Some recommended supplementary texts:

- Christensen, R. (1997). *Log-linear Models and Logistic Regression (2^{nd} Edition)*. Springer-Verlag.
- Agresti, A. (2007). *An Introduction to Categorial Data Analysis*, 2^{nd} Edition. John Wiley & Sons Canada.
- Le, C. (1998). *Applied Categorical Data Analysis*. John Wiley & Sons.
- Stokes, M.E., Davis, C.S., and Koch, G.C. (2000). *Categorical Data Analysis Using the SAS System (2^{nd} Edition)*. SAS Publishing.
- Thompson, L.A. (2006). *S-Plus (and R) Manual to Accompany Agresti's "Categorical Data Analysis (2002) 2^{nd} Edition"*. Available at https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf.

1.4.1 Tests for a binomial probability π

Let $Y \sim \text{bin}(n, \pi)$.

The likelihood is

$$\mathcal{L}(\pi) = \begin{pmatrix} n \\ y \end{pmatrix} \pi^y (1-\pi)^{n-y}$$

and the log-likelihood is

$$L(\pi) = \log \begin{pmatrix} n \\ y \end{pmatrix} + y \log \pi + (n-y) \log(1-\pi).$$

So

$$L'(\pi) = \frac{y}{\pi} - \frac{n-y}{1-\pi}.$$

Solving for π gives the MLE $\hat{\pi} = y/n$, the number of successes out of the total number of trials.

Taking the 2^{nd} derivative of $L(\pi)$ gives

$$L''(\pi) = -\frac{y}{\pi^2} - \frac{n-y}{(1-\pi)^2},$$

and so

$$-E(L''(\pi)) = E\left(\frac{Y}{\pi^2} + \frac{n-Y}{(1-\pi)^2}\right) = \frac{n\pi}{\pi^2} + \frac{n-n\pi}{(1-\pi)^2} = \frac{n}{\pi(1-\pi)}.$$

The large sample result is then

$$\hat{\pi} = \frac{Y}{n} \stackrel{\bullet}{\sim} N\left(\pi, \frac{\pi(1-\pi)}{n}\right).$$

See Section 1.3.2.

Let's consider $H_0 : \pi = \pi_0$ where π_0 is fixed and known (e.g. $H_0 : \pi = 0.5$.)

The **Wald** test plugs in the MLE $\hat{\pi} = y/n$ for the unknown π in the large sample variance:

$$\hat{\pi} = \frac{Y}{n} \overset{\bullet}{\sim} N\left(\pi, \frac{\hat{\pi}(1-\hat{\pi})}{n}\right).$$

Recall that $\mathrm{se}(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$.

So then

$$Z_W = \frac{\hat{\pi} - \pi_0}{\mathrm{se}(\hat{\pi})} = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} \overset{\bullet}{\sim} N(0,1)$$

when H_0 is true. Squaring, $W = Z_W^2 \overset{\bullet}{\sim} \chi_1^2$.

The **Score** test plugs the null value into the large sample variance

$$\hat{\pi} = \frac{Y}{n} \overset{\bullet}{\sim} N\left(\pi, \frac{\pi_0(1-\pi_0)}{n}\right).$$

So then

$$Z_S = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \overset{\bullet}{\sim} N(0,1)$$

when H_0 is true. Squaring, $S = Z_S^2 \overset{\bullet}{\sim} \chi_1^2$.

Evaluating the log-likelihood at the unconstrained MLE gives

$$L_1 = L(\hat{\pi}) = \log \begin{pmatrix} n \\ y \end{pmatrix} + y \log \hat{\pi} + (n - y) \log(1 - \hat{\pi}).$$

Under the constraint $H_0 : \pi = \pi_0$, the log-likelihood is simply

$$L_0 = L(\pi_0) = \log \begin{pmatrix} n \\ y \end{pmatrix} + y \log \pi_0 + (n - y) \log(1 - \pi_0),$$

(there are no parameters left to maximize the constrained likelihood under!) and so the **LRT**, plugging Y in for y,

$$L = -2(L_0 - L_1) = 2 \left(Y \log \frac{\hat{\pi}}{\pi_0} + (n - Y) \log \frac{1 - \hat{\pi}}{1 - \pi_0} \right) \overset{\bullet}{\sim} \chi^2_1$$

when H_0 is true.

