Nonparametric tests

Dr. Yen-Yi Ho

Department of Statistics, University of South Carolina

Outline

- One-Sample Test
  - Sign Test
  - Signed Rank Test
- Two-Sample Test
  - Wilcoxon Rank-Sum Test
  - Permutation Test

If data do not come from a normal population (and if the sample is not large), we cannot use a t-test. One useful approach to creating test statistics is through the use of *rank statistics*.

Resampling methods provide alternative approaches for testing simple hypotheses and obtaining confidence intervals. For example, the t approach can be used with a permutation test to test $H_0 : \mu_1 = \mu_2$ versus any of the alternatives, *regardless of whether the data are normal or not*. This is covered in Section 16.9 (pp. 712–716).

Non-parametric methods

- Many non-parametric methods convert raw values to ranks and then analyze ranks
- In case of ties, mid-rank are used, e.g. if the raw data were 105 120 120 121 $\rightarrow$ ranks would be 1 2.5 2.5 4

Parametric Test	Nonparametric Counterpart
1-sample t	Sign test, Wilcoxon signed-rank
2-sample t	Wilcoxon 2-sample rank-sum (Mann-Whitney-Wilcoxon Test)
k-sample ANOVA	Kruskal-Wallis
Pearson *r*	Spearman ρ

Sign test for one population

The sign test assumes the data come from from a continuous distribution with model

$$Y_i = \eta + \epsilon_i, \ i = 1, \ldots, n.$$

- η is the population median and ϵ_i follow an unknown, continuous distribution.
- Want to test $H_0 : \eta = \eta_0$ where η_0 is known versus one of the three common alternatives: $H_a : \eta < \eta_0$, $H_a : \eta \neq \eta_0$, or $H_a : \eta > \eta_0$.

Sign test

- Test statistic is $B^* = \sum_{i=1}^{n} I\{y_i > \eta_0\}$, the number of $y_1, \ldots, y_n$ larger than η_0.

Sign test

- Test statistic is $B^* = \sum_{i=1}^{n} I\{y_i > \eta_0\}$, the number of $y_1, \ldots, y_n$ larger than η_0.
- Under H_0, $B^* \sim \text{bin}(n, 0.5)$.

Sign test

- Test statistic is $B^* = \sum_{i=1}^{n} I\{y_i > \eta_0\}$, the number of $y_1, \ldots, y_n$ larger than η_0.
- Under H_0, $B^* \sim \text{bin}(n, 0.5)$.
- Reject H_0 if B^* is "unusual" relative to this binomial distribution.

Question: How would you form a "large sample" test statistic from B^*? You would not need to do that here, but this is common with more complex test statistics with non-standard distributions.

Example: eye relief data

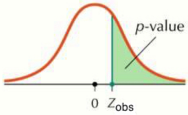
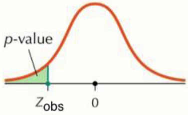
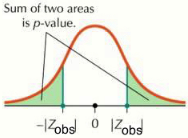
- Data are time in minutes that a drug takes to relieve $n = 20$ irritated eyes, measured redness.
- Rao (1998) page 178.
- $H_a : \eta = 5$ versus $H_a : \eta \neq 5$

Data:

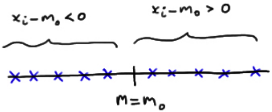
0.4	4.6	2.2	1.2	4.5	5.7	8.0	2.1	4.8	3.0
8.8	11.4	1.3	1.4	2.1	1.3	12.5	2.4	4.6	2.8

The P-value

P-value is: the probability of observing a result as or more extreme than you observed (in the direction toward the alternative hypothesis) by chance alone assuming H_0 is true.

