PBIO 504

Non-Parametric Tests

Textbook chapters 13.3, 13.4, 17.3, 21.2

Review

- Recall from previous lectures that many statistical tests assume that the continuous variables are normally distributed.
- This applies to T-tests, ANOVA, correlation, and linear regression.
- Recall that we also have to estimate certain population **parameters** from the information available in the study sample (e.g. mean and standard deviation).

- These tests are usually called **parametric tests**.

What do we do if the data are not normally distributed?

- There are tests available that have fewer assumptions.
- In particular, there is a group of tests that do not require normality, and so they have no need to estimate parameters such as mean and standard deviation.
- Therefore they are called ***nonparametric tests***

(***aka parameter-free or distribution-free tests***)

Nonparametric Tests

Applicable when the outcome does not follow a normal distribution. Examples:

- Outcome is an ordinal variable or a rank
- There are extreme values (outliers)
- Outcome variable corresponds to imprecise measurement, has clear limits of detection (e.g. assays cannot always measure the presence of specific quantities above or below certain limits)

Nonparametric Tests

- Many nonparametric procedures are based on rank transformed data (“rank tests”).

- Follow the same steps as for the parametric methods:
  - State the null and alternative hypotheses.
  - Calculate the test statistic and its corresponding p-value, and then reject or not the null hypothesis according to the preselected alpha level.

Spearman Correlation

- This is the nonparametric version of the Pearson correlation for bivariate observations.
- Recall that the usual Pearson correlation coefficient can vary from -1 to +1.

- The null hypothesis is that $\rho_s=0$
- The Spearman correlation coefficient r_s has the same range of values and the same null hypothesis, but it is based on the **ranks** of the data and not the original data themselves.
- Also known as the rank correlation coefficient.

Spearman Correlation Coefficient

$$r = \frac{\sum (R_X - \bar{R}_X)(R_Y - \bar{R}_Y)}{\sqrt{\sum (R_X - \bar{R}_X)^2 (R_Y - \bar{R}_Y)^2}}$$

Where R_X is the rank of the variable X and R_Y is the rank of the variable Y, and $R_{X\ (bar)}$ is the mean rank of the variable X and $R_{Y\ (bar)}$ is the mean rank of the variable Y.

Note: See next slide for a shortcut formula for easier calculations.

- The Spearman method is robust to outliers.

