Comparing means (or medians) between 2 groups

1

Flow-chart for popularly used statistical tests

Q1,Univariate /Mutivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)	Q6, No. of groups	Q7,sample size	Valid Tests
Univariate	Difference	Independent (un-paired)	Continuous (Normal)	2		Student's t-test
				>2		One-way ANOVA
			Continuous (Non-normal)/ Ordered categorical	2		Mann-Whitney U test
				>2		Kruskal-Wallis H test
			Nominal	2	≦20	Fisher's exact test
				≥2	≥20	Chi-square test
			Time to Event			Log-Rank test (Kaplan-Meier plot)
		Dependent (paired)	Continuous (Normal)	2		Paied-t test
				>2		Repeated measured ANOVA Mixed effect Regression
			Continuous (Non-normal)/ Ordered categorical	2		Wilcoxon signed-rank test
				>2		Friedman test
			Nominal	2		McNemar's test
	Correlation		Continuous (Normal)			Pearson's correlation (r)
			Continuous (Non-normal)/ordered			Spearman's correlation (rs)
			Nominal (2 levels)	2		Spearman/Kappa (Agrreement)
Multivariable		Independent (un-paired)	Continuous (Normal residulas)			Linear Regression
			Continuous (Non-normal residulas)			Linear Regression*
			Ordered categorical			Ordered Logistic Regression
			Nominal (2 levels)			Binary Logistic Regression
			Nominal (>2)			Multinomial Logistic Regression
			Time to Event			Cox Proportional Hazard Regression
		Dependent (paired)	Continuous (Normal residulas)			Linear Mixed Effect Regression
			Continuous (Non-normal residulas)			Linear Mixed Effect Regression*
			Ordered categorical			Generalized Estimation Equation (GEE)
			Nominal (2 levels)			Generalized Estimation Equation (GEE)

* Transform outcome variables for normalizing residuals

Created based on Publishing Your Medical Research Paper, by Daniel Byrne, Williams and Wilkins (1998)

Motivating Example: 2-arm RCT of Type 2 diabetes patients

The American Journal of Medicine (2005) 118, 276–284

ELSEVIER

THE AMERICAN JOURNAL of MEDICINE®

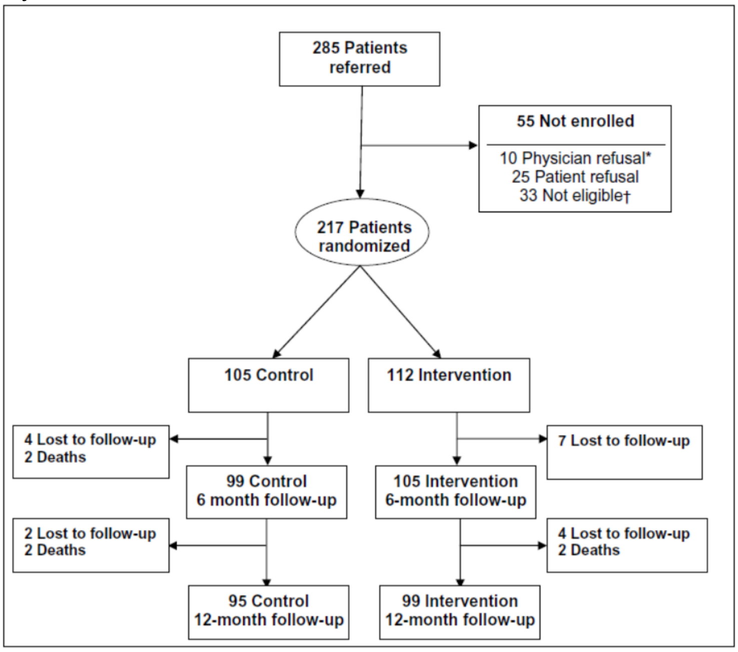
CLINICAL RESEARCH STUDY

A randomized trial of a primary care–based disease management program to improve cardiovascular risk factors and glycated hemoglobin levels in patients with diabetes

Russell L. Rothman, MD, MPP,[a,b] Robb Malone, PharmD, CDE,[c] Betsy Bryant, PharmD, CDE,[c] Ayumi K. Shintani, PhD, MPH,[a,b] Britton Crigler,[c] Darren A. Dewalt, MD,[c] Robert S. Dittus, MD, MPH,[a,f] Morris Weinberger, PhD,[d,e] Michael P. Pignone, MD, MPH[c]

