Module 9: Nonparametric Statistics

Statistics (OA3102)

Professor Ron Fricker

Naval Postgraduate School

Monterey, California

Reading assignment: WM&S chapter 15.1-15.6

Goals for this Lecture

- Discuss advantages and disadvantages of nonparametric tests
  - General two-sample shift model
- Nonparametric tests for paired data
  - Sign test
  - Wilcoxon signed-rank test
    - Small and large sample variants
- Nonparametric tests for two-samples of independent data
  - Wilcoxon rank sum test
  - Mann-Whitney U test

Challenges in Hypothesis Testing

- Some experiments give responses they defy exact quantification
  - Rank the "utility" of four weapons systems
    - Gives an ordering, but can be impossible to say things like "System A is twice as useful as B"
  - Compare two LVS maintenance programs
    - If the data clearly do not fit the assumptions of the (parametric) tests we have learned, what to do?
- Nonparametric tests may be the solution

Parametric vs. Nonparametric

- Parametric hypothesis testing:
  - Statistic distribution are specified (often normal)
  - Often follows from Central Limit Theorem, but sometimes CLT assumptions don't fit/apply
- Nonparametric hypothesis testing:
  - Does not assume a particular probability distribution
    - Often called "distribution free"
  - Generally based on ordering or order statistics

Advantages of Nonparametric Tests

- Tests make less stringent demands on the data
  - E.g., they require fewer assumptions
    - Usually require independent observations
    - Sometimes assume continuity of the measure
- Can be more appropriate:
  - When measures are not precise
  - For ordinal data where scale is not obvious
  - When only ordering of data is available

Disadvantages of Nonparametric Tests

- They may "throw away" information
  - E.g., Sign tests only looks at the signs (+ or -) of the data, not the numeric values
  - If the other information is available and there is an appropriate parametric test, that test will be more powerful
- The trade-off:
  - Parametric tests are more powerful *if the assumptions are met*
  - Nonparametric tests provide a more general result *if they are powerful enough to reject*

A General Two-Sample Shift Model

- Consider two independent samples of data, $X_1,\ldots,X_{n_1}$ and $Y_1,\ldots,Y_{n_2}$, taken from normal populations with means μ_X and μ_Y and equal variances
- Then we may wish to test

$$H_0 : \mu_X - \mu_Y = 0 \text{ vs. } H_a : \mu_X - \mu_Y \neq 0$$

- This is a two-sample parametric shift (or location) model
  - Parametric as the distribution is specified (normal)
  - All is known except μ_X and μ_Y (and perhaps σ^2)

Now, Generalizing to a Nonparametric Shift Model

- Let $X_1,...,X_{n_1}$ be a random sample from a population with distribution function $F(x)$
- Let $Y_1,...,Y_{n_2}$ be a random sample from a population with distribution function $G(y)$
- Consider testing the hypotheses that the two distributions are the same,

  $$H_0 : F(z) = G(z) \text{ vs. } H_a : F(z) \neq G(z)$$

  where the form of the distributions is unspecified
  - A nonparametric approach now clearly required

Generalizing to a Nonparametric Shift Model (continued)

- Notice that the hypotheses

  $$H_0 : F(z) = G(z) \text{ vs. } H_a : F(z) \neq G(z)$$

  are very broad
  - It just says the two distributions are different
- Often experimenters want to test something more specific, such as the distributions differ by location
  - E.g., $G(y) = \Pr(Y \leq y) = \Pr(X \leq y - \theta) = F(y - \theta)$
  - See Figure 15.2 in the text for an illustration

Generalizing to a Nonparametric Shift Model (continued)

- Throughout the rest of the module, a two-sample shift (or location) model means:
  - $X_1, \ldots, X_{n_1}$ is a random sample from $F(x)$, and
  - $Y_1, \ldots, Y_{n_2}$ is a random sample from $G(y)=F(y-\theta)$ for some unknown value θ
- For the two-sample shift model, we can then think of the hypotheses as

$$H_0 : \theta = 0 \text{ vs. } H_a : \theta \neq 0$$

  - Can also test for alternatives $H_a : \theta < 0$ or $H_a : \theta > 0$

Introduction to the Sign Test for a Matched Pairs Experiment

- Suppose there are n pairs of observations in the form (X_i, Y_i)
- We wish to test the hypothesis that the distribution of the Xs and Ys is the same except perhaps for the location
- One of the simplest nonparametric tests is called the sign test
  - Idea: Define $D_i = X_i - Y_i$. Then under the null hypothesis, the probability that D_i is positive is 0.5