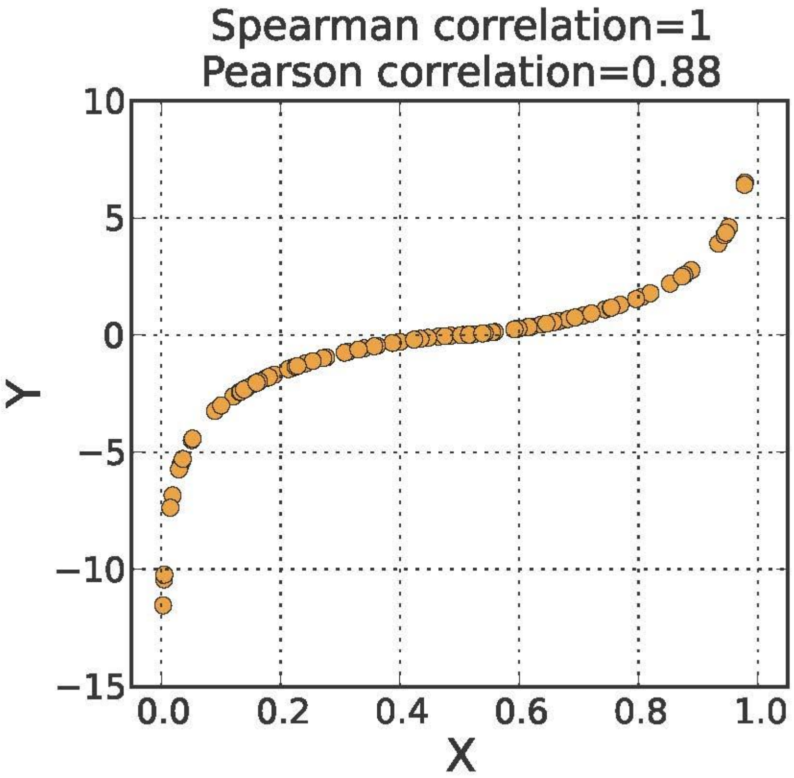
Ranking the Data

- Start by lining up the raw data values of the X and Y variables, assign the ranks from smallest to largest in X, and then do the same separately for Y.
- Then calculate the Spearman correlation coefficient for the set of paired ranks as:

$$r_s = 1 - \frac{6 \,\mathrm{sum}_i(d_i^2)}{n(n^2-1)} \qquad \text{where } i = 1,\ldots,n$$

where d_i are the differences in the X and Y ranks for each (X,Y) pair, n is the total sample size.

Comparison of the two coefficients



Note that the relationship here is monotonic. If we replace data points with their ranks, we'll get a straight line, perfect correlation.

Hence the rank correlation coefficient is one.

How Do We Handle Ties in Ranks

- The Spearman correlation requires resolution of tied values: when two or more data points have the same original value, they are assigned an "average rank".
- *Example: if two data points would have received ranks of 10 and 11, respectively, had they been unique values and not identical values, they will each be assigned a rank of 10.5 (i.e. the average of the ranks).*
- The next higher data point gets a rank of 12.
- Same procedure for more than two equal values.

Cigarette and Blood Nicotine Example

- *A study evaluated the linear relationship between blood nicotine concentration (nmol/liter) and the nicotine yield (mg) of the cigarette smoked, in 10 volunteers.*
- The raw data from these 10 subjects are shown on the next slide, along with their ranks.
- Note that the ranks of the blood nicotine values (X) are determined **separately** from the ranks of the nicotine yield in the cigarette (Y).
- <u>The statistical question</u>: is there a correlation in their ranks?

Ranks of X and Y

X (blood nicotine)	Y (nicotine of cigarette)
185.7 (2)	1.51 (8)
197.3 (5)	0.96 (3)
204.2 (8)	1.21 (6)
199.9 (7)	1.66 (10)
199.1 (6)	1.11 (4)
192.8 (3)	0.84 (2)
207.4 (9)	1.14 (5)
183.0 (1)	1.28 (7)
234.1 (10)	1.53 (9)
196.5 (4)	0.76 (1)

