Curtin University

STAT1006

REGRESSION & NONPARAMETRIC INFERENCE

WEEK 2

Dr Darfiana Nur

31.07.2017

Aims of this lecture

1. Parametric vs Nonparametric

2. t-test paired samples: A revision (Moore et al, 2021, Chapter 7)

3. Wilcoxon Signed Rank Test: Paired samples (Moore et al, 2021, Chapter 15)

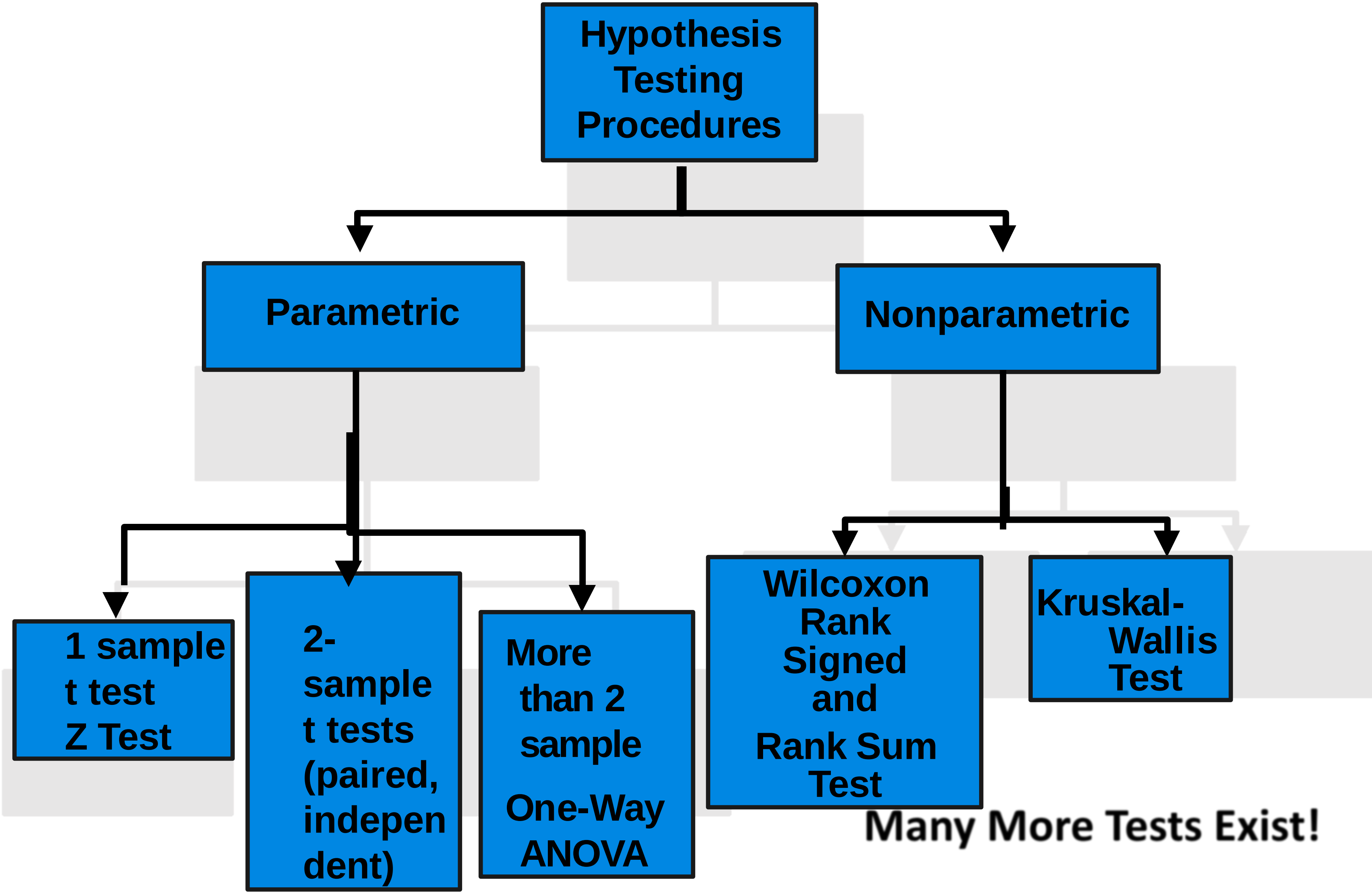
BREAK 5 mins

4. Wilcoxon Rank Sum Test: Two sample (Moore et al, 2021, Chapter 15)

    4.1 Review t-test for two sample (Moore et al, 2021, Chapter 7)

    4.2 Wilcoxon Rank Sum Test: Two sample

Aim 1 Hypothesis Testing Procedures



Many More Tests Exist!

