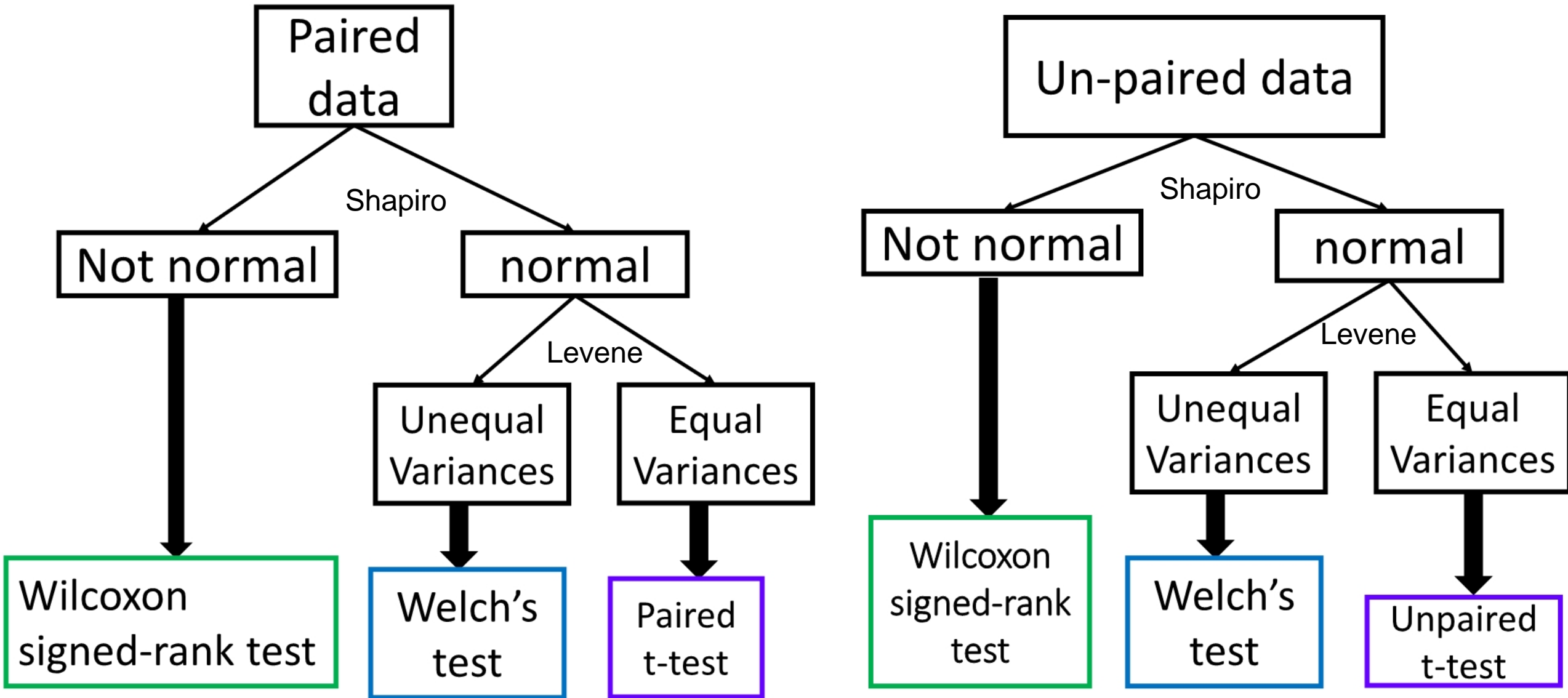# Statistical Methodology for Software Engineering

Hadas Lapid, PhD

Contents

Comparison of two paired & unpaired group means - Extensions

Decision tree for comparison of two sample means

Comparison of paired sample means using t-test

Assumptions:
- Equal group lengths
- No significant outliers
- the difference of pairs follow a normal distribution

If normal: perform paired t-test
Otherwise: perform Wilcoxon signed-rank test
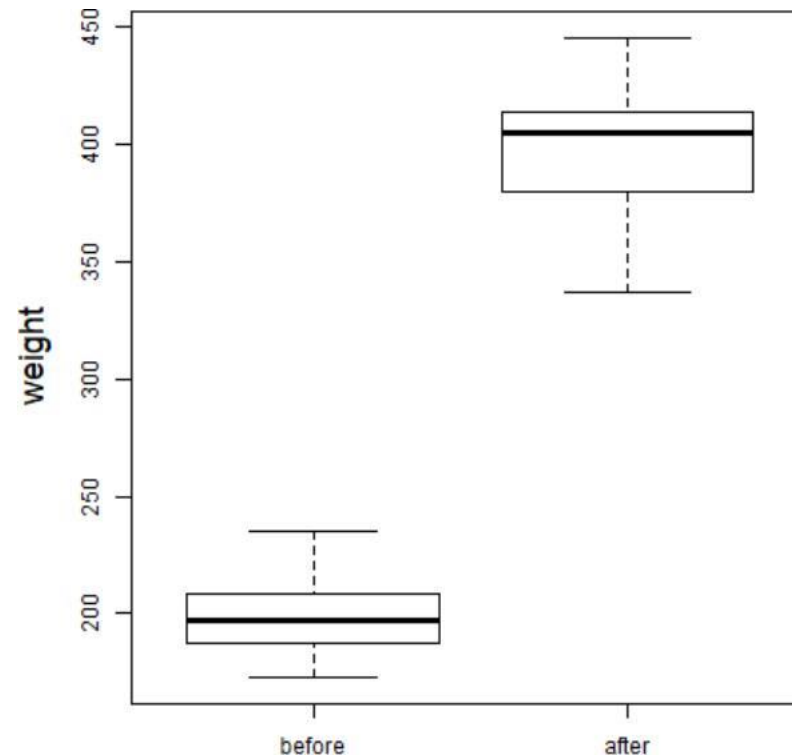
paired t-test
Equal group lengths

