

Summary Sheet for Binary Outcomes (One Sample and Two Samples)

1. One-sample test and confidence intervals

(a) One-sample test for binomial proportions.

- Observed data: $X \sim \text{binomial}(n, p)$.
- Parameter of interest: population proportion (μ).
- Hypotheses:
 - $H_0 : p = p_0$.
 - $H_1 : p \neq p_0$.
- Test statistic: $Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$.
- Exact binomial probability (small n) or normal approximation (large n).
- Sample R commands:


```
> binom.test(3, 10, p=.4)
```

Exact binomial test

```
data: 3 and 10
number of successes = 3, number of trials = 10, p-value = 0.7492
alternative hypothesis: true probability of success is not equal to 0.4
95 percent confidence interval:
 0.06673951 0.65245285
sample estimates:
probability of success
                0.3
```

(b) Confidence intervals for binomial proportions.

i. Clopper-Pearson ("exact") interval.

- Based on cumulative probabilities of binomial distribution.
- Guarantees minimum coverage of $1 - \alpha$ for all p ("exact").
- Can be (very) conservative due to discrete nature of binomial distribution.
- Sample R commands:

```
> library(Hmisc)
> binconf(3, 10, alpha=.01, method="exact")
PointEst      Lower      Upper
      0.3 0.03700722 0.735114
```

ii. Normal approximation (asymptotic) interval.

- Based on normal approximation to binomial distribution (CLT).
- Inversion of Wald test: $\{p \mid z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})/n}} < z_{1-\alpha/2}\}$.
- Formula: $\hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
- Poor performance for small numbers of trials and/or extreme proportions, e.g. p close to 0 or 1.
- Sample R commands:

```
> library(Hmisc)
> binconf(3,10,alpha=.01,method="asymptotic")
PointEst      Lower      Upper
0.3 -0.07327313 0.6732731
```

iii. Wilson score interval (recommended).

- Based on normal approximation to binomial distribution (CLT).
- Inversion of score test: $\{p \mid z_{\alpha/2} < \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} < z_{1-\alpha/2}\}$.
- Formula:

$$\frac{\hat{p} + \frac{1}{2n}z_{1-\alpha/2} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{1-\alpha/2}^2}{4n^2}}}{1 + \frac{1}{n}z_{1-\alpha/2}^2}$$

- Good performance even for small numbers of trials and/or extreme probabilities.
- Sample R commands:

```
> library(Hmisc)
> binconf(3,10,alpha=.01,method="wilson")
PointEst      Lower      Upper
0.3 0.07956632 0.6799753
> binconf(3,10,alpha=.01)
PointEst      Lower      Upper
0.3 0.07956632 0.6799753
```

2. Two-sample test and confidence intervals.

(a) Two-sample test for binomial proportions.

- Observed data: $X_1 \sim \text{binomial}(n_1, p_1)$, $X_2 \sim \text{binomial}(n_2, p_2)$.
 - Can be written as 2×2 contingency table:

	Event	No Event	Total
Group 1	a	b	$n_1 = a + b$
Group 2	c	d	$n_2 = c + d$
Total	$a + c$	$b + d$	$N = a + b + c + d$

- Hypotheses:
 - $H_0 : p_1 = p_2$.
 - $H_1 : p_1 \neq p_2$.
- Test statistic: $X^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi_1^2$ under H_0 .
- Sample R commands:

```
> tmp=cbind(c(15,40),c(55,96))
> tmp
      [,1] [,2]
[1,]  15  55
[2,]  40  96
> chisq.test(tmp,correct=F)
```

Pearson's Chi-squared test

```
data: tmp
X-squared = 1.5049, df = 1, p-value = 0.2199
```

(b) Measures of effect and confidence intervals.

i. Risk difference ($p_1 - p_2$).

- Confidence interval based on normal approximation to binomial distribution (CLT).

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

ii. Risk ratio or relative risk ($RR = p_1/p_2$).

- Confidence interval based on normal approximation to distribution of $\log \hat{RR}$.

- Formula: $\log \hat{RR} \pm z_{1-\alpha/2} \sqrt{\frac{1-\hat{p}_1}{n_1\hat{p}_1} + \frac{1-\hat{p}_2}{n_2\hat{p}_2}}$.

iii. Odds ratio ($OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$).

- Probability of an event can be equivalently expressed as odds, $p/(1-p)$.
- Odds ratio is the ratio of the odds in the first group to the odds in the second group.
- For rare events, odds ratio is very close to risk ratio.
- Odds ratio can be estimated from both prospective and retrospective (case-control) studies.
- Confidence interval based on normal approximation to distribution of $\log \hat{OR}$.