Nonparametric tests

- Useful when sample sizes are small and/or when distributional assumptions cannot be made
- Requires no distributional assumptions
  - But distribution-free does *not* mean *assumption-free*
  - Main assumptions are that:
    - Observations are independent
    - Observations are from a continuous population
    - (Additional assumptions sometimes required)
  - Mean (μ) is replaced by the median ($\tilde{\mu}$ or η) as the parameter of interest
- Can be less efficient when a parametric model is appropriate, but much more efficient when it is not

Nonparametric tests – general procedure

- In parametric tests, we
  - Specify a parametric model, e.g., $y_i \sim N(\mu, \sigma^2)$
  - Formulate null ($H_0: \mu = \mu_0$) and alternative (e.g., $H_A: \mu \neq \mu_0$) hypotheses
  - Using the data, calculate the value of an appropriate test statistic under H_0, and then determine whether the observed value of the test statistic is ‘unusual’ given its distribution
- In nonparametric testing, we
  - Formulate null ($H_0: \tilde{\mu} = \tilde{\mu}_0$) and alternative (e.g., $H_A: \tilde{\mu} \neq \tilde{\mu}_0$) hypotheses
  - Rank the observations in some way (different depending on test and number of samples) and calculate a test statistic based on the sum of ranks
  - Determine whether the test statistic is ‘unusual’ under H_0, given its distribution

NONPARAMETRIC METHODS

Setting	Parametric Test	Rank Test
One sample/paired comparison	One-sample/paired t-test	Wilcoxon signed-rank
Two independent samples	Two-sample t-test	Wilcoxon rank-sum
Several independent samples	One-way ANOVA	Kruskal-Wallis

AIM 2 Paired *t*-test: A revision

- Applies when observations are made
  - on the same person or object
  - often before and after some experimental treatment has been applied, or on matched pairs (e.g. twins, brothers, etc)

- The "samples" are not independent because two measurements made on the same person/object are very closely related and likely to be linked to each other.

- In these cases, we use the paired data to test the difference in the two population means.

- The variable studied becomes $X_{\text{difference}} = (X_1 - X_2)$, and

H_0: $\mu_{\text{diff}}=0$; H_A: $\mu_{\text{diff}} \neq 0$ (or <0, or >0) where $\mu_{\text{diff}} = \mu_1 - \mu_2$

Conceptually, this is no different from tests on one population.

Example 1 Fat content of meat

- A food science laboratory evaluated two different methods of determining the fat content of meat (%).

- They were interested in whether a proposed cheaper method would give results that were consistent with the traditional expensive method.

- Each of 8 random meat samples was divided into two halves.

- The halves were randomly allocated to either method 1 or method 2. The results and summary statistics are:

Sample	1	2	3	4	5	6	7	8
Method1	23.1	23.2	26.5	26.6	27.1	48.3	40.5	25.0
Method2	24.7	23.6	27.1	27.4	27.4	48.3	40.8	24.9
Diff (2-1)	1.6	0.4	0.6	0.8	0.3	0.0	0.3	-0.1

$$n_1 = 8 \qquad \bar{x}_1 = 30.0 \qquad s_1 = 9.23$$

$$n_2 = 8 \qquad \bar{x}_2 = 30.5 \qquad s_2 = 8.99$$

$$\bar{x}_{\text{diff}} = 0.49 \qquad s_{\text{diff}} = 0.54$$

n_1 and n_2: the sample sizes of Method 1 and 2

$\bar{x}_1, \bar{x}_2, \bar{x}_{diff}$: the sample means of Method 1 and 2, the mean difference

s_1, s_2, s_{diff} : the sample standard deviation of Method 1 and 2,
the sample s.d of the differences

STEP 1

$\mu_{diff} = \mu_2 - \mu_1$

H_0: $\mu_{diff} = 0$ H_A: $\mu_{diff} \neq 0$

$\bar{x}_{diff}$ sample means difference

STEP 2 Test statistic

$$t = \frac{\bar{x}_{\text{diff}} - \mu_{\text{diff}}}{\frac{s_{\text{diff}}}{\sqrt{n}}} = \frac{0.49 - 0}{\frac{0.54}{\sqrt{8}}} = 2.57$$

STEP 3 Sampling distribution

To find the p-value, use Student's t distribution with 7 degrees of freedom (df = n-1=8-1, n is number of differences)

STEP 4 By the table or R

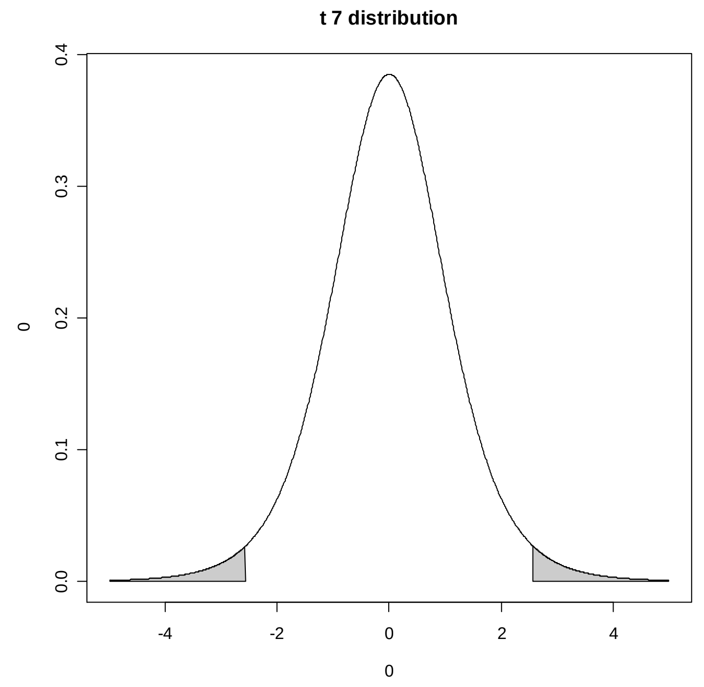
By Excel:

$$\text{p}-\text{value} = P(t_7 > 2.574) + P(t_7 < -2.574)$$
$$= 2\ P(t_7 < -2.574)$$
$$= 2 * T.DIST(-2.574, 7, TRUE)$$
$$= 2 \times 0.018 = 0.0368 \approx 0.037$$

Using the Table D:

Locate 2.574 within df=7

We obtain 2.517 < 2.574 < 2.998 correspond to the upper probability 0.01 < upper < 0.02 such that 0.02 < p-value < 0.04



STEPS 5-6 Decision and Conclusion:

Decision: p-value=0.037 < 0.05. Reject Ho.

