Probabilistic Modeling and Statistical Computing Fall 2015

November 24, 2015

CI for Proportions

Observe n trials of a Bernoulli random variable.
$X =$ Number of successes.

- Opinion poll: Should marijuana be legal?
- Opinion poll: Is the country headed in the wrong direction?
- School testing: Does a sixth-grader pass math proficiency tests?
- Online marketing: Does a Web user click through on my ad?

Maximum likelihood estimate for success probability p is $\hat{p} = \frac{X}{n}$. Make a CI for p.

Examples from This Week

Do you approve of the way Congress does its job?

- Gallup poll Nov. 2015: 1021 adults were surveyed by phone. 113 said that they approved.
- Gallup poll Dec. 2014: 805 adults were surveyed by phone. 121 said that they approved.

Does this represent a decrease?

50 % landline respondents in 2014, 40 % landline respondents in 2015. Does that matter?

Examples from This Week

Do you approve of the way Congress does its job?

- Gallup poll Nov. 2015: 1021 adults were surveyed by phone. 113 said that they approved.
- Gallup poll Dec. 2014: 805 adults were surveyed by phone. 121 said that they approved.

Does this represent a decrease?

50 % landline respondents in 2014, 40 % landline respondents in 2015. Does that matter?

Ex.: Confidence Interval for a Proportion

Data model: $X \sim B(n, p)$, where n is the number of trials (e.g. persons asked), p is the unknown success probability, and X is the number of successes. Use $\boxed{\hat{p} = \frac{X}{n}}$ = sample proportion.

By the CLT, approximately $\hat{p} \sim N(p, \frac{p(1-p)}{n})$ and therefore

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)\,.$$

Confidence Interval for a Proportion

Let z^* be such that $\mathcal{P}(-z^* \leq Z \leq z^*) = \alpha$, if $Z \sim N(0,1)$.

```
zstar <- qnorm((1+alpha)/2)
```

We know that

$$\mathcal{P}\left(-z^* \leq Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z^*\right) \approx \alpha$$

Rework this to obtain inequalities for p.

Method I: Score Confidence Interval

Solve the algebraic equations for p.

Score Confidence Interval

The confidence interval is $p \in (B_-, B_+)$ where

$$B_{\pm} = \frac{\hat{p} + {z^*}^2/n \pm z^* \sqrt{\hat{p}(1-\hat{p})/n + {z^*}^2/(4n^2)}}{1 + {z^*}^2/n}$$

This is implemented in **R** as `prop.test()`.

If n is very large, ${z^*}^2/n$ *is small and we obtain*

Method II: Wald Confidence Interval

In the inequalities $-z^* \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \leq z^*$*, replace the* $p(1-p)$ *term with its estimate* $\hat{p}(1-\hat{p})$.

Wald Confidence Interval

The confidence interval is $p \in (B_-, B_+)$ where

$$B_{\pm} = \hat{p} \pm z^* \sqrt{\hat{p}(1-\hat{p})/n}$$

Easy to implement for hand calculations.

If np *and* $n(1-p)$ *are small, both are wrong.*

Differences of Proportions

- Two independent samples, $X_1 \sim B(n_1, p_1)$, $X_2 \sim B(n_2, p_2)$
- Find a CI for $p_1 - p_2$
- **R** also does this with `prop.test()`

Approval rating for Congress: Does the drop from 15% to 11% approval rating represent a true decline?

Ex.: Hypothesis Test for Mean

Consider a normal distribution $N(\mu, \sigma^2)$ with known σ^2 and unknown μ.

- Null hypothesis: $H_0 : \mu = \mu_0$ given
- Alternative $H_a : \mu = \mu_1$ or $\mu > \mu_0$ or $\mu \neq \mu_0$ or $\mu_1 \leq \mu \leq \mu_2$ etc.
- Given a sample, are the data consistent with H_0 or do they favor H_a?

Possible decisions: Reject H_0 (there is evidence against H_0) or do not reject (not enough evidence against H_0).

Innocent until proven guilty.

The Exact Null Distribution

Use the sample mean $\bar{X}$ as test statistic. **If H_0 is true, then** $\bar{X} \sim N(\mu_0, \frac{\sigma^2}{n})$. Therefore,

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

The observed sample mean is $\bar{x}$.

