Lecture 4

Nicholas Reich

Independence

Definition:
Two categorical variable are independent iff

$$\pi_{ij} = \pi_{i+}\pi_{+j},\ \forall\ i \in \{1, 2, ..I\} \text{ and } j \in \{1, 2, ..J\}$$

or

$$\mathbb{P}(X = i, Y = j) = \mathbb{P}(X = i)\mathbb{P}(Y = j)$$

Independence implies that the conditional distribution reverts to marginal distribution

$$\pi_{j|i} = \frac{\pi_{ij}}{\pi_{i+}} = \frac{\pi_{i+}\pi_{j+}}{\pi_{i+}} = \pi_{j+}$$

or under the independence assumption

$$\mathbb{P}(Y = j|\ X = i) = \mathbb{P}(Y = j)$$

Testing for independence (Two-way contigency table)

- Under $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}, \forall\ i, j$, the expected cell counts are

$$\mu_{ij} = n\pi_{i+}\pi_{+j}$$

- Usually π_{i+} and π_{+j} are unknown. Their MLEs are

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

- Estimated expected cell counts are

$$\hat{\mu}_{ij} = n\hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}n_{+j}}{n}$$

- Pearson χ^2 statistic:

$$X^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} = \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\mu^2}$$

- $\hat{\mu}_{ij}$ requires estimating π_{i+} and π_{+j} which have degrees of freedom $I-1$ and $J-1$, respectively. Notice the constraints $\sum_i \pi_{i+} = \sum_j \pi_{+j} = 1$
- The degrees of freedom is

$$(IJ = 1) - (I-1) - (J-1) = (I-1)(J-1)$$

- X^2 is asymptotically $\chi^2_{(I-1)(J-1)}$
- It is helpful to look at the residuals

$$\{\frac{(O-E)^2}{E}\}$$

The residuals can give useful information about where model is well or not

Measure of Diagnostic Tests

	Diagnosis	
Disease Status	$+$	-
D	π_{11}	π_{12}
$\overline{D}$	π_{21}	π_{22}

- Sensitivity: $\mathbb{P}(+|D) = \frac{\pi_{11}}{\pi_{1+}}$
- Specificity: $\mathbb{P}(-|\overline{D}) = \frac{\pi_{22}}{\pi_{2+}}$
- An ideal diagnostic test has high Sensitivity, Specificity

Example:

Disease Status	Diagnosis +	Diagnosis -
D	0.86	0.14
$\overline{D}$	0.12	0.88

- Sensitivity $= 0.86$
- Specificity $= 0.88$

However, from the clinical point, sensitivity and specificity do not provide useful information. So we introduce Positive predictive value and Negative predictive value

- Positive predictive value (PPV) $= \mathbb{P}(D|+) = \frac{\pi_{11}}{\pi_{+1}}$
- Negative predictive value (NPV) $= \mathbb{P}(\overline{D}|-) = \frac{\pi_{22}}{\pi_{+2}}$
- Relationship between PPV and sensitivity:

$$
\begin{aligned}
\text{PPV} = \mathbb{P}(D|+) &= \frac{\mathbb{P}(D \cap +)}{\mathbb{P}(+)} \\
&= \frac{\mathbb{P}(+|D)\mathbb{P}(D)}{\mathbb{P}(+|D)\mathbb{P}(D) + \mathbb{P}(+|\overline{D})\mathbb{P}(\overline{D})} \\
&= \frac{\mathbb{P}(+|D)\mathbb{P}(D)}{\mathbb{P}(+|D)\mathbb{P}(D) + (1 - \mathbb{P}(-|\overline{D}))\mathbb{P}(\overline{D})} \\
&= \frac{\text{Sensitivity} \times \text{Prevalence}}{\text{Sensitivity} \times \text{Prevalence} + (1 - \text{Specificity}) \times (1 - \text{Prevalence})}
\end{aligned}
$$

The same example:

	Diagnosis	
Disease Status	$+$	-
D	0.86	0.14
$\overline{D}$	0.12	0.88

- If the the prevalence $\mathbb{P}(D) = 0.02$
- PPV $= \frac{0.86 \times 0.02}{0.86 \times 0.02 + 0.12 \times 0.98} \approx 13\%$
- Notice:

$$\text{PPV} \neq \frac{\pi_{11}}{\pi_{11} + \pi_{21}}$$

- This is only true when $\frac{n_1}{n_1 + n_2}$ equals the disease prevalence

Comparing two groups

We first consider 2×2 tables. Suppose that the response variable Y has two categories: success and failure. The explanatory variable X has two categories, group 1 and group 2, with fixed sample sizes in each group.

