# Inference for Binomial Proportions

*Chapter 8, Lab 1: Solutions*

*OpenIntro Biostatistics*

**Topics**

- Inference for a single proportion
- Inference for the difference of two proportions

This lab introduces methods of inference for binomial proportions, as a parallel to the one-sample and two-sample methods for means introduced in Chapters 4 and 5. In the setting where a binary outcome is recorded for a single group of participants, inference about the binomial probability of success provides information about a population proportion $p$. Just as inference can be done for the difference of two population means, inference can also be done in the setting of comparing two population proportions $p_1$ and $p_2$.

The material in this lab corresponds to Sections 8.1 and 8.2 in *OpenIntro Biostatistics*.

## Introduction

*Inference for a single proportion*

Suppose that $X$ is binomial with parameters $n$ (total number of trials) and $p$, the population parameter of success, where $x$ represents the number of successes. Inference about $p$ is based on the sample proportion $\hat{p}$, where $\hat{p} = x/n$; $\hat{p}$ is the point estimate of $p$.

Inference for $p$ can be made using the normal approximation to the binomial, or directly using the binomial distribution.

- *Inference with the normal approximation*
  - The sampling distribution of $\hat{p}$ is approximately normal when 1) the sample observations are independent, 2) $np \geq 10$, $n(1-p) \geq 10$.[1] Under these conditions, the sampling distribution of $\hat{p}$ is approximately normally distributed with mean $p$ and standard deviation $\sqrt{\frac{p(1-p)}{n}}$. For confidence intervals, substitute $\hat{p}$ for $p$; for hypothesis testing, substitute $p_0$ for $p$.
  - The approximate two-sided 95% confidence interval for $p$ is given by

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

  - The test statistic $z$ for the null hypothesis $H_0 : p = p_0$ based on a sample size of $n$ is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{(p_0)(1-p_0)}{n}}}$$

[1] The second condition is commonly referred to as the **success-failure condition**, since it can be effectively restated as the number of successes is greater than 10 and the number of failures is greater than 10.

- *Inference with exact methods*
  - Confidence intervals and $p$-values based on the binomial distribution are best calculated via R.
  - The logic behind calculating a $p$-value from the binomial distribution: Let $X$ be a binomial random variable with parameters $n$ and $p_0$, where $\hat{p} = x/n$ and $x$ is the observed number of events. For a test of $H_0 : p = p_0$ versus $H_A : p \neq p_0$, the $p$-value equals $2 \times P(X \geq x)$.

*Inference for the difference of two proportions*

The normal model can be applied to $\hat{p}_1 - \hat{p}_2$ if the sampling distribution for each sample proportion is nearly normal, and if the samples are independent random samples from the relevant populations and independent of each other.

Each sample proportion approximately follows a normal model when $n_1 p_1$, $n_1(1-p_1)$, $n_2 p_2$, and $n_2(1-p_2)$ are all $\geq 10$. To check success-failure in the context of a confidence interval, use $\hat{p}_1 and \hat{p}_2$.

The standard error of the difference in sample proportions is

$$\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}.$$

For hypothesis testing, an estimate of $p$ is used to compute the standard error of $\hat{p}_1 - \hat{p}_2$: $\hat{p}$, the weighted average of the sample proportions $\hat{p}_1$ and $\hat{p}_2$,

$$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{x_1 + x_2}{n_1 + n_2}.$$

To check success-failure in the context of hypothesis testing, check that $\hat{p}n_1$ and $\hat{p}n_2$ are both $\geq 10$.

**Inference for a single proportion**

1. Advanced melanoma is an aggressive form of skin cancer that until recently was almost uniformly fatal. In rare instances, a patient's melanoma stopped progressing or disappeared altogether when the patient's immune system successfully mounted a response to the cancer. Those observations led to research into therapies that might trigger an immune response.

   A 2013 report in the New England Journal of Medicine by Wolchok, et al. reported the results of a study in which patients were treated concurrently with two new therapies, nivolumab and ipilimumab. Of 52 patients, 21 experienced an immune response.

   a) What is the estimate of the population probability of an immune response after concurrent treatment with nivolumab and ipilimumab, $\hat{p}$?

