VII. Sample Size Calculations

A. Preliminaries

1. We’ll do power for 1-sided tests
   a. Conceptually easier (as we shall see)
   b. Get power for 2-sided tests by doubling $\alpha$

B. Exactly:

1. Select smallest $C$ such that $P_{0}[T \geq C] \leq \alpha$
2. Power is $P_{A}[T \geq C]$.

C. Approximately,

1. Suppose
   a. $H_{0}: T \sim \mathcal{N}(\mu_{0}, \sigma_{0}^{2})$
   b. $H_{A}: T \sim \mathcal{N}(\mu_{A}, \sigma_{A}^{2})$

2. Critical value: $C$
   a. Reject $H_{0}$ if $(T - \mu_{0})/\sigma_{0} \geq z_{\alpha}$
   b. $1 - \Phi(z_{\alpha}) = \alpha$
   c. Reject $H_{0}$ if $T \geq \mu_{0} + \sigma_{0}z_{\alpha}$
   d. $C = \mu_{0} + \sigma_{0}z_{\alpha}$

3. Power is $P_{A}[T \geq C] = \Phi((\mu_{A} - \mu_{0} - \sigma_{0}z_{\alpha})/\sigma_{A})$
Lecture 7

4. Sample size:
   a. Assume that $\sigma_0 = \tau_0/\sqrt{n}$, $\sigma_A = \tau_A/\sqrt{n}$.
   b. Require power $1 - \beta$
      i. Typically, $0.8 = 80\%$.
   c. Then $-z_\beta = (\mu_0 + \sigma_0 z_\alpha - \mu_A)/\sigma_A$.
      i. $(\sigma_A z_\beta + \sigma_0 z_\alpha) = \mu_A - \mu_0$
      ii. $(\tau_A z_\beta + \tau_0 z_\alpha)/\sqrt{n} = \mu_A - \mu_0$
      iii. $(\tau_A z_\beta + \tau_0 z_\alpha)/(\mu_A - \mu_0) = \sqrt{n}$
      iv. $n = (\tau_A z_\beta + \tau_0 z_\alpha)^2/(\mu_A - \mu_0)^2$
      v. When $\tau_A = \tau_0$, $n = \tau_0^2(z_\beta + z_\alpha)^2/(\mu_A - \mu_0)^2$

5. Example:
   a. Null probability $0.5$, alternative probability $0.6$, one-sided test
      size $\alpha = 0.025$, power $1 - \beta = 0.8$
   b. $z_\alpha = 1.96$, $z_\beta = 0.84$.
   c. Need $X_+ = \frac{(\sqrt{0.6 \times 4} z_{0.8} - z_{0.025 \times 0.5})^2}{(0.6 - 0.5)^2} = 194$ individuals.

D. Often use variance stabilizing transformation for power
   1. Suppose the parameter of interest is $\mu = E[T]$
   2. Sometimes $\text{Var}[T]$ is a function of $\mu$. 
3. Look for transformation $g(T)$ of test statistic $T$ so that

$$\text{Var}[g(T)] \text{ does not depend on } \mu$$

4. $\text{Var}[g(T)] \approx \text{Var}[g(\mu) + g'(\mu)(T - \mu)] = g'(\mu)^2 \text{Var}[T]$. 

5. Find $g$ so that $g'(\mu) = 1/\text{Var}[T]$. 

6. Ex., Poisson $\text{Var}[T] = \mu$, so $g'(\mu) = 1/\sqrt{\mu}$, $g(\mu) = 2\sqrt{\mu}$. 

7. Power is approximate

   a. Better approximation for Poisson uses fact that when
   
   $$X \sim P(\mu) \text{ then } \text{Var}[\sqrt{X}] \approx \mu \times (\frac{1}{2}\mu^{-1/2})^2 = \frac{1}{4}.$$ 

   b. Better approximation for binomial uses fact that
   
   $$\text{arcsin}(\sqrt{X_1/X_+}) \sim N(\text{arcsin}(\sqrt{Q_1/(Q_0 + \varsigma Q_1)}), 1/(4X_+))$$ 

   i. $Q_+$ is exponential of offset (or offset before you take log). 

   ii. $\frac{d}{dx} \text{arcsin}(x) = 1/\sqrt{1 - x^2}$ 

   iii. $\frac{d}{dx} \text{arcsin}(\sqrt{x}) = 1/(2\sqrt{x}\sqrt{1 - x})$ See Figures 9 and 10. 

