Mathematical Biostatistics Boot Camp 2: Lecture 4, Two Sample Binomial Tests

Brian Caffo

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Johns Hopkins University

April 22, 2013

Table of contents

Motivation

- Consider a randomized trial where 40 subjects were randomized (20 each) to two drugs with the same active ingredient but different expedients
- Consider counting the number of subjects with side effects for each drug

	Side Effects	None	total
Drug A	11	9	20
Drug B	5	15	20
Total	16	14	40

Hypothesis tests for binomial proportions

- Consider testing $H_0 : p = p_0$ for a binomial proportion
- The **score** test statistic
$$\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$
follows a Z distribution for large n
- This test performs better than the Wald test
$$\frac{\hat{p} - p_0}{\sqrt{\hat{p}(1 - \hat{p})/n}}$$

Inverting the two intervals

- Inverting the Wald test yields the Wald interval

$$\hat{p} \pm Z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

- Inverting the Score test yields the Score interval

$$\hat{p}\left(\frac{n}{n+Z_{1-\alpha/2}^2}\right) + \frac{1}{2}\left(\frac{Z_{1-\alpha/2}^2}{n+Z_{1-\alpha/2}^2}\right)$$

$$\pm Z_{1-\alpha/2}\sqrt{\frac{1}{n+Z_{1-\alpha/2}^2}\left[\hat{p}(1-\hat{p})\left(\frac{n}{n+Z_{1-\alpha/2}^2}\right) + \frac{1}{4}\left(\frac{Z_{1-\alpha/2}^2}{n+Z_{1-\alpha/2}^2}\right)\right]}$$

- Plugging in $Z_{1-\alpha/2} = 2$ yields the Agresti/Coull interval

Example

- In our previous example consider testing whether or not Drug A's percentage of subjects with side effects is greater than 10%
- $H_0 : p_A = .1$ verus $H_A : p_A > .1$
- $\hat p = 11/20 = .55$
- Test Statistic
$$\frac{.55 - .1}{\sqrt{.1 \times .9/20}} = 6.7$$
- Reject, pvalue $= P(Z > 6.7) \approx 0$

Exact binomial tests

- Consider calculating an exact P-value
- What's the probability, under the null hypothesis, of getting evidence as extreme or more extreme than we obtained?

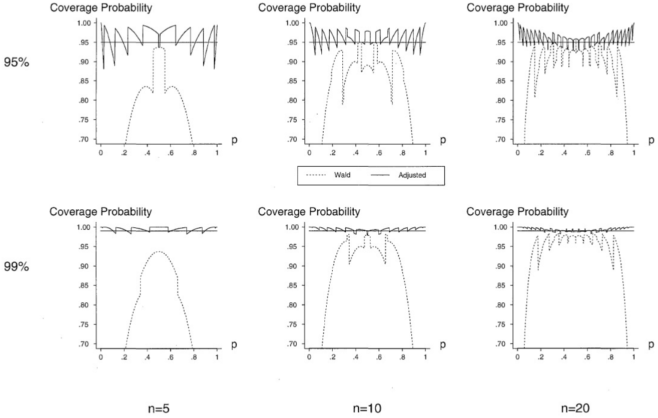
$$P(X_A \geq 11) = \sum_{x=11}^{20} \binom{20}{x} .1^x \times .9^{20-x} \approx 0$$

- `pbinom(10, 20, .1, lower.tail = FALSE)`
- `binom.test(11, 20, .1, alternative = "greater")`

Notes on exact binomial tests

- This test, unlike the asymptotic ones, guarantees the Type I error rate is less than desired level; sometimes it is much less
- Inverting the exact binomial test yields an exact binomial interval for the true proprotion
- This interval (the Clopper/Pearson interval) has coverage greater than 95%, though can be very conservative
- For two sided tests, calculate the two one sided P-values and double the smaller

Wald versus Agresti / Coull[1]



[1] Taken from Agresti and Caffo (2000) TAS

Comparing two binomials

- Consider now testing whether the proportion of side effects is the same in the two groups
- Let $X \sim \mathrm{Binomial}(n_1, p_1)$ and $\hat{p}_1 = X / n_1$
- Let $Y \sim \mathrm{Binomial}(n_2, p_2)$ and $\hat{p}_2 = Y / n_2$
- We also use the following notation:

$n_{11} = X$	$n_{12} = n_1 - X$	$n_1 = n_{1+}$
$n_{21} = Y$	$n_{22} = n_2 - Y$	$n_2 = n_{2+}$
n_{+1}	n_{+2}	

Comparing two proportions

- Consider testing $H_0 : p_1 = p_2$
- Versus $H_1 : p_1 \neq p_2$, $H_2 : p_1 > p_2$, $H_3 : p_1 < p_2$
- The score test statstic for this null hypothesis is

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

  where $\hat{p} = \frac{X+Y}{n_1+n_2}$ is the estimate of the common proportion under the null hypothesis
- This statistic is normally distributed for large n_1 and n_2.

