II. Inference on a Single Categorical Variable:

A: 1.1

A. Course investigates description and analysis of variables that are categorical
1. i.e., variables that name categories
2. Ex.,
   a. Healthy/Sick
   b. Young/Old
   c. Strongly Agree/Agree/Indifferent/Disagree/Strongly Disagree
   d. Alabama/Alaska/.../Wyoming
   A: 1.1.2
3. Categories might be ordered or unordered
   a. Careful: can get fooled
      i. Ex., zip codes
   b. Key question will you (and how will you) use information about ordering
   A: 1.1.1
4. May be used either as explanatory (independent) or response (dependent)
   A: 1.2

II. Inference on a Single Categorical Variable:

A. Distributional Assumptions:
   A: 3.3

1. Poisson
   a. Suppose that there are \( K \) categories (numbered from zero)
   b. Let \( X_k \) be the (random) number in category \( k \)
   c. Let \( x_k \) be a potential value for \( X_k \)

2. mixing distribution does not depend on unknown parameter
3. Then perform inference based on experiment we see
   A: 1.3
C. Special Case: \( K = 2 \).
1. \( X_1|X_+ \sim N(X_+ \pi_1, X_+ \pi_1 \pi_0) \).
2. \( p \)-value
   \[ 2 \times \min (P[X_0 \geq x_0|x_+ = X_+], P[X_0 \leq x_0|x_+ = X_+]) \]
   a. If \( x_0 = 0 \), then \( P[X_0 \geq x_0] = 1 \): Never determines \( p \)-value
   b. If \( x_0 = x_+ \), then \( P[X_0 \leq x_0] = 1 \): Never determines \( p \)-value
   c. Using normal approximation,
   \[ \approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+}{\sqrt{\pi_0 \pi_1 x_+}} \right), \Phi \left( -\frac{x_0 + \pi_0 x_+}{\sqrt{\pi_0 \pi_1 x_+}} \right) \right) \]
   d. To properly account for probability at \( x_0 \), add \( \pm \frac{1}{2} \) to numerator to make absolute value smaller. See Figure 1.
   \[ \approx 2 \min \left( \Phi \left( \frac{x_0 - \pi_0 x_+ + \frac{1}{2}}{\sqrt{\pi_0 \pi_1 x_+}} \right), \Phi \left( \frac{-x_0 + \pi_0 x_+ + \frac{1}{2}}{\sqrt{\pi_0 \pi_1 x_+}} \right) \right) \] (1)

3. Get CI for \( \pi \) using
   a. Exact confidence bound
   i. Lower bound \( \pi_L \) satisfies \( P_\pi [X_0 \geq x_0] = 0.025 \)
   ii. Upper bound \( \pi_U \) satisfies \( P_\pi [X_0 \leq x_0] = 0.025 \)
   iii. Vertical line has probability .95 for any value of parameter

d. Number of individuals in each category is given by a Poisson distribution with intensity \( \lambda_k \)
   i. \( P[X_k = x_k] = \exp(-\lambda_k) \lambda_k^{x_k}/x_k! \)
e. Assume entries are independent
f. Well known result: \( X_i \sim \mathcal{P}(\lambda_i), X_j \sim \mathcal{P}(\lambda_j) \Rightarrow X_i + X_j \sim \mathcal{P}(\lambda_i + \lambda_j) \)
i. Obvious extension to more summands
   A: 1.2.1-1.2.2

2. Multinomial:
   a. Notation: Quantities with a + for the subscript indicate summation over that subscript.
   b. Conditional on observing \( X_+ = x_+ \),
   \[ P[X_0 = x_0, \ldots, X_{K-1} = x_{K-1}|X_+ = x_+] = \prod_{k=0}^{K-1} \frac{\exp(-\lambda_k) \lambda_k^{x_k}/x_k!}{\exp(-\lambda_+) \lambda_+^{x_+}/x_+!} = \prod_{k=0}^{K-1} \frac{\pi_k x_k^{x_k}}{x_0! \cdots x_{K-1}!} \]
   for \( \pi_k = \lambda_k/\lambda_+ \)
   c. This distribution depends on the \( \lambda_0, \ldots, \lambda_{k-1} \) only through the ratios \( \lambda_k/\lambda_+ \).
d. Equivalent to taking each Poisson arrives to any category, and dividing them among bins according to probabilities \( \pi_k \).
e. Special Case: \( K = 2 \),
   \[ X_0, X_+ \sim \mathcal{N}(X_+ \pi_0, X_+ \pi_1 \pi_0) \).
B. Conditionality principal: If
1. data arises from random mixture of experiments
   a. Here indexed by \( X_+ \)

Fig. 1: Demonstration of Continuity Correction

Area approximated without cont. corr.

