then
$$\hat{\theta}_{MH} = 1.105$$

Models for Matched Pairs – McNemar Test Comparing Marginal Proportions

What is McNemar Test? The chi-squared test for a comparison of two dependent proportions is called the *McNemar test.*

Every method that we have looked at so far has assumed that subjects in the different treatment groups are independent. Here, we consider two samples that have a natural pairing between each subject in one sample and a subject in the other sample.

- In this case, we say that the samples are matched pairs and are statistically dependent.

The most common way for this to happen is if the two samples contain the same subjects.

- Longitudinal studies: same subjects are observed over time

- Surveys that observe two or more similar response variables (Y_1 and Y_2)

For example, you could ask a random sample of people two Yes/No questions and record the number of people who answer Yes/No to each question in a 2×2 table.

- Link to class survey, two yes/no questions

- If the two questions are similar, we would expect a strong association in the answers since it's the same people answering both questions

Example: Chemotherapy Treatments for Breast Cancer

Suppose that we are looking to compare two different chemotherapy treatments for breast cancer and plan to set up the experiment so that the two treatment groups are as comparable as possible on other prognostic indicators.

- In Treatment A, a patient gets chemo perioperatively (i.e. around the time of the surgery - within 1 week of the mastectomy) and for an additional 6 months afterwards.

- In Treatment B, chemo is only given perioperatively

The easiest way to do set up this is to match two people based on the prognostic indicators (such as age and clinical condition) and then randomly assign one to the treatment group and the other to the control group.

The response variable is survival status at the end of 5 years.

Treatment	Survive 5 years	Die within 5 years	Total
A	526	95	621
B	515	106	621
Total	1041	201	1242

For Treatment A: $\text{P(Survival)} = \frac{526}{621} = 0.847$

For Treatment B: $\text{P(Survival)} = \frac{515}{621} = 0.829$ Only a small difference!

A chi-squared test of independence gives:

$\chi^2 = 0.59$ with $v = 1$ df p-value=0.441 Not significant!

However, this test is valid only for two samples which are independent! So we need to come up with an alternate approach.

In matched pairs, "individual people" are not the unit of analysis like in all of the other studies up to this point. Instead, it's the "matched pair" that is the unit of analysis. So let's look at our data in terms of the paired data:

	Treatment B Outcome		
Treatment A Outcome	Survive 5 years	Die within 5 years	Total
Survive 5 years	510	16	526
Die Within 5 Years	5	90	95
Total	515	106	621

- 90 pairs where both patients die within 5 years

- 510 pairs in which both patients survive 5 years

- 16 pairs where Treatment A survived, but Treatment B died

- 5 pairs where Treatment B survived, but Treatment A died

It's easy to see the dependence between the two samples:

$$\text{P}(B\,\text{survives}\,|\,A\,\text{survives}) = \frac{510}{526} = 0.970$$

$$\text{P}(B\,\text{survives}\,|\,A\,\text{dies}) = \frac{5}{95} = 0.053$$

If the samples were independent, then these probabilities should be the same. The dependence makes it clear that a chi-square test is not valid here.

Notice that there are:

$600 = 90 + 510$ pairs where the outcomes agree

$21 = 16 + 5$ pairs where the outcomes disagree.

Definition: A *concordant* pair is a matched pair in which the outcome is the same for both members of the pair.

Definition: A *discordant* pair is a matched pair in which the outcome is different for the members of the pair.

- 600 concordant pairs and 21 discordant pairs.

The number of concordant pairs doesn't tell us anything about the difference between the treatment groups, so let's instead focus on the discordant pairs.

Types of Discordant Pairs:

- Type A: Treatment A has the event, but Treatment B does not.
- Type B: Treatment B has the event, but Treatment A does not.

So for example, if our event is "survive 5 years", there are $n_A = 16$ Type A and $n_B = 5$ Type B discordant events.