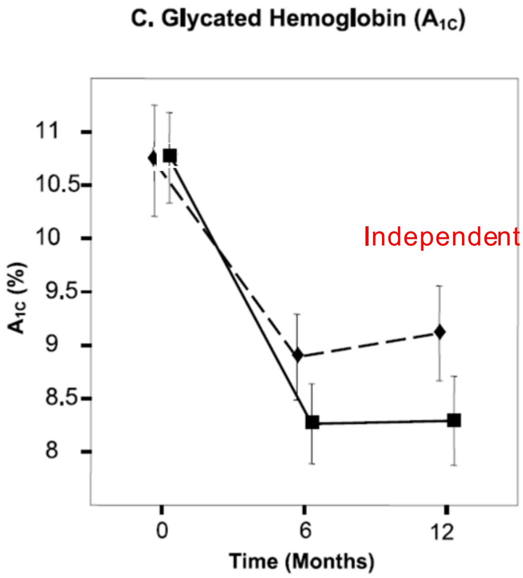
3

Study Flow of *Rothman et al.*



4

MED101x
Introduction to Applied Biostatistics



Independent Sample T-test

5

Flow-chart for popularly used statistical tests

Q1,Univariate /Mutivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)	Q6, No. of groups	Q7,sample size	Valid Tests
Univariate	Difference	Independent (un-paired)	Continuous (Normal)	2		Student's t-test
				>2		One-way ANOVA
			Continuous (Non-normal) / Ordered categorical	2		Mann-Whitney U test
				>2		Kruskal-Wallis H test
			Nominal	2	<20	Fisher's exact test
				≧2	≧20	Chi-square test
			Time to Event			Log-Rank test (Kaplan-Meier plot)
		Dependent (paired)	Continuous (Normal)	2		Paied-t test
				>2		Repeated measured ANOVA Mixed effect Regression
			Continuous (Non-normal) / Ordered categorical	2		Wilcoxon signed-rank test
				>2		Friedman test
			Nominal	2		McNemar's test
	Correlation		Continuous (Normal)			Pearson's correlation (r)
			Continuous (Non-normal) / ordered			Spearman's correlation (rs)
			Nominal (2 levels)	2		Spearman/Kappa (Agrreement)
Multivariable		Independent (un-paired)	Continuous (Normal residulas)			Linear Regression
			Continuous (Non-normal residulas)			Linear Regression*
			Ordered categorical			Ordered Logistic Regression
			Nominal (2 levels)			Binary Logistic Regression
			Nominal (>2)			Multinomial Logistic Regression
			Time to Event			Cox Proportional Hazard Regression
		Dependent (paired)	Continuous (Normal residulas)			Linear Mixed Effect Regression
			Continuous (Non-normal residulas)			Linear Mixed Effect Regression*
			Ordered categorical			Generalized Estimation Equation (GEE)
			Nominal (2 levels)			Generalized Estimation Equation (GEE)

* Transform outcome variables for normalizing residuals

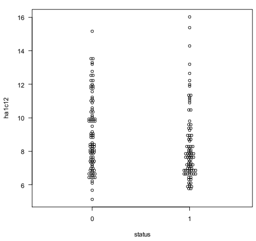
Created based on Publishing Your Medical Research Paper, by Daniel Byrne, Williams and Wilkins (1998)

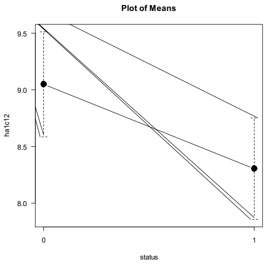
Independent Sample t-test (Student's t-test)

7

Example: In the RCT of pharmacy based intervention of Type 2 diabetes (DiabetesRCT180.xlsx), compare the difference in the improvement in HbA1c at12 month) between the intervention and control group.

H_o: Mean of Ha1c in the intervention group (M1)= Mean of ha1c in the control group (M0).
H_1: Mean of Ha1c in the intervention group (M1) ≠ Mean of ha1c in the control group (M0).

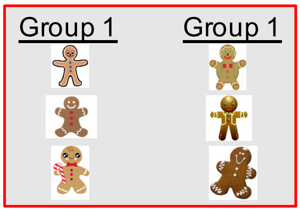




$$\text{Test statistics} = \frac{M1 - M0}{SE\,(M1 - M0)}$$

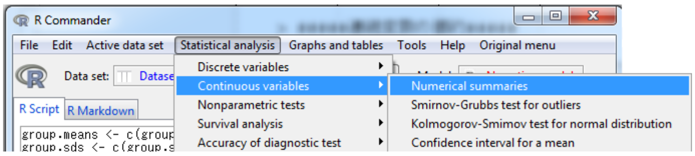
Larger the test statistics, smaller p-value

Difference in means is large => smaller p-value
Data variation is smaller => smaller p-value



8

Descriptive data:




```
> #####Numerical Summaries#####

      Mean       SD   0% 25% 50%  75% 100%  data:n
0 9.048315 2.192711 5.1 7.3 8.4 10.5 15.2     89
1 8.305495 2.148765 5.8 6.9 7.6  8.9 16.0     91
```

$$test\ statistics = \frac{M1 - M0}{SE(M1 - M0)} = \frac{M1 - M0}{\sqrt{\left(\frac{s^2}{n_1} + \frac{s^2}{n_0}\right)}}$$

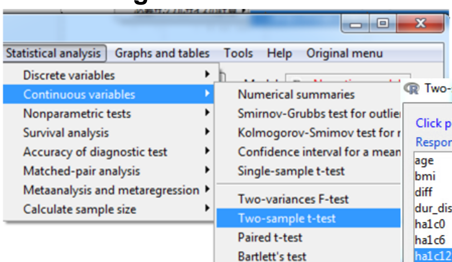
Where

$$s = \sqrt{\left\{\frac{(n_1 - 1){s_1}^2 + (n_0 - 1){s_0}^2}{n_1 + n_0 - 2}\right\}}$$

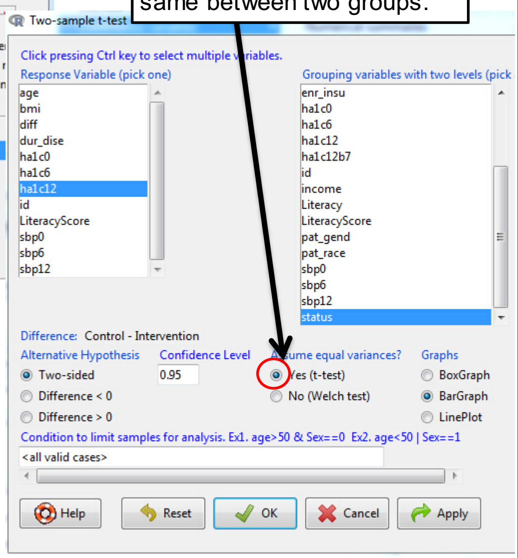
S_1=SD for the intervention group, S_0=SD for the control group
n_1=sample size of the intervention group, n_0=sample size of the control group

