CI for population proportions:

1. One sample problem $(n > 30)$

2. Two-sample problem with $(n_1 > 30, n_2 > 30)$

3. CI for $\sigma^2$:
   - The assumption of normality
   - Chi-square distribution with d.f. = n-1

4. Exercises

Homework: Ch. 4, Sec. 5
CONFIDENCE INTERVAL FOR PROPORTIONS

SPECIAL ONE-SAMPLE MODEL

When the outcomes are categorical, which is the case when we observe traits or characteristics of the experimental units, the basic data will be counts. In the simplest case, there are only two categories, for example, male / female, defective / not defective.

Very often we will call the category of interest success, $S$, and the other failure, $F$.

We can represent the outcome as a random variable $X$ that takes the value 1 or 0
according to whether or not $S$ occurs.

Assume that $P(X = 1) = p$ and

$P(X = 0) = 1 - p$.

The mean of $Y$:

$$E(X) = 1p + 0(1 - p) = p$$

Variance of $Y$:

$$\text{var}(X) = E(X^2) - [E(X)]^2 =$$

$$= 0^2 (1 - p) + 1^2 p - p^2 = p(1 - p).$$

Standard deviation

$$sd(X) = \sqrt{p(1 - p)}.$$
Now let $X_i = 1$ if the $i$-th trial is an $S$ and $X_i = 0$ otherwise.

Each $X_i$ has the same distribution as $X$ and are independent. The total number of successes is $Y = X_1 + X_2 + ... + X_n$ and $\hat{p} = Y/n = \bar{Y}$ is the sample mean. It is easy to see, that mean of $(\hat{p}) = p$ and $sd(\hat{p}) = \sqrt{p(1-p)/n}$

* Point estimators of $p$ and $p(1-p)$:

$\hat{p} = \bar{Y}$

$\hat{p}(1 - \hat{p})$

From the CLT it follows that the sampling distribution
of \( \hat{p} \) is approximately normal with mean \( p \) and

\[
sd = \sqrt{p(1 - p)/n}, \text{ so}
\]

\[
Z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}}
\]

is nearly standard normal \( N(0, 1) \).
The standard error of the estimator \( \hat{p} \)

\[
S.E.(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}
\]

and estimated standard error :

\[
S.E.(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}
\]

**INFERENCES ABOUT POPULATION PROPORTIONS**

**ONE SAMPLE PROBLEM**

Large sample \((1 - \alpha)\)% CI for \( p \) :

\[
\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \leq p \leq \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.
\]

Here \( z_{\alpha/2} \) is the \( \alpha/2 \) upper quantile of the standard normal distribution \( N(0,1) \).
TWO SAMPLE PROBLEM

Let \( \hat{p}_1 = \frac{Y_1}{n_1} \) and \( \hat{p}_2 = \frac{Y_2}{n_2} \) be the
point estimators of the population proportions \( p_1 \) and \( p_2 \), respectively.

Large sample \((1 - \alpha)\%\) CI for \( p_1 - p_2 \):

\[
\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}
\]

here \( z_{\alpha/2} \) -is the \( \alpha/2 \) upper quantile

of the standard normal distribution \( N(0, 1) \).
CONFIDENCE INTERVAL
FOR STANDARD DEVIATION

Assume that $X_1, \ldots, X_n$ is a sample from

a normal population $N(\mu, \sigma^2)$

with $\mu$ and $\sigma^2$ unknown. Let us consider

the case when sample size $n$ is small ($n \leq 30$).

To construct confidence interval for $\sigma^2$, note that

* the statistic

$$
\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{(n - 1)S^2}{\sigma^2}
$$

has a chi-square distribution with d.f. = $n$ - 1.
The chi-square distribution is the basic distribution for constructing the CI for \( \sigma^2 \).

For example, to construct 95\% CI for \( \sigma^2 \) we have to divide the probability .05 equally between two tails of the chi-square distribution and determine the upper .025 and lower .025 points of chi-square distribution from the Table C.5, i.e.

\[
P\left( \chi^2_{.975} \leq \frac{(n - 1)S^2}{\sigma^2} \leq \chi^2_{.025} \right) = .95.\]
This probability may be rewritten in the following way:

\[ P\left( \frac{(n - 1)S^2}{\chi^2_{0.025}} \leq \sigma^2 \leq \frac{(n - 1)S^2}{\chi^2_{0.975}} \right) = .95. \]

The CI for \( \sigma \) can be obtained by taking the square root of the endpoints of the interval

\[ S\sqrt{\frac{n - 1}{\chi^2_{0.025}}} \leq \sigma \leq S\sqrt{\frac{n - 1}{\chi^2_{0.975}}}. \]

In general, the \((1 - \alpha)100\%\) CI for \( \sigma \) has the form:

\[
S\sqrt{\frac{n - 1}{\chi^2_{\alpha/2}}} \leq \sigma \leq S\sqrt{\frac{n - 1}{\chi^2_{1-\alpha/2}}}.\]
CONFIDENCE INTERVALS
FOR PROPORTIONS

Problem 1. One tire manufacturer found that after 5000 miles, \( y = 32 \) of \( n = 200 \) steel belted tires selected at random were defective. Find an approximate 99 percent confidence interval for \( p \), the proportion of defective tires in the total production.

Problem 2. To test two different training methods, 200 workers were divided at random into two groups of 100 each. At the end of the training program there were \( y_1 = 62 \) and \( y_2 = 74 \) successes, respectively. Find an approximate 90% Confidence Interval for \( p_1 - p_2 \), the difference of the true proportions of success.

Problem 3. The number of days to maturity was recorded for 25 plants grown from seeds of a single stock. The sample mean and sample standard deviation were

\[
\bar{x} = 68.4, \ s = 6.5
\]

(a) Construct a 95% CI for \( \sigma \).