Lecture 7: Non-parametric Comparison of Location

GENOME 560

Doug Fowler, GS (dfowler@uw.edu)

Review

- How can we set a confidence interval on a proportion?

What do we mean by nonparametric?

Types of Data – A Review

	Nominal scale
Logical/ math operations × ÷	✗
+ −	✗
< >	✗
= ≠	✓
Examples: *Dichotomous and non-dichotomous* Variable name (data values)	*Dichotomous:* Gender (male vs. female) *Non-dichotomous:* Nationality (American/Chinese/etc)

Nominal data are differentiated based on name only

Types of Data – A Review

	Nominal scale	**Ordinal** scale
Logical/ math operations × ÷	✗	✗
+ −	✗	✗
< >	✗	✓
= ≠	✓	✓
Examples: *Dichotomous and non-dichotomous* Variable name (data values)	*Dichotomous:* Gender (male vs. female) *Non-dichotomous:* Nationality (American/Chinese/etc)	*Dichotomous:* Health (healthy vs. sick), Truth (true vs. false), Beauty (beautiful vs.ugly) *Non-dichotomous:* Opinion ('completely agree'/ 'mostly agree'/ 'mostly disagree'/ 'completely disagree'

Ordinal data are also differentiated by name but they can be rank ordered (1st, 2nd, etc)

Types of Data – A Review

		Nominal scale	**Ordinal** scale	**Interval** scale
Logical/ math operations	× ÷	✗	✗	✗
	+ −	✗	✗	✓
	< >	✗	✓	✓
	= ≠	✓	✓	✓
Examples: *Dichotomous and non-dichotomous* Variable name (data values)		*Dichotomous:* Gender (male vs. female) *Non-dichotomous:* Nationality (American/Chinese/etc)	*Dichotomous:* Health (healthy vs. sick), Truth (true vs. false), Beauty (beautiful vs.ugly) *Non-dichotomous:* Opinion ('completely agree'/ 'mostly agree'/ 'mostly disagree'/ 'completely disagree')	Date (from 1457 BC to AD 2013) Latitude (from +90° to –90°) Temperature in Celsius degrees (from 10°C to 20°C)

Interval data allow for degree of difference to be calculated, but not a ratio

Types of Data – A Review

	Nominal scale	**Ordinal** scale	**Interval** scale	**Ratio** scale
Logical/ math operations × ÷	✗	✗	✗	✓
+ −	✗	✗	✓	✓
< >	✗	✓	✓	✓
= ≠	✓	✓	✓	✓
Examples: *Dichotomous and non-dichotomous* Variable name (data values)	*Dichotomous:* Gender (male vs. female) *Non-dichotomous:* Nationality (American/Chinese/etc)	*Dichotomous:* Health (healthy vs. sick), Truth (true vs. false), Beauty (beautiful vs.ugly) *Non-dichotomous:* Opinion ('completely agree'/ 'mostly agree'/ 'mostly disagree'/ 'completely disagree')	Date (from 1457 BC to AD 2013) Latitude (from +90° to –90°) Temperature in Celsius degrees (from 10°C to 20°C)	Age (from 0 to 99 years) Temperature in Kelvin (from 10K to 20K)

Ratio data are a ratio between the magnitude of a quantity and a unit magnitude of the same type

When Would We Use NP Methods?

- As we have seen, parametric methods are appropriate when the data are interval or ratio data where assumptions (e.g. normality) can be verified

When Would We Use NP Methods?

- As we have seen, parametric methods are appropriate when the data are interval or ratio data where assumptions (e.g. normality) can be verified

- In many other cases, NP methods are appropriate

When Would We Use NP Methods?

- As we have seen, parametric methods are appropriate when the data are interval or ratio data where assumptions (e.g. normality) can be verified

- In many other cases, NP methods are appropriate
  - Data are counts/frequencies of different outcomes
  - Data are on an ordinal scale

When Would We Use NP Methods?

- As we have seen, parametric methods are appropriate when the data are interval or ratio data where assumptions (e.g. normality) can be verified

- In many other cases, NP methods are appropriate
  - Data are counts/frequencies of different outcomes
  - Data are on an ordinal scale
  - Assumptions for parametric test not met or cannot be verified
  - Shape of the distribution from which sample is drawn is unknown
  - When sample sizes are small

When Would We Use NP Methods?