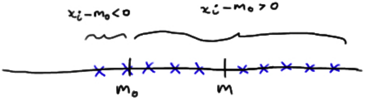
Type of hypothesis test	p-Value is tail area associated with Z_{obs}
Right-tailed test $H_0: \mu \leq \mu_0$ versus $H_a: \mu > \mu_0$ p-value $= P(Z \geqslant Z_{obs})$ Area to right of Z_{obs}	
Left-tailed test $H_0: \mu \geq \mu_0$ versus $H_a: \mu < \mu_0$ p-value $= P(Z \leqslant Z_{obs})$ Area to left of Z_{obs}	
Two-tailed test $H_0: \mu = \mu_0$ versus $H_a: \mu \neq \mu_0$ p-value $= P(Z \geqslant \lvert Z_{obs}\rvert) + P(Z \leqslant -\lvert Z_{obs}\rvert)$ $= 2 \cdot P(Z \geqslant \lvert Z_{obs}\rvert)$ Sum of the two tail areas.	

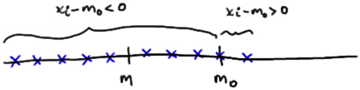
P-value for The Sign Test



$$B^{+} = \sum I(y > m_0), \qquad B^{-} = \sum I(y \leq m_0)$$

$H_0 : m = m_0$ vs $H_a : m > m_0$ $\qquad$ $H_0 : m = m_0$ vs $H_a : m < m_0$





P-value $= P(B^{-} \leq B^{-}_{obs})$ $\qquad$ P-value $= P(B^{+} \leq B^{+}_{obs})$

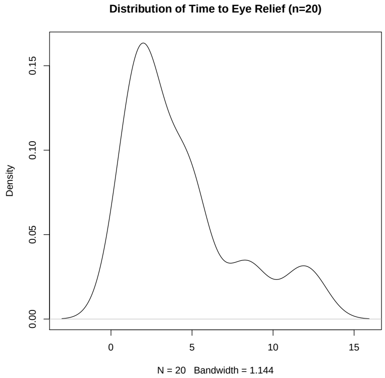
$$H_0 : m = m_0 \quad \text{vs} \quad H_a : m \neq m_0$$

$$P\text{-value} = 2 * P[B_{\text{min}} \leq min(B^{+}_{obs}, B^{-}_{obs})]$$

R code

```
eye<-c(0.4, 4.6, 2.2, 1.2, 4.5, 5.7, 8.0, 2.1,
4.8, 3.0, 8.8, 11.4, 1.3, 1.4, 2.1, 1.3, 12.5,
2.4, 4.6, 2.8)
plot(density(eye))

B<-sum(eye >= 5)
binom.test(B, length(eye), 0.5, alternative="less")
```



Wilcoxon signed rank test

- Again, test $H_0 : \eta = \eta_0$. However, this method *assumes a symmetric pdf* around the median η.
- Test statistic built from *ranks* of $\{|y_1 - \eta_0|, |y_2 - \eta_0|, \ldots, |y_n - \eta_0|\}$, denoted $R_1, \ldots, R_n$.
- The signed rank for observation i is

$$R_i^+ = \left\{ \begin{array}{cc} R_i & y_i > \eta_0 \\ 0 & y_i \leq \eta_0 \end{array} \right\}.$$

- The "signed rank" statistic is $W^+ = \sum_{i=1}^n R_i^+$.
- If W^+ is large, this is evidence that $\eta > \eta_0$.
- If W^+ is small, this is evidence that $\eta < \eta_0$.

Singed Rank Test

i	X_i	$X_i - 5$	$\lvert X_i - 5\rvert$	Rank (R_i)	Signed Rank	Z_i
1	0.40	-4.60	4.60	18.0	-18.0	0
2	4.60	-0.40	0.40	2.5	-2.5	0
3	2.20	-2.80	2.80	9.0	-9.0	0
4	1.20	-3.80	3.80	16.0	-16.0	0
5	4.50	-0.50	0.50	4.0	-4.0	0
6	5.70	0.70	0.70	5.0	5.0	1
7	8.00	3.00	3.00	12.0	12.0	1
8	2.10	-2.90	2.90	10.5	-10.5	0
9	4.80	-0.20	0.20	1.0	-1.0	0
10	3.00	-2.00	2.00	6.0	-6.0	0
11	8.80	3.80	3.80	17.0	17.0	1
12	11.40	6.40	6.40	19.0	19.0	1
13	1.30	-3.70	3.70	14.5	-14.5	0
14	1.40	-3.60	3.60	13.0	-13.0	0
15	2.10	-2.90	2.90	10.5	-10.5	0
16	1.30	-3.70	3.70	14.5	-14.5	0
17	12.50	7.50	7.50	20.0	20.0	1
18	2.40	-2.60	2.60	8.0	-8.0	0
19	4.60	-0.40	0.40	2.5	-2.5	0
20	2.80	-2.20	2.20	7.0	-7.0	0