9

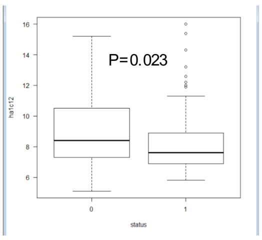
Performing Student's t-test



Select this if variance (SD) of the outcome variable is the same between two groups.




```
> summary.ttest
             Mean       SD    P-value
status=0  9.048315 2.192711  0.0229
status=1  8.305495 2.148765
```



10

Performing Student's t-test

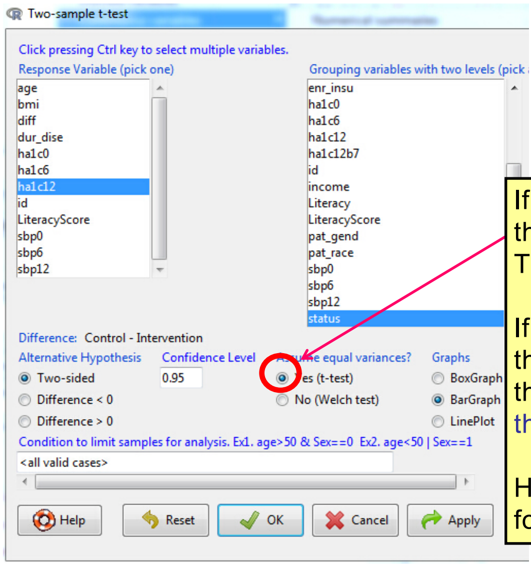
```
Two Sample t-test

data:  ha1c12 by factor(status)
t = 2.2955, df = 178, p-value = 0.02287
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1042454 1.3813948
sample estimates:
mean in group 0 mean in group 1
      9.048315        8.305495
```

This confidence interval is for the difference in means of group 1 and 0.
Unfortunately R output does not compute this for you.
9.048-8.305=0.743.

Thus mean difference in Ha1c between the groups is 0.743 95%CI=(0.104, 1.38) p=0.023

11

Performing Student's t-test : Checking equal variances



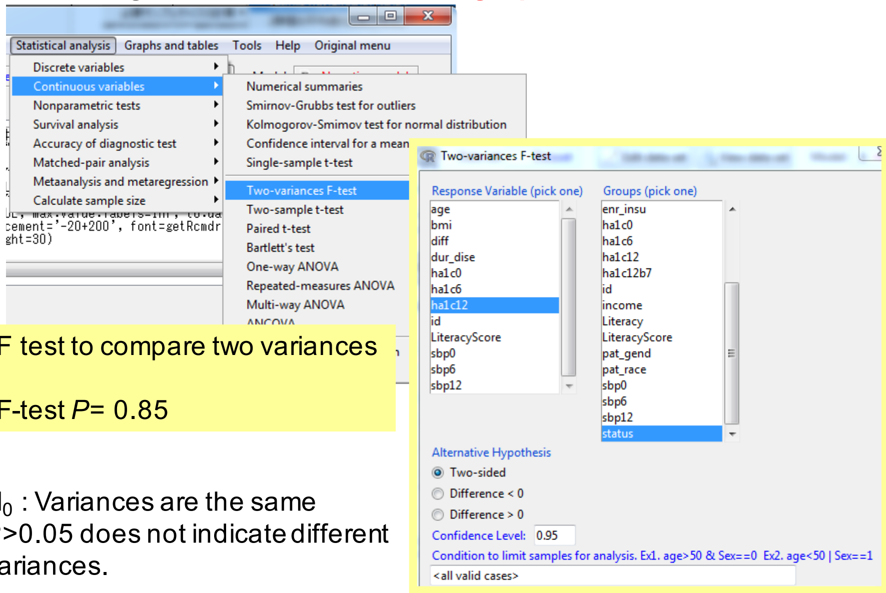
If variance is the same between the groups, use formula on Slide9, Thus click on Yes.

If variance is not the same between the groups, use a different formula, thus click on No (the formula not in these slides).

How to judge Yes or No in the following slide.

12

Performing Student's t-test : Checking equal variances



F test to compare two variances

F-test P= 0.85

H_0 : Variances are the same
P>0.05 does not indicate different variances.

13

Assumption of Student's (independent sample) t-test.

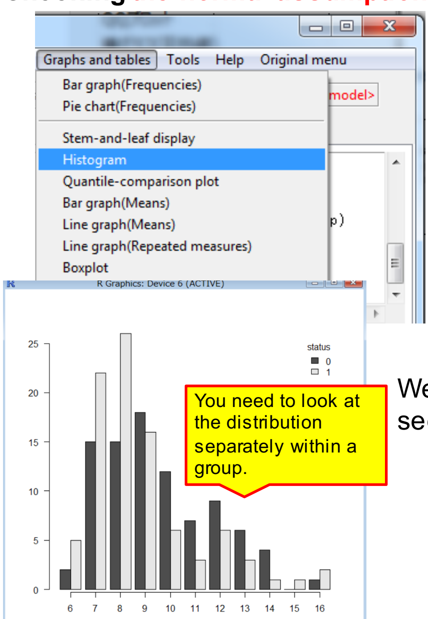
1. Distribution of outcome variable is normal within a group.
2. All observations are independent.