This small p-value indicates evidence against H_0, i.e. the data provides evidence that there is a difference between the two different methods of determining the fat content of meat (%).

Assumptions for paired *t*-test

1. The observations are differences from paired or matched samples (i.e. dependent samples)

2. The differences can be considered a random sample from a population of differences

3. The sample mean difference is Normally distributed
   either the underlying data is Normal, or
   the sample size is large enough for the Central Limit Theorem to apply

In Class Exercise 1

Which type of test? One sample, paired samples, two samples?

1. Comparing vitamin content of bread immediately after baking vs. 3 days later (the same loaves are used on day one and 3 days later).

2. Comparing vitamin content of bread immediately after baking vs. 3 days later (tests made on independent loaves).

3. Average fuel efficiency for 2005 vehicles is 21 miles per gallon. Is average fuel efficiency higher in the new generation "green vehicles"?

4. Is blood pressure altered by use of an oral contraceptive? Comparing a group of women not using an oral contraceptive with a group taking it.

5. Review insurance records for dollar amount paid after fire damage in houses equipped with a fire extinguisher vs. houses without one. Was there a difference in the average dollar amount paid?

Aim 3.1 The Wilcoxon Signed Rank Test for Matched Pairs

1. Draw an SRS of size n from a population for a matched pairs study.
2. Suppose we test Ho: $\eta_d=0$ where $\eta_d=\eta_1 - \eta_2$, where η_1 and η_2 are the medians
3. Take the differences $Xd=X_1-X_2$ in responses within pairs.
4. Remove zero differences, if any (i.e $Xd= 0$).
5. Rank the absolute values of these differences.
6. The sum W^+ of the ranks for the **positive** differences is the **Wilcoxon signed rank statistic**.

The Wilcoxon Signed Rank Test

7. If zero differences or ties in ranks exist, calculate the P-value for the Wilcoxon signed rank statistic using the Normal approximation with the continuity correction. Otherwise, use the exact method (*psignrank* or *wilcox.test (…exact=TRUE)*)

If the distribution of the responses is not affected by the different treatments within pairs, then W^+ has mean: $\mu_{W^+} = \frac{n(n+1)}{4}$

and standard deviation: $\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$

Under Ho, the sampling distribution of $z = \frac{(W^+ - \mu_{W^+})}{\sigma_{W^+}}$ is approximately Standard Normal (see next slides)

8. The **Wilcoxon signed rank test** rejects the hypothesis that there are no systematic differences within pairs when the rank sum W^+ is far from its mean.

Ties in the Wilcoxon Signed Rank Test

- Ties among the absolute differences are handled by assigning average ranks.
- A tie *within* a pair creates a difference of zero. Because these are neither positive nor negative, we drop such pairs from our sample.
- Ties within pairs simply reduce the number of observations, but ties among absolute differences complicate finding a *P*-value. In this case, we utilize statistical software.

Wilcoxon Signed Rank Test Statistic W

- The test statistic is an observation (w) of the random variable W.

The exact sampling distribution of W: Under null hypothesis, W follows a specific distribution which is in a very complicated form, depending on ties and rank.

W can be compared to a critical value from a reference table. We use "p-value" for making a decision.

R calculates the p-value from the exact distribution under two conditions:

- n is less than 50 AND
- No ties

Otherwise, Normal approximation with the continuity correction is being used.

Normal Approximation for Large Samples

When $n \geq 15$, the sampling distribution of W_+ (or W_-) approaches the normal distribution with mean

$$\mu_{W_+} = \frac{n(n+1)}{4} \quad \text{and variance} \quad \sigma^2_{W_+} = \frac{n(n+1)(2n+1)}{24}.$$

Therefore, when n exceeds the largest value in Table A.17, the statistic

$$Z = \frac{W_+ - \mu_{W_+}}{\sigma_{W_+}}$$

can be used to determine the critical region for our test.

Wilcoxon Signed Rank Test

- We should take pause here to recall that a paired test is just a single sample test about differences (covered last week).

  We now wish to calculate our test statistic which is an observation (*w*) of the random variable W.

We note:

$$W_{+} = \sum positive\ ranks$$

$$W_{-} = \sum negative\ ranks$$

$$W = min(W+,\ W-)$$

TEST	**Exact** *R only, no manual calculation*	**Approximate (with ties)** *R or Manually*
Wilcoxon Signed Rank (One sample or Paired)	psignrank() (discrete) wilcox.test(..exact=TRUE) *(No ties AND n is less than 50)*	Calculate the p-value using Normal approximation with continuity correction wilcox.test(..exact=FALSE)
Wilcoxon Rank Sum (Two sample)	pwilcox() (discrete) wilcox.test(..exact=TRUE) *(No ties AND n is less than 50)*	Calculate the p-value using Normal approximation with continuity correction wilcox.test(...exact=FALSE)

Signed Rank Test: Example 2

Here are the golf scores of 12 members of a college women's golf team in two rounds of tournament play. **Difference=Round 2 – Round 1**

Player	1	2	3	4	5	6	7	8	9	10	11	12
Round 2	94	85	89	89	81	76	107	89	87	91	88	80
Round 1	89	90	87	95	86	81	102	105	83	88	91	79
Difference	5	-5	2	-6	-5	-5	5	-16	4	3	-3	1

1. Assumption: Draw an SRS of size *n* from a population for a matched pairs study.

2. **STEP 1**

H_0: Scores have the same distribution in Rounds 1 and 2

H_a: Scores are systematically lower or higher in Round 2.(two-sided)

Steps 2-4. Compare **absolute values** of the differences between before and after results.