Continued

- This interval does not have a closed form inverse for creating a confidence interval (though the numerical interval obtained performs well)
- An alternate interval inverts the Wald test

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

- The resulting confidence interval is

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

Continued

- As in the one sample case, the Wald interval and test performs poorly relative to the score interval and test
- For testing, always use the score test
- For intervals, inverting the score test is hard and not offered in standard software
- A simple fix is the Agresti/Caffo interval which is obtained by calculating $\tilde{p}_1 = \frac{x+1}{n_1+2}$, $\tilde{n}_1 = n_1 + 2$, $\tilde{p}_2 = \frac{y+1}{n_2+2}$ and $\tilde{n}_2 = (n_2 + 2)$
- Using these, simply construct the Wald interval
- This interval does not approximate the score interval, but does perform better than the Wald interval

Example

- Test whether or not the proportion of side effects is the same for the two drugs
- $\hat{p}_A = .55$, $\hat{p}_B = 5/20 = .25$, $\hat{p} = 16/40 = .4$
- Test statistic
$$\frac{.55 - .25}{\sqrt{.4 \times .6 \times (1/20 + 1/20)}} = 1.61$$
- Fail to reject H_0 at .05 level (compare with 1.96)
- P-value $P(|Z| \geq 1.61) = .11$

Wald versus Agresti /Caffo[2]

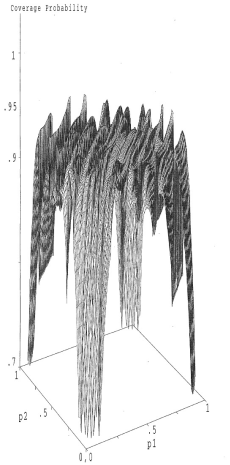


Figure 7. Coverage probabilities for 95% nominal Wald confidence interval as a function of p1 and p2, when n1 = n2 = 10.

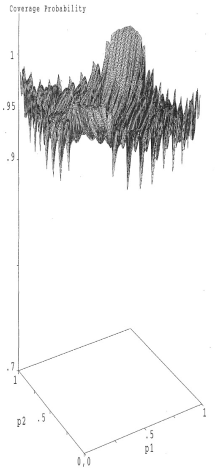


Figure 8. Coverage probabilities for 95% nominal adjusted confidence interval (adding t = 4 pseudo observations) as a function of p1 and p2, when n1 = n2 = 10.

[2] Taken from Agresti and Caffo (2000) TAS

Wald versus Agresti /Caffo[3]

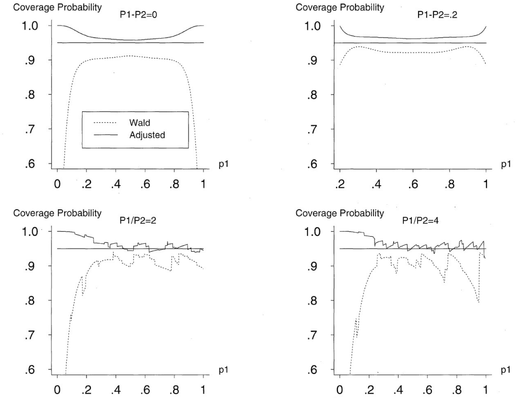


Figure 6. Coverage probabilities for nominal 95% Wald and adjusted confidence intervals (adding t = 4 pseudo observations) as a function of p1 when p1 − p2 = 0 or .2 and when p1/p2 = 2 or 4, for n1 = n2 = 10.

[3] Taken from Agresti and Caffo (2000) TAS

Bayesian and likelihood inference for two binomial proportions

- Likelihood analysis requires the use of profile likelihoods, or some other technique and so we omit their discussion
- Consider putting independent Beta(α_1, β_1) and Beta(α_2, β_2) priors on p_1 and p_2 respectively
- Then the posterior is

$$\pi(p_1, p_2) \propto p_1^{x+\alpha_1-1}(1-p_1)^{n_1+\beta_1-1} \times p_2^{y+\alpha_2-1}(1-p_2)^{n_2+\beta_2-1}$$

- Hence under this (potentially naive) prior, the posterior for p_1 and p_2 are independent betas
- The easiest way to explore this posterior is via Monte Carlo simulation

```
x <- 11; n1 <- 20; alpha1 <- 1; beta1 <- 1
y <-  5; n2 <- 20; alpha2 <- 1; beta2 <- 1
p1 <- rbeta(1000, x + alpha1, n - x + beta1)
p2 <- rbeta(1000, y + alpha2, n - y + beta2)
rd <- p2 - p1
plot(density(rd))
quantile(rd, c(.025, .975))
mean(rd)
median(rd)
```

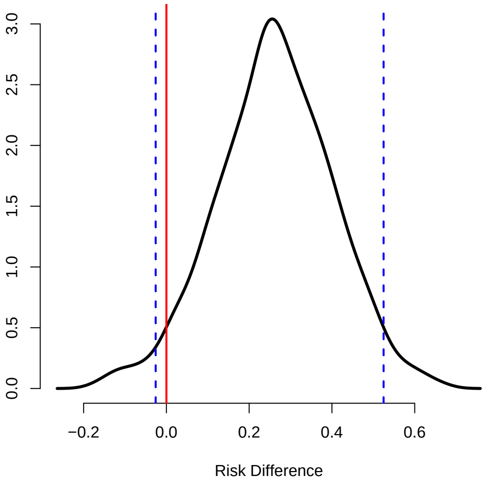
- The function `twoBinomPost` on the course web site automates a lot of this
- The output is

```
Post mn rd (mcse) = -0.278 (0.004)
Post mn rr (mcse) = 0.512 (0.007)
Post mn or (mcse) = 0.352 (0.008)

Post med rd        =  -0.283
Post med rr        =  0.485
Post med or        =  0.288

Post mod rd        =  -0.287
Post mod rr        =  0.433
Post mor or        =  0.241

Equi-tail rd       =  -0.531 -0.008
Equi-tail rr       =  0.195 0.98
Equi-tail or       =  0.074 0.966
```


0.0
0.5
1.0
1.5
2.0
2.5
3.0
−0.2
0.0
0.2
0.4
0.6
Risk Difference