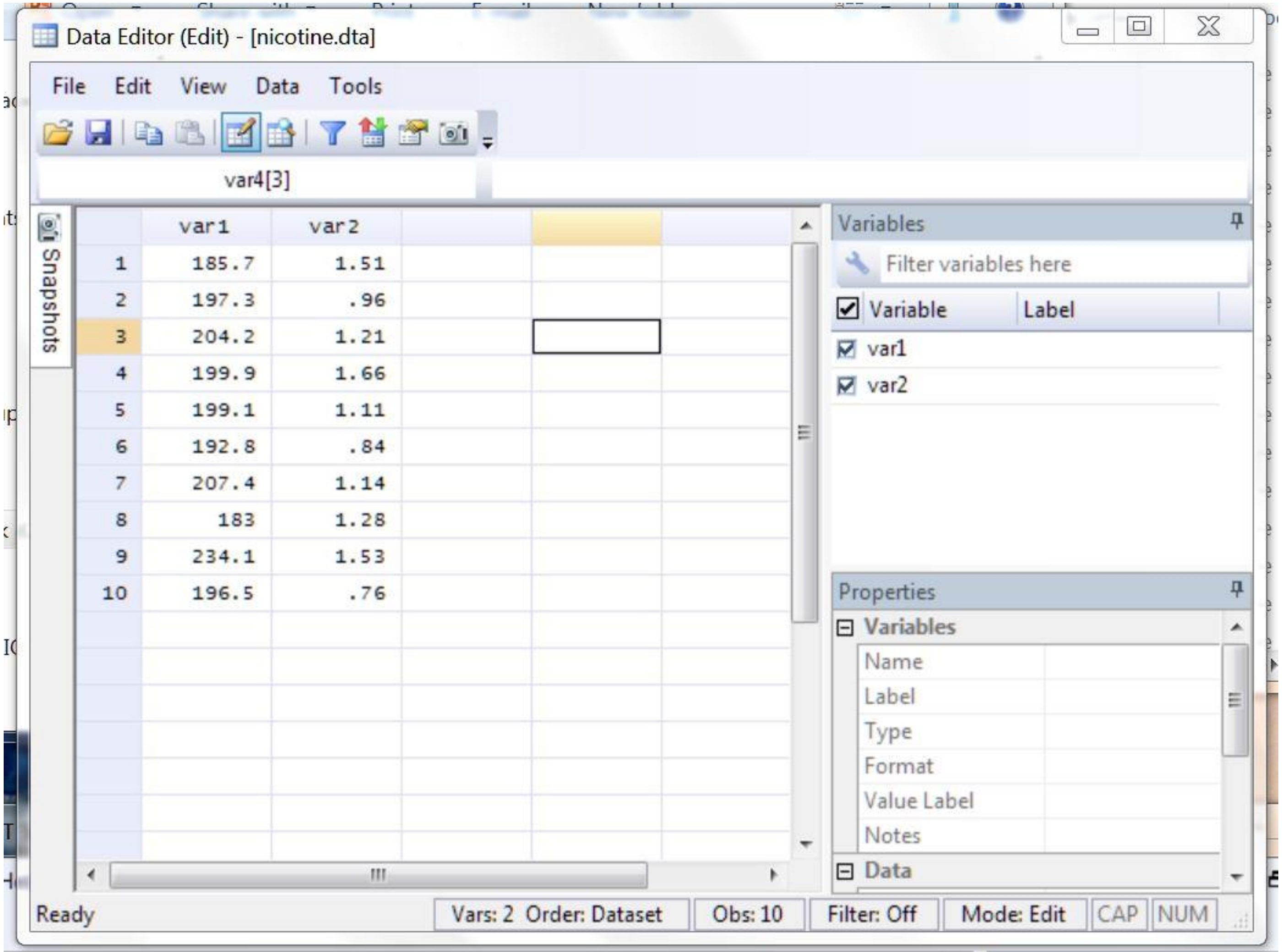
To compute r_s, we have to square the differences in these ranks. So,

$$r_s = 1 - \frac{6\,[\,(2-8)^2 + (5-3)^2 + (8-6)^2 + \dots + (4-1)^2}{10\,(10^2 - 1\,)} = 1 - \frac{6\,[120]}{10\,(99)} = 0.27$$

We would interpret this as a weakly positive correlation between the 2 variables. The p-value from the test will allow us to decide whether to reject H_0.

Data – Data Editor – Data Editor (Edit)

Then cut and paste the two columns from Excel



Statistics – Nonparametric Analysis – Hypothesis Test-Spearman

Then choose and submit

```
. spearman var1 var2, stats(rho obs p)

 Number of obs =      10
Spearman's rho =       0.2727

Test of Ho: var1 and var2 are independent
    Prob > |t| =       0.4458

.
```

Conclusion: based on this p-value we do not have enough evidence to reject the null hypothesis that the two variables are not correlated.

Note that Stata says that the two variables are independent, which is a strong statement. The null hypothesis is actually no correlation.

Wilcoxon Rank Sum Test

- This is a nonparametric version of the two sample T-test, wherein we are interested in comparing two independent groups on the distribution of a continuous or ordinal variable.
- Recall that he null hypothesis for the T-test is that $mean_1 = mean_2$
- In the Wilcoxon test, we will use the ranks of the data, rather than the data values themselves, similar to the Spearman test.
- **Here we will test H_0: $distribution_1 = distribution_2$**

Note: *This test is aka Mann Whitney Wilcoxon*

Assignment of Ranks

- The two groups are combined, and then all data points are sorted from lowest to highest and assigned ranks (1,2,3, etc…) from lowest to highest.

