

2 Binomial Test (Section 2.1 of H&W):The dichotomous data problem

2.1 Binomial Test

2.1.1 Data

Dichotomous outcomes of independent Bernoulli trials having constant probability of success p .
Want to make inference about p .

2.1.1.1 Example 1

Success prob p for certain condition (A) is established by previous studies to be 15%. We believe for the new condition (B), the success probability is higher. We observe 6 success out of 7 trials for the new condition. Do this new condition have the same probability of success as the original condition? (condition A and B can be: control vs treatment, old drug vs new drug etc, old therapy and new therapy)

2.1.2 Assumptions:

Check the following!

1. the outcome of each trial is either success or fail (T or F, H or T,...)
2. the prob of success p remains constant
3. the n trials are independent

Want to test $H_0: p=p_0$, for some p_0 b/w 0 and 1.

Let $B = \#$ of success

2.1.3 Test Procedure (One-sided upper-tail test)

To test H_0 vs $H_1: p > p_0$, at the α level of significance,

"Reject H_0 if $B \geq b_\alpha$; do not reject otherwise"

Where b_α is chosen to make type I error prob = α .

$\text{Prob}(H_0 \text{ is true but reject } H_0) = \text{Prob}(B \text{ has the success probability } p_0 \text{ but } B \geq b_\alpha) = P_{p_0}(B \geq b_\alpha) = \alpha$

b_α is the upper α percentile point of the binomial distribution with sample size n and success prob p_0 , $\text{bin}(n, p_0)$

$P_{p_0}(\dots)$ is probability of the event .. when the true success probability is p_0 .

Prob(...) is a generic probability of the event ..

P-value = the lowest significance level at which we can reject with the observed data b

= Prob(H_0 is true but we observe the data 'as extreme' as what we have)
 for $\alpha \geq P\text{-value}$, b will be significant. For $\alpha < P\text{-value}$, b will NOT be significant.
 = Prob(H_0 is true but $B \geq b$) for upper tail binomial case

Q Why P-value:

Convey more information. (OK. We reject. But how significant? Just barely or VERY significant?)

Power of a test under a particular alternative:

Prob(Correctly rejecting H_0)
 = 1- Prob(Incorrectly accepting H_0)
 = 1- Prob(Type II error)
 = 1- Prob(H_0 is not true but accept H_0)

Table

	Accept H_0	Reject H_0
H_0 true	.	Type I error (false alarm)
H_0 false	Type II error	Power of the test

2.1.4 Example

Suppose $n=8$ and test $H_0: p=.4$

H_0 versus $p > .4$.

$P_{.4}(B \geq b) = \text{Prob}(\# \text{ of successes } \geq b \text{ if true success probability is } .4)$

(make room for c table below)

c										
b	0	1	2	3	4	5	6	7	8	
p	1.0000	0.9832	0.8936	0.6846	0.4059	0.1737	0.0498	0.0085	0.0007	

What is $\alpha=.0085$ upper-tail test? Reject when $B \geq b_{.0085}=7$ (Critical region = {7,8})

What is $\alpha=.0498$ upper-tail test? Reject when $B \geq b_{.0498}=6$ (Critical region = {6, 7, 8})

What is $\alpha=.1737$ upper-tail test? Reject when $B \geq b_{.1737}=5$ (Critical region = {5, 6, 7, 8})

At $\alpha = .0085$ and $.0498$ and $.1737$ (and) [making a table-like]

- Is $B=8$ significant at each level? Yes, Yes, Yes, (Yes...)
- Is $B=7$ significant at each level? Yes, Yes, Yes, (Yes...)
- Is $B=6$ significant at each level? No, Yes, Yes, (Yes,...)
- Is $B=5$ significant at each level? No, No, Yes, (Yes...)
- Is $B=4$ significant at each level? No, No, No, (...)
- Is $B=3, 2, 1$ significant at each level? No, No, No, (...)

As α increases:

- Q the critical constant b_α decreases [draw a line above]
- Q it is easier to reject H_0 (critical region grows)
- Q increase the power, or decrease the probability of a type II error [but there's no free lunch. Type I error increase! Tradeoff b/w Type I and II errors]

Q What is the p-value for data observed to be $B=7$.
 Significant at $\alpha = .0085$ and $.0498$ and $.1737$
 Simply specifying " $B=7$ is significant at $.0498$ is not enough!"

Explain lower-tail test and two-sided test

H_0 versus $p < .4$.

$$P_{.4}(B \leq c) = \text{Prob}(\# \text{ of successes} \leq c \text{ if true success probability is } .4)$$

$$= 1 - \text{Prob}(\# \text{ of successes} \geq c+1 \text{ if true success probability is } .4)$$

$$= 1 - P_{.4}(B \geq c+1)$$

$$P_{.4}(B \leq 0) = 1 - P_{.4}(B \geq 1)$$

Q $P_{.4}(B \leq 8) = 1$

c	0	1	2	3	4	5	6	7	8
0	0.0168	0.1064	0.3154	0.5941	0.8263	0.9502	0.9915	0.9993	1

What is $\alpha = .1064$ test? Reject when $B \leq c_{.1064} = 1$ (Critical region = {0, 1})

Q What's the level of the two-sided test which rejects when $B \geq 6$ or $B \leq 1$?
 $\alpha = \alpha_1 + \alpha_2 = P_{.4}(B \geq 6) + P_{.4}(B \leq 1) = .0498 + .1064 = .1562$.