Absolute Value Difference	1	2	3	3	4	5	5	5	5	5	6	16
Rank	1	2	**3.5**	3.5	5	**8**	8	**8**	8	**8**	**11**	**12**

Signed Rank Test: Example 2

Absolute Value	1	2	3	3	4	5	5	5	5	5	6	16
Rank	1	2	**3.5**	3.5	5	**8**	8	**8**	8	**8**	**11**	**12**

6. The test statistic is the sum of the ranks of the negative differences *(highlighted in blue)* and positive differences. This is the Wilcoxon signed rank statistic. Its value here is $W^- = 3.5 + 8 + 8 + 8 + 11 + 12 = 50.5$; $W+=27.5$ (purple)

To compute the *P*-value correctly, we must also consider the alternative hypothesis, which is "scores are systematically higher for Round 2."

What values of W^+ would favor this alternative?

> **Large values would,** because large values of W^+ indicate either that (1) there are unusually *many* cases where the Round 2 score is larger than the Round 1 score, or (2) there are *quite a few large differences* where the Round 2 score is larger than the Round 1 score.

STEP 2. Its value here $W+=1+2+3.5+5+8+8=27.5$

$W^*=\min(50.5, 27.5)=27.5$

Wilcoxon Signed Rank Test: Example 2

7. **STEP 3.** Given the ties, we calculate the P-value for the Wilcoxon signed rank statistic using the Normal approximation with the continuity correction. n=12

$\mu_{W^+} = \frac{n(n+1)}{4}$ =12(13)/4=39 $\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$ = 12.74755

As this is a two-sided test, we calculate the p-value as:

P-value=2 P(W ≤ 27.5) (Note that 27.5 is to the left of 39)

≈ 2P(W < 27.5 + 0.5) (continuity correction)

Using the Normal approximation with continuity correction:

Z_W+= ((W+ + 0.5) - mu_W+)/ sd(W+) = (28-39)/12.7476 =-0.8629

STEP 4. P-value≈ 2 P(Z< -0.8629) =2*pnorm(-0.8629)=0.3881.

8. **STEPS 5-6.** As the p-value = 0.3881 is large, then we can conclude that there is insufficient evident that scores are systematically lower or higher in Round 2

Signed Rank Test in R: Example 2

```
> wilcox.test(Round2, Round1, paired = TRUE, alternative = "two.sided")
        Wilcoxon signed rank test with continuity correction
data:  Round2 and Round1
V = 27.5, p-value = 0.3843
alternative hypothesis: true location shift is not equal to 0
Warning message:
In wilcox.test.default(Round2, Round1, paired = TRUE, alternative = "two.sided") :
cannot compute exact p-value with ties

> wilcox.test(Round2, Round1, paired = TRUE, alternative = "two.sided",
exact=FALSE)
        Wilcoxon signed rank test with continuity correction
data:  Round2 and Round1
V = 27.5, p-value = 0.3843
alternative hypothesis: true location shift is not equal to 0
```

Program	*P*-value
JMP	$P = 0.388$
Minitab	$P = 0.388$
SPSS	$P = 0.363$

All three programs suggest the same conclusion: we *fail* to reject the null hypothesis, and we conclude that the scores in the two rounds could have equal distributions.

Wilcoxon Rank Sum and Signed Rank Tests in R

Performs one- and two-sample Wilcoxon tests on vectors of data; the latter is also known as 'Mann-Whitney' test.

If only x is given, or if both x and y are given and paired is TRUE, a Wilcoxon signed rank test of the null that the distribution of x (in the one sample case) or of x - y (in the paired two sample case) is symmetric about mu is performed.

```
## One sample
wilcox.test(x, y = NULL,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = FALSE, exact = NULL, correct = TRUE,
        conf.int = FALSE, conf.level = 0.95, ...)
```

Wilcoxon Rank Sum and Signed Rank Tests in R

If both x and y are given and paired is FALSE, a Wilcoxon rank sum test is carried out. In this case, the null hypothesis is that the distributions of x and y differ by a location shift of mu and the alternative is that they differ by some other location shift (and the one-sided alternative "greater" is that x is shifted to the right of y).

```
## Paired samples
wilcox.test(x, y ,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = TRUE, exact = NULL, correct = TRUE,
        conf.int = FALSE, conf.level = 0.95, ...)
## Two independent samples
wilcox.test(x, y ,
        alternative = c("two.sided", "less", "greater"),
        mu = 0, paired = FALSE, exact = NULL, correct = TRUE,
        conf.int = FALSE, conf.level = 0.95, ...)
```

Density, distribution, etc. functions of the Wilcoxon Signed Rank Statistic in *R*

- `dsignrank(x, n, log = FALSE)`

Density for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size *n*.

- `psignrank(q, n, lower.tail = TRUE, log.p = FALSE)`

Distribution function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size *n*.