Hypothesis Test for Mean

Assume that $\mu > \mu_0$ if H_a is true.
Then $\bar{x} > \mu_0$ is possible evidence in favor of H_a.
If a larger $\bar{x}$ were observed, this would be stronger evidence in favor of H_a.
And $\bar{x} < \mu_0$ cannot be evidence in favor of H_a.

p-value for one-sided alternative

$$p = \mathcal{P}(Z \geq \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}})$$

Hypothesis Test for Single Mean

Assume that $\mu \neq \mu_0$ if H_a is true.
Then large $|\bar{x} - \mu_0|$ is possibly strong evidence against H_0 in favor of H_a.
And small $|\bar{x} - \mu_0|$ is weak evidence against H_0.

$$
\begin{aligned}
\mathcal{P}(|\bar{X} - \mu_0| \geq |\bar{x} - \mu_0|) &= \mathcal{P}\left(\frac{|\bar{X} - \mu_0|}{\sigma/\sqrt{n}} \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right) \\
&= \mathcal{P}\left(|Z| \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right) \\
&= 2\mathcal{P}\left(Z \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}}\right)
\end{aligned}
$$

Two-sided Alternative

p-value for two-sided alternative

$$p = 2\mathcal{P}(Z \geq \frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}})$$

where $Z \sim N(0, 1)$.

This is twice the p-value for the one-sided alternative $\mu > \mu_0$, if in fact $\bar{x} > \mu_0$.
It is also twice the p-value for $\mu < \mu_0$ if $\bar{x} < \mu_0$ is observed.

Critical Region, Critical Value

The test statistic is $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$. Suppose we decide before hand to reject H_0 whenever $p < \alpha$, where $\alpha =$ **significance level**.

Critical Region

If $H_a : \mu > \mu_0$, let $C =$ **critical value** be such that $\mathcal{P}(Z \geq C) = \alpha$. Reject H_0 whenever $\boxed{z > C}$.

If $H_a : \mu \neq \mu_0$, let $C =$ be such that $\mathcal{P}(Z \geq C) = \alpha/2$. Reject H_0 whenever $\boxed{|z| > C}$.

Critical region $z > C$ **or** $|z| > C$.

Unknown Variance

Assume that $\mu > \mu_0$ if H_a is true.
If we do not know σ^2, we replace it with the sample variance and must use T-distributions.

p-value for one-sided alternative $\mu > \mu_0$

$$p = \mathcal{P}(T \geq \frac{\bar{x} - \mu_0}{s/\sqrt{n}})$$

where $T \sim$ T-distribution with $n - 1$ d.o.f. and s is the sample standard deviation.

Same Thing with Critical Region

Assume that $\mu > \mu_0$ if H_a is true.

Test Statistic: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$

Critical value for significance level α:
$\mathcal{P}(T \geq C) = \alpha$
where $T \sim$ T-distribution with $n - 1$ d.o.f.

Reject H_0 whenever $\boxed{t > C}$.

Hypothesis Test for Single Proportion

Example: Toss a coin 20 times. Observe "heads" in 14 tosses. Is this evidence that $\mathcal{P}(heads) > \frac{1}{2}$?

$$H_0 : p = 1/2 \qquad H_a : p > 1/2$$

p-value for one-sided alternative $p > p_0$

Observe x successes in n trials.

$$p = \mathcal{P}(X \geq x)$$

where $X \sim B(n, p_0)$.

Same Thing with Critical Region

Assume that $p > p_0$ if H_a is true.

Test Statistic: x = number of successes

Critical value for significance level α:
$\mathcal{P}(X \geq C) = \alpha$
where $X \sim B(n, p_0)$

Reject H_0 whenever $\boxed{x > C}$.
There is no approximation here.
Can we always find C exactly?
Better to compute the p-value and go from there.

Same Thing with Critical Region

Assume that $p > p_0$ if H_a is true.

Test Statistic: x = number of successes

Critical value for significance level α:
$\mathcal{P}(X \geq C) = \alpha$
where $X \sim B(n, p_0)$

Reject H_0 whenever $\boxed{x > C}$.
There is no approximation here.
Can we always find C exactly?
Better to compute the p-value and go from there.

Same Thing with Critical Region

Assume that $p > p_0$ if H_a is true.

Test Statistic: x = number of successes

Critical value for significance level α:
$\mathcal{P}(X \geq C) = \alpha$
where $X \sim B(n, p_0)$

Reject H_0 whenever $\boxed{x > C}$.
There is no approximation here.
Can we always find C exactly?
Better to compute the p-value and go from there.

Same Thing with Critical Region

Assume that $p > p_0$ if H_a is true.

Test Statistic: x = number of successes

Critical value for significance level α:
$\mathcal{P}(X \geq C) = \alpha$
where $X \sim B(n, p_0)$

Reject H_0 whenever $\boxed{x > C}$.
There is no approximation here.
Can we always find C exactly?
Better to compute the p-value and go from there.

Test vs. Confidence Interval

- The p-value is computed from the hypothetical distribution under H_0.
- *This is easy for binomial distributions.*
- *A permutation approach is appropriate in the general case.*
- A CI is computed from the actual data, without hypothesizing a distribution.
- *This is complicated for binomial distributions and requires an approximation.*
- *A bootstrap approach is appropriate in the general case.*

Hypothesis Tests for Differences

. . . of means, of proportions, . . .