Pretty useless.. Use normal approximation

POWER CONSIDERATION. Consider two tests for $H_1: p > .4$: (revisit above)

T1 for level $\alpha = .0085$: rejects if $B \geq 7$

T2 for level $\alpha = .0498$: rejects if $B \geq 6$

Suppose in fact that the alternative $p = .5$ is true.

$$R_1 = \text{power of T1} = P_{.5}(B \geq 7) = .0352$$

$$R_2 = \text{power of T2} = P_{.5}(B \geq 6) = .1445$$

Probability of type II errors are 1-the above = $.9648$ and $.8555$.

T1 has lower type I error but higher type II error prob.

2.1.4.1A toy example for one-sided test

Example 1: $H_0: p = .15$, $H_1: p > .15$, $B = 6$, $n = 7$, for $\alpha = .0121$, $b_{.0121} = 4$.

Thus, reject H_0 when $B \geq 4$. Thus, we reject it. The p-value $P_{.15}(B \geq 6) = .0001$.

p-value: the smallest significance level at which we can reject H_0 in favor of the alternative with our observed value B .

2.1.5 Test Procedure (One-sided lower-tail test)

To test H_0 vs $H_2 : p < p_0$, at the α level of significance,

"Reject H_0 if $B \leq c_\alpha$; do not reject otherwise"

Where c_α is chosen to make type I error prob = α .

$$P_0(B \leq c_\alpha) = \alpha$$

b_α is the lower α percentile point of the binomial distribution with sample size n and success prob p_0 , $\text{bin}(n, p_0)$

* For the special case of testing $p=1/2$, $c_\alpha = n - b_\alpha$.

2.1.6 Test Procedure (Two-sided test)

To test H_0 vs $H_3 : p \neq p_0$, at the α level of significance,

"Reject H_0 if $B \geq b_\alpha$ or $B \leq c_\alpha$; do not reject otherwise"

Where b_α, c_α are chosen to make type I error prob = α .

$$P_0(B \geq b_\alpha \text{ or } B \leq c_\alpha) = \alpha$$

b_α is the lower α percentile point, c_α is the lower α percentile point, and $\alpha_1 + \alpha_2 = \alpha$.

2.1.7 Large sample approximation

When H_0 is true,

$$E_{p_0}(B) = np_0$$

$$\text{Var}_{p_0}(B) = np_0(1-p_0)$$

The standardized version of B

$$B^* = (B - E_{p_0}(B)) / \sqrt{\text{Var}_{p_0}(B)}$$

Has an "asymptotic" $N(0,1)$ distribution as n tends to infinity.

Let z_α be the upper α percentile point of the standard normal $N(0,1)$ distribution.

If $Z \sim N(0,1)$, it is such value that $\text{Prob}(Z \geq z_\alpha) = \alpha$.

The normal approximation to the three procedures are:

Reject H_0 if $B^* \geq z_\alpha$; do not reject otherwise

Reject H_0 if $B^* \leq -z_\alpha$; do not reject otherwise

For $\alpha_1 = \alpha_2 = \alpha/2$, Reject H_0 if $|B^*| \geq z_{\alpha/2}$; do not reject otherwise

2.1.7.1 Example

$H_0: p=1/3$ and the one-sided alternative $p > 1/3$

Out of $n=50$ trials, there were 25 correct selections.

$$B^* = 2.5$$

Q Can we reject it at the level $\alpha=.05$? $z_{.05}=1.645$ Yes.

Q What are other levels that we can reject H_0 as well?

$P(Z>2.5) = .0062$. For all α larger than this, we can reject H_0 , since for all $\alpha>.0062$, z_α is smaller than 2.5.

POWER COMPUTATION. (GRAD ONLY)

After doing normal approx.

Consider level $\alpha=.05$ test. (reject the null if $B^* > 1.645$)

What's the power if in fact $p=.6$?

$$\begin{aligned} P_6(B^*>1.645) &= P_6(B-n/3/(n1/32/3)^{1/2} > 1.645) = P_6(B>...) \\ &= P_6(B-n.6/(n.6.4)^{1/2} >...) \\ &= P(Z>-2.27) = P(Z>-2.27) = .9884 \end{aligned}$$

Dealing with $p>.5$.

When $p>.5$, Consider $C=n-B$ (= # of failures). Then $C \sim \text{bin}(n, 1-p)$. $1-p$ = prob of 'failure'

$P(B>b) = P(C<=n-b)$ where c

If $n=10$, $p=0.7$, $b=8$, Want $P(B>=8)$. $B=$

$$P(B>=8) = P(C<=2) = 1 - P(C>=3) = 1 - .6172 = .3828$$

Also for $p=.5$, $P(B>b) = P(B<=n-b)$.

2.2 Estimating the probability of success

Q. Ever wondered why don't we reject why we reject if B is too large or too small?

The estimator of p , associated with the statistic B is

$$\hat{p} = B/n$$

Variance(\hat{p})= $p(1-p)/n$ (Good exercise if you know the theory. Skip it otherwise.)

Standard dev = $sd(\hat{p}) = \sqrt{\text{var}(\hat{p})}$

Q What's the problem? Don't know p ! Substitute \hat{p} in place of p . Get

$sd\text{-hat}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n}$ also called Standard dev.