- `qsignrank(p, n, lower.tail = TRUE, log.p = FALSE)`

Quantile function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size *n*.

- `rsignrank(nn, n)`

Random generation function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size *n*.

Psignrank function

- **psignrank(q, n, lower.tail = TRUE, log.p = FALSE)**

`Distribution function for the distribution of the Wilcoxon Signed Rank statistic obtained from a sample with size` *n*.

  - *q*: the vector of quantiles
  - *n*: number(s) of observations in the sample(s). A positive integer, or a vector of such integers.
  - *lower.tail :* logical; if TRUE (default), probabilities are P[X ≤ x], otherwise, P[X > x].
  - *log.p*: logical; if TRUE, probabilities p are given as log(p).

- The range of W is between 0 and n(n+1)/2

Example 3 (WMMY9e, Ex. 16.7)

The table to the right gives measurements of systolic blood pressure of 16 runners measured before and after a paced 8 km jog. Use the signed-rank test at a 0.05 level of significance to test the null hypothesis that jogging increases the median systolic blood pressure by 8 points against an alternative that the increase is less than 8 points.

	before	after
1	158	164
2	149	158
3	160	163
4	155	160
5	164	173
6	138	147
7	163	167
8	159	169
9	165	174
10	145	147
11	150	156
12	161	164
13	132	133
14	155	161
15	146	153
16	159	170

Example 3: Solution

- Assumption: Draw an SRS of size *n* from a population for a matched pairs study. In this case, the same subjects, before and after.
- **Step 1**. Test $H_0: \tilde{\mu} = 8$ against $H_{01}: \tilde{\mu} < 8$, one-sided test

$\tilde{\mu}$ be the population median systolic blood pressure among runners

In this context, we want to test the null hypothesis of a median increase of 8 points against an alternative that the increase in the median is less than 8 points. Test at a significance level of $\alpha = 0.05$

- **Steps 2-4 (next slide)** After calculating the differences $Z_i = Y_i - X_i$, where X_i is the measurement before running and Y_i is the measurement after running, we subtract 8 from each Z_i
- Rank the absolute values of the $Z_i - 8$
- If there are zeros, we discard them and reduce the sample size accordingly; if there are ties, we take the *average* of the ranks

Example 3: Solution

Z - 8	-2	1	-5	-3	1	1	-4	2	1	-6	-2	-5	-7	-2	-1	3
abs(Z - 8)	2	1	5	3	1	1	4	2	1	6	2	5	7	2	1	3
Ranks	7.5	3	13.5	10.5	3	3	12	7.5	3	15	7.5	13.5	16	7.5	3	10.5

Median of Z_i is 6

- **Step 2.** The test statistic, n=16

$$w_+ = 3 + 3 + 3 + 7.5 + 3 + 10.5 = 30$$

- **Steps 3 and 4**. Using the approximate p-value with continuity correction:

$\mu_{W^+} = \frac{n(n+1)}{4}$=16(17)/4=68; $\sigma_{W^+} = \sqrt{\frac{n(n+1)(2n+1)}{24}}$=19.3391

Using the continuity correction: $z_{W^+} =$((W+ + 0.5) - mu_W+)/ sd(W+)

$p(W \leq 30)$≈P(Z < ((30 +0.5)-68)/19.3391)=P(Z < -1.939)=0.026245.

- **Steps 5-6.** As the p-value is smaller than the significance level 5% then we reject H_0. We conclude that the median increase is less than 8 points

The Wilcoxon signed-rank statistic $(n = 16)$ – Using R for Step 7

R calculates the p-value from the exact distribution under two conditions:

- n is less than 50 AND
- no ties

This case has ties, the approximation is being used.

