Hypothesis Testing, Power, Sample Size and Confidence Intervals (Part 2)

B.H. Robbins Scholars Series

June 23, 2010

Outline

- Comparing two proportions
  - Z-test
  - χ^2-test
  - Confidence Interval
  - Sample size and power
- Relative effect measures
- Summary example

Overview

- Compare dichotomous independent variable (predictor) with a dichotomous outcome
  - Independent variables: treatment/control, exposed/not exposed
  - Outcome variables: Diseased/not diseased, Yes/no

Recall from last time

- General form of the test statistics (T- or Z-) are

$$\frac{QOI - constant}{SE}$$

- QOI: quantity of interest (Sample average)
- Constant: A value that is consistent with H_0 (population mean under H_0)
- SE (Standard error): Square root of the variance of the numerator (e.g., uncertainty) under H_0
- How big is the signal (numerator) relative to the noise (denominator)

Normal theory test

- Two independent samples

	Sample 1	Sample 2
Sample size	n_1	n_2
Population probability of event	p_1	p_2
Sample proportion of events	$\hat{p}_1$	$\hat{p}_2$

- We want to test $H_0: \; p_1 = p_2 = p$ versus $H_A: \; p_1 \neq p_2$
- This is equivalent to testing $H_0: \; p_1 - p_2 = 0$ versus $H_A: \; p_1 - p_2 \neq 0$

Normal theory test

- $QOI = \hat{p}_1 - \hat{p}_2$
- To obtain the test statistic, we need the variance of $\hat{p}_1 - \hat{p}_2$
- From last time: $Var(\hat{p}) = \hat{p} \cdot (1 - \hat{p})/n$
- Note: Variance of a difference is equal to the sum of the variances (if they are uncorrelated)
- Variance of $\hat{p}_1 - \hat{p}_2$

$$Var(\hat{p}_1 - \hat{p}_2) = \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}$$

but under H_0 (where $p_1 = p_2 = p$)

$$Var(\hat{p}_1 - \hat{p}_2) = p(1 - p) \cdot \left[\frac{1}{n_1} + \frac{1}{n_2}\right]$$

Normal theory test

- $Var(\hat{p}_1 - \hat{p}_2)$ is estimated by using

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

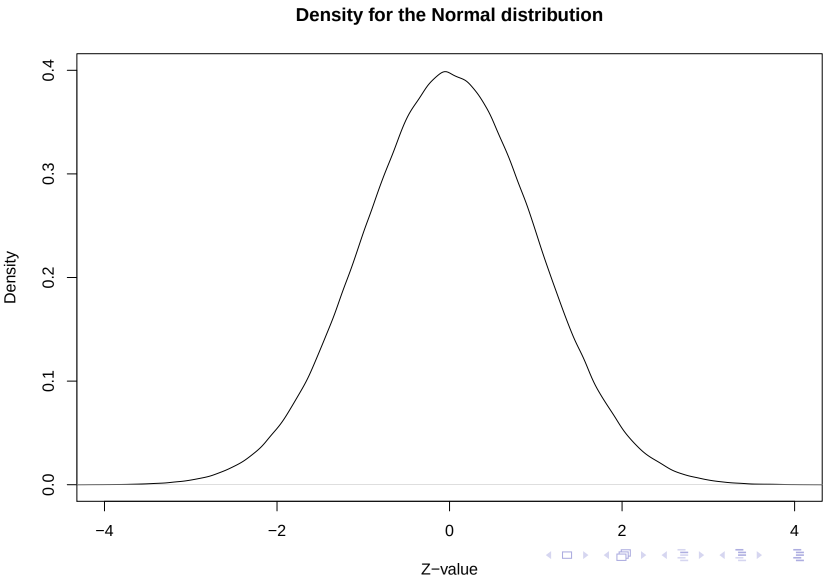
  which is the pooled estimate of the probability under $H_0 : p_1 = p_2$
- Z-test statistic is given by

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1 - \hat{p}) \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}}$$

  which has a normal distribution under H_0 if $n_i \hat{p}_i$ are large enough
- We look to see if this z-value is far out in the tails of the standard normal distribution

Normal theory test: Recall again what we are doing...