Example: For the triangle test data, $\hat{p} = B/n = 25/50 = .5$.

$$SD\text{-hat}(\hat{p}) = \sqrt{.5(.5)/50} = .0707$$

2.2.1 Sample size determination

Suppose we want to choose the sample size n so that \hat{p} is within a distance D of p with probability $1-\alpha$. We want

$$P_p(-D < \hat{p} - p < D) = 1-\alpha$$

$$\text{LHS} = P_p(-D/sd(\hat{p}) < \hat{p} - p/sd < +D/sd(\hat{p}))$$

$$\sim P_p(-D/sd(\hat{p}) < Z < D/sd(\hat{p}))$$

Equate $D/sd(\hat{p}) = z_{\alpha/2}$ to make it $1-\alpha$.

$$\text{We get } D/\sqrt{p(1-p)/n} = z_{\alpha/2}$$

Solving for n , we get $n = (z_{\alpha/2})^2 p(1-p)/D^2$.

Again, the problem is that we don't know p . It is maximized at $p=1/2$, we compute it for the conservative value $\frac{1}{2}$. (For other p -value, the margin will be tighter!) Thus we get $N = (z_{\alpha/2})^2 / 4D^2$

Q To get 2% error within 95% confidence interval, what's the sample size we need?

$$1.96^2 / 4 / 0.02^2 = 2,401$$

How about 1% error?

$$1.96^2 / 4 / 0.01^2 = 9,604$$

How about 99% confidence interval?

$$2.57^2 / 4 / 0.02^2 = 4,128$$

2.3 A confidence interval for the probability of success

Clopper-Pearson Procedure

For a two-sided confidence interval for p , with confidence coefficient at least $1-\alpha$, obtain the values $p_L(\alpha)$ and $p_U(\alpha)$ from Table A.3. The $1-\alpha$ confidence interval for p is $(p_L(\alpha), p_U(\alpha))$

$P(p_L(\alpha) < p < p_U(\alpha)) \geq 1-\alpha$.

Very simple.

Large-sample approximation

For large n , they can be approximated by

$$p_L(\alpha) = \hat{p} - z_{\alpha/2} \text{sd-hat}(\hat{p})$$

$$p_U(\alpha) = \hat{p} + z_{\alpha/2} \text{sd-hat}(\hat{p})$$

Q For $n=7$ and $B=6$, for $\alpha=.02$, they are .3566, .9986. That's 98% interval.

What are intervals.

99% .3151, .9993

95% .4213, .9964

Q Which are longest? 99%.

HW: 4-7, 13, 17, 20. For grad, #8 too. (Due Next Monday; will be graded!)

2.3.1.1 Example: Vioxx of Merck

Q recent newspaper article about Vioxx of Merck

The risk was small - 15 cases of heart attack, stroke or blood clots per thousand people over three years compared with 7.5 such events per thousand patients taking a placebo.

Let $H_0: p=0.0075$

$H_1: p>0.0075$

The article doesn't have # of patients. But suppose $n = 1000$.

$B=15$.

$P\text{-value} = P_{0.0075}(B \geq 15) = 1 - P_{0.0075}(B \leq 14)$

Can we use the table? **No**

Do R-variation

P-value = $1 - \text{pbinom}(14, 1000, 0.0075) = 0.00998601$

2.3.2 R variation

Limitations of the table-based method

1. only have n up to 10
2. can't work with p other than those on the table. Vioxx example
3. Tables are cumbersome to look up
4. Can't automate (Suppose you have to do it for 1,000 outcomes!)

Why R (command line interface!) instead of minitab, statexact, spss, (some are point click) etc...

1. industrial strength
2. free
3. room to grow
4. Can't use it unless you know what you're doing (command line interface)

Given $B \sim \text{bin}(n, p)$

$\text{pbinom}(b, n, p) = \Pr(B \leq b)$

$\text{pbinom}(b_1:b_2, n, p) = (\Pr(B \leq b_1), \Pr(B \leq b_1), \dots, \Pr(B \leq b_2))$

Example: Suppose $n=8$ and test $H_0: p=.4$. Under H_0 , we can let $B \sim \text{Bin}(8, 0.4)$,

To get

$P(B \geq 7) = 1 - P(B \leq 6)$: $1 - \text{pbinom}(6, 8, 0.4) = 0.00851968$

$P(B \geq 6) = 1 - P(B \leq 5)$: $1 - \text{pbinom}(5, 8, 0.4) = 0.04980736$

$P(B \geq 5) = 1 - P(B \leq 4)$: $1 - \text{pbinom}(4, 8, 0.4) = 0.1736704$

Can just type

$1 - \text{pbinom}(6:4, 8, 0.4) = 0.00851968 \ 0.04980736 \ 0.17367040$

or,

$\text{round}(1 - \text{pbinom}(6:4, 8, 0.4), 4)$

$[1] \ 0.0085 \ 0.0498 \ 0.1737$

Given $Z \sim N(0, 1)$

$\text{pnorm}(x) : \Pr(Z < x)$

$\text{qnorm}(p) : x \text{ such that } \Pr(Z < x) = p$

$\text{qnorm}(1 - \alpha) : x \text{ such that } \Pr(Z > x) = \alpha$ (**WHY?** $\Pr(Z < x) = 1 - \alpha \Rightarrow \alpha = 1 - \Pr(Z < x) = \Pr(Z > x)$) aka z_α

$\text{qnorm}(1 - 0.025) = \text{qnorm}(0.975) = z_{0.025} = 1.959964$

$\text{qnorm}(0.995) = z_{0.005} = 2.575829$

Want to know more about those functions? Run $?rnorm, ?rbinom$ etc

3 The one-sample location problem

Interested in 'location (median) of population' draw

Two types of data:

1. paired replicates data: pairs of 'pretreatment' and 'posttreatment' observations; interested in a shift in location due to the application of the 'treatment'.
2. one-sample data: observations from a single population about whose location we wish to make inference.

