March 16 — Confidence intervals for proportions

1. Suppose \( x \) is the number of successes in \( n \) trials of a Bernoulli process with probability of success \( \pi \). Then \( x \) has the binomial distribution with parameters \( n \) and \( \pi \).

2. \( \hat{\pi} = p = x/n \) is an unbiased estimator for \( \pi \).

3. The distribution of \( p \) has mean \( \pi \) and variance \( \pi(1 - \pi)/n \).

4. For large \( n \), the central limit theorem then implies that the following random variable is approximately normal with mean 0 and variance 1:

\[
\frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}
\]

so a 95% confidence interval can be found by solving

\[-1.96 < \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} < 1.96
\]

for an interval of form \( a(p) < \pi < b(p) \).

5. Three different confidence intervals based on this:

   (a) Sloppy - replace \( \pi \) in the denominator by \( p \) (some books do this)

   (b) R (\texttt{prop.test}): solve the equation directly for an inequality of form \( a(p) < \pi < b(p) \). (Use the quadratic equation.)

   (c) R (with continuity correction, \texttt{prop.test} default): solve the equation after realizing that \( x \) is really between \( x - \frac{1}{2} \) and \( x + \frac{1}{2} \).

6. Exact confidence intervals are also available in R (\texttt{binom.test}). These are called exact but are not really exact. They do have the property however that a 95% confidence interval is at least a 95% confidence interval.

7. Confidence intervals for differences in proportions can also be computed by the same approximations (R \texttt{prop.test}).

8. Determining sample sizes to ensure given confidence levels. (See Gallup poll.)

**Homework**

1. Read Devore and Farnum, Section 7.3, pages 303–306.

2. Do problems 7.22, 24, 26, 28.