	Response Y		
Explanatory X	Success	Failure	Row Total
group 1	$n_{11} = x_1$	$n_{12} = n_1 - x_1$	n_1
group 2	$n_{21} = x_2$	$n_{22} = n_1 - x_2$	n_2

The goal is to compare the probability of an outcome (success) of Y across the two levels of X. Assume: $X_1 \sim bin(n_1, \pi_1), X_2 \sim bin(n_2, \pi_2)$

- difference of proportions
- relative risk
- odds ratio

Difference of Proportions

	Response Y		
Explanatory X	Success	Failure	Row Total
group 1	$n_{11} = x_1$	$n_{12} = n_1 - x_1$	n_1
group 2	$n_{21} = x_2$	$n_{22} = n_1 - x_2$	n_2

- The difference of proportions of successes is: $\pi_1 - \pi_2$
- Comparison on failures is equivalent to comparison on successes:

$$(1 - \pi_1) - (1 - \pi_2) = \pi_2 - \pi_1$$

- Difference of proportions takes values in $[-1, 1]$

- The estimate of $\pi_1 - \pi_2$ is $\hat{\pi}_1 - \hat{\pi}_2 = \frac{n_{11}}{n_1} - \frac{n_{21}}{n_2}$
- the estimate of the asymptotic standard error:

$$\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2) = [\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} - \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}]^{1/2}$$

- The statistic for testing $H_0 : \pi_1 = \pi_2$ vs. $H_a : \pi_1 \neq \pi_2$

$$Z = (\hat{\pi}_1 - \hat{\pi}_2)/\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2)$$

  which follows a standard normal distribution (normal + normal = normal)
- The CI is given by

$$(\hat{\pi}_1 - \hat{\pi}_2) \pm Z_{\alpha/2}\hat{\sigma}(\hat{\pi}_1 - \hat{\pi}_2)$$

Relative Risk

- Definition

$$r = \pi_1/\pi_2$$

- Motivation: The difference between $\pi_1 = 0.010$ and $\pi_2 = 0.001$ is more noteworthy than the difference between $\pi_1 = 0.410$ and $\pi_2 = 0.401$. The "relative risk" (0.010/0.001=10, 0.410/0.401=1.02) is more informative than "difference of proportions" (0.009 for both).
- The estimate of r is

$$\hat{r} = \hat{\pi}_1/\hat{\pi}_2$$

- The estimator converges to normality faster on the log scale.
- The estimator of $\log r$ is

$$\log \hat{r} = \log \hat{\pi}_1 - \log \hat{\pi}_2$$

The asymptotic standard error of $\log \hat{r}$

$$\hat{\sigma}(\log \hat{r}) = (\frac{1-\pi_1}{\pi_1 n_1} + \frac{1-\pi_2}{\pi_2 n_2})^{1/2}$$

- Delta method: If $\sqrt{n}(\hat{\beta} - \beta_0) \to N(0, \sigma^2)$, then

$$\sqrt{n}(f(\hat{\beta}) - f(\beta_0)) \to N(0, [f'(\beta_0)]^2\sigma^2)$$

for any function f satisfying the condition that $f'(\beta)$ exists
- Here $\beta = \pi_1$ or π_2 and $f(\beta) = \log(\pi_1)$ or $\log(\pi_1)$

- The CI for $\log \hat{r}$ is

$$[\log \hat{r} - Z_{1-\alpha/2}\hat{\sigma}(\log \hat{r}), \log \hat{r} + Z_{1-\alpha/2}\hat{\sigma}(\log \hat{r})]$$

- The CI for $\hat{r}$ is

$$[\exp\{\log \hat{r} - Z_{1-\alpha/2}\hat{\sigma}(\log \hat{r})\}, \exp\{\log \hat{r} + Z_{1-\alpha/2}\hat{\sigma}(\log \hat{r})\}]$$