- The ranks are then summed in each of the two groups, with sample sizes n_s and n_L (n_s denotes the smaller of the 2 groups).

- The Wilcoxon test statistic, W_s, is the sum of the ranks in the smaller group. If the p-value of W_s is less than 0.05, we would reject the null hypothesis and conclude that the two distributions are different.

How Do We Handle Ties in Ranks?

- The Wilcoxon Rank Sum Test requires resolution of tied values: when two or more data points have the same original value, they are assigned an "average rank".
- *Example, if the two data points would have received ranks of 10 and 11, respectively, had they been unique values and not identical values, they will each be assigned a rank of 10.5 (i.e. the average of the ranks).*
- The next highest data point gets a rank of 12.

Note: *Ties dealt with same as before for Spearman*

Wilcoxon Test Formula

- The test is actually a form of the Z (or T) test

$$Z\ (or\ T) = (W_s - \mu_w)\ /\ \sigma_w$$

- W_s is the sum of ranks from the smaller group
- $\mu_w = n_s(n_s + n_L + 1)/2$
- where n_s is the sample size of the smaller group and n_L is the sample size of the larger group, and

$$\sigma_w = \sqrt{n_s n_L (n_s + n_L + 1)/12}$$

- *We will calculate the test by hand and using STATA*

Example of Wilcoxon Test

- *Example: A study on the effects of cigarette smoking on sleep patterns recorded the time (minutes) that it took volunteer smokers (m=12) and non-smokers (n=15) to fall asleep. Do the data show that smokers tend to fall asleep differently than non-smokers?*

- Here are the data (minutes to sleep):

Smokers (S): 69.3 56.0 22.1 47.6 53.2 48.1 23.2 13.8 52.7 34.4 60.2 43.8

Non-smokers (NS): 28.6 25.1 26.4 34.9 29.8 28.4 38.5 30.2 30.6 31.8 41.6 21.1 36.0 37.9 13.9

Pool all the data points, sort them from smallest to largest, and rank them.

Observed value	**13.8**	**13.9**	**21.1**	**22.1**	**23.2**	**25.1**	**26.4**	**28.4**	**28.6**
Group	**S**	**NS**	**NS**	**S**	**S**	**NS**	**NS**	**NS**	**NS**
Rank	1	2	3	4	5	6	7	8	9
Observed value	**29.8**	**30.2**	**30.6**	**31.8**	**34.4**	**34.9**	**36.0**	**37.9**	**38.5**
group	**NS**	**NS**	**NS**	**NS**	**S**	**NS**	**NS**	**NS**	**NS**
Rank	10	11	12	13	14	15	16	17	18
Observed value	**41.6**	**43.8**	**47.6**	**48.1**	**52.7**	**53.2**	**56.0**	**60.2**	**69.3**
Group	**NS**	**S**	**S**	**S**	**S**	**S**	**S**	**S**	**S**
Rank	19	20	21	22	23	24	25	26	27

Calculate W_s

- Next we have to calculate W_s, which in this example is the sum of the ranks of the smokers, the smaller of the 2 groups.
- **W_s = 1+4+5+14+20+21+22+23+24+25+26+27= 212**
- μ_s = 12*(12+15+1)/2=168

$$\sigma_w = \sqrt{12*15(12+15+1)/12} = \sqrt{420} = 20.5$$

$$Z\ (or\ T)\ = (W_s - \mu_w)\ /\sigma_w = (212-168)/20.5 = 2.147$$

- The p = 0.03 (STATA next slide). Since p-value < 0.05 we reject the null hypothesis and conclude that smokers differ from non smokers with respect to time to fall asleep.