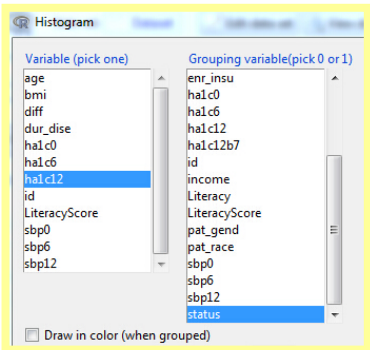
14

**How to deal with a violation of normality assumption:
Log-transforming outcome variables**

15

Checking the normal assumption



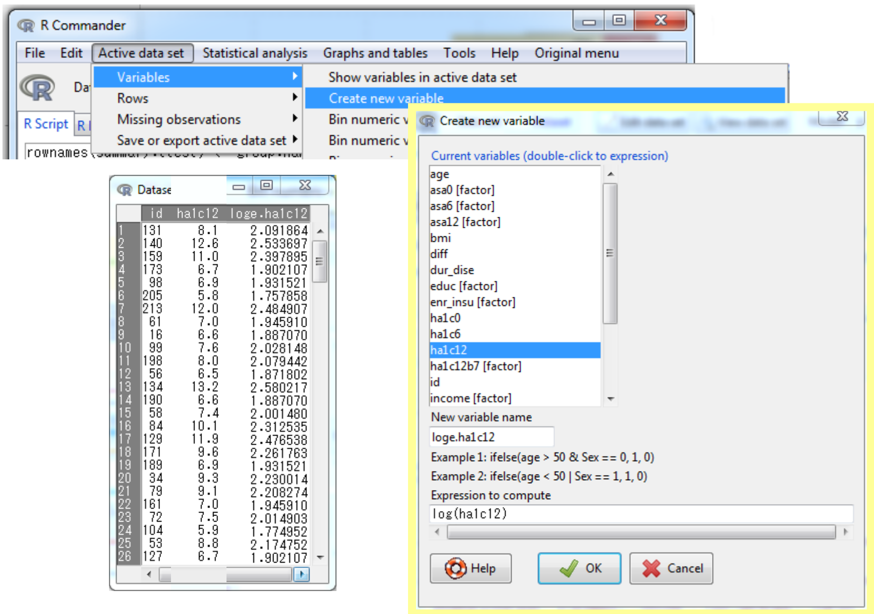


Well, obviously the normality assumption seems violated, thus t-test may not be valid.

16

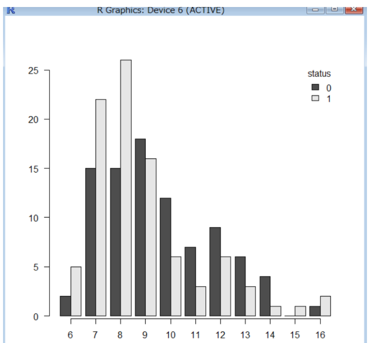
MED101x
Introduction to Applied Biostatistics

How to deal with a violation of normality assumption (1): Log-transforming the outcome variable

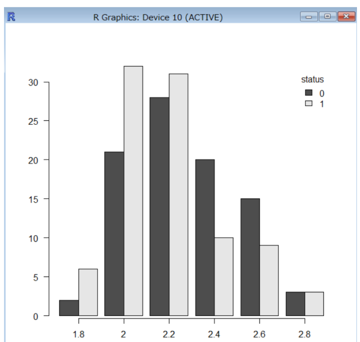


17

Checking for assumption for the model using Log (e) transformed outcome variable, Ha1c12



Ha1c12
Before Log(e) transformation



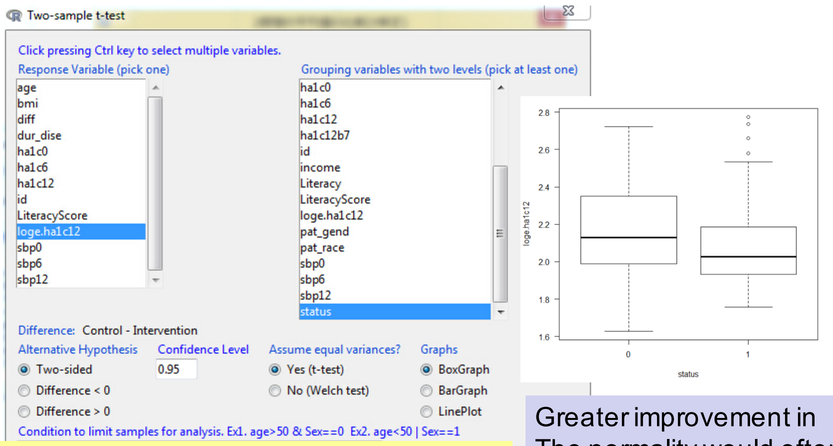
Ha1c12
After Log(e) transformation

18

How to deal with a violation of normality assumption:
Performing student's t-test using the log-transformed outcome variable

19

Solution to the example: student's t-test for log (e) transformed outcome variable, Ha1c



> summary.ttest

	Mean	SD	P-value
status=0	2.174374	0.2378323	0.0151
status=1	2.088704	0.2305144	

Greater improvement in The normality would often lead to gaining analytical power, resulting in smaller p-value.