- Under $H_0: \; p_1 = p_2$
  1. Draw a sample from the target population
  2. Calculate $\hat{p}_1$, $\hat{p}_2$, $\hat{p}$, and then z
  3. Save your z value
  4. Go back to 1
- Repeat again and again.
- What does this distribution of z values look like?


Density for the Normal distribution
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
Z−value

Example

- Do women in the population who are less than 30 at first birth have the same probability of breast cancer as those who are at least 30.
- Case-control study data

	With BC	Without BC
Number of subjects (n_1 and n_2)	3220	10245
Number of subjects ≥ 30	683	1498
Sample proportions ($\hat{p}_1$ and $\hat{p}_2$)	0.212	0.146

Example

- Pooled probability:

$$\frac{683 + 1498}{3220 + 10245} = 0.162$$

- Calculate variance under H_0

$$Var(\hat{p}_1 - \hat{p}_2) = \hat{p}(1 - \hat{p}) \left[\frac{1}{n_1} + \frac{1}{n_2} \right] = 5.54 \times 10^{-5}$$

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{Var(\hat{p}_1 - \hat{p}_2)} = 0.00744$$

- Test statistic

$$z = \frac{0.212 - 0.146}{0.00744} = 8.85$$

- Two tailed p-value is effectively 0

χ^2 test

- If Z$\sim$ $N(0,1)$ then $Z^2 \sim \chi^2$ with 1 d.f. (e.g., testing a single difference against 0)
- The data just discussed can be shown in a two by two table

	With BC	Without BC	total
Age $\geq$ 30	683	1498	2181
Age $<$ 30	2537	8747	11284
Total	3220	10245	13465

χ^2 test

- Two by two table:

	Disease	No disease	total
Exposed	a	b	a+b
Not Exposed	c	d	c+d
Total	a+c	b+d	N=a+b+c+d

- Like the other test statstics, the χ^2 examines the difference between what is observed and what we expected to observe if H_0 is true

$$\chi^2_1 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(Observed_{ij} - Expected_{ij})^2}{Expected_{ij}}$$

- $Observed_{ij}$: is the observed frequency in cell (i, j)
- $Expected_{ij}$: is the expected cell frequency if H_0 is true.

χ^2 test

- Under the null hypothesis, the rows and the columns are independent of one another
  - Having age < 30 or age ≥ 30 provides no information about presence/absence of BC
  - Presence / absence of BC provides no information about age
- Under Independence: $Expected_{ij} = \frac{Row\ \ i\ \ total \times\ \ Column\ \ j\ \ total}{grand\ \ total}$
- $Expected_{21} = \frac{11284 \times 3220}{13465} = 2698.4$

χ^2 test

- Observed

	With BC	Without BC
Age ≥ 30	683	1498
Age < 30	2537	8747

- Expected

	With BC	Without BC
Age ≥ 30	522	1659
Age < 30	2698	8586

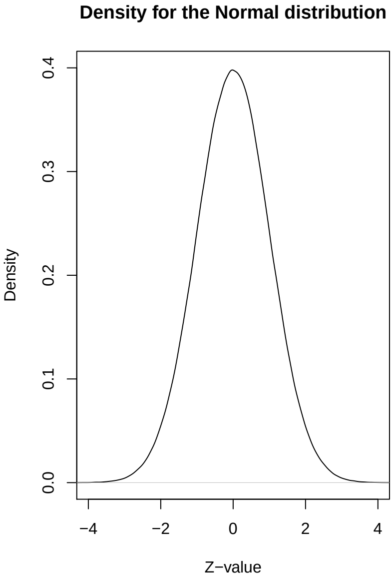
$$\chi_1^2 = \frac{(683-522)^2}{522} + \frac{(1498-1659)^2}{1659} + \ldots$$