   iv. $\mu_0 = \text{arcsin}(\sqrt{Q_1/(Q_0 + Q_1)})$, $\mu_A =$ 
   
   $\text{arcsin}((\varsigma\sqrt{Q_1/(Q_0 + \varsigma Q_1)}), \sigma_A = \sigma_0 = \sqrt{1/(4X_+})$ 

8. Exponential family models:

   a. Suppose that $T$ has probabilities or mass function
\[ \exp(t \tau - \mathcal{K}(\tau) - c(t)) \]

i. Upper left corner of 2 \times 2 table fits, if \( \tau \) is log odds ratio

ii. If independent addends have this pattern, then so does sum.

b. Calculate \( \mathcal{K}(\tau) = \log(\mathbb{E}_0[\exp(\tau T)]) \)

c. Differentiating once,
\[
\mathcal{K}'(\tau) = \frac{d}{d\tau} \mathbb{E}[\exp(\tau T)] / \mathbb{E}_0[\exp(\tau T)]
\]
\[
= \mathbb{E}_0 \left[ \frac{d}{d\tau} \exp(\tau T) \right] / \mathbb{E}_0[\exp(\tau T)]
\]
\[
= \mathbb{E}_0[T \exp(\tau T)] / \mathbb{E}_0[\exp(\tau T)] = \mathbb{E}_\tau[T]
\]

d. Differentiating again,
Fig. 10: Variance Stabilizing Transformation

\[ K''(\tau) = \frac{d}{d\tau} \left[ \frac{E_0 [T \exp(\tau T)]}{E_0 [\exp(\tau T)]} \right] \]

\[ = \frac{d}{d\tau} E_0 [T \exp(\tau T)] \frac{E_0 [\exp(\tau T)]^2}{E_0 [\exp(\tau T)]} - \frac{d}{d\tau} E_0 [\exp(\tau T)] E_0 [T \exp(\tau T)] \]

\[ = \frac{E_0 [T^2 \exp(\tau T)]}{E_0 [\exp(\tau T)]} - \frac{E_0 [T \exp(\tau T)] E_0 [T \exp(\tau T)]}{E_0 [\exp(\tau T)]^2} \]

\[ = E_\tau [T^2] - E_\tau [T]^2 = \text{Var}_\tau [T] \]

e. So \( \frac{d}{d\tau} E_\tau [T] = \text{Var}_\tau [T] \)

f. If \( H_0 : \tau = 0, \ H_A : \tau = \theta \), then for small \( \theta \), \( \sigma_0^2 \approx \sigma_A^2 \),

\[ \mu_A - \mu_0 \approx \sigma_0^2 \theta \]
Lecture 8

g. Power is $\Phi(\sigma_0\theta - z_\alpha)$

9. Mantel–Haenszel example:
   a. $T = \text{sum of upper right corners}$
   b. $\sigma_0 = \sqrt{\sum_i \frac{X_{1+}^i X_{0+}^i + X_{+1}^i X_{+0}^i}{X_{++}^i X_{+++}(X_{++}^i - 1)}}$
   c. $X_{0+}, X_{1+}, \text{and } X_{++}$ fixed in advance.
   d. $X_{+j}$ should be replaced by $X_{++}\pi_{+j}$

10. Mantel-Haentzel Example: Henhouse data set
   a. Six labs, and expect 9 control and 9 treatment chicks per lab
      i. So $X_{0+}^i = X_{1+}^i = 9$
   b. Expect null proportions of abnormalities to be $1/9 \text{ to } 6/9$.
      i. So null column totals are $X_{+0}^i = (2, 4, 6, 8, 10, 12)$, $X_{+1}^i = (16, 14, 12, 10, 8, 6)$.
   c. Null variance is $9 \times 9 \times (2 \times 16 + 4 \times 14 + 6 \times 12 + 8 \times 10 + \ldots )/(18 \times 18 \times 17) = 5.76$, and SE is 2.40
   d. Power for detecting log odds ratio of .5 is $\Phi(2.40 \times .5 - 1.96) = .224$.
   e. For 80% power, need log odds ratio satisfying $2.27\tau - 1.96 = .84 \text{ or } \tau = (0.84 + 1.96)/2.27$.
   f. If you want 80% power with log odds ratio .5, need
\[ \sigma_0 \times 0.5 - 1.96 = 0.84, \text{ or } \sigma_0 = (0.84 + 1.96)/0.5 = 5.6. \]

Since \( \sigma_0 \) is approximately proportional to the square root of the number of chicks per group per lab, you need \( 9 \times (5.6/2.27)^2 \) chicks per group in each lab.