3.1 Paired replicates analysese by way of signed ranks

Data: $2n$ observations, two obs on each of n subjects. Subject i , X_i (before) and Y_i , $i=1, \dots, n$

Assumptions:

1. the differences $Z_i = Y_i - X_i$, $i=1, \dots, n$ are mutually independent
2. Each Z_i comes from a continuous distribution (not necessarily the same one), that is symmetric about a common median θ . The parameter θ is called the treatment effect. If F_i represents the d.f. for Z_i , this requires $F_i(\theta+t) + F_i(\theta-t) = 1$ for every t and i .

Q check student's understanding of 'continuous distribution'

Q Implication of these assumptions?

1. We don't require X_i and Y_i be independent. In most cases they don't.
2. Both X_i and Y_i need to be continuous for Z_i to be continuous.
3. If X_i follows some distribution and $Y_i = X_i + \theta$, the symmetry assumption is satisfied automatically. If not only the location, but also the shape of the distribution changes, the assumption is violated. Example? After some treatment, a patient can have much wider variability in certain response. Then the symmetry of Z_i won't hold. Plots.

3.2 Distribution-free Wilcoxon Signed Rank test (H&W 3.1) [wilcox.test]

Hypothesis: $H_0: \theta = 0$. Interpretation? 'zero shift in location due to the treatment'. Each distribution (possibly different) for the differences is symmetrically distributed about 0.

Procedure: Compute WSR statistic T^+ :

1. Form the absolute values $|Z_1|, \dots, |Z_n|$
2. Order them from least to greatest.
3. Let R_i be the rank of $|Z_i|$.
4. Define indicator $\psi_i = 1(Z_i > 0)$.
5. Produce the positive signed rank of Z_i , $R_1\psi_1, R_2\psi_2, \dots, R_n\psi_n$.
6. $T^+ = \sum_{i=1, \dots, n} R_i\psi_i$

Table-form. make room for the large sample approximation on the right of the table:

One-sided upper-tail test: to test H_0 vs $H_1: \theta > 0$ at the α level of significance, Reject H_0 if $T^+ \geq t_{\alpha}$; do not reject otherwise.

Where the constant t_{α} is chosen to make $\Pr(\text{type I error}) = \alpha$. See Table A.4.

One-sided lower-tail test: to test H_0 vs $H_1: \theta < 0$ at the α level of significance, Reject H_0 if $T^+ \leq n(n+1)/2 - t_{\alpha}$; do not reject otherwise.

Two sided test: to test H_0 vs $H_1: \theta \neq 0$ at the α level of significance, Reject H_0 if $T^+ \geq t_{\alpha/2}$ or $T^+ \leq n(n+1)/2 - t_{\alpha/2}$; do not reject otherwise.

Two-sided symmetric test with $\alpha/2$ probability in each tail of the null distribution of T^+ .

* why $n(n+1)/2 - t_{\alpha}$ for the lower-tail test? The distribution of T^+ is symmetric about the mean $n(n+1)/4$. $P(T^+ < \text{mean} - t) = P(T^+ > \text{mean} + t)$, let $u = \text{mean} - t$, then $t = \text{mean} - u$. LHS = $P(T^+ < u) = P(T^+ > \text{mean} + (\text{mean} - u)) = P(T^+ > 2 * \text{mean} - u)$. this is $P(T^+ > n(n+1)/2 - u)$ for Wilcoxon.

Convert upper-tail probability to lower tail probability when

Draw the distribution; then it's much easier

Similarly, $P(T^+ > u) = P(T^+ < 2 * \text{mean} - u)$ no need

* why such test? If $\theta > 0$, there will tend to be a large proportion of positive Z differences and they will tend to have larger absolute values

* to test $H_0: \theta = \theta_0$, for some specified non-zero θ_0 , proceed as above with $Z'_i = Z_i - \theta_0$
Use the cholesterol drug example below.

* Large sample approximation: First need to know their expectation and variance.

$$E_0(T^+) = n(n+1)/4$$

$$\text{var}_0(T^+) = n(n+1)(2n+1)/24$$

When H_0 is true, the standardized version

$T^* = (T^+ - E_0(T^+)) / \sqrt{\text{var}_0(T^+)} \sim N(0,1)$ asymptotically as n tends to infinity. The normal approximation the the above are:

Reject H_0 if $\{ T^* \geq z_{\alpha}; T^* \leq -z_{\alpha}; |T^*| \geq z_{\alpha/2}; \}$ add to the above proecrueses

Ties and zero absolute Z_i s:

Zero Z_i s: remove them as long as there are not many of them.

Tied $|Z_i|$ s: assign each observations in a tied group the average of the integer ranks that are associated with the tied group. Again, it's OK as long as there are not many of them.

Recall:

$$P_0(T^+ \leq x) = P_0(T^+ \geq n(n+1)/2 - x)$$

3.2.1.1 Example: Hamilton Depression Scale Factor IV

Show the data. Test H_0 vs $H_1: \theta < 0$ at $\alpha = 0.049$. $n = 9$. $t_{0.049} = 37$. $n(n+1)/2 = 45$.