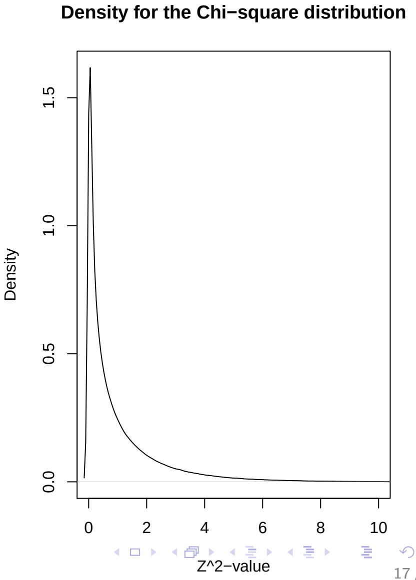
χ^2 test

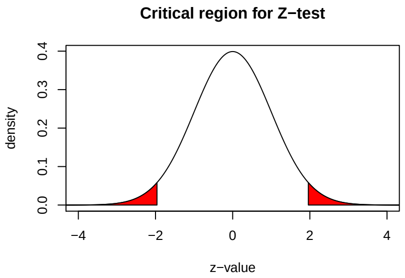
- For 2 by 2 tables the χ^2 test statistic can also be written

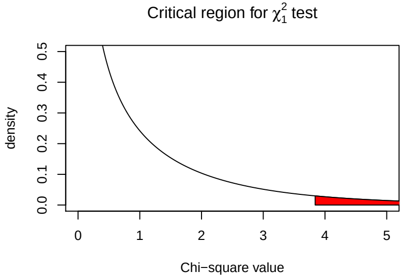
$$\frac{N(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$$

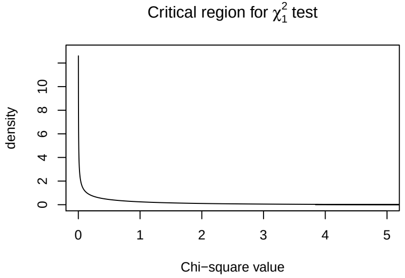
- In this example, $\chi^2_1 = 78.37$
- This χ^2 value is z^2 from earlier.
- The two sided critical value for the χ^2_1 test is 3.84
- Note that even though we're doing a 2-tailed test we only use the right tail in the χ^2 test
- We've squared the difference when computing the statistic and so the sign is lost
- This is called the ordinary Pearson's χ^2 test


Density for the Normal distribution
Density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
Z−value


Density for the Chi−square distribution
Density
0.0
0.5
1.0
1.5
0
2
4
6
8
10
Z^2−value


Critical region for Z−test
density
0.0
0.1
0.2
0.3
0.4
−4
−2
0
2
4
z−value


Critical region for χ^2_1 test
density
0.0
0.1
0.2
0.3
0.4
0.5
0
1
2
3
4
5
Chi−square value


Critical region for χ^2_1 test
density
0
2
4
6
8
10
0
1
2
3
4
5
Chi−square value

Fisher's Exact Test

- Meant for small sample sizes but is a conservative test (e.g., reject H_0 less often than you could or should)
- Pearson's χ^2 test work fine often even when expected cell counts are less than 5 (contrary to popular belief)

Confidence Interval

- An approximate $1-\alpha$ two-sided ci is

$$\hat{p}_1 - \hat{p}_2 \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

- $z_{1-\alpha/2}$ is the critical value for the normal distribution (1.96 when $\alpha = 0.05$).
- The confidence limits for the number of patients needed to treat (NNT) to save one event is given by the reciprocal of the two confidence limits.

Confidence Interval: Physicians Health Study

- Five-year randomized study of whether regular aspirin intake reduces mortality due to CVD
- 11037 randomized to receive daily aspirin dose and 11043 randomized to placebo
- Let's only consider incidence of an MI over the 5 year time frame

	With MI	Without MI	total
Placebo	198	10845	11043
Aspirin	104	10933	11037
Total	302	21778	22080

Confidence Interval Example: Physicians Health Study

$$
\begin{aligned}
\hat{p}_1 &= 198/11043 = 0.0179 \\
\hat{p}_2 &= 104/11037 = 0.0094 \\
\hat{p}_1 - \hat{p}_2 &= 0.0085 \\
(LCI, UCI) &= (0.0054, 0.0116)
\end{aligned}
$$

- 1.79/0.94 percent chance of having an MI in 5 years on placebo/medication
- The difference (CI) is 0.85 percent (0.54,1.16).
- Number needed to treat (NNT) to prevent an MI is the inverse of these quantities
- The interval $(86.44, 183.64)$ contains the NNT to prevent one MI (with 95 percent confidence).