Compare two means μ_1, μ_2.
Typical $H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$
Typical $H_a : \mu_1 > \mu_2$ or $\mu_1 - \mu_2 > 0$
This is an example of a one-sided alternative.

Compare two probabilities p_1, p_2.
Typical $H_0 : p_1 = p_2$ or $p_1 - p = 0$
Typical $H_a : p_1 > p_2$ or $p_1 - p_2 > 0$

`t.test()` and `prop.test()`

Hypothesis Tests for Differences

. . . of means, of proportions, . . .

Compare two means μ_1, μ_2.
Typical $H_0 : \mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$
Typical $H_a : \mu_1 > \mu_2$ or $\mu_1 - \mu_2 > 0$
This is an example of a one-sided alternative.

Compare two probabilities p_1, p_2.
Typical $H_0 : p_1 = p_2$ or $p_1 - p = 0$
Typical $H_a : p_1 > p_2$ or $p_1 - p_2 > 0$

`t.test()` and `prop.test()`

Example: Gallup Poll

What is the null distribution?
The assumption under H_0 is that both samples come from the same population with unknown p.
CLT approximation for the two sample proportions:

$$\hat{p}_1 \sim N\left(p, \frac{p(1-p)}{n_1}\right), \quad \hat{p}_2 \sim N\left(p, \frac{p(1-p)}{n_2})\right)$$

Test statistic

$$
\begin{aligned}
\hat{p}_1 - \hat{p}_2 &\sim N\left(0, \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2})\right) \\
&= N\left(0, p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})\right)
\end{aligned}
$$

Pooled estimate for p:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Use this instead of p. This is another approximation!

Example: Birth Weights in Texas

Type I Error

This error occurs when H_0 is rejected although it should not be.

- A new drug appears to be better than an old one although it isn't.
- An accused person appears to be guilty although he is innocent.
- A person who feels ill is declared well although she isn't.

P-value = $\mathcal{P}(\textit{type I error})$**, assuming** H_0

Type II Error

This error occurs when H_a is not detected (H_0 is not rejected) rejected although it should be.

- A new drug appears to as good as an old one although it better.
- An accused person is let go although he is guilty.
- A person who feels ill is sent to the hospital although she is well.

$\mathcal{P}(type\ II\ error)$ **is computed assuming** H_a

Power

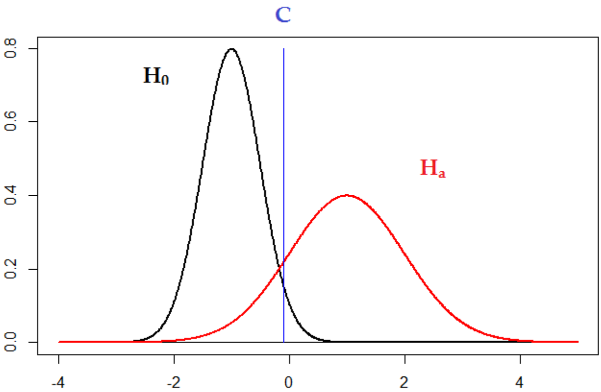
This is $1 - \mathcal{P}(\textit{type II error})$. It is the probability that the alternative hypothesis will be detected, assuming it is true.

- Detect that a new drug is indeed better than an old one.
- Convict a guilty criminal.
- A person who feels ill but is in fact well is sent home.

Power is computed assuming H_a

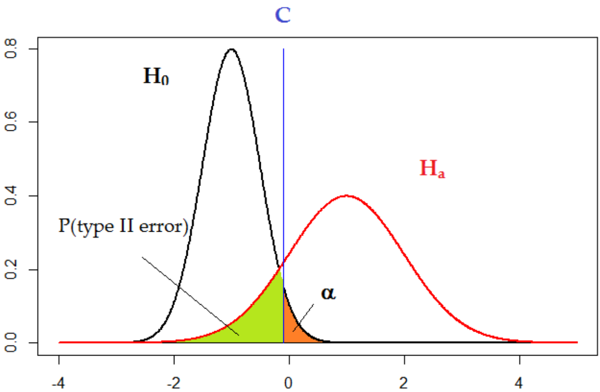
Type I Error vs. Type II error

Distribution of test statistic under H_0 and H_a.
The critical value C is at the blue line.



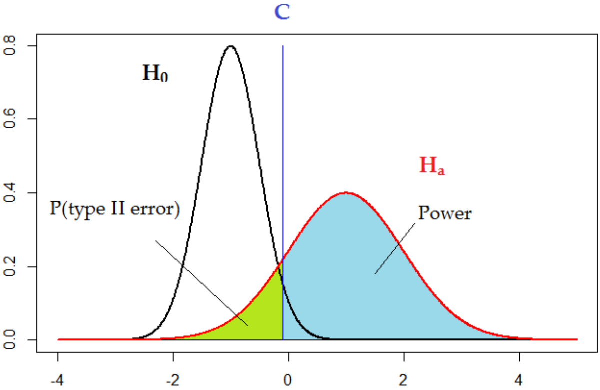
Type I Error vs. Type II error

Significance level α and corresponding type II error probability



Type I Error vs. Type II error

Type II error probability and power for this significance level.