Reject if $T^* \leq 45 - 37 = 8$

```

      X      Y      Z      R psi
[1,] 1 1.83 0.878 -0.952 0.952 8 0 0
[2,] 2 0.50 0.647  0.147 0.147 3 1 3
[3,] 3 1.62 0.598 -1.022 1.022 9 0 0
[4,] 4 2.48 2.050 -0.430 0.430 4 0 0
[5,] 5 1.68 1.060 -0.620 0.620 7 0 0
[6,] 6 1.88 1.290 -0.590 0.590 6 0 0
[7,] 7 1.55 1.060 -0.490 0.490 5 0 0
[8,] 8 3.06 3.140  0.080 0.080 2 1 2
[9,] 9 1.30 1.290 -0.010 0.010 1 0 0
> T <- sum(R*psi)
> T
[1] 5

```

Observed $T^* = 5$. Can reject.

Q What is the P-value?

$P_0(T^* \leq 5) = P_0(T^+ \geq 9 \cdot 10/2 - 5) = P_0(T^+ \geq 40) = .020$

Walsh sums and estimation, confidence intervals

Hodges-Lehmann estimator of θ

```

Z <- sort(Z)
> Z
[1] -1.022 -0.952 -0.620 -0.590 -0.490 -0.430 -0.010  0.080  0.147
> tmp <- outer(Z, Z, "+")/2
> tmp[lower.tri(tmp)] <- NA
> tmp
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]  [,9]
[1,] -1.022 -0.987 -0.821 -0.806 -0.756 -0.726 -0.516 -0.471 -0.4375
[2,]      NA -0.952 -0.786 -0.771 -0.721 -0.691 -0.481 -0.436 -0.4025
[3,]      NA      NA -0.620 -0.605 -0.555 -0.525 -0.315 -0.270 -0.2365
[4,]      NA      NA      NA -0.590 -0.540 -0.510 -0.300 -0.255 -0.2215
[5,]      NA      NA      NA      NA -0.490 -0.460 -0.250 -0.205 -0.1715
[6,]      NA      NA      NA      NA      NA -0.430 -0.220 -0.175 -0.1415
[7,]      NA      NA      NA      NA      NA      NA -0.010  0.035  0.0685
[8,]      NA      NA      NA      NA      NA      NA      NA  0.080  0.1135
[9,]      NA      NA      NA      NA      NA      NA      NA      NA  0.1470
> sort(tmp)
[1] -1.0220 -0.9870 -0.9520 -0.8210 -0.8060 -0.7860 -0.7710 -0.7560 -0.7260
[10] -0.7210 -0.6910 -0.6200 -0.6050 -0.5900 -0.5550 -0.5400 -0.5250 -0.5160
[19] -0.5100 -0.4900 -0.4810 -0.4710 -0.4600 -0.4375 -0.4360 -0.4300 -0.4025
[28] -0.3150 -0.3000 -0.2700 -0.2550 -0.2500 -0.2365 -0.2215 -0.2200 -0.2050
[37] -0.1750 -0.1715 -0.1415 -0.0100  0.0350  0.0685  0.0800  0.1135  0.1470
>
median(tmp)
[1] -0.46

```

Q 96% CI for θ .

With $1 - \alpha = 96\%$, $\alpha = 4\%$ and Table shows that $t_{\alpha/2} = t_{.02} = 40$. Thus,

$C_{.04} = 9(9+1)/2 + 1 - 40 = 6$

$\theta_L = W^{(6)} = -.786$

$$\theta_U = W^{(40)} = -.010$$

Out of 45 values, they are sixth smallest and largest $(Z_i + Z_j)/2$.

Large sample approximation

$$C_{.04} \sim [9*(9+1)/4] - 2.05 \{ (9*(9+1)*(2*9+1)/24) \}^{1/2} = 5.2$$

Pick 5 (that is a conservative approach. **Why?**)

$$(\theta_L, \theta_U) = (W^{(5)}, W^{(41)}) = (-.806, .035)$$

3.2.1.2 Example: Cholesterol drug

Suppose we're developing a new drug for lowering the cholesterol level.

If there hasn't been any drug before, a good hypotheses for the effectiveness of the new drug will be $H_0: \theta=0$ vs $H_1: \theta < 0$.

If there is an existing drug which lowers the cholesterol level by 10 on the average, and we want to improve upon the old drug (benchmarking!), a more suitable hypotheses are:

$H_0: \theta = -10$ vs $H_1: \theta < -10$. This is a good example for non-zero θ_0 .

Explaining the idea behind the method based on $Z'_i = Z_i - \theta_0$ turned out to be rather tricky to explain. Again, simple numerical example helps:

X_i 's: 200.5, 190, 210.5.

Y_i 's: 190, 185, 195

That'd be what one would expect from the old drug.

Illustrate that $Z'_i - \theta_0$ is centered around 0.

$$n=3$$

$$H_1: \theta < 0$$

$$Z_i$$
's = -10.5, -5, -15.5

$$T^+ = 0!$$

$$n(n+1)/2 = 6$$

$$P\text{-value} = P(T^+ \leq 0) = P(T^+ \geq 6) = .125$$

(not that significant...) **Why?** (we need more samples!)

$$H_1: \theta < 10$$

$$Z_i$$
's = -0.5, 5, -5.5

$$R = 1, 2, 3$$

$$P_i = 0, 1, 0$$

$$T^+ = 2$$

$$P\text{-value} = P(T^+ \leq 2) = P(T^+ \geq 4) = .375$$

3.2.2 Using R to compute it

```
X <- c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
Y <- c(0.878, 0.647, 0.598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29)
wilcox.test(Y-X, alternative='less')
t.test(Y-X, alternative='less')
```

Learn to use ?wilcox.test etc.