Sample size and power for comparison two proportions

- Power increases when
  - Sample size (n_1 and n_2) increases
  - As $n_1/n_2 \to 1$
  - As $\Delta = |p_1 - p_2|$ increases
- Power calculation (http://statpages.org/#Power)
  - Current therapy: 50% infection-free at 24 hours.
  - New therapy: 70% infection-free
  - Randomly assign $n_1 = 100$ subjects to receive standard care and $n_2 = 100$ to receive new therapy.
  - What is the power to detect a significant difference between the two therapies ($\alpha = 0.05$)?
  - $p_1 = 0.5$, and $p_2 = 0.7$, $n_1 = n_2 = 100$
  - Power is 0.83.

Sample size and power for comparison two proportions

- Required sample size decreases when
  - As $n_1/n_2 \to 1$
  - As $\Delta = |p_1 - p_2|$ increases
- Required sample size depends on both p_1 and p_2
- Example: Number of subjects needed to detect a 20 percent decrease in the probability of colorectal cancer if baseline probability of cancer is 0.15 percent

$$p_1 = 0.0015,\ p_2 = 0.8 \times p_1 = 0.0012,$$
$$\alpha = 0.05,\ \beta = 0.2,\ \ n_1 = n_2 \sim 235145$$

Relative effect measures

- So far, we have discussed absolute risk differences
- Measures of relative effect include risk ratios and odds ratios

$$RR = \frac{p_2}{p_1}$$

$$OR = \frac{p_2/(1-p_2)}{p_1/(1-p_1)}$$

- RRs are easier to interpret than ORs but they have problems (e.g., a risk factor that doubles your risk of lung cancer cannot apply to a subject having a risk of 50 percent)
- ORs can apply to any subject
- Testing $H_0: \ p_1 = p_2$ is equivalent to $H_0: \ OR = 1$
- There are formulas for computing CIs for odds ratios, although we usually compute CIs for ORs by anti-logging CIs for log ORs based on logistic regression

Summary Example

- Consider patients who undergo CABG surgery
- Study question: Do emergency cases have a surgical mortality rate that is different from that of non-emergent cases.
- Population probabilities
  - p_1 : Probability of death in patients with emergency priority
  - p_2 : Probability of death in patients with non-emergency priority
- Statistical Hypotheses
  - $H_0: \ p_1 = p_2$ or $H_0: \ OR = 1$
  - $H_A: \ p_1 \neq p_2$ or $H_0: \ OR \neq 1$

Summary Example: Power

- Prior Research shows that just over 10 percent of non-emergent surgeries result in death
- Researcher want to be able to detect a 3-fold increase in risk in death
- For every 1 emergency priority, we expect 10 non-emergency
- $p_1 = 0.3$, $p_2 = 0.1$, $\alpha = 0.05$ and power=0.90
- Calculate the sample sizes (done using PS software) for these and other combinations of p_1 and p_2

(p_1, p_2)	(0.3, 0.1)	(0.4, 0.2)	(0.03, 0.01)	(0.9, 0.7)
n_1	40	56	589	40
n_2	400	560	5890	400

Summary Example: Data

- In hospital mortality for emergency or other surgery

	Outcome	
Priority	Dead	alive
Emergency	6	19
Other	11	100

- $\hat{p}_1 = \frac{6}{25} = 0.24$
- $\hat{p}_2 = \frac{11}{111} = 0.10$
- $\hat{p}_1 - \hat{p}_2 = 0.14$
- $95\% CI = (-0.035, 0.317)$
- p-values:
  - Fisher exact test: 0.087
  - Pearson χ^2: 0.054

Summary Example: Interpretation

- Compare the odds of death in the emergency group $\frac{p_1}{1-p_1}$ versus the other group $\frac{p_2}{1-p_2}$

$$OR_{1,2} = \frac{0.24}{0.76} / \frac{0.1}{0.9} = 2.87$$

- Emergency cases are estimated to have 2.87 fold increase in odds of death during surgery compared to non-emergency cases with $95\% CI: \ [0.95, 3.36]$