20

```
> (res <- t.test(loge.ha1c12~factor(status), alternative='two.sided', conf.level=0.95,
+     var.equal=TRUE, data=Dataset))
         Two Sample t-test

data:  loge.ha1c12 by factor(status)
t = 2.4541, df = 178, p-value = 0.01509
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.0167815 0.1545582
sample estimates:
mean in group 0 mean in group 1
       2.174374        2.088704
```

This confidence interval is for the difference in means of group 1 and 2.

2.174374 -2.088704
=0.08567
Exp(0.08567)= 1.089 = 9.05/ 8.30
=Ratio of mean Ha1c12 between groups
=mean Ha1c12 in the control group is 1.089 times of Mean Ha1c12 in the intervention group.

Flow-chart for popularly used statistical tests

Q1,Univariate /Mutivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)	Q6, No. of groups	Q7,sample size	Valid Tests
Univariate	Difference	Independent (un-paired)	Continuous (Normal)	2		Student's t-test
				>2		One-way ANOVA
			Continuous (Non-normal)/ Ordered categorical	2		Mann-Whitney U test
				>2		Kruskal-Wallis H test
			Nominal	2	<20	Fisher's exact test
				≥2	≥20	Chi-square test
			Time to Event			Log-Rank test(Kaplan-Meier plot)
		Dependent (paired)	Continuous (Normal)	2		Paied-t test
				>2		Repeated measured ANOVA Mixed effect Regression
			Continuous (Non-normal)/ Ordered categorical	2		Wilcoxon signed-rank test
				>2		Friedman test
			Nominal	2		McNemar's test
	Correlation		Continuous (Normal)			Pearson's correlation (r)
			Continuous (Non-normal)/ordered			Spearman's correlation (rs)
			Nominal (2 levels)	2		Spearman/Kappa (Agrreement)
Multivariable		Independent (un-paired)	Continuous (Normal residulas)			Linear Regression
			Continuous (Non-normal residulas)			Linear Regression*
			Ordered categorical			Ordered Logistic Regression
			Nominal (2 levels)			Binary Logistic Regression
			Nominal (>2)			Multinomial Logistic Regression
			Time to Event			Cox Proportional Hazard Regression
		Dependent (paired)	Continuous (Normal residulas)			Linear Mixed Effect Regression
			Continuous (Non-normal residulas)			Linear Mixed Effect Regression*
			Ordered categorical			Generalized Estimation Equation (GEE)
			Nominal (2 levels)			Generalized Estimation Equation (GEE)

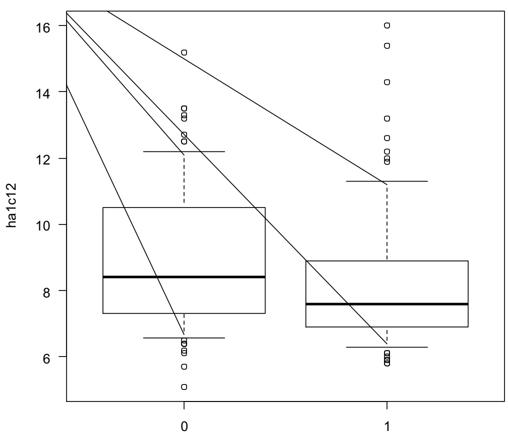
* Transform outcome variables for normalizing residuals

Created based on Publishing Your Medical Research Paper, by Daniel Byrne, Williams and Wilkins (1998)

How to deal with a violation of normality assumption (2):
Use non-parametric test

Mann-Whitney U test
Wilcoxon Rank-Sum test

Comparing two independent distributions of continuous outcome:
Wilcoxon Rank-Sum test / Mann-Whitney U test



Mann-Whitney U test / Wilcoxon rank-sum test

id	status	ha1c12	rank_ha1c12
131	0	8.1	93.0
159	0	11.0	149.0
16	0	6.6	24.5
198	0	8.0	88.0
56	0	6.5	21.0
134	0	13.2	172.5
84	0	10.1	139.0
129	0	11.9	161.0
79	0	9.1	121.0
161	0	7.0	47.0
72	0	7.5	67.5
53	0	8.8	113.0
127	0	6.7	30.0
174	0	6.4	17.5
⋮	⋮	⋮	⋮
140	1	12.6	170.0
173	1	6.7	30.0
98	1	6.9	39.5
205	1	5.8	3.5
213	1	12.0	163.5
61	1	7.0	47.0
99	1	7.6	73.0
190	1	6.6	24.5
58	1	7.4	63.5
171	1	9.6	127.5
189	1	6.9	39.5
34	1	9.3	123.0
104	1	5.9	6.0
28	1	11.0	149.0

(1) Combine the data from the two groups, and order the values from the lowest to the highest.

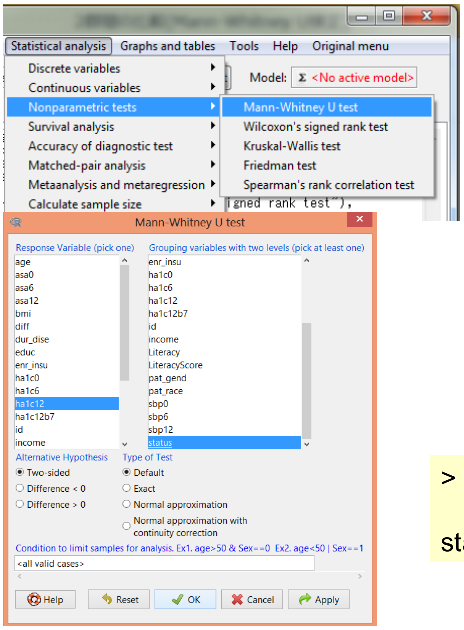
(2) Assign ranks to the individual values from the lowest rank to the highest rank.