Sign Test for Matched Pairs

- Let $p = \Pr(X > Y)$
- The null hypothesis is $H_0\!: \; p = 1/2$
- The test statistic is $M = \#\left(D_i > 0\right)$
- Three possible alternative hypotheses and tests:

Alternative Hypothesis	Rejection Region	
$H_a\!: \; p > 1/2$	$M \geq c$	(upper-tailed test)
$H_a\!: \; p < 1/2$	$M \leq c$	(lower-tailed test)
$H_a\!: \; p \neq 1/2$	$M \geq n - c \text{ or } \mathrm{M} \leq c$	(two-tailed test)

Example 15.1

- Number of defective electrical fuses for each of two production lines recorded daily for 10 days
- Is there sufficient evidence to say that one line produces more defectives than the other?

- Write out the hypotheses:

Example 15.1 (continued)

- Now, calculate the test statistic:

Day	A	B
1	172	201
2	165	179
3	206	159
4	184	192
5	174	177
6	142	170
7	190	182
8	169	179
9	161	169
10	200	201

Example 15.1 (continued)

- And now determine the rejection region for a test of level $0.05 < \alpha < 0.1$

Example 15.1 (continued)

- And, finally, conduct the test
  - What do you conclude?

Example 15.2

- Find the p-value for the test in Example 15.1

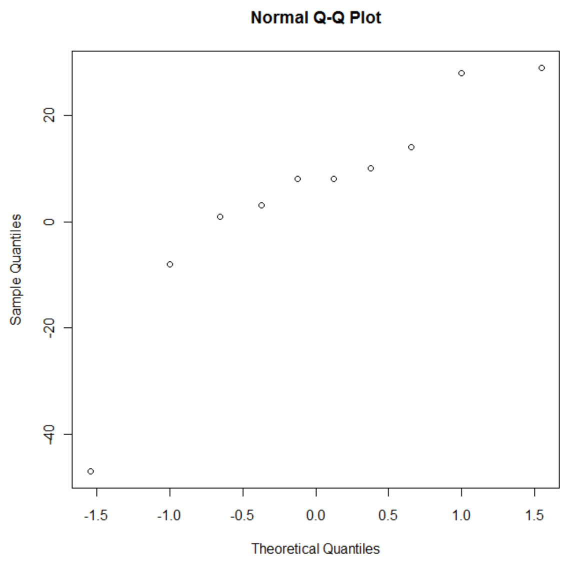
In R

- This is simply a binomial problem
  - We're asking the question: What's the chance of seeing only 2 successes out of 10 trials if $p=0.5$?
  - Use the **`binom.test`** function:

```
> binom.test(2,10,0.5)

        Exact binomial test

data:  2 and 10
number of successes = 2, number of trials = 10, p-value = 0.1094
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.02521073 0.55609546
sample estimates:
probability of success
                   0.2
```

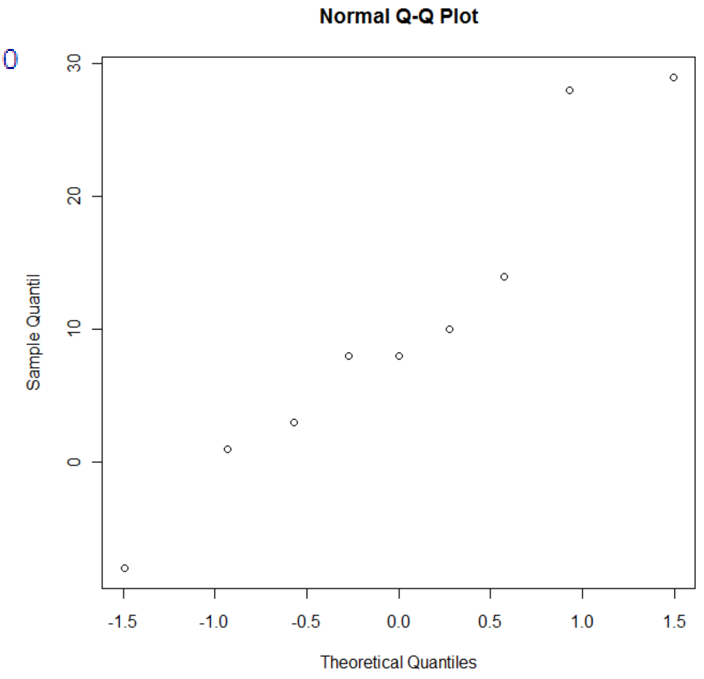
An Aside:
A Parametric Test, the Paired t-Test

```
> A.data <- c(172,165,206,184,174,142,190,169,161,200)
> B.data <- c(201,179,159,192,177,170,182,179,169,201)
> D <- B.data-A.data
> qqnorm(D)
> t.test(A.data, B.data, paired=T)

        Paired t-test

data:  A.data and B.data
t = -0.6798, df = 9, p-value = 0.5137
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -19.90633  10.70633
sample estimates:
mean of the differences
                   -4.6

> t.test(D)

        One Sample t-test

data:  D
t = 0.6798, df = 9, p-value = 0.5137
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -10.70633  19.90633
sample estimates:
mean of x
      4.6
```