- Formula: $\log \hat{OR} \pm z_{1-\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$.

iv. Sample R commands:

```
> library(Epi)
> #twoby2 uses second row as reference category
> #can only be used with 2x2 tables
> twoby2(tmp,alpha=.05)
2 by 2 table analysis:
```

```
-----
Outcome      : Col 1
Comparing    : Row 1 vs. Row 2

      Col 1 Col 2      P(Col 1) 95% conf. interval
Row 1     15   55         0.2143    0.1335    0.3256
Row 2     40   96         0.2941    0.2237    0.3760

                                     95% conf. interval
      Relative Risk:  0.7286    0.4337    1.2238
      Sample Odds Ratio: 0.6545    0.3317    1.2916
      Conditional MLE Odds Ratio: 0.6559    0.3074    1.3458
      Probability difference: -0.0798   -0.2089    0.0697

      Exact P-value: 0.2475
      Asymptotic P-value: 0.2217
-----
```

```
> library(epitools)
> #riskratio and oddsratio use first row as reference category
> #can be used with rX2 tables
```

```
> riskratio(tmp,method='wald',conf.level=.99,rev='columns')
$data
```

Predictor	Outcome		
	Disease2	Disease1	Total
Exposed1	55	15	70
Exposed2	96	40	136
Total	151	55	206

```
$measure
```

Predictor	risk ratio with 99% C.I.		
	estimate	lower	upper
Exposed1	1.000000	NA	NA
Exposed2	1.372549	0.6942195	2.713682

```
$p.value
```

Predictor	two-sided		
	midp.exact	fisher.exact	chi.square
Exposed1	NA	NA	NA
Exposed2	0.2245576	0.2474667	0.2199135

```
$correction
```

```
[1] FALSE
```

```
attr(,"method")
```

```
[1] "Unconditional MLE & normal approximation (Wald) CI"
```

```
> oddsratio(tmp,method='wald',conf.level=.99,rev='columns')
```

```
$data
```

Predictor	Outcome		
	Disease2	Disease1	Total
Exposed1	55	15	70
Exposed2	96	40	136
Total	151	55	206

```
$measure
```

Predictor	odds ratio with 99% C.I.		
	estimate	lower	upper
Exposed1	1.000000	NA	NA
Exposed2	1.527778	0.6253371	3.732555

```
$p.value
```

Predictor	two-sided		
	midp.exact	fisher.exact	chi.square
Exposed1	NA	NA	NA
Exposed2	0.2245576	0.2474667	0.2199135

```
$correction
```

```
[1] FALSE
```

```
attr(,"method")
```

```
[1] "Unconditional MLE & normal approximation (Wald) CI"
```

```

> tmp2=cbind(c(12,34,56,78),c(87,65,43,21))
> tmp2
      [,1] [,2]
[1,]  12  87
[2,]  34  65
[3,]  56  43
[4,]  78  21
> riskratio(tmp2,method='wald',conf.level=.99,rev='columns')
$data
      Outcome
Predictor Disease2 Disease1 Total
  Exposed1      87      12     99
  Exposed2      65      34     99
  Exposed3      43      56     99
  Exposed4      21      78     99
  Total         216     180    396

$measure
      risk ratio with 99% C.I.
Predictor estimate      lower      upper
  Exposed1 1.000000      NA      NA
  Exposed2 2.833333 1.294162  6.203069
  Exposed3 4.666667 2.242079  9.713209
  Exposed4 6.500000 3.196064 13.219384

$p.value
      two-sided
Predictor midp.exact fisher.exact  chi.square
  Exposed1      NA      NA      NA
  Exposed2 2.119365e-04 3.409493e-04 2.137762e-04
  Exposed3 1.992317e-11 3.617778e-11 4.547294e-11
  Exposed4 0.000000e+00 3.072554e-22 4.517879e-21

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"
> oddsratio(tmp2,method='wald',conf.level=.99,rev='columns')
$data
      Outcome
Predictor Disease2 Disease1 Total
  Exposed1      87      12     99
  Exposed2      65      34     99
  Exposed3      43      56     99
  Exposed4      21      78     99
  Total         216     180    396

$measure
      odds ratio with 99% C.I.

```

Predictor	estimate	lower	upper
Exposed1	1.000000	NA	NA
Exposed2	3.792308	1.448434	9.929068
Exposed3	9.441860	3.652609	24.406861
Exposed4	26.928571	9.759189	74.304121

\$p.value

Predictor	two-sided		
	midp.exact	fisher.exact	chi.square
Exposed1	NA	NA	NA
Exposed2	2.119365e-04	3.409493e-04	2.137762e-04
Exposed3	1.992317e-11	3.617778e-11	4.547294e-11
Exposed4	0.000000e+00	3.072554e-22	4.517879e-21

\$correction

[1] FALSE

attr(,"method")

[1] "Unconditional MLE & normal approximation (Wald) CI"