$$W = \sum Z_i R_i$$

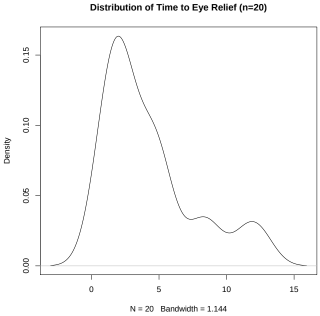
Discussion

	t-test	Signed Rank	Sign
Assumptions	Large sample or normal	Symmetric	None
Statistical Power	+++	++	+

- Both the sign test and the signed-rank test can be used with paired data (e.g. we could test whether the median difference is zero).
- **When to use what?** Use t-test when data are approximately normal, or in large sample sizes. Use sign test when data are highly skewed, etc. Use signed rank test when data are approximately symmetric but non-normal (e.g. heavy or light-tailed, multimodal yet symmetric, etc.)

Note: The sign test and signed-rank test are more flexible than the t-test because they require less strict assumptions, but the t-test has more power when the data are approximately normal.

Eye relief



Which of the three tests (t-test, sign, signed rank) is most appropriate?

Mann-Whitney-Wilcoxin test (pp. 795–796)

The Mann-Whitney test assumes

$$Y_{11}, \ldots Y_{1n_1} \stackrel{iid}{\sim} F_1 \text{ independent } Y_{21}, \ldots, Y_{2n_2} \stackrel{iid}{\sim} F_2,$$

where F_1 is the cdf of data from the first group and F_2 is the cdf of data from the second group. The null hypothesis is $H_0 : F_1 = F_2$, i.e. that the distributions of data in the two groups are identical.

The alternative is commonly taken to be $H_1 : F_1 \neq F_2$. One-sided tests can also be performed.

Building the test statistic

The Mann-Whitney test is intuitive. The data are

$$y_{11}, y_{12}, \ldots, y_{1n_1} \text{ and } y_{21}, y_{22}, \ldots, y_{2n_2}.$$

For each observation j in the first group count the number of observations in the second group c_j that are smaller; ties result in adding 0.5 to the count.

Assuming H_0 is true, on average half the observations in group 2 would be above Y_{1j} and half would be below if they come from the same distribution. That is $E(c_j) = 0.5n_2$.

The sum of these guys is $U = \sum_{j=1}^{n_1} c_j$ and has mean $E(U) = 0.5n_1 n_2$. The variance is a bit harder to derive, but is $\text{Var}(U) = n_1 n_2(n_1 + n_2 + 1)/12$.

Large sample inference

Something akin to the CLT tells us

$$Z_0 = \frac{U - E(U)}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}} \dot{\sim} N(0, 1),$$

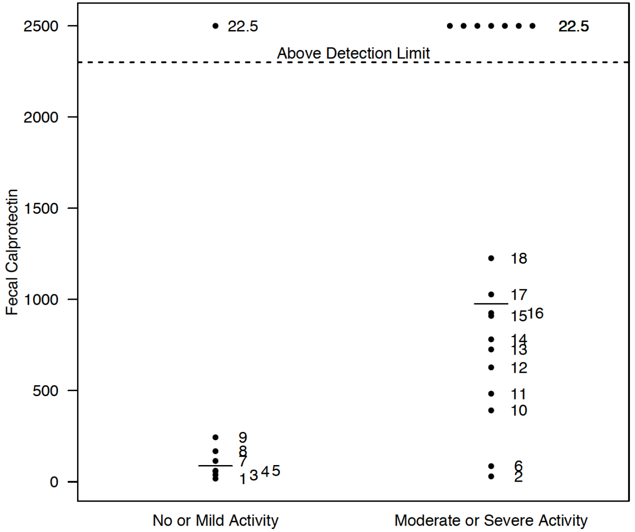
when H_0 is true. Seeing a U far away from what we expect under the null gives evidence that H_0 is false; U is then standardized as usual (subtract off then mean we expect under the null and standardize by an estimate of the standard deviation of U).