STATA output

```
. ranksum var1, by(var2)

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

        var2 |      obs    rank sum    expected
-------------+---------------------------------
          NS |       15         166         210
           S |       12         212         168
-------------+---------------------------------
    combined |       27         378         378

unadjusted variance       420.00
adjustment for ties         0.00
                        ----------
adjusted variance         420.00

Ho: var1(var2==NS) = var1(var2==S)
             z =  -2.147
    Prob > |z| =   0.0318

.
```

Conclude $\mathbf{H_a}$: the distributions of time to fall asleep (var1) for smokers and non smokers are different. (Note that var2 has to be the grouping variable, the first variable is of interest)

Wilcoxon Signed Rank Test

- This is the nonparametric version of the paired t-test for matched or paired data

H_0 : the two distributions are the same

-We calculate the difference for each pair.
-Rank the absolute values of the differences.
-Assign signs to the ranks based on the signs of the differences.

Wilcoxon Signed Rank Test

TABLE 13.2
Reduction in forced vital capacity (FVC) for a sample of patients with cystic fibrosis

paired design: FVC from the same patient with/out the drug

Subject	Reduction in FVC (ml)		Difference	Rank	Signed Rank	
	Placebo	Drug				
1	224	213	11	1	1	
2	80	95	−15	2		−2
3	75	33	42	3	3	
4	541	440	101	4	4	
5	74	−32	106	5	5	
6	85	−28	113	6	6	
7	293	445	−152	7		−7
8	−23	−178	155	8	8	
9	525	367	158	9	9	
10	−38	140	−178	10		−10
11	508	323	185	11	11	
12	255	10	245	12	12	
13	525	65	460	13	13	
14	1023	343	680	14	14	
					86	−19

Calculate the Test Statistic

- H_0 : the two distributions are the same
- Calculate the sum of positive ranks, and sum of the negative ranks. Denote the smaller sum by T. Under H_0, we expect ~equal # of + and – ranks.