Constructing Tests

Recall maximum likelihood estimators: general method for constructing estimation formulae, often leads to identification of good statistics ("data reduction").

Is there a similar approach for constructing hypothesis tests?

Need to balance two objectives: significance level (should be kept small) and power (should be large).

Simple Hypotheses

Assume that both H_0 and H_a consist of single, well specified distributions.

- $H_0 : N(\mu_0, \sigma_0^2)$ vs. $H_a : N(\mu_1, \sigma_1^2)$
- Exponential distribution. $H_0 : \lambda = \lambda_0$ vs. $H_a : \lambda = \lambda_1$
- Multinomial distribution: $H_0 : (p_1, \ldots, p_n)$ vs. $H_a : (q_1, \ldots, q_n)$

Find a test of H_0 against H_a that has given significance level α and maximum power.
Find a test statistic and a critical value!

Likelihood Ratio

Assume that both H_0 and H_a consist of single, well specified distributions with pdf's or pmf's $f_0(x), f_a(x)$. The **likelihood function** for H_0 is

$$L_0(x_1, \ldots, x_n) = f_0(x_1) \times \cdots \times f_0(x_n)$$

and similarly for H_a.

Likelihood Ratio

For a sample $(x_1, \ldots, x_n)$, the likelihood ratio is

$$T = \frac{L_0(x_1, \ldots, x_n)}{L_a(x_1, \ldots, x_n)} = \frac{f_0(x_1) \times \cdots \times f_0(x_n)}{f_a(x_1) \times \cdots \times f_a(x_n)}$$

Likelihood Ratio

The likelihood ratio is

$$T = \frac{L_0(x_1, \ldots, x_n)}{L_a(x_1, \ldots, x_n)} = \frac{f_0(x_1) \times \cdots \times f_0(x_n)}{f_a(x_1) \times \cdots \times f_a(x_n)}$$

Interpretation: If T is small, the sample is more likely to come from the alternative distribution. If T is large, the sample is more likely to come from the null distribution.

Likelihood Ratio Test

Likelihood Ratio Test: Given a critical value C, reject H_0 if $T < C$. The significance level is $\mathcal{P}(T < C|H_0)$.

Neyman - Pearson Lemma

Of all tests of H_0 versus H_a with given significance level α, the likelihood ratio test has the largest power $\mathcal{P}(T > C|H_a)$ (the lowest type II error probability).

This tells one how to find a test statistic. It does not tell us how to find the critical value C or the power.

Example: Exponential Distribution

Consider data coming from an exponential distribution.

$H_0 : \lambda = \lambda_0$ versus $H_a : \lambda = \lambda_a > \lambda_0$

Given a sample $(x_1, \ldots, x_n)$.
Likelihood function for H_0:

$$L_0(x_1, \ldots, x_n) = \lambda_0^n e^{-\lambda_0 x_1} e^{-\lambda_0 x_2} \ldots e^{-\lambda_0 x_n} = \lambda_0^n e^{-\lambda_0 \sum_i x_i}$$

and similarly for H_a

Example: Exponential Distribution

Consider data coming from an exponential distribution.
Likelihood ratio for this case:

$$T = \frac{L_0(x_1, \ldots, x_n)}{L_a(x_1, \ldots, x_n)} = \frac{\lambda_0^n e^{-\lambda_0 \sum_i x_i}}{\lambda_a^n e^{-\lambda_a \sum_i x_i}}$$

$$= \left(\frac{\lambda_0}{\lambda_a}\right)^n e^{(-\lambda_0 + \lambda_a) \sum_i x_i}$$

Reject H_0 if $T < C$, where C depends on α. This means reject H_0 if $\tilde{T} = \sum_i x_i < c_1$, since T depends only on $\tilde{T}$.

Where are we now? What is left?

The **likelihood ratio test** uses the test statistic $\tilde{T} = \sum_i x_i$ and rejects H_0 if $\tilde{T}$ is small, $\tilde{T} < c_1$.

Need to find a relation between critical value c_1 and desired significance level α.

To do this, need the distribution of $\tilde{T}$ if H_0 is true.

This can be done analytically or by a simulation.

Fact: If H_0 is true, then $\tilde{T}$ has a Γ distribution, shape parameter $\alpha = n$, rate parameter λ_0.

```
c1 <- qgamma(alpha, shape = n, rate
= lambda0, lower.tail = F)
```