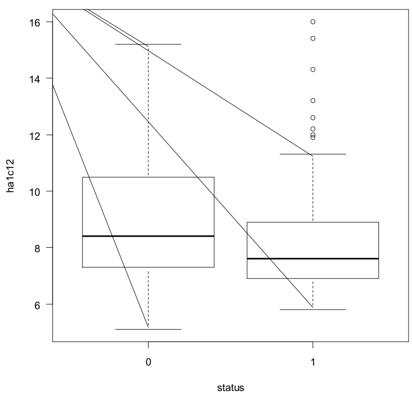
(3) If a group of observations has the same value, then compute the range of ranks for the group, and then assign the average rank for each observation in the group.

Then sum of these ranks is compared between groups.

25

Wilcoxon Rank-Sum test / Mann-Whitney U test: Basic Concept (2)

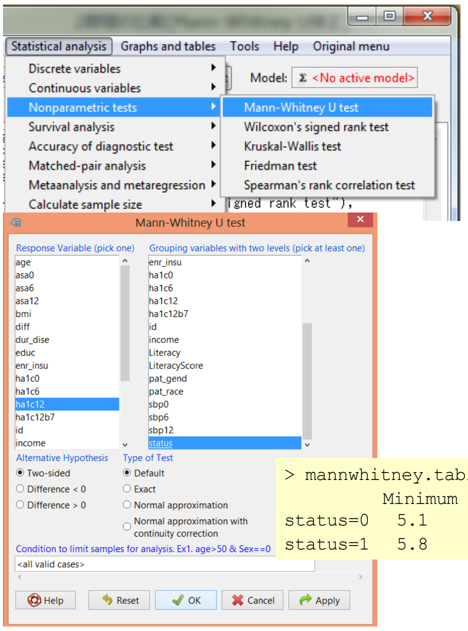


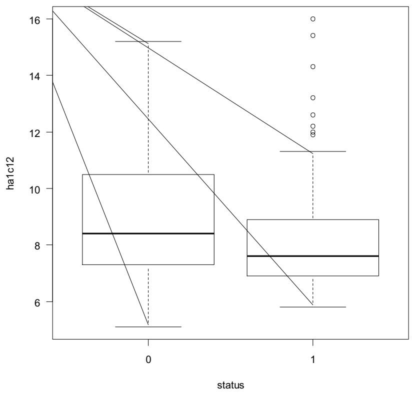



```
> mannwhitney.table
       p.value
status 0.00764
```

26

Wilcoxon Rank-Sum test / Mann-Whitney U test: Basic Concept (2)






```
> mannwhitney.table
         Minimum   25%   Median  75%  Maximum    p.value
status=0   5.1     7.3    8.4    10.5   15.2     0.00764
status=1   5.8     6.9    7.6     8.9   16.0
```

27

Flow-chart for popularly used statistical tests

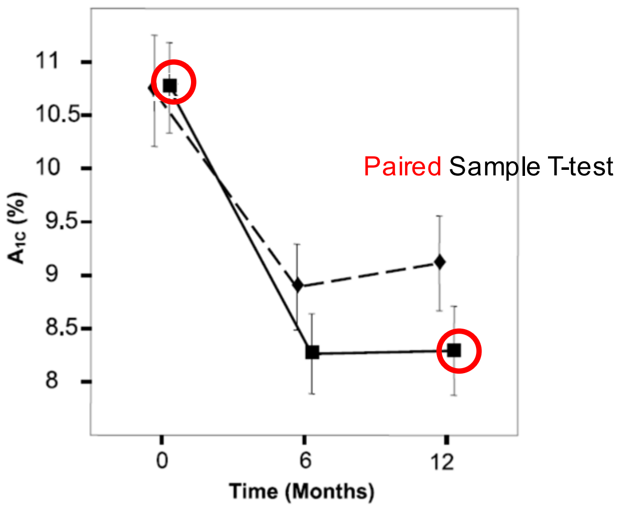
Q1,Univariate /Mutivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)	Q6, No. of groups	Q7,sample size	Valid Tests
Univariate	Difference	Independent (un-paired)	Continuous (Normal)	2		Student's t-test
				>2		One-way ANOVA
			Continuous (Non-normal)/ Ordered categorical	2		Mann-Whitney U test
				>2		Kruskal-Wallis H test
			Nominal	2	<20	Fisher's exact test
				≧2	≧20	Chi-square test
			Time to Event			Log-Rank test(Kaplan-Meier plot)
		Dependent (paired)	Continuous (Normal)	2		Paied-t test
				>2		Repeated measured ANOVA Mixed effect Regression
			Continuous (Non-normal)/ Ordered categorical	2		Wilcoxon signed-rank test
				>2		Friedman test
			Nominal	2		McNemar's test
	Correlation		Continuous (Normal)			Pearson's correlation (r)
			Continuous (Non-normal)/ordered			Spearman's correlation (rs)
			Nominal (2 levels)	2		Spearman/Kappa (Agrreement)
Multivariable		Independent (un-paired)	Continuous (Normal residulas)			Linear Regression
			Continuous (Non-normal residulas)			Linear Regression*
			Ordered categorical			Ordered Logistic Regression
			Nominal (2 levels)			Binary Logistic Regression
			Nominal (>2)			Multinomial Logistic Regression
			Time to Event			Cox Proportional Hazard Regression
		Dependent (paired)	Continuous (Normal residulas)			Linear Mixed Effect Regression
			Continuous (Non-normal residulas)			Linear Mixed Effect Regression*
			Ordered categorical			Generalized Estimation Equation (GEE)
			Nominal (2 levels)			Generalized Estimation Equation (GEE)

* Transform outcome variables for normalizing residuals

Created based on Publishing Your Medical Research Paper, by Daniel Byrne, Williams and Wilkins (1998)

Paired t-test

C. Glycated Hemoglobin (A_{1C})



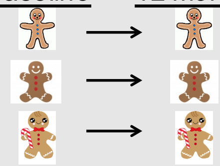
Example (slide 6): Randomized Controlled Clinical Trial.