      The sample proportion $\hat{p}$ is given by $x/n = 21/52 = 0.404$.

```
#use r as a calculator
x = 21
n = 52
p.hat = x/n
p.hat
```

```
## [1] 0.4038462
```

b) Evaluate whether the assumptions for using the normal approximation are met.

The outcome of one patient should not affect that of other patients, so independence is reasonable. The success-failure condition is satisfied since $n\hat{p} = (52)(0.404) = 21 > 10$ and $n(1-\hat{p}) = (52)(1-0.404) = 31 > 10$.

```
#use r as a calculator
n*p.hat
```

```
## [1] 21
```

```
n*(1 - p.hat)
```

```
## [1] 31
```

c) Calculate an approximate 95% confidence interval for $p$ and interpret the interval in context of the data.

The standard error is estimated as $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.404)(1-0.404)}{52}}$.

Thus, the 95% confidence interval is

$$0.404 \pm 1.96(0.068) \rightarrow (0.27, 0.54)$$

```
#use r as a calculator
se = sqrt((p.hat*(1-p.hat))/n)
z.star = qnorm(0.975)
m = z.star*se
p.hat - m; p.hat + m
```

```
## [1] 0.2704837
```

```
## [1] 0.5372086
```

d) Confirm the answer to part c) using `prop.test( )`.

The result from `prop.test( )` is (0.27, 0.55), which is very similar to the answer from part c).

```
#use prop.test
prop.test(x = 21, n = 52)$conf.int
```

```
## [1] 0.2731269 0.5487141
## attr(,"conf.level")
## [1] 0.95
```

2. Suppose that out of a cohort of 120 patients with stage I lung cancer at the Dana-Farber Cancer Institute (DFCI), 80 of the patients survive at least 5 years. National Cancer Institute statistics indicate that the 5-year-survival probability for stage I lung cancer patients nationally is 0.60. Do the data collected from the 120 patients support the claim that the DFCI population with stage I lung cancer has a different 5-year-survival probability than the national population? Let $\alpha = 0.05$.

   a) State the null and alternative hypotheses.

      The null hypothesis is that the proportion in the DFCI population is equal to the national statistic, $H_0 : p = 0.60$. The alternative hypothesis is that the proportion in the DFCI population is different from the national statistic, $H_A : p \neq 0.60$.

   b) Evaluate whether the assumptions for using the normal approximation are met. If so, conduct the hypothesis test based on the normal approximation.

      It is reasonable to assume independence; knowing cancer status of one patient is not likely informative about the cancer patient of another. The success-failure condition is met: $np_0 = (120)(0.60) = 72$ and $n(1 - p_0) = (120)(0.40) = 48$.

      The $p$-value is 0.16, which is greater than $\alpha = 0.05$. There is insufficient evidence to reject the null hypothesis that the 5-year survival probability in the DFCI population is different from 0.60.

```
#use prop.test
prop.test(x = 80, n = 120, p = 0.60)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  80 out of 120, null probability 0.6
## X-squared = 1.9531, df = 1, p-value = 0.1623
## alternative hypothesis: true p is not equal to 0.6
## 95 percent confidence interval:
##  0.5740262 0.7484656
## sample estimates:
##         p
## 0.6666667
```

   c) Compare the results in part b) to those from conducting the hypothesis test based on exact binomial methods.

      The result from `prop.test( )` is (0.27, 0.55), which is very similar to the answer from part c). As sample size increases, the normal approximation improves such that the results from the $z$-test and exact binomial become very close.

```
#use binom.test
binom.test(x = 80, n = 120, p = 0.60)
```

```
##
##  Exact binomial test
##
## data:  80 and 120
```

```
## number of successes = 80, number of trials = 120, p-value = 0.1619
## alternative hypothesis: true probability of success is not equal to 0.6
## 95 percent confidence interval:
##  0.5748217 0.7500649
## sample estimates:
## probability of success
##              0.6666667
```

3. In 2009, the FDA Oncology Drug Advisory Committee (ODAC) recommended that the drug Avastin be approved for use in glioblastoma, a form of brain cancer. Tumor shrinkage after taking a drug is called a response; out of 85 patients, 24 exhibited a response. Historically, response probabilities for brain cancer drugs were approximately 0.05, or about 5%. Assess whether there is evidence that the response probability for Avastin is different from previous drugs. Report both a $p$-value and 95% confidence interval; summarize your findings.