Odds Ratio

- Odds in group 1:
$$\phi_1 = \frac{\pi_1}{(1 - \pi_1)}$$
- Interpretation: $\phi 1 = 3$ means a success is three times as likely as a failure in group 1
- Odds ratio:
$$\theta = \frac{\phi_1}{\phi_2} = \frac{\pi_1 \; /(1 - \pi_1)}{\pi_2 \; /(1 - \pi_2)} \sim \chi^2$$
- Interpretation: $\theta = 4$ means the odds of success in group 1 are four times the odds of success in group 2

- The estimate is
$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}}$$
- $\log(\hat{\theta})$ converge to normality much faster than $\hat{\theta}$
- An estimate of asymptotic standard error for $\log(\hat{\theta})$ is
$$\hat{\sigma}(\log\hat{\theta}) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

This formula can be derived using the Delta method

Recall $\log \hat{\theta} = \log(\hat{\pi}_1) - \log(1 - \hat{\pi}_1) - \log(\hat{\pi}_2) + \log(1 - \hat{\pi}_2)$

First, $f(\beta) = \log(\hat{\pi}_1) - \log(1 - \hat{\pi}_1)$

$$\sigma = \frac{\pi_1(1-\pi_1)}{n_1}, \quad f'(\beta) = \frac{1}{\pi_1} + \frac{1}{1-\pi_1}$$

$$[f'(\beta)]^2\sigma^2 = \frac{1}{n_1\pi_1} + \frac{1}{n_1(1-\pi_1)}$$

The estimate is $\frac{1}{n_{11}} + \frac{1}{n_{12}}$

Similar, when $f(\beta) = \log(\hat{\pi}_2) - \log(1 - \hat{\pi}_2)$

- The Wald CI for $\log \hat{\theta}$ is

$$\log \hat{\theta} \pm Z_{\alpha/2} \hat{\sigma}(\log \hat{\theta})$$

- Exponentiation the endpoints provides a confidence interval for $\hat{\theta}$

Relationship between Odds Ratio and Relative Risk

- A large relative risk does not imply large odds ratio
- From the definitions of relative risk and odds ratio, we have

$$\theta = \frac{\pi_1}{\pi_2} \frac{1-\pi_2}{1-\pi_1} = \text{relative risk} \times \frac{1-\pi_2}{1-\pi_1}$$

- When probabilities π_1 and π_2 (the risk in each row group)are both very small, then the second ratio above ≈ 1. Thus

$$\text{odds ratio} \approx \text{relative risk}$$

- This means when relative risk is not directly estimable, e.g., in case-control studies, and the probabilities π_1 and π_2 are both very small, the relative risk can be approximated by the odds ratio.

Case-Control Studies and Odds Ratio

Consider the case-control study of lung cancer:

	Lung Cancer	
Smoker	Cases	Controls
Yes	688	650
No	21	59
Total	709	709

- People are recruited based on lung cancer status, therefore $\mathbb{P}(Y = j)$ is known. However $\mathbb{P}(X = i)$ is unknown
- Conditional probabilities $\mathbb{P}(X = i | Y = j)$ can be estimated
- Conditional probabilities $\mathbb{P}(Y = j | X = i)$ cannot be estimated
- Relative risk and difference of proportions cannot be estimated

- Odds can be estimated:

$$\begin{aligned}
\text{Odds of lung cancer among smoker} &= \frac{\mathbb{P}(\text{Case}|\text{Smoker})}{\mathbb{P}(\text{Control}|\text{Smoker})} \\
&= \frac{\mathbb{P}(\text{Case} \cap \text{Smoker})\mathbb{P}(Smoker)}{\mathbb{P}(\text{Control} \cap \text{Smoker})\mathbb{P}(Smoker)} \\
&= \frac{\mathbb{P}(\text{Case} \cap \text{Smoker})}{\mathbb{P}(\text{Control} \cap \text{Smoker})} \\
&= 688/650 = 1.06
\end{aligned}$$

- Odds is irrelevant to the probability of being a smoker
- Odds ratio can also be estimated:

$$\theta = \frac{\mathbb{P}(X=1|Y=1)\mathbb{P}(X=2|Y=2)}{\mathbb{P}(X=1|Y=2)\mathbb{P}(X=2|Y=1)} = 2.97$$

Supplementary: Review of the Delta Method

The Delta method builds upon the Central Limit Theorem to allow us to examine the convergence of the distribution of a function g of a random variable X.