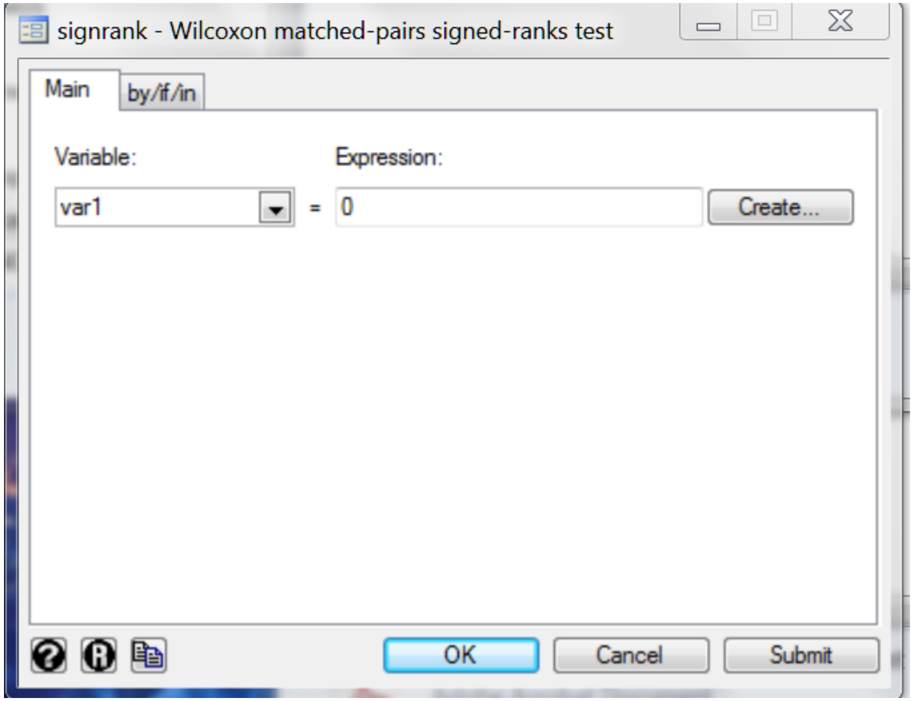
$$Z_T = (T - \mu_T) / \sigma_T,$$

$$\text{where } \mu_T = n(n+1)/4, \sigma_T = \sqrt{n(n+1)(2n+1)/24}$$

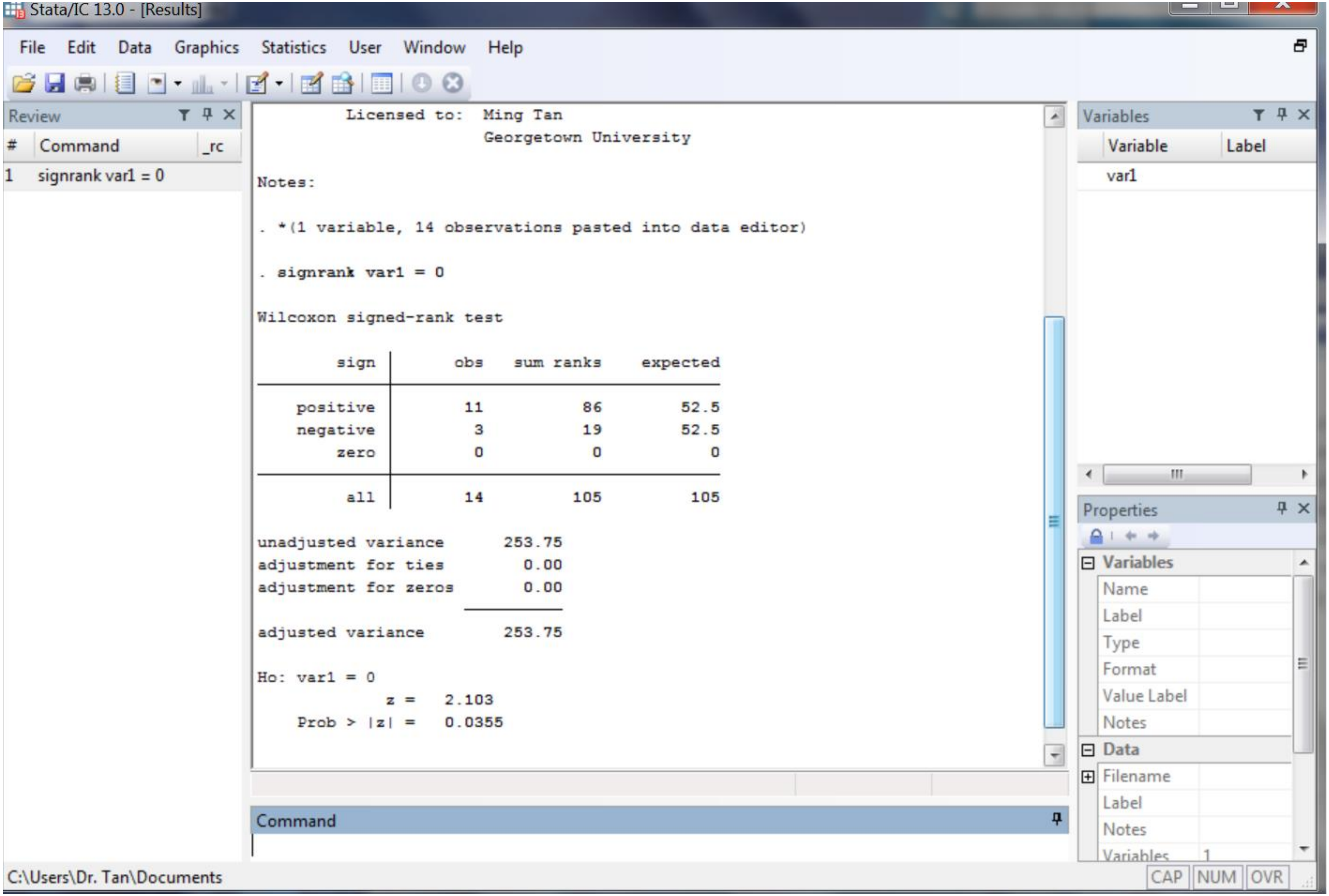
- The z-value=2.10, and p-value = 0.036 < 0.05
- In conclusion, there are different distributions of reduction in FVC for Placebo and Drug.