   Test $H_0 : p = 0.05$ against $H_A : p \neq 0.05$. Let $\alpha = 0.05$.

   Check assumptions. The independence assumption is reasonable in the context of a drug trial; participants are typically recruited independently. The success-failure condition is not met, since under the null hypothesis, only $(85)(0.05) = 4.25$ successes are expected. Exact binomial methods should be used.

   The $p$-value is highly significant ($p = 2.67 \times 10^{-12}$); there is sufficient evidence to reject the null hypothesis that the population response probability for Avastin is equal to 0.05. The observed response probability is 0.28, which suggests the population response probability to Avastin is higher than 0.05. With 95% confidence, the interval (0.19, 0.39) captures the population response probability (i.e., the response probability in a hypothetical population of glioblastoma patients treated with Avastin).

```
binom.test(x = 24, n = 85, p = 0.05)
```

```
##
##  Exact binomial test
##
## data:  24 and 85
## number of successes = 24, number of trials = 85, p-value = 2.674e-12
## alternative hypothesis: true probability of success is not equal to 0.05
## 95 percent confidence interval:
##  0.1900101 0.3904038
## sample estimates:
## probability of success
##              0.2823529
```

**Inference for the difference of two proportions**

4. The use of screening mammograms for breast cancer has been controversial for decades because the overall benefit on breast cancer mortality is uncertain. A 30-year study to investigate the effectiveness of mammograms versus a standard non-mammogram breast cancer exam was conducted in Canada with 89,835 female participants.[2] Each woman was randomized to receive either annual mammograms or standard physical exams for breast cancer over a 5-year screening period.

   By the end of the 25 year follow-up period, 1,005 women died from breast cancer. The results are summarized in the following table.[3]

| | Death from breast cancer? | |
|---|---|---|
| | Yes | No |
| Mammogram Group | 500 | 44,425 |
| Control Group | 505 | 44,405 |

   a) Calculate $\hat{p}_1$ and $\hat{p}_2$, the two sample proportions of interest.

      The two sample proportions of interest are the proportion of breast cancer deaths in the mammogram group and the proportion of breast cancer deaths in the control group: $\hat{p}_M = 500/(500+44425) = 0.0111$, $\hat{p}_C = 505/(505+44405) = 0.0112$.

```
#use r as a calculator
x = c(500, 505)
n = c(500 + 44425, 505 + 44405)
p.hat.vector = x/n
p.hat.vector
```

```
## [1] 0.01112966 0.01124471
```

   b) Analyze the results; do the data suggest that annual mammography results in a reduction in breast cancer mortality relative to standard exams? Be sure to check the assumptions for using the normal approximation.

      Since the participants were randomly assigned to each group, the groups can be treated as independent, and it is reasonable to assume independence of patients within each group.

      The pooled proportion $\hat{p}$ is

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = 0.0112$$

      The success-failure condition is met; the expected number of successes is about 503 in both groups (and expected number of failures is naturally much larger, given that both $\hat{p}$ is less than 0.50 and $n$ is very large).

      Test $H_0 : p_M = p_C$ against $H_A : p_M \neq p_C$. Let $\alpha = 0.05$.

[2] Miller AB. 2014. *Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ* 348 (2014): g366.

[3] During the 25 years following the screening period, each woman was screened for breast cancer according to the standard of care at her health care center.

The two-sided $p$-value is 0.895. Since $p > \alpha$, there is insufficient evidence to reject the null hypothesis; the observed difference in proportions of breast cancer deaths is reasonably explained by chance. The results do not suggest that annual mammography results in a reduction in breast cancer mortality relative to standard exams.

```
#use r as a calculator
p.hat.pooled = sum(x)/sum(n)
p.hat.pooled
```

```
## [1] 0.01118718
```

```
#check success-failure
n*p.hat.pooled
```

```
## [1] 502.5839 502.4161
```

```
n*(1 - p.hat.pooled)
```

```
## [1] 44422.42 44407.58
```

```
#conduct inference
prop.test(x = x, n = n)$p.val
```

```
## [1] 0.8948174
```

c) Calculate and interpret a 95% confidence interval for the difference in proportions of deaths from breast cancer. Be sure to check the assumptions for using the normal approximation.