- As we have seen, parametric methods are appropriate when the data are interval or ratio data where assumptions (e.g. normality) can be verified

- In many other cases, NP methods are appropriate
  - Data are counts/frequencies of different outcomes
  - Data are on an ordinal scale
  - Assumptions for parametric test not met or cannot be verified
  - Shape of the distribution from which sample is drawn is unknown
  - When sample sizes are small
  - There are outliers/extreme values rendering the mean unhelpful

Goals

- Comparing the median of one sample to a given value – sign test
- Comparing the median of one sample to a given value – Wilcoxon signed rank test
- Assigning confidence intervals using non-parametric methods

Sign Test

- Used to test if a sample median M is equal to some hypothesized median M_0

Sign Test

- Used to test if a sample median *M* is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of *n* observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0

$$H_0 : M = M_0$$
$$H_A : M \neq M_0$$

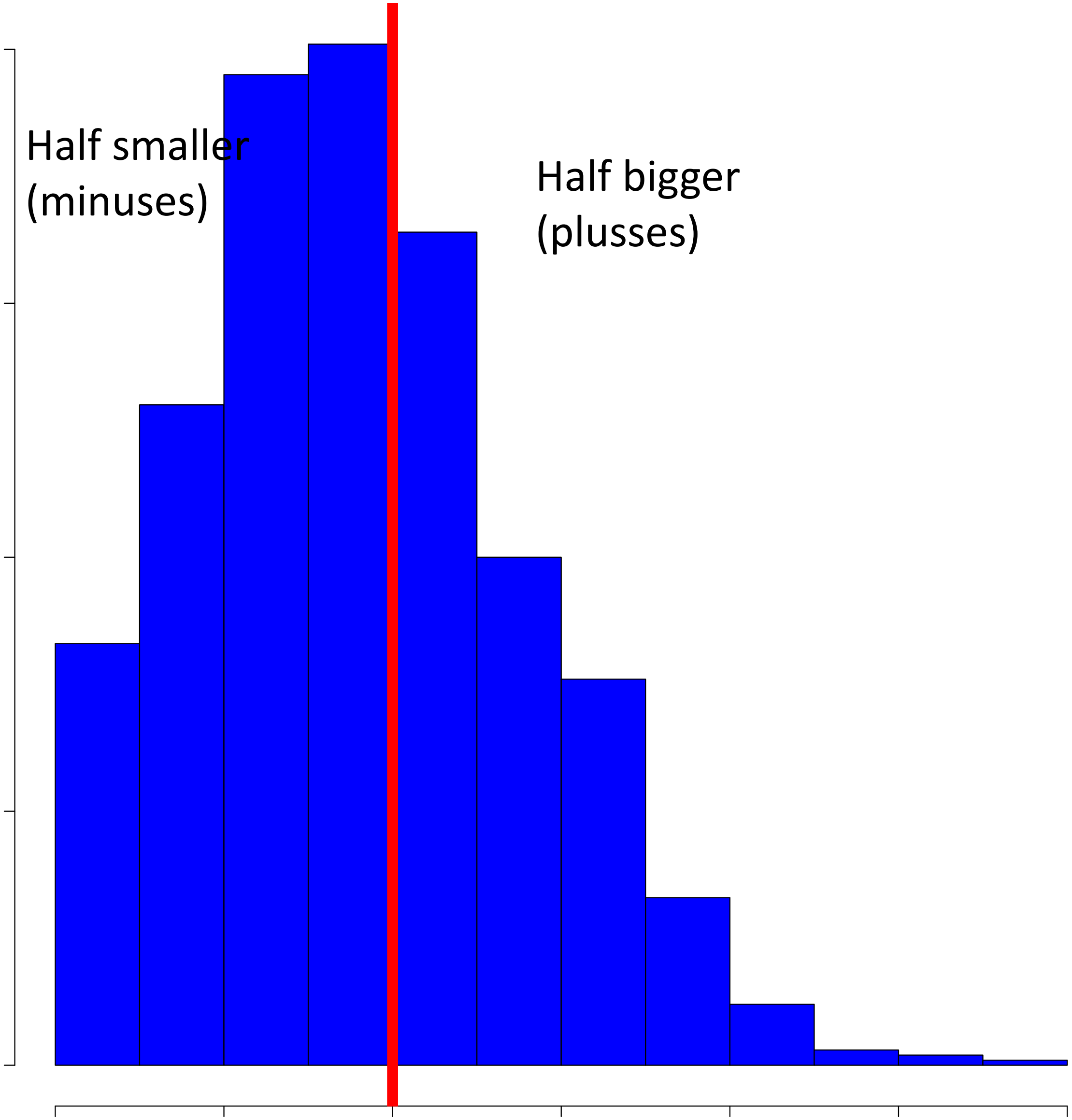
Sign Test

- Used to test if a sample median M is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of n observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0

$$S = \text{number of plus signs} \qquad H_0 : M = M_0$$
$$X_1 - M_0, ..., X_n - M_0 \qquad H_A : M \neq M_0$$

- The sign test statistic, S, is the number of plus signs among the differences

Sig



- Us … some hy…
- Th … ample of n ob … l scale about ha … than M_0

$$S \ldots = M_0$$
$$X \ldots \neq M_0$$

- Super simple idea: half should be bigger and half smaller!