It is not too complicated to derive the Delta method in the univariate case. We need to use Slutsky's Theorem along the way; it will be helpful to first review ideas of convergence in order to better understand where Slutsky's Theorem fits into the derivation.

Delta Method: Convergence of Random Variables

Consider a sequence of random variables $X_1, X_2, \ldots, X_n$, where the distribution of X_i may be a function of of i.

Let $F_n(x)$ be the CDF for X_n and $F(x)$ be the CDF for X. It is said that X_n **converges in distribution** to X, written $X_n \rightarrow^d X$, if $\lim_{n\rightarrow\infty}[F_n(x) - F(x)] = 0$ for all x where $F(x)$ is continuous.

It is said that X_n **converges in probability** to X, written $X_n \rightarrow^p X$ if $\lim_{n\rightarrow\infty}[X_n - X] = 0$.

Note that if $X_n \rightarrow^p X$, then $F_n(x) \rightarrow^d F(x)$, since $F_n(x) = P(X_n \leq x)$ and $F(x) = P(X \leq x)$. (This is not a proof, but an intuition. The Wikipedia article on convergence has a nice proof.)

Delta Method: Slutsky's Theorem and First-Order Taylor Approximation

Slutsky's Theorem tells us that if some random variable X_n converges in distribution to X and some random variable Y_n converges in probability to c, then $X_n + Y_n$ converges in distribution to $X + c$ and $X_n Y_n$ converges in distribution to cX.

Recall that the **first-order Taylor approximation** of a function g centered at u can be written as $g(x) = g'(u)(x-u) + g(u) + R(x)$, where $R(x) = \sum_{i=2}^{n} g^{(i)}(u)\frac{(x-u)^i}{i!}$.

Delta Method: Hand-wave-y Derivation

Suppose we know that $\sqrt{n}(X_n - \theta) \to^d N(0, \sigma^2)$ and we are interested in the behavior of some function $g(X_n)$ as $n \to \infty$. If $g'(\theta)$ exists and is not zero, we can write $g(X_n) \approx g'(\theta)(X_n - \theta) + g(\theta)$ using Taylor's approximation:

$$g(X_n) = g'(\theta)(X_n - \theta) + g(\theta) + \sum_{i=2}^{\infty} g^{(i)}(\theta)\frac{(X_n - \theta)^i}{i!}$$

Delta Method: Hand-wave-y Derivation

Some manipulation gives:

$$\sqrt{n}g(X_n) = \sqrt{n} * g'(\theta)(X_n - \theta) + \sqrt{n} * g(\theta) + \sqrt{n} * \sum_{i=2}^{\infty} g^{(i)}(\theta)\frac{(X_n - \theta)^i}{i!}$$

or, using the definition of R from the previous slide,

$$\sqrt{n}(g(X_n) - g(\theta)) = \sqrt{n} * g'(\theta)(X_n - \theta) + \sqrt{n} * R(X_n)$$

Delta Method: Hand-wave-y Derivation

Since $g'(\theta)$ is a constant with respect to n and $\sqrt{n}(X_n - \theta) \rightarrow^d N(0, \sigma^2)$, we have

$$g'(\theta)\sqrt{n}(X_n - \theta) \rightarrow^d N(0, \sigma^2(g'(\theta))^2)$$

.

It can be shown that the remainder term $R(X_n) \rightarrow^p 0$ (see the Stephens link from McGill below for a proof).

We now have the necessary setup to apply Slutsky's Theorem, and we can conclude that

$$\sqrt{n}(g(X_n) - g(\theta)) \rightarrow^d N(0, \sigma^2(g'(\theta))^2)$$

.

Delta Method: References

- http://www.stat.rice.edu/~dobelman/notes_papers/math/TaylorAppDeltaMethod.pdf
- https://en.wikipedia.org/wiki/Convergence_of_random_variables
- http://www.stat.cmu.edu/~larry/=stat325.01/chapter5.pdf
- https://en.wikipedia.org/wiki/Slutsky%27s_theorem
- http://www.math.mcgill.ca/dstephens/OldCourses/556-2007/Math556-19-AsympNormal.pdf