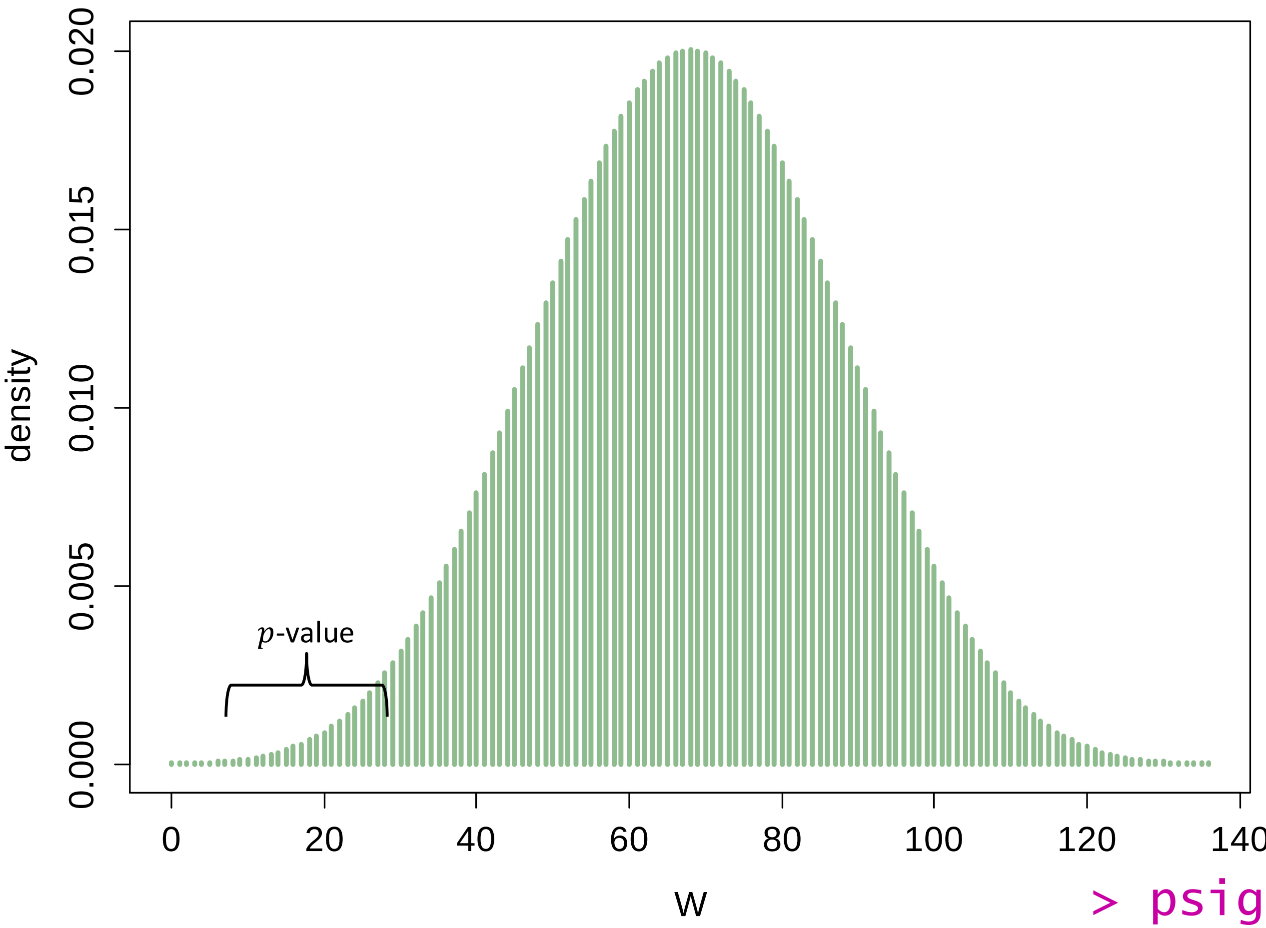
In R: Normal approximation with continuity correction

```
> wilcox.test(Z, alternative = "less", mu = 8, exact = FALSE)

        Wilcoxon signed rank test with continuity correction

data: Z
V = 30, p-value = 0.02562
alternative hypothesis: true location is less than 8
```

Distribution of signed-rank statistic $(n = 16)$ – Exact method (psignrank)




```
# plot signed-rank
#distribution for n = 16
#(range is 0 – 16*17/2)
> UL <- 16*17/2
> plot(0:UL,
dsignrank(0:UL, 16),
type = "h", col =
"darkseagreen", lwd = 3,
     ylab = "density",
xlab = "W")
```

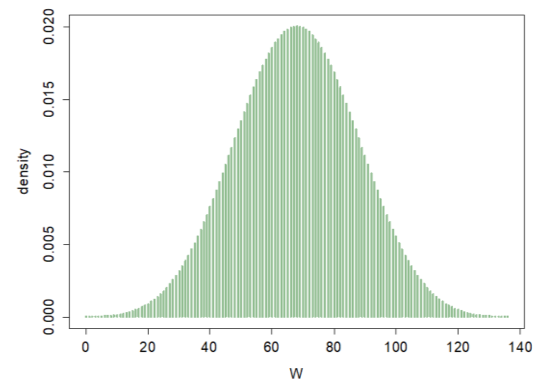
```
> psignrank(q = 30, n = 16)
[1] 0.02532959
```

What about a two-sided alternative?

- **Steps 2, 3 (calculate test stat)** are the same as for the one-sided test.
- **Step 1.** Test $H_0: \tilde{\mu} = 8$ against $H_{01}: \tilde{\mu} \neq 8$, two-sided test
- **Step 4 p-value** is as follows.

If we have framed a two-sided alternative ($H_A: \tilde{\mu} \neq 8$) at the same overall level of significance, then the p-value associated with a two-sided test is in this case

$$p(W \le w_+) + p(W \ge w_-)$$



or, because of symmetry, $\mathrm{p-value} = 2 \times p(W \le w_+)$

The p-value is now

p-value=$2 \times p(W \le w_+)$

≈ 2 P(Z < ((30 +0.5)-68)/19.3391)

=2 P(Z < -1.939)= 2(0.0262) = 0.0524.

- **Steps 5 and 6.** As the p-value is greater than the significance level 5% then we do not reject H_0. We conclude that the median increase is 8 points.
- If using w_-, calculate $w_- = 16 \times 17/2 - w_+$= 136-30=106.
- If using the (discrete) distribution of W using R:

```
> psignrank(q = 30, n = 16)
[1] 0.02532959

> psignrank(105, n = 16, lower.tail = FALSE)
[1] 0.02532959
```

Aim 4.2 Wilcoxon Rank Sum Test Introduction

- The most commonly used methods for inference about the means of quantitative response variables assume that the variables in question have Normal distributions in the population or populations from which the data were drawn.

- In practice, few variables have true Normal distributions, but our methods have been **robust** (not sensitive to moderate deviations from Normality).

- If the data are clearly not Normal, then using the methods such as t-tests will yield inaccurate results. Other approaches must be investigated.

Hypotheses for Wilcoxon Tests

- The Wilcoxon rank sum test compares any two continuous distributions, whether or not they have the same shape, by testing hypotheses that can be stated as

  ***H_0*: the two distributions are the same**

  ***H_a*: one has values that are systematically larger**

- These hypotheses are "nonparametric" because they do not involve any specific parameter such as the mean or median.