A p-value can be computed as usual as well as a CI.

Note: This test essentially boils down to replacing the observed values with their ranks and carrying out a simple pooled t-test!

Example: Fecal calprotectin

- Fecal calprotectin being evaluated as a possible biomarker of disease severity
- Calprotectin measured in 26 subjects, 8 observed to have no/mild activity by endoscopy
- Calprotectin has upper detection limit at 2500 units
  - A type of missing data, but need to keep in analysis



22.5
22.5
Above Detection Limit
Fecal Calprotectin
2500
2000
1500
1000
500
0
18
17
15 16
14
13
12
11
10
9
8
7
1 3 4 5
6
2
No or Mild Activity
Moderate or Severe Activity

2-Sample Example (Continued)

- Study Question: Are calprotectin levels different in participants with no or mild activity compared to subjects with moderate or severe activity?
- Statement of the null hypothesis

2-Sample Example (Continued)

- Study Question: Are calprotectin levels different in participants with no or mild activity compared to subjects with moderate or severe activity?
- Statement of the null hypothesis
  - H_0: Populations with no/mild activity have the same distribution of calprotectin as populations with moderate/severe activity ($F_{no/mild} = F_{moderate/severe}$)

2-Sample Example (Continued)

- R output

```
Wilcoxon rank sum test
data: calpro by endo2
W = 23.5, p-value = 0.006257
alternative hypothesis: true location shift is n
```

- A common interpretation: People with moderate/severe activity have higher **median** fecal calprotectin levels than people with no/mild activity ($p = 0.006$).

Confidence Intervals for the median difference

Bootstrap

- General method, not just for medians
- Non-parametric, does not assume symmetry
- Iterative method that repeatedly samples from the original data
- Algorithm:
  1. sample **with replacement** from sample 1 and sample 2
  2. Calculate the difference in medians, save result
  3. Repeat Steps 1 and 2, $B=1000$ times
- A (naive) 95% CI is given by the 2.5^{th} and 97.5^{th} quantile of 1000 bootstrap median differences.

Permutation

Null distribution: Distribution of the test statistic when the null hypothesis is true.
Idea: generate the null distribution by random shuffling group label.

$$H_0 : \mu_1 = \mu_2$$

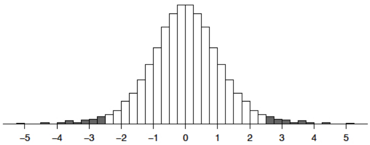
Observed Data

Group	g1	g1	g1	g2	g2	g2
Value	-1.928	-0.047	-1.572	-0.33	1.003	-0.305

Permuted Data

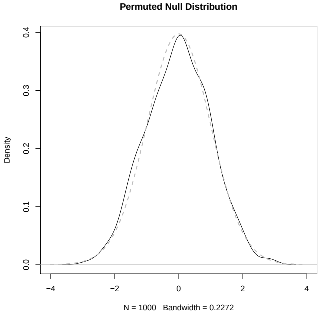
Group*	g2*	g2*	g1*	g2*	g1*	g1*
Value	-1.928	-0.047	-1.572	-0.33	1.003	-0.305

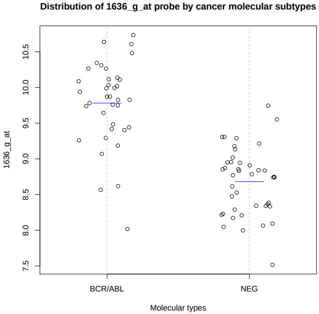
Randomly assign the group labels $\rightarrow T^*$



P-value = $\Pr(|T^*| \geq |T_{obs}|)$

Permutation Test





Permutation Test is A Good Friend

Good: Do not assume distribution for the test statistic
Bad: Computational intense (longer computation time)

Reference: Mewhort et al. Randomization tests and the unequal-N/unequal-variance problem (2009). Behavior Research Methods: 41: 664