In all three cases, an approximate $\alpha = 0.05$ significance test of $H_0 : \pi = \pi_0$ is carried out by computing W, S, or L and rejecting if the test statistic is larger than the quantile corresponding to 0.05 right tail probability from a χ^2_1 distribution, i.e. larger than $\chi^2_1(0.05) = 3.84$.

Confidence intervals are obtained by inverting the test statistics; read Section 1.4.2.

1.4.3 where for art thou, vegetarians?

Out of $n = 25$ students, $y = 0$ were vegetarians. Assuming binomial data, the 95% CIs found by inverting the Wald, score, and LRT tests are

Wald	$(0, 0)$
score	$(0, 0.133)$
LRT	$(0, 0.074)$

The Wald interval is particularly troublesome. Why the difference? for small or large (true, unknown) π the normal approximation for the distribution of $\hat{\pi}$ is pretty bad in small samples.

A solution is to consider the *exact* sampling distribution of $\hat{\pi}$ rather than a normal approximation.

1.4.4 Exact inference

An exact test proceeds as follows.

Under $H_0 : \pi = \pi_0$ we know $Y \sim \text{bin}(n, \pi_0)$. Values of $\hat{\pi}$ far away from π_0, or equivalently, values of Y far away from $n\pi_0$, indicate that $H_0 : \pi = \pi_0$ is unlikely.

Say we reject H_0 if $Y < a$ or $Y > b$ where $0 \le a < b \le n$. Then we set the type I error at α by requiring $P(\text{reject } H_0 | H_0 \text{ is true}) = \alpha$. That is,

$$P(Y < a | \pi = \pi_0) = \frac{\alpha}{2} \text{ and } P(Y > b | \pi = \pi_0) = \frac{\alpha}{2}.$$

However, since Y is discrete, the best we can do is *bounding* the type I error by choosing a as large as possible such that

$$P(Y < a|\pi = \pi_0) = \sum_{i=0}^{a-1} \left(\begin{array}{c} n \\ i \end{array} \right) \pi_0^i (1 - \pi_0)^{n-i} < \frac{\alpha}{2},$$

and b as small as possible such that

$$P(Y > b|\pi = \pi_0) = \sum_{i=b+1}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) \pi_0^i (1 - \pi_0)^{n-i} < \frac{\alpha}{2}.$$

For example, when $n = 20$, $H_0 : \pi = 0.25$, and $\alpha = 0.05$ we have

$$P(Y < 2|\pi = 0.25) = 0.024 \text{ and } P(Y < 3|\pi = 0.25) = 0.091,$$

so $a = 2$. Also,

$$P(Y > 9|\pi = 0.25) = 0.014 \text{ and } P(Y > 8|\pi = 0.25) = 0.041,$$

so $b = 9$. We reject $H_0 : \pi = 0.25$ when $Y < 2$ or $Y > 9$. The type I error is bounded: $\alpha = P(\text{reject } H_0|H_0 \text{ is true}) \leq 0.05$, but in fact this is conservative, $P(\text{reject } H_0|H_0 \text{ is true}) = 0.024 + 0.014 = 0.038$.

Nonetheless, this type of exact test can be inverted to obtain exact confidence intervals for π. However, the actual coverage probability is *at least* as large as $1 - \alpha$, but typically more. So the procedure errs on the side of being conservative (CI's are bigger than they need to be). Read Section 1.4.4 (pp. 18-20) for more discussion if interested.

To obtain the 95% CI from inverting the score test, and from inverting the exact (Clopper-Pearson) test:

```
> out1 <- prop.test(x=0,n=25,conf.level=0.95,correct=F)
> out1$conf.int
[1] 0.0000000 0.1331923
attr(,"conf.level") [1] 0.95
> out2 <- binom.test(x=0,n=25,conf.level=0.95)
> out2$conf.int
[1] 0.0000000 0.1371852
attr(,"conf.level") [1] 0.95
```

Comment: I would recommend, wherever possible, the use of exact tests instead of approximate tests.

1.5 inference for multinomial parameters

Assume $\mathbf{n} \sim \text{mult}(n, \boldsymbol{\pi})$ where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_c)$ and $\mathbf{n} = (n_1, \ldots, n_c)$.

1.5.1 MLE estimation

A bit of calculus (p. 21) yields the MLE

$$\hat{\boldsymbol{\pi}} = \left(\frac{n_1}{n}, \frac{n_2}{n}, \ldots, \frac{n_c}{n}\right).$$

The sample proportion of trials falling into category j is the MLE of π_j for all $j = 1, \ldots, c$ categories (intuitive!)