3.3 An estimator Associated with WSR Statistics (Hodges-Lehmann Estimator)

Goal: Want to estimate the treatment effect θ .

Procedure:

1. Form $M=n(n+1)/2$ averages $(Z_i+Z_j)/2$ for $i \leq j=1, \dots, n$
2. Estimate θ by $\theta\text{-hat} = \text{median}((Z_i+Z_j)/2, i \leq j=1, \dots, n)$

Formally, let $W^{(1)} \leq \dots \leq W^{(M)}$ denote the ordered values. The median is given either as $W^{(k+1)}$ if $M=2k+1$ and $(W^{(k)} + W^{(k+1)})/2$ if $M=2k$.

Cholesterol data example (continued)

Z_i s = -10, -5, -15

'Walsh' averages: -10, $(-10-5)/2$, $(-10-15)/2$, -5, $(-5-15)/2$, -15
= -10, -7.5, -12.5, -5, -10, -15

ordering, we get -15, -12.5, -10, -10, -7.5, -5

(We have ties, but let's talk about it later)

$\theta\text{-hat} = (-10-10)/2 = -10$.

Motivator: Subtracting $\theta\text{-hat}$ from each Z_i 's will make the corresponding statistics T^+ based on $Z_i - \theta\text{-hat}$ distributed around $n(n+1)/4$

In-Sensitivity to gross error: the estimator is relatively insensitive to outliers. Compare the values $\theta\text{-hat}$ and the sample mean of Z with the original values and with -15 changed to -150.

= -10, -7.5, -70*, -5, -70*, -150*

ordering, we get -150*, -70*, -70*, -10,

Such estimator is called robust.

Ties and zero absolute Z_i s: do not discard them. Zero Z values contain important information about the magnitude of the treatment effect.

3.4 A distribution-free confidence interval based on WSR test (Tukey)

Procedure:

For a symmetric two-sided CI for θ , with confidence coefficient $1-\alpha$,

1. Obtain $t_{\alpha/2}$ of the null distribution of T^+ (Table A.4)
2. Set $C_\alpha = n(n+1)/2 + 1 - t_{\alpha/2}$.

3. The $100(1-\alpha)\%$ CI is (θ_L, θ_U) with
- $\theta_L = W^{(C_\alpha)}$ and
 - $\theta_U = W^{(M+1-C_\alpha)} = W^{(t_{\alpha/2})}$
 - where $M=n(n+1)/2$ and $W^{(1)} \leq \dots \leq W^{(M)}$ are ordered Walsh sums.

We have

$$P_\theta(\theta_L < \theta < \theta_U) = 1-\alpha \text{ for all } \theta.$$

Example

3.4.1 Large sample approximation

For large n , we approximate C_α by

$$C_\alpha \approx \frac{n(n+1)}{4} - z_{\alpha/2} \left\{ \frac{n(n+1)(2n+1)}{24} \right\}^{1/2} \text{ (or an integer close to it)}$$

CI and Two-sided test: The $100(1-\alpha)\%$ CI for θ consists of those θ_0 values for which the two-sided α -level test of $\theta = \theta_0$ does not reject the null hypothesis.

Confidence bounds: If interested in making $1-\alpha$ one-sided confidence bounds for θ , obtain the upper α th (not $\alpha/2$ th!) percentile point t_α of the null distribution of T^* . Set

$$C_\alpha^* = \frac{n(n+1)}{2} + 1 - t_\alpha$$

LOWER : The $100(1-\alpha)\%$ lower CB θ_L^* for θ (cf. $H_1: \theta > 0$) is

$$(\theta_L^*, \infty) = (W^{(C_\alpha^*)}, \infty)$$

We have

$$P_\theta(\theta_L^* < \theta < \infty) = 1 - \alpha \text{ for all } \theta.$$

UPPER : $100(1-\alpha)\%$ upper CB θ_U^* for θ (cf. $H_1: \theta < 0$) is

$$(-\infty, \theta_U^*) = (-\infty, W^{(M+1-C_\alpha^*)}) = (-\infty, W^{(t_\alpha)})$$

We have

$$P_\theta(-\infty < \theta < \theta_U^*) = 1 - \alpha \text{ for all } \theta.$$

They are related to the acceptance regions of the one-sided WSR tests.

Zero and Tied absolute Z_j s: Don't discard. (same as estimation. Different from test)

3.5 Sign (rather than Signed Ranks)

Paired replicate analyses by ways of signs.

Data: we obtain $2n$ observations, two observations on each of n subjects (blocks, patients etc)

Assumptions

1. (B1) The differences $Z_i = Y_i - X_i$, $i=1, \dots, n$ are mutually independent.
2. (B2) Each Z_i comes from a continuous population that has a common median θ , referred to as the unknown treatment effect. **Formal Definition?**

Q Are these assumptions stronger or weaker than those for WSR?

Assumption B2 is less stringent than A2. Advantage of sign test. **(why?)**

We can further weaken it to B': $P(Z_i < 0) = P(Z_i > 0) = 1/2$.