Additional area represented by cont. corr.

Potential Binomial Value

iv. Hence horizontal line has same coverage
v. Lower confidence bound is generated by upper quantile and \textit{vice versa}
vi. Figure 2 shows construction for a generic family
of continuous distributions in which quantile is non-decreasing in the parameter.

vii. Figure 3 shows the construction for the binomial family. Note the role of discreteness. [Mark b R]

Fig. 2: Confidence Interval Example for Continuous Distributions

\[ \beta(\theta, 4) \]  

\[ \theta \]

\[ \text{Example} \]

\[ \text{Table 1: .95 Confidence Intervals for } X_+ = 10 \]

\[ \begin{array}{cccccc}
0 & 1 & 2 & 3 & 4 & 5 \\
\text{Normal Lower} & 0.000 & 0.086 & 0.048 & 0.016 & 0.096 & 0.190 \\
\text{Normal Upper} & 0.000 & 0.2589 & 0.4479 & 0.584 & 0.704 & 0.810 \\
\text{Exact Lower} & 0.000 & 0.0025 & 0.0252 & 0.067 & 0.122 & 0.187 \\
\text{Exact Upper} & 0.309 & 0.4450 & 0.5561 & 0.653 & 0.738 & 0.813 \\
\end{array} \]

b. Can be expressed in terms of \( F \) distribution upper quantile

\[ \frac{X_0}{(X_0 + X_1 + 1)F_{n/2}(2X_0 + 2, 2X_1)} \]

\[ X_0 + X_1 + 1 \]

(Clopper and Pearson, 1934).

i. Let \( G_{a,b}(y) = \int_0^y x^{a-1}(1-x)^b dx \)

\[ a > 0, \quad b > 0 \quad \text{for } a \neq -1. \]

ii. Suppose \( a > 0 \). Then

\[ G_{a+1,b}(y) = y^a(1-y)^b(a+b)! \]

\[ \frac{a+b}{(a-1)!(b-1)!} 

+ \int_0^y x^{a-1}(1-x)^b dx \frac{(a+b)!}{(a-1)!(b!)!} 

= - \left( \frac{a+b}{b} \right) y^a(1-y)^b + G_{a,b+1}(y). \]

iii. Then

\[ G_{n-b,b+1}(y) - G_{n-b+1,b}(y) = \binom{n}{b} y^{n-b}(1-y)^b. \]

iv. Also,

\[ G_{n,1}(y) = n \int_0^y x^{n-1} dx = y^n. \]

v. By induction,

\[ G_{n-b,b+1}(y) = \sum_{a=0}^b \binom{n}{a} y^{n-a}(1-y)^a. \]

vi. Finish using relationship between Beta and F distributions.

c. Normal approx. \( \pi \in X_1/X_+ \pm 1.96 \sqrt{\frac{X_0X_1}{(X_1+X_0)^2}} \)

i. Problem if \( X_0 = 0 \)

ii. Less obvious problem for small \( X_1 + X_0 \)

d. Better Normal approx.: See homework.

A: 3.5.2

4. Testing \( H_0 : \beta = \beta^0 \) vs \( H_A : \beta \neq \beta^0 \) using the likelihood function

a. \( \beta \approx N(\beta, \ell''(\beta)^{-1}) \)

i. \( \hat{\beta} \) approximately unbiased.

ii. \( (\hat{\beta} - \beta)^{\top} [\ell''(\hat{\beta})] (\hat{\beta} - \beta) \) is Wald test statistic

b. Wald Test

i. Let \( \hat{\beta} = \left( \hat{\psi}, \hat{\phi} \right) \) maximize \( \ell \) over \( \beta \)

ii. Maximize \( \ell \) with \( \psi = 0 \) and \( \phi \) unconstrained.

iii. Compare 2× difference to \( \chi^2 \)

• d.f. is length of \( \psi \)

c. Wald Test

i. Let \( \beta = \left( \hat{\psi}, \hat{\phi} \right) \) maximize \( \ell \) over \( \beta \)

b. Can be expressed in terms of \( F \) distribution upper quantile

\[ \frac{X_0}{(X_0 + X_1 + 1)F_{n/2}(2X_0 + 2, 2X_1)} \]

(Clopper and Pearson, 1934).