The Aside (continued)

```
> D
 [1]  29  14 -47   8   3  28  -8  10   8   1
> qqnorm(D[D>-40])
> t.test(D[D>-40])

        One Sample t-test

data:  D[D > -40]
t = 2.5722, df = 8, p-value = 0.03301
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
  1.069369 19.597298
sample estimates:
mean of x
 10.33333
```



Issues and Variants

- The sign test is actually testing whether the <u>medians</u> of the distributions is equal
- What to do with ties in the sign test?
  - Just delete them and decrement n appropriately
- What if n is large (i.e., $n > 25$ or 30)?
  - Can use the large sample approximation to the binomial with

$$Z = \frac{M - np}{\sqrt{np(1-p)}} = \frac{M - n/2}{\sqrt{n}/2}$$

Sign Test for Large Samples ($n > 25$)

- Let $p = \Pr(X > Y)$
- The null hypothesis is $H_0\colon\ p = 1/2$
- The test statistic is $Z = \left(M - n/2\right)\big/ 0.5\sqrt{n}$
- Three possible alternative hypotheses and tests:

Alternative Hypothesis	Rejection Region for Level α Test	
$H_a\colon\ p > 1/2$	$z \geq z_\alpha$	(upper-tailed test)
$H_a\colon\ p < 1/2$	$z \leq -z_\alpha$	(lower-tailed test)
$H_a\colon\ p \neq 1/2$	$z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$	(two-tailed test)

Wilcoxon Signed-Rank Test

- One- or two-sided test for the hypotheses of the means of a paired sample: (X_i, Y_i)
  - Unlike the sign test, here we also use the information contained in the magnitude of the differences, $D_i = Y_i - X_i$, i = 1,…,n
  - I.e., we'll use the ranks of the absolute values of the differences in the test, not just the signs
- Hypotheses:
  - H_0: the distributions of the Xs and Ys are identical
  - H_a: the population distributions differ in location (two-tailed) or population distribution for Xs is shifted to the right (one-tailed)

Signed-Rank Methodology

- To conduct the test:
  - For n matched pairs, one observation from each population (X_i, Y_i), define $D_i = Y_i - X_i$
  - Compute the signed ranks: R_i=sign(D_i) R(|D_i|)
    - R(|D_i|) is the rank of |D_i| among the n D_is
    - Give tied observations the average rank
- If doing the calculation by hand, build a table:

i	X	Y	$D_i = X_i - Y_i$	|D_i|	R(|D_i|)	R_i=sign(D_i) R(|D_i|)
1						
2						
•						

•
•

The Test Statistic

- For a one-sided test:
  - To test if the Xs are shifted to the right of the Ys, use $T=T^-$, the sum of the negative signed ranks
  - To test if the Ys are shifted to the right of the Xs, use $T=T^+$, the sum of the positive signed ranks
- For a two-sided test, the test statistic is $T=\min(T^+,T^-)$, the minimum of either the sum of the positive or negative signed ranks

The Rejection Region

- Use Table 9:

Table 9 Critical Values of T in the Wilcoxon Matched-Pairs, Signed-Ranks Test; $n = 5(1)50$