3.6 A distribution-free sign test (Fisher) (H&W 3.4)

Hypothesis: $H_0: \theta = 0$.

Procedure: Let $B = (\# \text{ of positive } Z_i \text{ s in the sample})$. More formally,

$$B = \sum_{i=1}^n \psi_i \text{ where } \psi_i = 1 \text{ iff } Z_i > 0.$$

1. One sided upper-tail test: To test H_0 vs $H_1: \theta > 0$ at the α level of significance, "Reject H_0 if $B \geq b_{\alpha, 1/2}$ "
2. One sided lower-tail test: To test H_0 vs $H_1: \theta < 0$ at the α level of significance, "Reject H_0 if $B \leq n - b_{\alpha, 1/2}$ "
3. Two-sided test: To test H_0 vs $H_1: \theta \neq 0$ at the α level of significance, "Reject H_0 if $B \geq b_{\alpha/2, 1/2}$ or $B \leq n - b_{\alpha/2, 1/2}$ "

Motivation of the test? Larger θ leads to...

3.6.1 Large sample approximation

Under H_0 , $B \sim \text{bin}(n, 1/2)$, thus standard $E_0(B)$, $\text{var}_0(B)$ and asymptotic normality applies. One obtains standardized version B^* of B in the usual manner and do one-sided tests and two-sided test based on B^* using normal table.

$$E_0(B) = \frac{n}{2}, \text{var}_0(B) = \frac{n}{4},$$

$$B^* = \frac{B - E_0(B)}{\sqrt{\text{var}_0(B)}} = \frac{B - (n/2)}{\sqrt{n/4}}$$

Ties: discard them.

Under H_0 , the distribution of B is symmetric.

$$P_0(B \leq x) = P_0(B \geq n - x) \text{ for } x=0, 1, \dots, n.$$

Power results: Consider the upper-tail α -level test of $H_0: \theta = 0$ vs $H_1: \theta > 0$.

When we have a common underlying distribution with median θ for the differences Z_i , the sign statistic $B \sim \text{binomial}(n, p_\theta)$, where $p_\theta = P_\theta(Z > 0) = 1 - F(0) > 1 - F(\theta) = \frac{1}{2}$, one can compute power as in binomial test.

Also, given target $p=P(Z>0)$, we can determine approximate sample size n so that the α -level one-sided test has approximate power $1-\beta$ against that particular alternative.

Power

3.6.2 Nonzero θ_0

Testing $\theta = \theta_0$ for some specified nonzero value θ_0 .

Recheck the assumption (how?)

Modify Z_i s (how?).

3.6.3 Example: Cholesterol data.

3.6.4 Example: Beak Clapping Counts

The data in Table 2 are a subset of the data obtained by Oppenheim (1968) in an experiment investigating light responsivity in chick embryos. The subjects were white leghorn chick embryos, and the behavioral response measured in the investigation was **beak-clapping** (i.e. the rapid opening and closing of the **beak** that occurs during the latter one-third of incubation in chick embryos). (Gottlieb (1965) had previously shown that changes in the rate of **beak-clapping** constituted a sensitive indicator of auditory responsiveness in chick embryos) The embryos were placed in a dark chamber 30 min. before the initiation of testing. Then ten 1-min readings were taken in the dark, and at the end of this 10-min. period, a single reading was obtained for a 1-min period of illumination. Table 2 gives the average number of claps per minute during the dark period (X_1) and the corresponding rate during the period of illumination (X_2) for 25 chick embryos. Estimate θ for the **beak clapping** data of Table 2.

Table 2. **Beak-Claping Counts** per Minute

Embryo i	(Dark Period) X_{1i}	(Illumination) X_{2i}
1	5.8	5
2	13.5	21
3	36.1	73
4	7.4	25
5	7.6	3
6	23	77
7	10.7	59
8	9.1	13
9	19.3	36
10	26.3	46
11	17.5	9
12	17.9	25

13	18.3	59
14	14.2	38
15	55.2	70
16	15.4	36
17	30	55
18	21.3	46
19	26.8	25
20	8.1	30
21	24.3	29
22	21.3	46
23	18.2	71
24	22.5	31
25	31.1	33

Source: R. W. Oppenheim (1968).

$H_0: \theta=0, H_1: \theta>0.$

$B=21, n=25.$

Reject if B is too large.

Can we reject at $\alpha=0.0216$?

What is P-value?

$P_{0.5}(B \geq 21) = .0005$

P-value from Large sample approximation:

$B^* = 3.40, P(Z > 3.40) = 0.0003.$

Estimate of θ .

Sort Z 's.

-8., -4.6, -1.8, -0.8, 1.9,
 3.9, 4.7, 7.1, 7.5, 8.5,
 14.8, 16.7, 17.6, 19.7, 20.6,
 21.9, 23.8, 24.7, 24.7, 25,
 36.9, 40.7, 48.3, 52.8, 54,

Order $n=25$ Z values. $N=25$, use $\tilde{\theta} = Z^{(13)} = 17.6$ for the treatment effect.

"A typical chick embryo of the type included in this study will produce 17.6 more beak-claps per minute during periods of illumination than during periods of darkness"

```
MTB > SET C1
DATA > ...
DATA > END
MTB > SINTERVAL .8922 C1
```

CI for θ

89.22% CI for $\theta, \alpha=.1078.$

For $n=25$, $p=1/2$, $b_{\alpha/2, 1/2} = 17$.

$$C_{.1078} = 25 + 1 - 17 = 9$$

$Z^{(9)} = 7.5$ and $Z^{(17)} = 23.8$ gives two endpoints of the CI.