Sign Test

- Used to test if a sample median *M* is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of *n* observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0

$$S = \text{number of plus signs}$$
$$X_1 - M_0, ..., X_n - M_0$$

$$H_0 : M = M_0$$
$$H_A : M \neq M_0$$

$$p = P(X_i > M_0) = 0.5$$

- Formalizing this idea… the chance of an observation being bigger than the median is 50%

Sign Test

- Used to test if a sample median M is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of n observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0

$$S = \text{number of plus signs}$$

$$H_0 : M = M_0$$
$$H_A : M \neq M_0$$

- How should the test statistic be distributed (i.e. how can we describe getting r plus signs from n observations)?

Sign Test

- Used to test if a sample median M is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of n observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0
- Assumptions:
  - Observations are assumed to be independent
  - Each difference comes from the same continuous population
  - Data are ordered (at least ordinal) such that the comparisons greater than, less than and equal to have meaning

Sign Test - Example

- Let's say that we transfected cells with GFP and RFP. Then, we examined them, scoring the GFP and RFP fluorescence in a continuous way.

$$RFP : [3.1, 2.5, 0.2, 5.0, 0.4, 2.2, 0.3, 1.6, 1.0, 0.5, 1.9]$$
$$GFP : [4, 5.2, 2.0, 0.5, 0.6, 3.1, 1.5, 1.7, 0.9, 0.3, 2.1]$$

- We want to know if the GFP and RFP signal intensity is the same in each cell

Sign Test - Example

- Let's say that we transfected cells with GFP and RFP. Then, we examined them, scoring the GFP and RFP fluorescence in a continuous way.

$$
\begin{aligned}
RFP &: [3.1, 2.5, 0.2, 5.0, 0.4, 2.2, 0.3, 1.6, 1.0, 0.5, 1.9] \\
GFP &: [4, 5.2, 2.0, 0.5, 0.6, 3.1, 1.5, 1.7, 0.9, 0.3, 2.1] \\
diff &: [-0.9, -2.7, -1.8, 4.5, -0.2, -0.9, -1.2, -0.1, 0.9, 0.3, -0.2]
\end{aligned}
$$

- Because the data are paired, we can do a one sample test on the difference

Sign Test - Example

- Let's say that we transfected cells with GFP and RFP. Then, we examined them, scoring the GFP and RFP fluorescence in a continuous way.

$$RFP : [3.1, 2.5, 0.2, 5.0, 0.4, 2.2, 0.3, 1.6, 1.0, 0.5, 1.9]$$
$$GFP : [4, 5.2, 2.0, 0.5, 0.6, 3.1, 1.5, 1.7, 0.9, 0.3, 2.1]$$
$$diff : [-0.9, -2.7, -1.8, 4.5, -0.2, -0.9, -1.2, -0.1, 0.9, 0.3, -0.2]$$

$$S = number\ of\ plus\ signs\ among\ the\ differences$$

$$S = 3$$

- To get *S*, we count the number of plus signs

Sign Test - Example

- Let's say that we transfected cells with GFP and RFP. Then, we examined them, scoring the GFP and RFP fluorescence in a continuous way.

$$RFP : [3.1, 2.5, 0.2, 5.0, 0.4, 2.2, 0.3, 1.6, 1.0, 0.5, 1.9]$$
$$GFP : [4, 5.2, 2.0, 0.5, 0.6, 3.1, 1.5, 1.7, 0.9, 0.3, 2.1]$$
$$diff : [-0.9, -2.7, -1.8, 4.5, -0.2, -0.9, -1.2, -0.1, 0.9, 0.3, -0.2]$$

$$S = number\ of\ plus\ signs\ among\ the\ differences$$

$$S = 3$$

$$\sum_{r=0}^{3} \binom{11}{r} 0.5^r * 0.5^{(11-r)} + \sum_{r=8}^{11} \binom{11}{r} 0.5^r * 0.5^{(11-r)}$$

- Then we calculate the probability of getting at most 3 or at least 11-3 = 8 plus signs amongst 11 observations (two tailed)

Sign Test - Example

- Let's say that we transfected cells with GFP and RFP. Then, we examined them, scoring the GFP and RFP fluorescence in a continuous way.

$$RFP: [3.1, 2.5, 0.2, 5.0, 0.4, 2.2, 0.3, 1.6, 1.0, 0.5, 1.9]$$
$$GFP: [4, 5.2, 2.0, 0.5, 0.6, 3.1, 1.5, 1.7, 0.9, 0.3, 2.1]$$
$$diff: [-0.9, -2.7, -1.8, 4.5, -0.2, -0.9, -1.2, -0.1, 0.9, 0.3, -0.2]$$

$$S = number\ of\ plus\ signs\ among\ the\ differences$$

$$S = 3$$

$$p = 0.227$$

- So, we conclude that the median of the differences is 0, and that the GFP/RFP signals are equal