DiabetesRCT180.xlsx

Assess if there is a change in HbA1c level between baseline and 12 months among patients in the intervention group.

Paired groups: A patient was measured twice at baseline and at 12 months follow-up. Two samples are said to be paired when each data point of the first sample is matched (or related) to a unique data point of the second sample.

Baseline 12 months



31

Paired t-test:

These 2 observations are related, since they came from the same person.

Ho: $\mu_d = 0$. Where μ_d = the difference between baseline and follow-up values

This is equivalent with one-sample test for mean which we learned at the last lecture!!!

id	status	ha1c0	ha1c12	diff
140	1	11.0	12.6	-1.60
173	1	15.2	6.7	8.50
98	1	9.0	6.9	2.09
205	1	9.7	5.8	3.89
213	1	14.3	12.0	2.30
61	1	8.9	7.0	1.89
99	1	11.1	7.6	3.50
190	1	11.6	6.6	5.00
58	1	11.6	7.4	4.20
171	1	9.8	9.6	0.19
189	1	8.7	6.9	1.79
34	1	11.3	9.3	2.00
104	1	9.8	5.9	3.90
28	1	10.6	11.0	-0.39
118	1	12.3	7.5	4.80
123	1	12.7	9.6	3.09

Mean diff > 0
$\updownarrow$
Reduction in Ha1c

Mean diff = 0
$\updownarrow$
No change in Ha1c

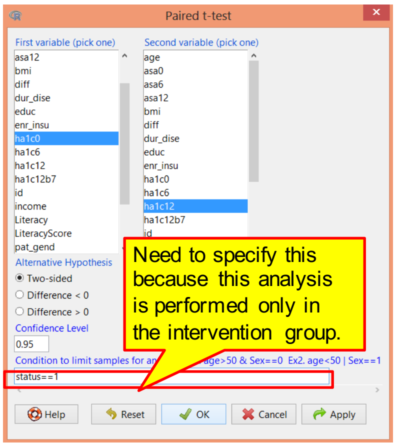
Mean diff < 0
$\updownarrow$
Increase in Ha1c

32

Performing paired t-test

Paired t-test
data: subset(Dataset, status == 1)$ha1c0 and subset(Dataset, status == 1)$ha1c12
t = 8.242, df = 90, p-value = 1.29e-12
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
1.879043 3.072605
sample estimates:
mean of the differences
2.475824

```
> summary.ttest
        mean       sd  p.value
ha1c0  10.781319 2.178323 1.29e-12
ha1c12  8.305495 2.148765
```

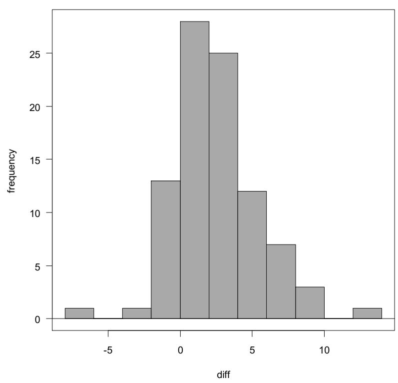


Mean reduction of Ha1c was 2.48 with 95% CI = (1.88, 3.07) p<0.0001. Thus the observed reduction was statistically significant at 2-sided significance level of 5%. Note that the 95% CI for the difference in means ***does not include 0***.

33

Important

Assumption of paired t-tests is:
The difference between the two related variables are normally distributed.



Remember the assumption of the paired t-test is that the difference between the two related variables are normally distributed. Thus we first create the difference variable.

34

Wilcoxon Signed-Rank test

35

Flow-chart for popularly used statistical tests

Q1,Univariate /Mutivariable	Q2, Difference /Correlation	Q3, Paired / related	Q4, Q5 Type of outcome (Normality)		Q6, No. of groups	Q7,sample size	Valid Tests
Univariate	Difference	Independent (un-paired)	Continuous (Normal)		2		Student's t-test
					＞2		One-way ANOVA
			Continuous (Non-normal)/ Ordered categorical		2		Mann-Whitney U test
					＞2		Kruskal-Wallis H test
			Nominal		2	≤20	Fisher's exact test
					≥2	≥20	Chi-square test
			Time to Event				Log-Rank test (Kaplan-Meier plot)
		Dependent (paired)	Continuous (Normal)		2		Paied-t test
					＞2		Repeated measured ANOVA Mixed effect Regression
			Continuous (Non-normal)/ Ordered categorical		2		Wilcoxon signed-rank test
					＞2		Friedman test
			Nominal		2		McNemar's test
	Correlation		Continuous (Normal)				Pearson's correlation (r)
			Continuous (Non-normal)/ordered				Spearman's correlation (rs)
			Nominal (2 levels)		2		Spearman/Kappa (Agrreement)
Multivariable		Independent (un-paired)	Continuous (Normal residulas)				Linear Regression
			Continuous (Non-normal residulas)				Linear Regression*
			Ordered categorical				Ordered Logistic Regression
			Nominal	(2 levels)			Binary Logistic Regression
				(＞2)			Multinomial Logistic Regression
			Time to Event				Cox Proportional Hazard Regression
		Dependent (paired)	Continuous (Normal residulas)				Linear Mixed Effect Regression
			Continuous (Non-normal residulas)				Linear Mixed Effect Regression*
			Ordered categorical				Generalized Estimation Equation (GEE)
			Nominal (2 levels)				Generalized Estimation Equation (GEE)