One-sided	Two-sided	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$P = .05$	$P = .10$	1	2	4	6	8	11
$P = .025$	$P = .05$		1	2	4	6	8
$P = .01$	$P = .02$			0	2	3	5
$P = .005$	$P = .01$				0	2	3
One-sided	**Two-sided**	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$P = .05$	$P = .10$	14	17	21	26	30	36
$P = .025$	$P = .05$	11	14	17	21	25	30
$P = .01$	$P = .02$	7	10	13	16	20	24
$P = .005$	$P = .01$	5	7	10	13	16	19

Example 15.3

- Because of the variations in ovens, two types of cake mix were tested in six different ovens
  - So, each oven was used to bake each type of mix (“A” and “B”)
  - It’s a paired experimental design (by oven)
- Using the Wilcoxon signed-rank test, test the hypothesis that there is no difference in the population distribution of cake densities between the two mixes

Example 15.3 (continued)

- Calculate the test statistic:

Oven (i)	Mix A	Mix B	$D_i = A_i - B_i$	$\lvert D_i \rvert$	R($\lvert D_i \rvert$)	R_i=sign(D_i) R($\lvert D_i \rvert$)
1	0.135	0.129				
2	0.102	0.120				
3	0.108	0.112				
4	0.141	0.152				
5	0.131	0.135				
6	0.144	0.163				

Example 15.3 (continued)

- Determine the test outcome

Table 9 Critical Values of *T* in the Wilcoxon Matched-Pairs, Signed-Ranks Test; $n = 5(1)50$

One-sided	Two-sided	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$P = .05$	$P = .10$	1	2	4	6	8	11
$P = .025$	$P = .05$		1	2	4	6	8
$P = .01$	$P = .02$			0	2	3	5
$P = .005$	$P = .01$				0	2	3

Large Sample Wilcoxon Signed-Rank Test

- When $n > 25$, can use normal approximation
- It turns out that

$$E\left(T^{+}\right) = n\left(n+1\right)/4$$

$$\text{Var}\left(T^{+}\right) = n\left(n+1\right)\left(2n+1\right)/24$$

- So we can use the test statistic

$$Z = \frac{T^{+} - E\left(T^{+}\right)}{\sqrt{\text{Var}(T^{+})}} = \frac{T^{+} - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

Sign Test for Large Samples ($n > 25$)

- The null hypothesis is H_0: population dist'ns the same
- The test statistic is $Z = \dfrac{T^+ - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$
- Three possible alternative hypotheses and tests:

Alternative Hypothesis	Rejection Region for Level α Test	
H_a: Xs to right of Ys	$z \geq z_\alpha$	(upper-tailed test)
H_a: Xs to left of Ys	$z \leq -z_\alpha$	(lower-tailed test)
H_a: locations differ	$z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$	(two-tailed test)

Wilcoxon Rank Sum Test

- Now, consider two independent samples of data, $X_1,...,X_{n_1}$ and $Y_1,...,Y_{n_2}$, where goal is to test whether population dist'ns are the same
- Idea: Pool the $n_1+n_2=n$ observations, rank them in order of magnitude, and then sum their ranks of the Xs and Ys
  - Under the null hypothesis (distributions are the same) the sum of the ranks should be about equal
  - If there is a location shift, one of the sums should be larger

Wilcoxon Rank Sum Test (cont'd)

- The hypotheses are like before:
  - H_0: the distributions of the Xs and Ys are identical
  - H_a: the population distributions differ in location
    - Either two-tailed or one tailed

- An equivalent test: Mann-Whitney U test
  - We'll get to that next…

Example 15.4

- Four measurements made for bacteria counts per volume for each of two types of cultures (“I” and “II”):

	I	II
1	27	32
2	31	29
3	26	35
4	25	28

- Is there sufficient evidence to indicate a difference in locations?

Example 15.4 (continued)

- Calculate the test statistic:

Data

	I	II
1	27	32
2	31	29
3	26	35
4	25	28

Ranks

	I	II
Rank Sum (W)		

Example 15.4 (continued)

- And now determine the rejection region

Example 15.4 (continued)

Example 15.4 (continued)

- And, finally, conduct the test
  - What do you conclude?