Sign Test

- Used to test if a sample median M is equal to some hypothesized median M_0
- The null hypothesis is that given a random sample of n observations measured on at least an ordinal scale about half are bigger than M_0 and half are smaller than M_0
- Zeroes are unsigned and removed (hence reducing n)
- Sensitive to too many zeros - if your sample contains many zeroes, increase measurement precision

Goals

- Comparing the median of one sample to a given value – sign test

- Comparing the median of one sample to a given value – Wilcoxon signed rank test

- Assigning confidence intervals using non-parametric methods

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful

- How can we take advantage of magnitude without resorting to making assumptions about how the data is distributed?

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful
- Like the sign test, it can be used to compare a single sample to a median or the medians of two paired samples

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful
- Like the sign test, it can be used to compare a single sample to a median or the medians of two paired samples

$$H_0 : M = M_o$$
$$H_1 : M \neq M_o$$

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful
- Like the sign test, it can be used to compare a single sample to a median or the medians of two paired samples

$$H_0 : M = M_o$$
$$H_1 : M \neq M_o$$

- First, compute the difference, recording the absolute value and the sign

$$for\ X_1 ... X_n : |X_1 - M_o| ... |X_n - M_o|, sgn(X_1 - M_o) ... sgn(X_1 - M_o)$$

Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful
- Like the sign test, it can be used to compare a single sample to a median or the medians of two paired samples

$$H_0 : M = M_o$$
$$H_1 : M \neq M_o$$

- First, compute the difference, recording the absolute value and the sign

$$for\ X_1 ... X_n : |X_1 - M_o| ... |X_n - M_o|, sgn(X_1 - M_o) ... sgn(X_1 - M_o)$$

- Next, rank each observation by the magnitude of the difference and compute W (R_i = rank of i^{th} observation)