＊Transform outcome variables for normalizing residuals

Created based on Publishing Your Medical Research Paper, by Daniel Byrne, Williams and Wilkins (1998)

Wilcoxon Signed Rank test
(Non parametric test for comparing paired data).

id	status	ha1c0	ha1c12	diff	Sign	Ranks on Absolute Difference
140	1	11.0	12.6	-1.60	-	3
173	1	15.2	6.7	8.50	+	16
98	1	9.0	6.9	2.09	+	7
205	1	9.7	5.8	3.89	+	11
213	1	14.3	12.0	2.30	+	8
61	1	8.9	7.0	1.89	+	5
99	1	11.1	7.6	3.50	+	10
190	1	11.6	6.6	5.00	+	15
58	1	11.6	7.4	4.20	+	13
171	1	9.8	9.6	0.19	+	1
189	1	8.7	6.9	1.79	+	4
34	1	11.3	9.3	2.00	+	6
104	1	9.8	5.9	3.90	+	12
28	1	10.6	11.0	-0.39	-	2
118	1	12.3	7.5	4.80	+	14
123	1	12.7	9.6	3.09	+	9

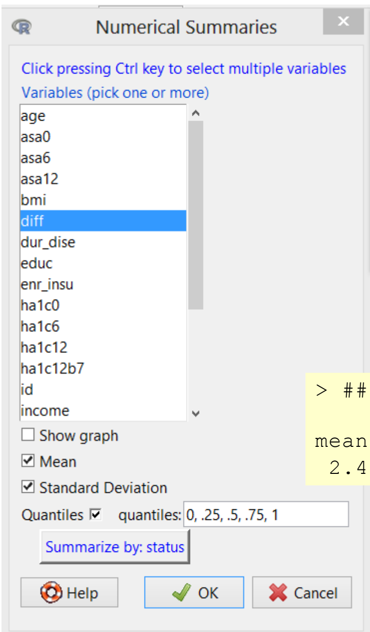
(1) Arrange the difference in order of absolute values.

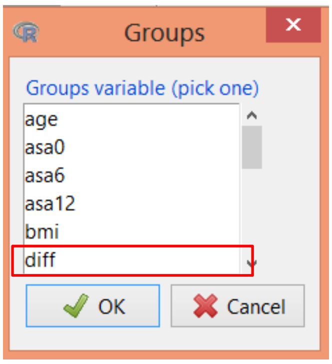
(2) Count the number of differences with the same absolute value.

(3) Ignore the observations where differences = 0, and rank the remaining observations from 1 for the observation with the lowest absolute value, up to n for the observation with the highest absolute value.

(4) If there is a group of several observations with the same absolute values (**ties**), then find the lowest rank in the range = 1 + R and the highest rank in the range = G + R, where R = the highest rank used prior to considering this group and G= the number of differences in the range of ranks for the group. Assign the **average rank** = (lowest rank in the range + highest rank in the range) / 2 as the rank for each difference in the group.

Rank-sum for “+” =16+7+11+8+5+10+15+…+9
Rank-sum for “-” = 3 +2

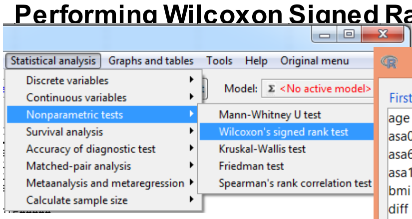


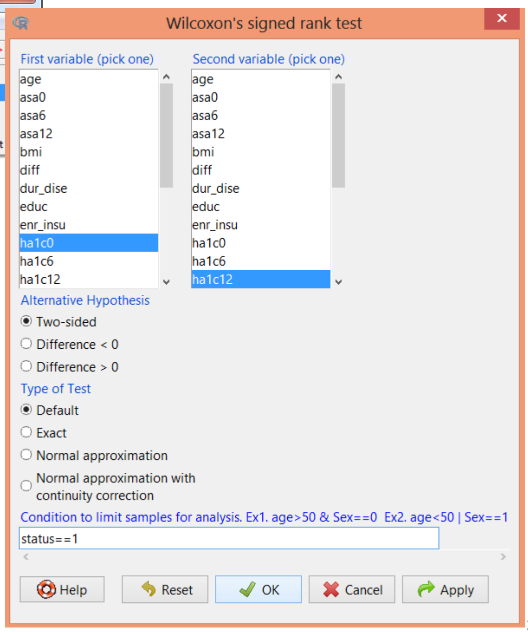



```
> #####Numerical summaries#####

mean         sd        0% 25% 50%  75% 100%  n
  2.475824 2.865556 -6.3 0.7 2.1 3.95 12.8 91
```

MED101x
Introduction to Applied Biostatistics

Performing Wilcoxon Signed Rank test





Median reduction in Ha1c was 2.1.
P< 0.05 thus we reject the Ho, indicating that there was a statistically significant reduction in median Ha1c from baseline to 12 months among patients in the intervention group at two sided-significance level of 5%.

39