Large sample approximation

$$C_{.1078} \approx 25/2 - 1.608 * (25/4)^{1/2} = 8.48$$

Conservatively set $C_{.1078} = 8$

The CI is given by $(Z^{(8)}, Z^{(18)}) = (7.1, 24.7)$

3.7 An estimator associated with the sign statistic

Procedure:

To estimate the treatment effect θ , use

$$\tilde{\theta} = \text{median}\{Z_i, 1 \leq i \leq n\}.$$

Use `median(x)` in R.

Motivation: $\tilde{\theta}$ is the statistic that makes $Z_i - \tilde{\theta}$ appear as a sample from a population with median 0.

Very simple.

Robust.

Somewhat inefficient. (but OK in many cases)

Don't discard zero Z_i s.

3.8 Distribution-free confidence interval based on the sign test

For two-sided CI for θ with confidence coefficient $1-\alpha$, obtain the upper $(\alpha/2)$ nd percentile point $b_{\alpha/2, \frac{1}{2}}$ of the null distribution B from table A.2.

Set $C_\alpha = n+1 - b_{\alpha/2, \frac{1}{2}}$. The $100(1-\alpha)\%$ CI for θ is

$(Z^{(C_\alpha)}, Z^{(n+1-C_\alpha)})$ where $Z^{(i)}$ s are ordered sample observations.

Large sample approximation is given by

$$C_\alpha \approx \frac{n}{2} - z_{\alpha/2} \left(\frac{n}{4} \right)^{1/2}$$

To be conservative, take C_α to be the largest integer that is less than or equal to the right-hand side of equation.

3.9 Confidence Bounds

For specified confidence coefficient $1-\alpha$, find the upper α th point $b_{\alpha, 1/2}$ of the null distribution B, set

$$C_\alpha = n+1 - b_{\alpha, 1/2}$$

The $100(1-\alpha)\%$

LOWER confidence bound : $(\theta_L, \infty) = (Z^{(C_\alpha)}, \infty)$

UPPER confidence bound: $(-\infty, \theta_U) = (-\infty, Z^{(n+1-C_\alpha)}) = (-\infty, Z^{(b_{\alpha,1/2})})$

For large n, approximate the integer C_α by

$$C_\alpha \approx \frac{n}{2} - z_\alpha \left(\frac{n}{4} \right)^{1/2}.$$

(Be conservative!)

3.10 One sample data

3.10.1 Procedure based on the signed rank statistic

Data: We obtain n observations Z_1, \dots, Z_n

Assumptions:

C1. The Z's are mutually independent

C2. Each Z comes from a population (not necessarily the same) that is continuous and symmetric about θ .

Use modified $Z_i' = Z_i - \theta_0$ for testing. Estimation and CI is the same.

3.10.2 Procedure based on the sign statistic

Data: We obtain n observations Z_1, \dots, Z_n

Assumptions:

D1. The Z's are mutually independent

D2. Each Z comes from the same continuous population with median θ .

Use modified $Z_i' = Z_i - \theta_0$ for testing. Estimation and CI is the same.

3.10.2.1 HW #2 Due Monday, 10/25

In Chapter 3, do #1, 2, 6, 7, 16; 18, 27, 36 (the first part only), 37, 41, 42

Do all by hand & simple calculator except: for #16, 42, use R or any other software including 'StatExact'.

#14, 22, 23 (Graduate students only)

3.11 Downloading data and run Wilcoxon Signed Rank test

How to use R to

- 1) Read in data
- 2) Compute exact p-value
- 3) Compute large-sample approximation p-value
- 4) Compute Hodges-Lehman estimator
- 5) Compute 100(1- α)% confidence interval

From the terminal,

```
x <- c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
y <- c(0.878, 0.647, 0.598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29)
wilcox.test(y - x, alternative = "less") # exact
wilcox.test(y - x, alternative = "less",
            exact = FALSE, correct = FALSE) # large sample approximation
wilcox.test(y-x, conf.int=TRUE, conf.level=.95) # CI
wilcox.test(y-x, conf.int=TRUE, conf.level=.95, alternative="less") #one-sided CI
```

Look at the table : depress.txt

Once we have data file prepared, we could do something like:

```
data <- read.table('depress.txt')
wilcox.test(data$y - data$x, alternative = "less")
wilcox.test(data$y - data$x, alternative = "less",
            exact = FALSE, correct = FALSE)
wilcox.test(data$y - data$x, conf.int=TRUE)
```

Output:

```
wilcox.test(data$y - data$x, alternative = "less") # The same.
```

```
Wilcoxon signed rank test
```

```
data: data$y - data$x
V = 5, p-value = 0.01953
alternative hypothesis: true mu is less than 0
```

```
> wilcox.test(data$y - data$x, alternative = "less",
+             exact = FALSE, correct = FALSE) # H&W large sample
```

```
Wilcoxon signed rank test
```

```
data: data$y - data$x
V = 5, p-value = 0.01908
alternative hypothesis: true mu is less than 0
```

```
>
```

```
Wilcoxon signed rank test
```

```
data: data$y - data$x
V = 5, p-value = 0.03906
alternative hypothesis: true mu is not equal to 0
95 percent confidence interval:
-0.786 -0.010
sample estimates:
(pseudo)median
-0.46
```