$$W = \sum [sgn(X_i - M_o) * R_i]$$

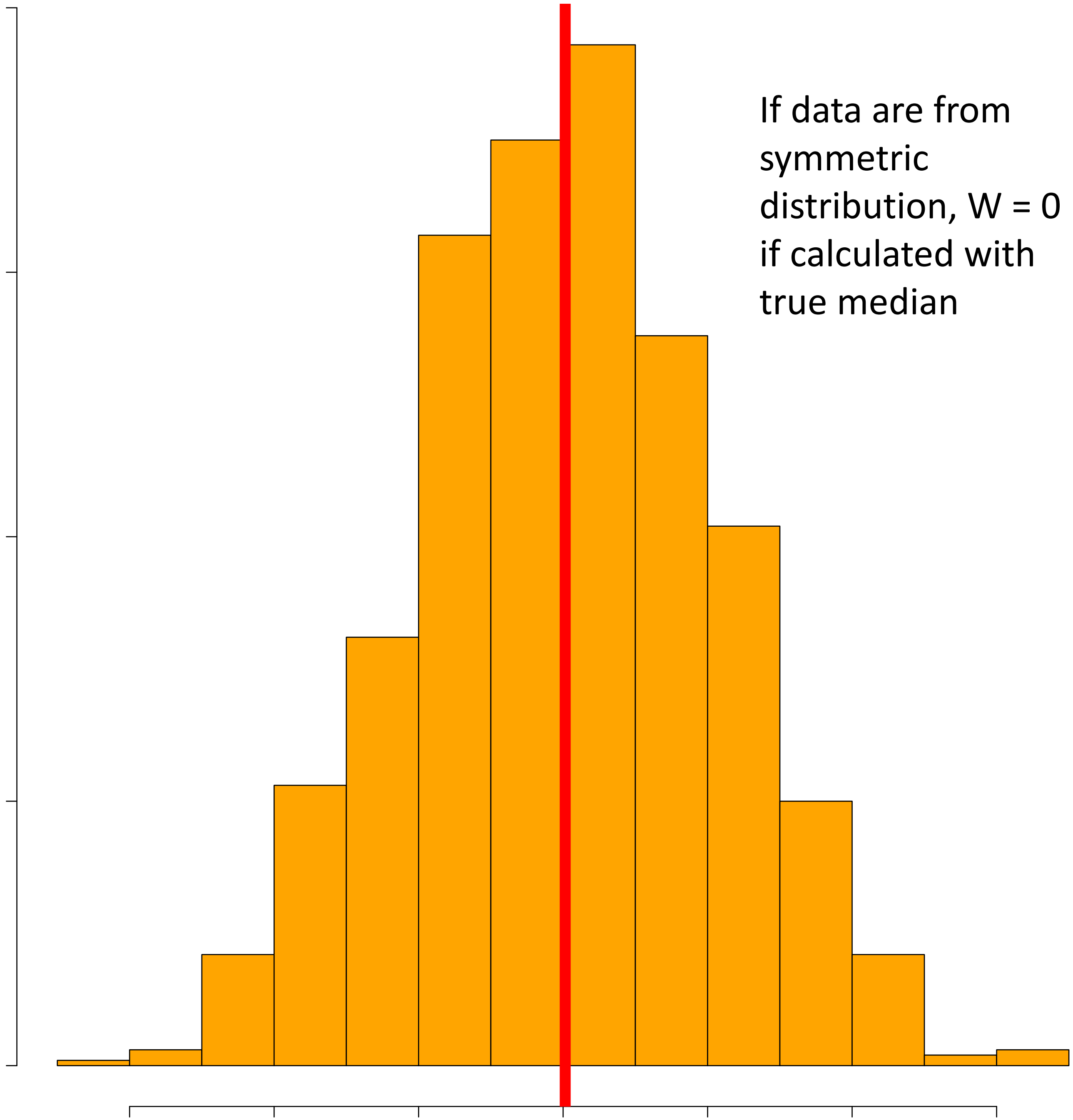
Wilc

- Sim ... es in a
- Like sam sam
- Firs valu

$$for\ X_1 \ldots \quad X_1 - M_o)$$

- Nex diff

$$W =$$



Wilcoxon Signed-Rank Test

- Similar to sign test, but takes advantage of magnitudes in addition to signs and is therefore more powerful
- Like the sign test, it can be used to compare a single sample to a median or the medians of two paired samples
- Assumptions:
  - Observations are independent
  - Observations are drawn from a continuous distribution
  - The observations can ordered
  - The distribution must be symmetric

Wilcoxon Signed-Rank Test - Example

- Let’s say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i
1	2.4
2	20
3	7.1
4	4.6
5	21.9
6	15.9
7	24.9
8	21.9
9	7.4
10	23.3

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i	$\lvert X_i - M_o \rvert$
1	2.4	18.1
2	20	0.5
3	7.1	13.4
4	4.6	15.9
5	21.9	1.4
6	15.9	4.6
7	24.9	4.4
8	21.9	1.4
9	7.4	13.1
10	23.3	2.8

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i	$\|X_i - M_o\|$	$sgn(X_i - M_o)$
1	2.4	18.1	-
2	20	0.5	-
3	7.1	13.4	-
4	4.6	15.9	-
5	21.9	1.4	+
6	15.9	4.6	-
7	24.9	4.4	+
8	21.9	1.4	+
9	7.4	13.1	-
10	23.3	2.8	+

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i	$\lvert X_i - M_o \rvert$	$sgn(X_i - M_o)$	R	R*sgn
2	20	0.5	-	1	-1
8	21.9	1.4	+	2.5	2.5
5	21.9	1.4	+	2.5	2.5
10	23.3	2.8	+	4	4
7	24.9	4.4	+	5	5
6	15.9	4.6	-	6	-6
9	7.4	13.1	-	7	-7
3	7.1	13.4	-	8	-8
4	4.6	15.9	-	9	-9
1	2.4	18.1	-	10	-10

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

Ties are awarded the midrank value

Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i	$\lvert X_i - M_o \rvert$	$sgn(X_i - M_o)$	R	R*sgn
2	20	0.5	-	1	-1
8	21.9	1.4	+	2.5	2.5
5	21.9	1.4	+	2.5	2.5
10	23.3	2.8	+	4	4
7	24.9	4.4	+	5	5
6	15.9	4.6	-	6	-6
9	7.4	13.1	-	7	-7
3	7.1	13.4	-	8	-8
4	4.6	15.9	-	9	-9
1	2.4	18.1	-	10	-10

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

$$W = -1 + 2.5 + 2.5 + 4 + 3 - 6 - 7 - 8 - 9 - 10 = -29$$

Distribution of W – A Simple Example

- Consider the case where n=3

Distribution of W – A Simple Example

- Consider the case where n=3
- Ranks must be 1, 2 and 3; the only question is which are positive and which are negative

Distribution of W – A Simple Example

- Consider the case where n=3
- Ranks must be 1, 2 and 3; the only question is which are positive and which are negative
- There are $2^3 = 8$ different ways of associating signs with ranks, each equally likely

Distribution of W – A Simple Example

- Consider the case where n=3
- Ranks must be 1, 2 and 3; the only question is which are positive and which are negative
- There are $2^3 = 8$ different ways of associating signs with ranks, each equally likely