The Mann-Whitney U Test

- As with the Wilcoxon rank sum test, this test is based on two independent samples of data, $X_1,\ldots,X_{n_1}$ and $Y_1,\ldots,Y_{n_2}$
- Again, the goal is to test whether population dist'ns are the same
- Idea: Order the n_1+n_2 observations and count the number of X observations that are smaller then each of the Y observations

Example

- From Example 15.4, the eight ordered observations are:

25	26	27	28	29	31	32	35
$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$y_{(1)}$	$y_{(2)}$	$x_{(4)}$	$y_{(3)}$	$y_{(4)}$

- So:
  - $u_1=3$ since there are three Xs before $y_{(1)}$
  - $u_2=3$ since there are three Xs before $y_{(2)}$
  - $u_3=4$ since there are three Xs before $y_{(3)}$
  - $u_4=4$ since there are three Xs before $y_{(4)}$
- And thus $U = u_1+u_2+u_3+u_4 = 3+3+4+4 = 14$

To Test U

- Use Table 8 to identify the rejection region
  - So, for example, RR={U: U$\leq$1} gives an α=0.0286 level one-sided test
  - For a two-sided test, RR={U: U$\leq$1 or U$\geq$4*4-1=15} Gives an α=2*0.0286= 0.0572 level two-sided test

$n_2 = 4$

	n_1			
U_0	1	2	3	4
0	.2000	.0667	.0286	.0143
1	.4000	.1333	.0571	.0286
2	.6000	.2667	.1143	.0571
3		.4000	.2000	.1000
4		.6000	.3143	.1714
5			.4286	.2429
6			.5714	.3429
7				.4429
8				.5571

- So, for the example, we fail to reject the hypothesis that the distributions are the same

Mann-Whitney U vs. Rank Sum Test

- Turns out the two tests are directly related:

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W$$

  where

  n_1 is the number of X observations

  n_2 is the number of Y observations

  W is the rank sum for the Xs

- So, first calculate the rank sums of the Xs and then calculate U

Some Notes

- U can take on values $0, 1, 2, \ldots, n_1 n_2$
  - It's symmetric about $n_1 n_2 / 2$
  - $\Pr(U \leq U_0) = \Pr(U \geq n_1 n_2 - U_0)$
- Table 8 is set up for $n_1 \leq n_2$
  - So, label the two sets of data appropriately
- Handle ties by averaging the ranks for the tied observations
  - E.g., if there are three tied observations due to receive ranks 3, 4, and 5, then give all three rank 4
  - Then the next observation gets rank 6

Mann-Whitney U Test

- The null hypothesis is H_0: population dist'ns the same
- The test statistic is $U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W$
- Three possible alternative hypotheses and tests:

Alternative Hypothesis	Rejection Region	
H_a: Xs to right of Ys	$U \le U_0$	(upper-tailed test)
H_a: Xs to left of Ys	$U \ge n_1 n_2 - U_0$	(lower-tailed test)
H_a: locations differ	$U \le U_0$ or $U \ge n_1 n_2 - U_0$	(two-tailed test)

Example 15.5

- Conduct the test using

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - W =$$

Example 15.6

- An experiment was conducted to compare the strengths of two types of kraft paper (i.e., cardboard)
  - Standard kraft paper
  - Paper treated with a chemical substance
- Test the hypothesis of no difference in the distributions of the strength of the papers versus the alternative that the treated paper tends to be stronger

Example 15.6 (continued)

- Calculate the test statistic:

	Standard, I	Treated, II
1	1.21 (2)	1.49 (15)
2	1.43 (12)	1.37 (7.5)
3	1.35 (6)	1.67 (20)
4	1.51 (17)	1.50 (16)
5	1.39 (9)	1.31 (5)
6	1.17 (1)	1.29 (3.5)
7	1.48 (14)	1.52 (18)
8	1.42 (11)	1.37 (7.5)
9	21.29 (3.5)	1.44 (13)
10	1.40 (10)	1.53 (19)
Rank Sum	W=85.5	

Example 15.6 (continued)

- And now determine the rejection region

Table 8 (*Continued*)