Definition: If $P(X = 1) = P(Y = 1)$, then there is said to be *marginal homogeneity.*

In other words, the two treatments are equally effective

Let's compute $P(X = 1) - P(Y = 1)$.

$$P(X = 1) - P(Y = 1) = (p_{11} + p_{12}) - (p_{11} + p_{21}) = p_{12} - p_{21}.$$

Marginal homogeneity is equivalent to there being an equal proportion of Type A and Type B discordant pairs. That is, $p_{12} = p_{21}$. In other words, equality of the off-diagonal elements of the table.

Let $n_D = n_A + n_B =$ total number of discordant pairs and define the p-value to be $p_A = P(\text{discordant pair is Type } A)$.

When the null is true, we expect similar values for n_A and n_B, so $p_A = 1/2$.

- Under the null, each of the n_D observations has an equal chance of belonging to n_A and n_B.
- The number of observations in each cell follows a binomial distribution with $p = 1/2$!

Note: If Treatment A is more effective than Treatment B (more Type A than Type B), then $p > 1/2$. If Treatment B is more effective, then $p < 1/2$.

Here is the summary in terms of a table:

	Y	
X	Y = 1	Y = 2
X = 1	# of Concordant Paris of Type A	# of Discordant Pairs of Type A (n_A)
X = 2	# of Discordant Pairs of Type B (n_B)	# of Concordant Pairs of Type B

McNemar Test for Correlated Proportions – Large Sample

For matched-pairs data with a binary response, a test of marginal homogeneity has null hypothesis that the treatments are equally effective:

$$H_0 : P(X = 1) = P(Y = 1), \quad \text{or equivalently } H_0 : p_A = 1/2$$

Against the two-sided alternative:

$$H_a : p_A \neq 1/2$$

Then, when n_D is large, the binomial distribution can be approximated by a normal distribution with mean $\frac{n_D}{2}$ and variance $n_D\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{n_D}{4}$.

Therefore, our test statistic equals:

$$Z = \frac{n_A - \frac{n_D}{2}}{\sqrt{\frac{n_D}{4}}} = \frac{n_A - n_B}{\sqrt{n_A + n_B}}$$

Adding in a correction for continuity and squaring Z so that it becomes chi-square with 1 df:

$$\chi^2 = \frac{\left(\left|n_A - \frac{n_D}{2}\right| - \frac{1}{2}\right)^2}{\frac{n_D}{4}}$$

Rejection Region: $\chi^2 > \chi^2_{\alpha,1}$

p-value: $\mathrm{P}(\chi^2 > \chi^2_{obs})$.

Note: The normal approximation holds as long as $n_D > 20$, since then we have 10 expected "success" and 10 expected "failure".

A large sample $100(1-\alpha)\%$ confident interval for the true difference of proportions is

$$(\hat{p}_{1+} - \hat{p}_{+1}) \pm Z_{\alpha/2}\,(\mathrm{SE}),$$

Where $\mathrm{SE} = \frac{1}{n}\sqrt{n_A + n_B - \frac{(n_A - n_B)^2}{n}}$. Note that $\hat{p}_{1+} - \hat{p}_{+1} = \frac{n_{12} - n_{21}}{n}$.

Exact Version of McNemar Test

If $n_D < 20$, we can't use the normal approximation to the binomial and instead should use the exact binomial formula. As before:

$$H_0 \,:\, P(X=1) = P(Y=1), \quad \text{or equivalently } H_0 : p_A = 1/2$$

Against the two-sided alternative:

$$H_a \,:\, p_A \neq 1/2$$

Calculate the exact p-value as follows:

(1) $2\sum_{k=0}^{n_A} \binom{n_D}{k}\left(\frac{1}{2}\right)^{n_D}$ when $n_A < n_D/2$.

(2) $2\sum_{k=n_A}^{n_D} \binom{n_D}{k}\left(\frac{1}{2}\right)^{n_D}$ when $n_A > n_D/2$.

(3) $p = 1$ if $n_A = n_D/2$

Note: We multiply the sums in (1) and (2) by 2 because the test is two-sided.