Positive Ranks	Negative Ranks	Value of W	Probability
None	1,2,3	-6	0.13
1	2,3	-4	0.13
2	1,3	-2	0.13
3	1,2	0	0.13
1,2	3	0	0.13
1,3	2	2	0.13
2,3	1	4	0.13
1,2,3	None	6	0.13

Distribution of W – A Simple Example

- Consider the case where n=3
- Ranks must be 1, 2 and 3; the only question is which are positive and which are negative
- There are $2^3 = 8$ different ways of associating signs with ranks, each equally likely

Positive Ranks	Negative Ranks	Value of W	Probability
None	1,2,3	-6	0.13
1	2,3	-4	0.13
2	1,3	-2	0.13
3	1,2	0	0.13
1,2	3	0	0.13
1,3	2	2	0.13
2,3	1	4	0.13
1,2,3	None	6	0.13

- Sum of ranks could range from $n(n+1)/2$ to $–n(n+1)/2$

Distribution of W – A Simple Example

- Consider the case where n=3
- Ranks must be 1, 2 and 3; the only question is which are positive and which are negative
- There are $2^3 = 8$ different ways of associating signs with ranks, each equally likely

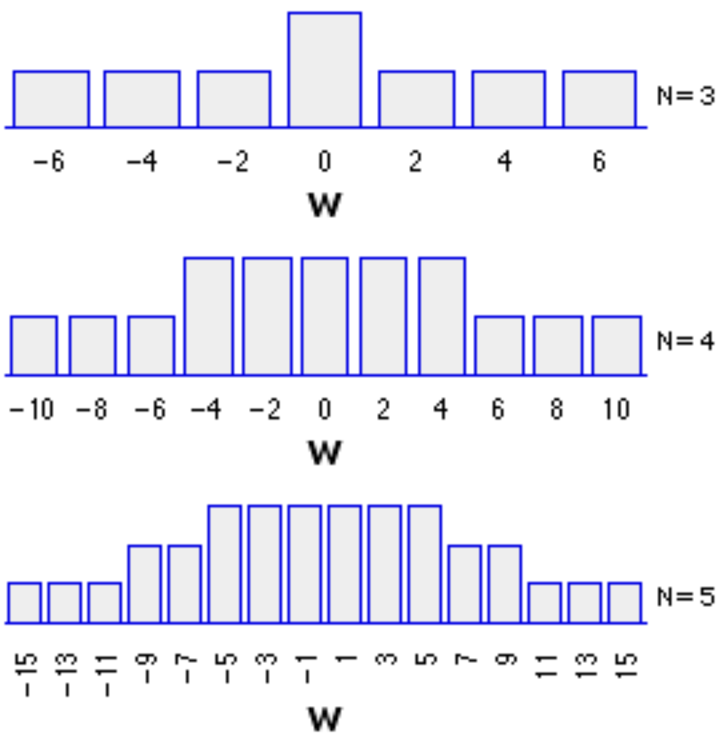
Positive Ranks	Negative Ranks	Value of W	Probability
None	1,2,3	-6	0.13
1	2,3	-4	0.13
2	1,3	-2	0.13
3	1,2	0	0.13
1,2	3	0	0.13
1,3	2	2	0.13
2,3	1	4	0.13
1,2,3	None	6	0.13