STATA – Data – Data Editor, cut and paste the differences from Excel

STATA - Statistics – Nonparametric Tests of Hypotheses - Wilcoxon matched pair signed rank
Input 0 under Expression (for the null hypothesis of equal distributions)



There is only one variable: the difference


Stata/IC 13.0 - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command _rc
1 signrank var1 = 0
Licensed to: Ming Tan
Georgetown University
Notes:
. *(1 variable, 14 observations pasted into data editor)
. signrank var1 = 0
Wilcoxon signed-rank test
sign | obs sum ranks expected
positive | 11 86 52.5
negative | 3 19 52.5
zero | 0 0 0
all | 14 105 105
unadjusted variance 253.75
adjustment for ties 0.00
adjustment for zeros 0.00
adjusted variance 253.75
Ho: var1 = 0
z = 2.103
Prob > |z| = 0.0355
Command
Variables
Variable Label
var1
Properties
Variables
Name
Label
Type
Format
Value Label
Notes
Data
Filename
Label
Notes
Variables 1
C:\Users\Dr. Tan\Documents
CAP NUM OVR

Kruskal-Wallis Test

•The Kruskal-Wallis test is the non-parametric version of one-way ANOVA for k indep samples.

•It does *not* require Normality or Equal Variance.

•It is based on rank transformed data and it tests if the distributions differ between groups.

Advantages and Disadvantages of Non-Parametric Tests

Advantages

- They don’t have all the restrictive assumptions of parametric tests.
- Ranks are relatively easy to handle, especially in small data sets.
- Ranks are less subject to measurement errors.
- Ranks can be used with both continuous and ordinal data.

Disadvantages

- In general, they require slightly larger sample size
- Have less power than parametric tests when sample size is small (i.e. higher type II error).
- Their hypotheses tend to be less specific.
- The use of ranks results in loss of information about the original variables.

Other Examples of Nonparametric Analysis

Survival Data Analysis

Log-rank test to compare survival distributions (estimated by Kaplan-Meier curves)

It was presented in the previous lecture

Non-Parametric Measures

- The following slides include examples of non parametric measures.

- These are ***non-parametric*** because **we do not assume any distribution**.

Measure of Agreement

This is different from correlation or association, which only indicate the same trend.

Agreement means x=y for every data point and it does imply correlation or association, but not the other way around.

		Observer 1 Positive	Observer 1 Negative
Observer 2	Positive	a 41	b 3
	Negative	c 4	d 27

$$\text{Overall Agreement} = \frac{a + d}{a + b + c + d} = 0.907\ (91\%)$$

Measuring Agreement Beyond Chance

The Kappa Statistic

$$\text{Kappa} = \frac{\text{Overall agreement - Agreement expected by chance}}{\text{1.0 - Agreement expected by chance}}$$

		Observer 1 Positive	Observer 1 Negative	
Observer 2	Positive	a = 41	b = 3	m_1=44
	Negative	c = 4	d = 27	m_2=31
		n_1=45	n_2=30	T=75

Agreement expected by chance = $\frac{26.4 + 12.4}{75} = 0.517$

Cell a:$(m_1/T * n_1/T) * T = (m_1/T)n_1 = 26.4$

For cell b: $(m_1/T)n_2 = 17.6$

For cell c: $(m_2/T)n_1 = 18.6$

For cell d: $(m_2/T)n_2 = 12.4$

$$\text{Kappa} = \frac{0.907 - 0.517}{1.0 - 0.517}$$

Kappa = 0.81

Interpretation of Kappa Values

- Excellent agreement: kappa from 0.75-1.0
- Good agreement: kappa from 0.40-0.74
- Poor agreement: kappa less than 0.40
- No agreement: kappa at or near 0