- If the two distributions have the same shape, the general hypotheses reduce to comparing medians.

The Wilcoxon Rank Sum Test

The **Wilcoxon rank sum test** rejects the hypothesis that the two populations have identical distributions when the rank sum *W* is far from its mean.

- If the two distributions are identical, then samples of the same size should have roughly the same number of small values, the same number of medium values, and the same number of large values. Thus, the *sums of ranks* for each sample should be roughly the same.
- Instead, if the sample sizes differ, then the *average* of the ranks should be similar for the two samples.

Wilcoxon Rank Sum Test

The null and alternate hypotheses of the Wilcoxon rank sum test:

H_0	H_1	Compute
	$\tilde{\mu}_1 < \tilde{\mu}_2$	u_1
$\tilde{\mu}_1 = \tilde{\mu}_2$	$\tilde{\mu}_1 > \tilde{\mu}_2$	u_2
	$\tilde{\mu}_1 \neq \tilde{\mu}_2$	u

The Wilcoxon Rank Sum Test

1. Draw an SRS of size n_1 from one population and draw an independent SRS of size n_2 from a second population. There are N observations in all, where $N = n_1 + n_2$.

2. We test

   H_0: the two distributions are the same

   Ha: one has values that are systematically larger

3. Rank all N observations, keeping track of which sample the data value comes from.

4. The sum W of the ranks for the **first sample** is the **Wilcoxon rank sum statistic substracted by n1(n1+1)/2**

$$u_1 = w_1 - \frac{n_1(n_1+1)}{2}$$

5. If the two populations have the same continuous distribution (under Ho), then W has the following mean and standard deviation. $\mu_W = \frac{n_1(N+1)}{2}$ $\sigma_W = \sqrt{\frac{n_1 n_2 (N+1)}{12}}$

The Wilcoxon Rank Sum Test

6. If **no ties and *N* is small**, use the exact method in R (either *pwilcox* or *wilcox.test*).

7. Approximate method. To calculate the P-value, we need to know the sampling distribution of the rank sum W when the null hypothesis is true.

(i) P-values for the Wilcoxon test are often based on the fact that the rank sum statistic W becomes approximately Normal as the two sample sizes increase.

$$z = \frac{W - \mu_W}{\sigma_W} = \frac{W - n_1(N+1)/2}{\sqrt{n_1 n_2 (N+1)/12}}$$

To a good approximation, *z* has a standard Normal distribution when the null hypothesis is true and the two sample sizes are relatively large.

(ii) A better approximation for the p-value may be found using a **continuity correction**. This correction acts as if each whole number occupies the entire interval from 0.5 below the number to 0.5 above it.

Wilcoxon Rank Sum Test: Example 4

In 1973, the American League adopted the designated hitter rule, which allows a substitute player to take the place of the typically poor-hitting pitcher. The National League has not adopted this rule. Does the American League produce baseball games with more hits? Here are the number of hits for eight games played on the same spring day.

League	Hits			
American	21	18	24	20
National	19	7	11	13

First, we rank the entire data set from low to high, keeping track of which sample the data value comes from.

Hits	7	11	13	18	19	20	21	24
Rank	1	2	3	4	5	6	7	8

Second, no ties. We can calculate the p-value using the exact method using R.

Then, we sum the ranks for one sample (w1) or the other. The test statistic (n1=4) is

***u1* = (4+6+7+8) – (n1(n1+1)/2)=25 – (4(4+1)/2)= 25-10=15**.

Wilcoxon Rank Sum Test: Example 4 using R

1. **Step 1 Hypotheses**

H_0: The two leagues have identical hits distribution.

H_a: The American League has more hits.

- American <-c(21, 18, 24, 20)
- > National<- c(19, 7, 11, 13)

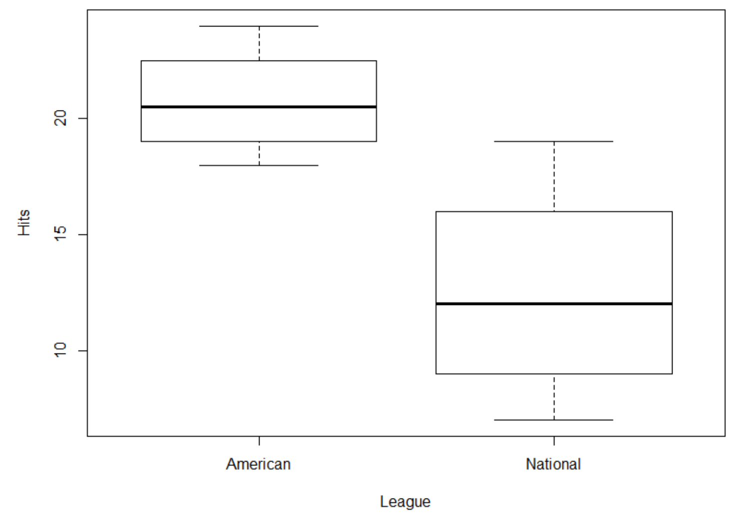
> wilcox.test(American, National, alternative="greater",paired=FALSE)

Wilcoxon rank sum test

data: American and National

W = 15, p-value = 0.02857

alternative hypothesis: true location shift is greater than 0



2. **Step 2 Test statistic** u1=15 (from R as W)