1.5.2 Pearson statistic for testing $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$

Old test; motivated by roulette, Karl Pearson introduced in 1900. Example of a score test.

When $H_0 : (\pi_1, \ldots, \pi_c) = (\pi_{01}, \ldots, \pi_{0c})$ is true then $E(n_j) = n\pi_{0j}$ (Section 1.2.2). Pearson's test statistic is

$$X^2 = \sum_{j=1}^{c} \frac{(n_j - n\pi_{0j})^2}{n\pi_{0j}}.$$

When $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$ is true n_j will be close to what's expected $n\pi_{0j}$ and the statistic will be small. When $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$ is false the statistic will be large (for fixed sample size n). In large samples $X^2 \stackrel{\bullet}{\sim} \chi^2_{c-1}$ (Section 1.5.4).

1.5.3 Testing Mendel's theories

We'll briefly discuss this in class.

1.5.5 Likelihood ratio χ^2

The LRT statistic for $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0$ is

$$G^2 = -2\left[\log\prod_{j=1}^{c}(\pi_{0j})^{n_j} - \log\prod_{j=1}^{c}(n_j/n)^{n_j}\right] = 2\sum_{j=1}^{c} n_j \log(n_j/n\pi_{j0}).$$

What does this statistic equal when $\hat{\pi}_j = \frac{n_j}{n} = \pi_{0j}$ for $j = 1, \ldots, c$?

Pearson's X^2 overall has better properties & can work well when n/c is as small as one if the elements of $\boldsymbol{\pi}_0$ are not highly dissimilar (close to 1 or 0). See discussion, bottom of p. 24. Note that an exact test is also possible for this hypothesis using the multinomial distribution.

1.5.6 Testing with estimated expected frequencies

Basic idea: extend Pearson's method to test a model $H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}_0(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ are parameters of a smaller-dimensional model. Once the model is fit through ML yielding $\hat{\boldsymbol{\theta}}$, the expected frequencies are $n\pi_{j0}(\hat{\boldsymbol{\theta}})$ to be used in (1.15). Construct X^2 as usual except $X^2 \overset{\bullet}{\sim} \chi^2_{c-1-p}$ where p is the dimension of $\boldsymbol{\theta}$.

Example: $n = 156$ calves were classified as one of "no pneumonia", "pneumonia, no secondary infection," or "pneumonia then secondary infection." We treat the data $\mathbf{n} = (n_1, n_2, n_3)$ as multinomial with probabilities $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$.

It is of interest to test that the probability of a calf getting pneumonia is equal to the conditional probability of a calf getting a secondary infection after getting pneumonia:

$$H_0 : \pi_2 + \pi_3 = \frac{\pi_3}{\pi_2 + \pi_3}.$$

This hypothesis restricts the parameter space from 2 dimensions $\boldsymbol{\beta} = (\pi_1, \pi_2)$ to just one. Let $\pi = \pi_2 + \pi_3$. Then under the constrained model $\pi_3 = \pi^2$. Also, we must have $\pi_1 = 1 - (\pi_2 + \pi_3) = 1 - \pi$. Finally, $\pi_2 = \pi(1 - \pi)$ (verify this!)

So $\boldsymbol{\theta} = \pi$ here and $p = 1$.

$\mathcal{L}(\pi) \propto (1-\pi)^{n_1}(\pi-\pi^2)^{n_2}(\pi^2)^{n_3}$ and calculus (p. 26) leads to the MLE

$$\hat{\pi} = \frac{2n_3 + n_2}{2n_3 + 2n_3 + n_1}.$$

For the data $\mathbf{n} = (63, 63, 30)$, $\hat{\pi} = 0.494$, the estimated probability of pneumonia under the model. Then

$$\begin{aligned} X^2 &= \frac{[63 - 156(1-0.494)]^2}{156(1-0.494)} + \\ &\quad \frac{[63 - 156(0.494 - 0.494^2)]^2}{156(0.494 - 0.494^2)} + \frac{[30 - 156(0.494^2)]^2}{156(0.494^2)} = 19.7. \end{aligned}$$

The p-value is $P(\chi^2_1 > 19.7) = 0.00001$.

An alternative test is an approximate Wald test using the delta method and large-sample normality of $(\hat{\pi}_2, \hat{\pi}_3)$.