$$P(W \leq -2) = 0.39$$

- Sum of ranks could range from n(n+1)/2 to –n(n+1)/2

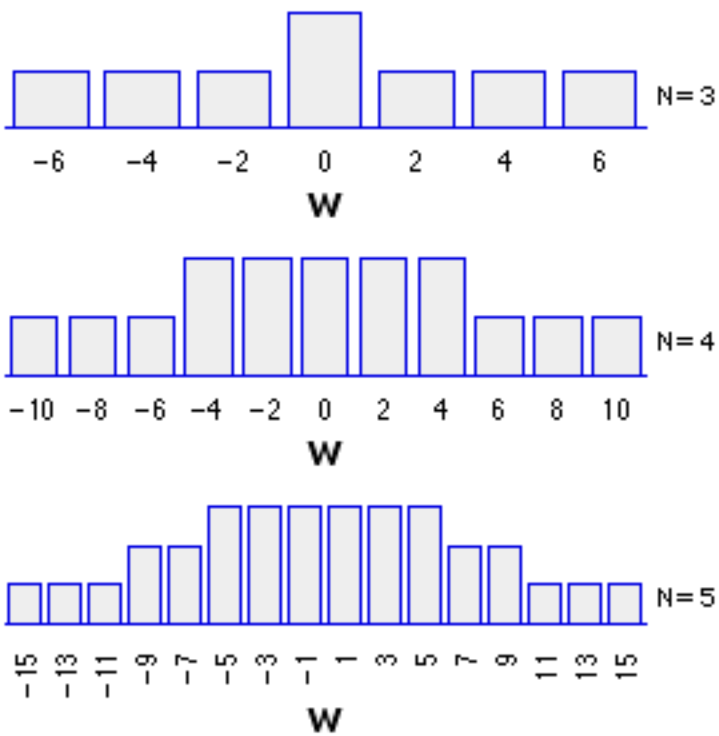
Distribution of W – A Simple Example

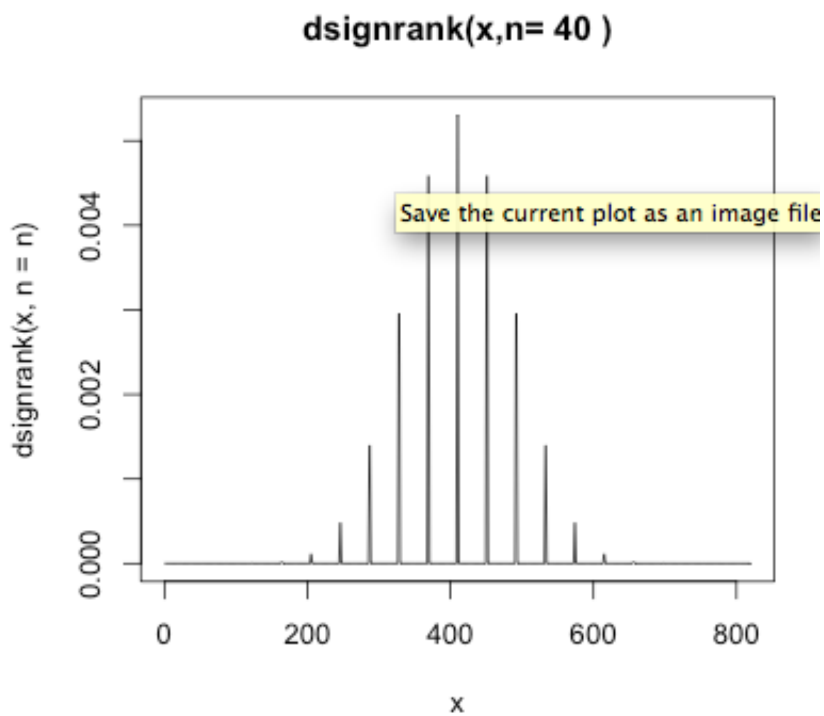
- The number of possibilities grows with n



Distribution of W

- The number of possibilities grows with n, and eventually becomes normally distributed
- So, table is used for n<10, normal approximation for n>10





Wilcoxon Signed-Rank Test - Example

- Let's say we have measured a transcript level in 10 cells. We believe the median level to be 20.5 copies/cell.

i	X_i	$\lvert X_i - M_o \rvert$	$sgn(X_i - M_o)$	R	R*sgn
2	20	0.5	-	1	-1
8	21.9	1.4	+	2.5	2.5
5	21.9	1.4	+	2.5	2.5
10	23.3	2.8	+	4	4
7	24.9	4.4	+	5	5
6	15.9	4.6	-	6	-6
9	7.4	13.1	-	7	-7
3	7.1	13.4	-	8	-8
4	4.6	15.9	-	9	-9
1	2.4	18.1	-	10	-10

$$H_0 : M = 20.5$$
$$H_1 : M \neq 20.5$$

$$p = 0.185$$

$$W = -1 + 2.5 + 2.5 + 4 + 3 - 6 - 7 - 8 - 9 - 10 = -29$$

Goals

- Comparing the median of one sample to a given value – sign test

- Comparing the median of one sample to a given value – Wilcoxon signed rank test

- Assigning confidence intervals using non-parametric methods

Confidence Interval Estimates for M

- How can we assign a confidence interval to the median?

Confidence Interval Estimates for M

- How can we assign a confidence interval to the median?

- For ordinal data?

Confidence Interval Estimates for M

- How can we assign a confidence interval to the median?
- For ordinal data?
- A 95% confidence interval for M would correspond to the range of values in a two-sided hypothesis test $M=M_0$ that would lead to acceptance of H_0 with $\alpha = 0.05$

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$X_1, X_2, X_3, ..., X_n$

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$$X_1, X_2, X_3, ..., X_n$$

- We would like to find k corresponding to the 95% CI such that:

$$X_k \leq M \leq X_{n-k+1}$$

- X_k is the k^{th} from the smallest observation; $X_{n\text{-}k+1}$ is the k^{th} from the largest

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$$X_1, X_2, X_3, ..., X_n$$

- We would like to find k corresponding to the 95% CI such that:

$$X_k \leq M \leq X_{n-k+1}$$

- Recall that this interval will contain the true median 95% of the time, on long-run sampling