Manually: ***u1*** **= (4+6+7+8) – (n1(n1+1)/2)=25 – (4(4+1)/2)= 25-10=15**

Steps 3 & 4. The exact method is used, as no ties: P-value = 0.02857 (from R)

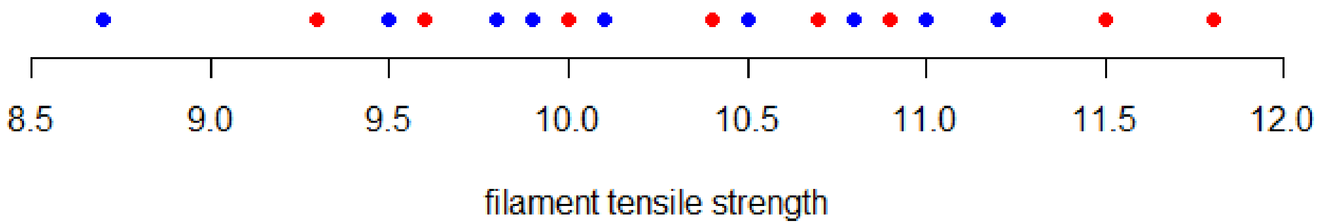
Steps 5 and 6. Decision and Conclusion Because the *P*-value is smaller than $\alpha = 0.05$, there is fairly strong evidence that the American League produces more hits per game.

Example 5. Ex. 16.18 (WMMY9e)

- A polymer filament is manufactured by two processes (1 and 2 respectively). Use the rank-sum test to determine if there is a difference between the median tensile strengths.

1	2
10.4	8.7
9.8	11.2
11.5	9.8
10.0	10.1
9.9	10.8
9.6	9.5
10.9	11.0
11.8	9.8
9.3	10.5
10.7	9.9

Example 5. Ex. 16.18 (WMMY9e)



2	**1**	**2**	**1**	**1**	**2**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**1**	**1**
8.7	9.3	9.5	9.6	9.8	9.8	9.8	9.9	9.9	10	10.1	10.4	10.5	10.7	10.8	10.9	11	11.2	11.5	11.8
1	2	3	4	6	6	6	8.5	8.5	10	11	12	13	14	15	16	17	18	19	20

- **Step 1** Hypotheses: $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ against $H_A: \tilde{\mu}_1 \neq \tilde{\mu}_2$ (two-sided)

- **Step 2** Test Statistic

$w_1 = 111.5, w_2 = 98.5$ and hence

$$u_1 = 111.5 - \frac{10(11)}{2} = 56.5 \text{ and } u_2 = 98.5 - \frac{10(11)}{2} = 43.5$$

Hence, $\min(u_1, u_2) = u = 43.5$, and

Example 5. Ex. 16.18 (WMMY9e)

2	**1**	**2**	**1**	**1**	**2**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**1**	**2**	**2**	**1**	**1**
8.7	9.3	9.5	9.6	9.8	9.8	9.8	9.9	9.9	10	10.1	10.4	10.5	10.7	10.8	10.9	11	11.2	11.5	11.8
1	2	3	4	6	6	6	8.5	8.5	10	11	12	13	14	15	16	17	18	19	20

- **Step 3** We can use the exact (pwilcox) or approximation methods (wilcox.test or the Normal approximation calculated manually)
- **Step 4**

P-value= 2 * pwilcox(q = 43.5, 10, 10) gives 0.63 (exact)

P-value= 0.6495 (approximate, see below)

- **Steps 5 and 6**

As the p-value is large, we do not reject the null hypothesis.

We conclude that there is no difference between the median tensile strengths.

```
> wilcox.test(Ex16.18$process.1,Ex16.18$process.2,alternative="two.sided", exact = FALSE)

        Wilcoxon rank sum test with continuity correction

data:  Ex16.18$process.1 and Ex16.18$process.2
W = 56.5, p-value = 0.6495
alternative hypothesis: true location shift is not equal to 0
medians
```

Exact vs Approximate: A summary

TEST	**Exact** *R only, no manual calculation*	**Approximate (with ties)** *R or Manually*
Wilcoxon Signed Rank (One sample or Paired)	psignrank() (discrete) wilcox.test(..exact=TRUE) *(No ties AND n is less than 50)*	Calculate the p-value using Normal approximation with continuity correction wilcox.test(..exact=FALSE)
Wilcoxon Rank Sum (Two sample)	pwilcox() (discrete) wilcox.test(..exact=TRUE) *(No ties AND n is less than 50)*	Calculate the p-value using Normal approximation with continuity correction wilcox.test(...exact=FALSE)

Rank, *t*, and Permutation Tests

- The big picture of Weeks 1-4 is to compare rank tests (Wilcoxon tests in Weeks 1-3) with traditional *t* procedures and also to the bootstrap and permutation tests (Week 4)
  - Rank tests .vs. traditional *t* procedures
    - Converting to ranks allows us to find exact sampling distributions when the null hypothesis is true. However, in practice the robustness of the *t* procedures means that we rarely encounter data that require nonparametric procedures.
    - Rank methods focus on significance tests, not confidence intervals.
    - Inference based on ranks is restricted to simple settings, whereas *t* procedures extend to more complicated situations, such as experimental design and multiple regression.
- We will learn about permutation tests, which are also nonparametric in nature. However, the calculation of the *P*-value is more complicated for permutation tests than for rank tests.
- Permutation tests have the advantage of flexibility and are available for complicated settings such as multiple regression.