i. Let \( G_{a,b}(y) = \int_0^y x^{a-1}(1-x)^b dx \)

\[ a > 0, \quad b > 0 \quad \text{for } a \neq -1. \]

ii. Suppose \( a > 0 \). Then

\[ G_{a+1,b}(y) = y^a(1-y)^b(a+b)! \]

\[ \frac{a+b}{(a-1)!(b-1)!} 

+ \int_0^y x^{a-1}(1-x)^b dx \frac{(a+b)!}{(a-1)!(b!)!} 

= - \left( \frac{a+b}{b} \right) y^a(1-y)^b + G_{a,b+1}(y). \]

iii. Then

\[ G_{n-b,b+1}(y) - G_{n-b+1,b}(y) = \binom{n}{b} y^{n-b}(1-y)^b. \]

iv. Also,

\[ G_{n,1}(y) = n \int_0^y x^{n-1} dx = y^n. \]

v. By induction,

\[ G_{n-b,b+1}(y) = \sum_{a=0}^b \binom{n}{a} y^{n-a}(1-y)^a. \]

vi. Finish using relationship between Beta and F distributions.

c. Normal approx. \( \pi \in X_1/X_+ \pm 1.96 \sqrt{\frac{X_0X_1}{(X_1+X_0)^2}} \)

i. Problem if \( X_0 = 0 \)

ii. Less obvious problem for small \( X_1 + X_0 \)

d. Better Normal approx.: See homework.

A: 3.5.2

4. Testing \( H_0 : \beta = \beta^0 \) vs \( H_A : \beta \neq \beta^0 \) using the likelihood function

a. \( \beta \approx N(\beta, \ell''(\beta)^{-1}) \)

i. \( \hat{\beta} \) approximately unbiased.

ii. \( (\hat{\beta} - \beta)^{\top} [\ell''(\hat{\beta})] (\hat{\beta} - \beta) \) is Wald test statistic

b. Wald Test

i. Let \( \beta = \left( \hat{\psi}, \hat{\phi} \right) \) maximize \( \ell \) over \( \beta \)

ii. Maximize \( \ell \) with \( \psi = 0 \) and \( \phi \) unconstrained.

iii. Compare 2× difference to \( \chi^2 \)

• d.f. is length of \( \psi \)

c. Wald Test

i. Let \( \beta = \left( \hat{\psi}, \hat{\phi} \right) \) maximize \( \ell \) over \( \beta \)
ii. \( \text{Var} \left[ \hat{\beta} \right] \approx I(\hat{\beta}) = [-\ell''(\hat{\beta})]^{-1} \)

iii. \( \text{Var} \left[ \hat{\psi} \right] \approx \) appropriate submatrix \( I^{11} \)

iv. Wald statistic is \( \hat{\phi}^T [I^{11}(\hat{\beta})]^{-1} \hat{\phi} \)

d. Score Test
   i. Let \( \hat{\phi} \) maximize \( \ell \) over \( \phi \) with \( \psi = 0 \)
   ii. Let \( \ell^1 \) be components of \( \ell' \) corresponding to \( \psi \)
   iii. Test is \( l^1(0, \eta)^T I^{11}(0, \eta) l^1(0, \eta) \sim \chi^2 \)

6. Score test for binomial variables
   a. Score statistic is derivative of log of likelihood evaluated at null hypothesis
   b. \( L(\pi_0) = \left( X_0 / X_0 \right) \pi_0 X_0 \left( 1 - \pi_0 \right) X_1 \)
   c. \( l(\pi_0) = \log(\left( X_0 / X_1 \right)) + \log(\pi_0) X_0 + \log(1 - \pi_0) X_1 \)
   d. \( l'(\pi_0) = \frac{X_0/\pi_0 - X_+ - X_0/(1 - \pi_0)}{(1 - \pi_0) n_0} = \frac{X_0 - \pi_0 X_+}{(1 - \pi_0) n_0} \)

7. Coverages shown in Figs. 4 and 5. [Mark e R]

B&D2: 3.4c–e

D. Multiple (\( K \)) Categories
1. How do exposure groups differ?
2. Wrong Solution
   a. Choose one group as baseline
      i. Usually the one with no exposure, if there is one
   ii. Be careful what you lump in here
   b. Calculate relative risks with respect to this group
   c. Calculate hypothesis tests
      i. for each pair
      ii. or against a baseline

Lecture 1

\[
\sum_{k=0}^{K-1} X_k \log(\lambda_k) - \log(\lambda_+) \] 

\[
\log(C). 
\]

c. Score statistics are
Fig. 6: Rejection Regions for Chi-Square Tests

\[
\begin{align*}
X_1 & \quad 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 \\
X_2 & \quad 0 & 2 & 4 & 6 & 8 & 10 & 12 & 14 \\
\end{align*}
\]

\(\alpha = 0.05\), 3 category multinomial, 13 items