$n_2 = 10$

	n_1									
U_0	1	2	3	4	5	6	7	8	9	10
0	.0909	.0152	.0035	.0010	.0003	.0001	.0001	.0000	.0000	.0000
1	.1818	.0303	.0070	.0020	.0007	.0002	.0001	.0000	.0000	.0000
2	.2727	.0606	.0140	.0040	.0013	.0005	.0002	.0001	.0000	.0000
3	.3636	.0909	.0245	.0070	.0023	.0009	.0004	.0002	.0001	.0000
4	.4545	.1364	.0385	.0120	.0040	.0015	.0006	.0003	.0001	.0001
5	.5455	.1818	.0559	.0180	.0063	.0024	.0010	.0004	.0002	.0001
6		.2424	.0804	.0270	.0097	.0037	.0015	.0007	.0003	.0002
7		.3030	.1084	.0380	.0140	.0055	.0023	.0010	.0005	.0002
8		.3788	.1434	.0529	.0200	.0080	.0034	.0015	.0007	.0004
9		.4545	.1853	.0709	.0276	.0112	.0048	.0022	.0011	.0005
10		.5455	.2343	.0939	.0376	.0156	.0068	.0031	.0015	.0008
11			.2867	.1199	.0496	.0210	.0093	.0043	.0021	.0010
12			.3462	.1518	.0646	.0280	.0125	.0058	.0028	.0014
13			.4056	.1868	.0823	.0363	.0165	.0078	.0038	.0019
14			.4685	.2268	.1032	.0467	.0215	.0103	.0051	.0026
15			.5315	.2697	.1272	.0589	.0277	.0133	.0066	.0034
16				.3177	.1548	.0736	.0351	.0171	.0086	.0045
17				.3666	.1855	.0903	.0439	.0217	.0110	.0057
18				.4196	.2198	.1099	.0544	.0273	.0140	.0073
19				.4725	.2567	.1317	.0665	.0338	.0175	.0093
20				.5275	.2970	.1566	.0806	.0416	.0217	.0116
21					.3393	.1838	.0966	.0506	.0267	.0144
22					.3839	.2139	.1148	.0610	.0326	.0177
23					.4296	.2461	.1349	.0729	.0394	.0216
24					.4765	.2811	.1574	.0864	.0474	.0262
25					.5235	.3177	.1819	.1015	.0564	.0315
26						.3564	.2087	.1185	.0667	.0376
27						.3962	.2374	.1371	.0782	.0446
28						.4374	.2681	.1577	.0912	.0526

Example 15.6 (continued)

- And, finally, conduct the test
  - What do you conclude?

Large Sample Mann-Whitney U Test

- When $n_1 > 10$ and $n_2 > 10$, can use normal approximation
- It turns out that

$$E(U) = n_1 n_2 / 2$$

$$\text{Var}(U) = n_1 n_2 (n_1 + n_2 + 1) / 12$$

- So we can use the test statistic

$$Z = \frac{U - E(U)}{\sqrt{\text{Var}(U)}} = \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1) / 12}}$$

Large Sample U Test ($n_1 > 10$, $n_2 > 10$)

- The null hypothesis is H_0: population dist'ns the same
- The test statistic is $Z = \dfrac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$
- Three possible alternative hypotheses and tests:

Alternative Hypothesis	Rejection Region for Level α Test	
H_a: Xs to left of Ys	$z \geq z_\alpha$	(upper-tailed test)
H_a: Xs to right of Ys	$z \leq -z_\alpha$	(lower-tailed test)
H_a: locations differ	$z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$	(two-tailed test)

Other Nonparametric Tests

- Sign tests exist for one-sample tests as well
  E.g., $H_0 : F(y_0) = p_0$ vs. $H_a : F(y_0) \neq p_0$
  - Common to test p_0=0.5; i.e., test the median
    - For symmetric distributions, equivalent to testing the mean
    - Can also test quartiles or any other percentile
- Also, signed-rank and rank sum tests for one sample
- Kolmogorov-Smirnov tests for distributions
- Kruskall-Wallis and Friedman tests for ANOVA
- Runs test for testing randomness

What We Covered in this Module

- Discussed advantages and disadvantages of nonparametric tests
  - Described the general two-sample shift model
- Nonparametric tests for paired data
  - Sign test
  - Wilcoxon signed-rank test
    - Small and large sample variants
- Nonparametric tests for two-samples of independent data
  - Wilcoxon rank sum test
  - Mann-Whitney U test

Homework

- WM&S chapter 15
  - Required: 4, 9, 13, 17, 23, 25, 27
  - Extra credit: None