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$$X_1, X_2, X_3, ..., X_n$$

- We would like to find k corresponding to the 95% CI such that:

$$X_k \leq M \leq X_{n-k+1}$$

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$$

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$$X_1, X_2, X_3, \ldots, X_n$$

- We would like to find k corresponding to the 95% CI such that:

$$X_k \leq M \leq X_{n-k+1}$$

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$$

$$\uparrow$$

$$M$$

Confidence Interval Estimates for M

- First, we arrange our data in order of relative magnitude

$$X_1, X_2, X_3, ..., X_n$$

- We would like to find k corresponding to the 95% CI such that:

$$X_k \leq M \leq X_{n-k+1}$$

$$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$$

$$X_k \longleftarrow \quad \uparrow \quad \longrightarrow X_{n-k+1}$$

$$M$$

$$k = 2$$

Use Sign Test Formalism to Find k

- Recall

$$S = number\ of\ plus\ signs\ among\ the\ differences$$

$$P(S = r) = \binom{n}{r} 0.5^r * 0.5^{(n-r)}$$

Use Sign Test Formalism to Find k

- Recall

$$S = number\ of\ plus\ signs\ among\ the\ differences$$

$$P(S=r) = \binom{n}{r} 0.5^r * 0.5^{(n-r)}$$

- For a 95% CI, k corresponds to the minimum value of S (i.e. number of plus signs) we could observe with P <0.95

Example

- Let’s say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff
-1.4
-0.6
-0.2
-0.9
-3.2
-3.2
-2.4
-0.7
-5.5
0.1
-0.1
-0.3

Example

- Let's say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff	Tdiff, sorted
-1.4	-5.5
-0.6	-3.2
-0.2	-3.2
-0.9	-2.4
-3.2	-1.4
-3.2	-0.9
-2.4	-0.7
-0.7	-0.6
-5.5	-0.3
0.1	-0.2
-0.1	-0.1
-0.3	0.1

Example

- Let's say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff	Tdiff, sorted	Rank
-1.4	-5.5	1
-0.6	-3.2	2
-0.2	-3.2	3
-0.9	-2.4	4
-3.2	-1.4	5
-3.2	-0.9	6
-2.4	-0.7	7
-0.7	-0.6	8
-5.5	-0.3	9
0.1	-0.2	10
-0.1	-0.1	11
-0.3	0.1	12

$$M = \frac{-0.9 + -0.7}{2} = -0.8$$

- Find the median value

Example

- Let's say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff	Tdiff, sorted	Rank
-1.4	-5.5	1
-0.6	-3.2	2
-0.2	-3.2	3
-0.9	-2.4	4
-3.2	-1.4	5
-3.2	-0.9	6
-2.4	-0.7	7
-0.7	-0.6	8
-5.5	-0.3	9
0.1	-0.2	10
-0.1	-0.1	11
-0.3	0.1	12

$$P(S = r) = \binom{n}{r} 0.5^r * 0.5^{(n-r)}$$

$$M = \frac{-0.9 + -0.7}{2} = -0.8$$

pbinom(2, 12, prob=0.5) = 0.019

pbinom(3, 12, prob=0.5) = 0.073

- Use the cumulative binomial distribution to find minimum number of plus signs – which would you pick?

Example

- Let's say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff	Tdiff, sorted	Rank
-1.4	-5.5	1
-0.6	-3.2	2
-0.2	-3.2	3
-0.9	-2.4	4
-3.2	-1.4	5
-3.2	-0.9	6
-2.4	-0.7	7
-0.7	-0.6	8
-5.5	-0.3	9
0.1	-0.2	10
-0.1	-0.1	11
-0.3	0.1	12

$$M = \frac{-0.9 + -0.7}{2} = -0.8$$

pbinom(2, 12, prob=0.5) = 0.019
pbinom(3, 12, prob=0.5) = 0.073

So, k = 2 is our (conservative) pick for the 95% CI

Example

- Let's say we measure the differences temperature of patients before and after surgery. We would like to place a 95% CI on the median

Tdiff	Tdiff, sorted	Rank
-1.4	-5.5	1
-0.6	-3.2	2
-0.2	-3.2	3
-0.9	-2.4	4
-3.2	-1.4	5
-3.2	-0.9	6
-2.4	-0.7	7
-0.7	-0.6	8
-5.5	-0.3	9
0.1	-0.2	10
-0.1	-0.1	11
-0.3	0.1	12

-3.2 to -0.2 give an ~95% confidence interval for the median

So, k = 2 is our (conservative) pick for the 95% CI

R Goals

- Scripts in R
  - Why?
  - How?

- Don’t worry; we will cover nonparametric tests in R next time…