Notations on a 2×2 Table

- Variables X and Y take two values each.
- Denote $P[X = i, Y = j] = \pi_{ij}$, $i = 1, 2$, $j = 1, 2$.
- Denote $\pi_{i+} = \sum_{i=1}^{I} \pi_{ij}$ and $\pi_{+j} = \sum_{i=1}^{I} \pi_{ij}$.
- Denote $P[X = i] = \pi_{i+}$ and $P[Y = j] = \pi_{+j}$.
- Let n_{ij} be the observed frequency for $X = i$, $Y = j$. Then, the MLE are

$$\hat{\pi}_{ij} = p_{ij} = \frac{n_{ij}}{n_{++}}$$

and

$$\hat{\pi}_{i+} = p_{i+} = \frac{n_{i+}}{n_{++}},\ \hat{\pi}_{+j} = p_{+j} = \frac{n_{+j}}{n_{++}}$$

where $n_{i+} = \sum_j n_{ij}$, $n_{+j} = \sum_i n_{ij}$, and $n_{++} = \sum_i \sum_j n_{ij}$.

Table 1: Two-Way Contingency Probability Table with $I = J = 2$

Row	Column 1	Column 2	Total
1	π_{11}	π_{12}	π_{1+}
2	π_{21}	π_{22}	π_{2+}
Total	π_{+1}	π_{+2}	1.0

Table 2: Two-Way Contingency observation Table with $I = J = 2$

	Column		
Row	1	2	Total
1	n_{11}	n_{12}	n_{1+}
2	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n_{++}

Examples

Table 3: Impeachment for Clinton in 1998

	Guilty		
Party	Yes	No	Total
Rep	50	5	55
Dem	0	45	45
Total	50	50	100

Table 4: Example for Oral Contraceptive Practice by Myocardial Infraction.

Oral Contraceptive	Myocardial Infraction		
Practice	Yes	No	Total
Used	23	34	57
Never Used	35	132	167
Total	58	166	224

Table 5: Aspirin and Heart Attack

	Heart Attack	
Aspirin	Yes	No
Yes	104	11037
No	189	11034

- Is the table independent.
- What changes from the first row to the second row.
- Old method: two-sample binomial comparison.

Definition of Odds Ratio

- The relative risk is defined by

$$\frac{\pi_{11}/\pi_{1+}}{\pi_{21}/\pi_{2+}}$$

for the comparison of the first row and the second row, which is estimated by

$$\frac{n_{11}/n_{1+}}{n_{21}/n_{2+}}$$

- The odds is defined by

$$\pi_{11}/\pi_{12}$$

given the first row and

$$\pi_{21}/\pi_{22}$$

given the second row which are estimated by n_{11}/n_{12} and n_{21}/n_{22} respectively.

- The odds ratio is defined by

$$\theta = \frac{\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}}.$$

- It can be estimated by

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{21}n_{12}}.$$

- Interpretation: odds ratio indicates the increase of the risks from the first row to the second row.

Interpretation

- Interpretation: odds ratio indicates the increase of the risks from the first row to the second row.
- If $\theta = 1$, X and Y are independent.

Test and CI

- The asymptotic variance

$$\hat{\sigma}^2_{\log(\hat{\theta})} = V[\log(\hat{\theta})] = \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}$$

- A z-test rejects H_0 of independence if

$$\left|\frac{\log(\hat{\theta})}{\hat{\sigma}_{\log(\hat{\theta})}}\right| > z_{\alpha/2}.$$

- A $(1-\alpha)100\%$ CI for $\log(\theta)$ is

$$\log(\hat{\theta}) \pm z_{\alpha/2}\hat{\sigma}_{\log(\hat{\theta})}.$$

- A $(1-\alpha)100\%$ CI for θ is the corresponding transformation as

$$\hat{\theta} e^{\pm z_{\alpha/2}\hat{\sigma}_{\log(\hat{\theta})}}$$

Two Sample Binomial Method

- Suppose $n_{11} \sim Bin(n_{1+}, p_1)$ and $n_{21} \sim Bin(n_{2+}, p_2)$.
- Then, $\hat{p}_1 = n_{11}/n_{1+}$ and $\hat{p}_2 = n_{21}/n_{2+}$.
- In addition, we also have

$$V(\hat{p}_1) = \frac{p_1(1-p_1)}{n_{1+}}$$

and

$$V(\hat{p}_2) = \frac{p_2(1-p_2)}{n_{2+}}.$$

- This implies

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_{1+}} + \frac{p_2(1-p_2)}{n_{2+}}.$$

- Assume asymptotic normality. A $(1-\alpha)100\%$ CI for $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_{1+}} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_{2+}}}$$

- The method can be used to test $H_0 : p_1 = p_2 = p$. It has two ways.
  - In the first way, we reject H_0 if the CI does not contain 0.
  - In the second way, we estimate p by $\hat{p} = (n_{11} + n_{21})/n_{++}$ and then estimate $V(\hat{p}_1 - \hat{p}_2)$ by

$$\hat{V}(\hat{p}_1 - \hat{p}_2) = \hat{p}(1-\hat{p})(\frac{1}{n_{1+}} + \frac{1}{n_{2+}}).$$

- The z value is defined by the ratio of the estimate of $p_1 - p_2$ and the estimate of its standard error.

Results

Table 6: absolute z-values in the tests for $H_0 : p_1 = p_2$ in the previous examples

Example	Odds Ratio	2-Sample Bin (I)	2-Sample Bin (II)
Impeachment	–	23.45	9.045
Contraceptive	2.84	2.69	2.89
Aspirin	4.87	4.94	4.94