The success-failure condition should be checked for each sample: from the data, the number of successes and failures are both well over 10 in each group.

The 95% confidence interval is (-0.0015, 0.0013). With 95% confidence, the difference in probability of death is within the interval (-0.15%, 0.13%); i.e., 0.15% lower in the mammogram group to 0.13% higher in the mammogram group. As expected from the large $p$-value, the confidence interval contains the null value 0.

```
#conduct inference
prop.test(x = x, n = n)$conf.int
```

```
## [1] -0.001512853  0.001282751
## attr(,"conf.level")
## [1] 0.95
```

5. Remdesivir is an antiviral drug previously tested in animal models infected with coronaviruses like SARS and MERS. As of May 2020, remdesivir had temporary approval from the FDA for use in severely ill COVID-19 patients and was the subject of numerous ongoing studies.

   A randomized controlled trial conducted in China enrolled 236 patients with severe COVID-19; 158 were assigned to receive remdesivir and 78 to receive a placebo. In the remdesivir group, 103 patients showed clinical improvement; in the placebo group, 45 patients showed clinical improvement.[4]

   a) Calculate $\hat{p}_1$ and $\hat{p}_2$, the two sample proportions of interest.

      The two sample proportions of interest are 0.652 and 0.577, the proportion of individuals in each treatment group that showed clinical improvement.

```
#use r as a calculator
x = c(103, 45)
n = c(158, 78)
p.hat.vector = x/n
p.hat.vector
```

```
## [1] 0.6518987 0.5769231
```

   b) Conduct a formal comparison of the clinical improvement rates and summarize your findings. Be sure to check the assumptions for using the normal approximation.

      Since the participants were randomly assigned to each group, the groups can be treated as independent, and it is reasonable to assume independence of patients within each group.

      The pooled proportion $\hat{p}$ is

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = 0.627$$

      The success-failure condition is met; the expected number of successes and failures are all larger than 10.

      Test $H_0 : p_1 = p_2$ against $H_A : p_1 \neq p_2$, where $p_1$ represents the population proportion of clinical improvement in COVID-19 patients treated with remdesivir and $p_2$ represents the population proportion of clinical improvement in COVID-19 patients treated with placebo. Let $\alpha = 0.05$. The $p$-value is 0.328, which is greater than $\alpha$; there is insufficient evidence to reject the null hypothesis of no difference. Even though the proportion of patients who experienced clinical improvement about 7% higher in the remdesivir group, this difference is not extreme enough to represent sufficient evidence that remdesivir is more effective than placebo.

```
#use r as a calculator
p.hat.pooled = sum(x)/sum(n)
p.hat.pooled
```

```
## [1] 0.6271186
```

[4] Wang, Y, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multi-centre trial. *Lancet* 395(10236). 16 May 2020.

```r
#check success-failure
n*p.hat.pooled
```

```
## [1] 99.08475 48.91525
```

```r
n*(1 - p.hat.pooled)
```

```
## [1] 58.91525 29.08475
```

```r
#conduct inference
prop.test(x = x, n = n)$p.val
```

```
## [1] 0.3284038
```

c) Report and interpret an appropriate interval estimate. Be sure to check the assumptions for using the normal approximation.

The success-failure condition should be checked for each sample: from the data, the number of successes and failures are both well over 10 in each group. The 95% confidence interval is (-0.067, 0.217); with 95% confidence, this interval captures the difference in population proportion of clinical mortality between COVID-19 patients treated with remdesivir and those treated with placebo. The interval contains 0, which is consistent with no statistically significant evidence of a difference. The interval reflects the lack of precision around the effect estimate that is characteristic of an insufficiently large sample size.

```r
prop.test(x = x, n = n)$conf.int
```

```
## [1] -0.06703113  0.21698245
## attr(,"conf.level")
## [1] 0.95
```