Example:

mice before and after treatment

id	before	after
1	187.2	429.5
2	194.2	404.4
3	231.7	405.6
4	200.5	397.2
5	201.7	377.9
6	235.0	445.8
7	208.7	408.4
8	172.4	337.0
9	184.6	414.3
10	189.6	380.3

Paired t-test
No significant outliers

$$\text{Outliers} \equiv \{y_i \in Y: \ |y_i| > (\bar{Y} + 3 \cdot s)\}$$
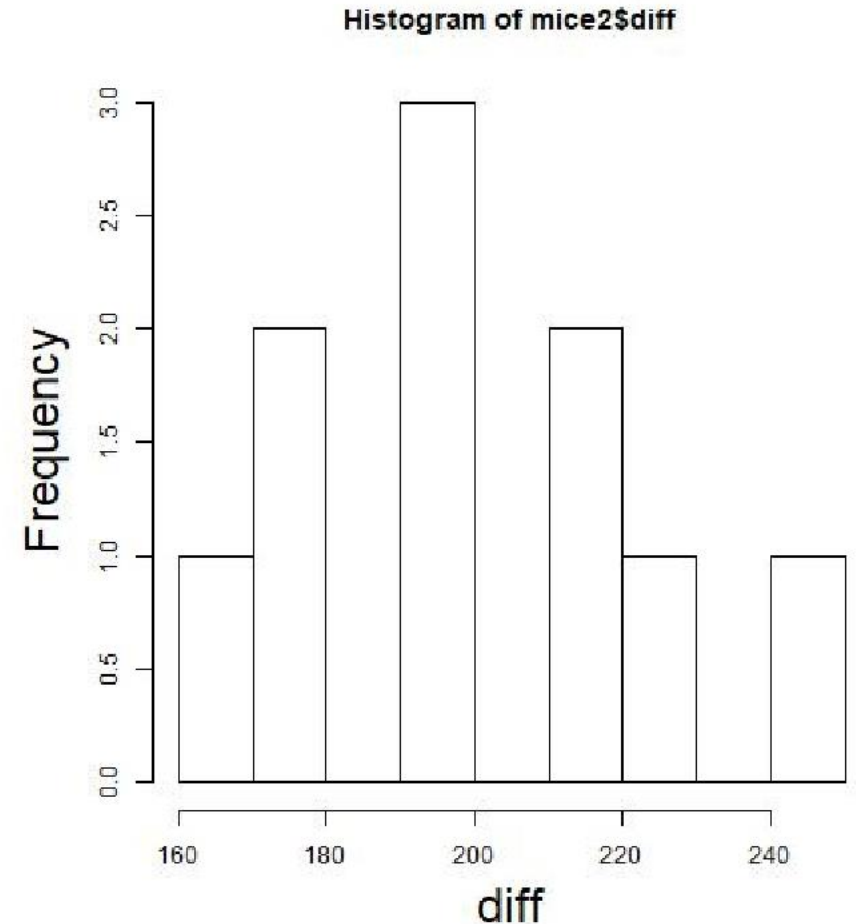
Group differences follow normal distribution

$$\Delta Y = y_1 - y_2 \sim N \quad \text{(Normal)}$$

$$H_0: \Delta Y \sim N$$
$$H_A: \Delta Y \neq N$$

Perform **Shapiro–Wilk test** on ΔY, demand $P_v > 0.05$

🔽 shapiro_h	list [4] (S3: htest)	List of length 4
▶ statistic	double [1]	0.9675122
p.value	double [1]	0.8668697
method	character [1]	'Shapiro-Wilk normality test'
data.name	character [1]	'mice2$diff'

(> 0.05)

Histogram of mice2$diff

$\Delta Y \neq N$: Paired Wilcoxon signed-rank test (non-parametric test)

H_0: $\mu_{before} = \mu_{after}$
H_A: $\mu_{before} \neq \mu_{after}$

If $P_v < 0.05$, group population means are different

wilcoxon_h_mice	list [7] (S3: htest)	List of length 7
statistic	double [1]	0
parameter	NULL	Pairlist of length 0
p.value	double [1]	1.082509e-05
null.value	double [1]	0
alternative	character [1]	'two.sided'
method	character [1]	'Wilcoxon rank sum test'
data.name	character [1]	'values by ind'

If $\Delta Y \sim N$ ➔ Perform Levene test
Test equality of group variances

	Df	F value	Pr(>F)
group	1	0.5266265	0.4773587
	18	NA	NA

$$H_0: \sigma_1 = \sigma_2$$
$$H_A: \sigma_1 \neq \sigma_2$$

If $P_v > 0.05$ variances are not different

Perform Paired test for normal distribution

Paired T-test

⊙ h_paired_mice	list [10] (S3: htest)	List of length 10
⊙ statistic	double [1]	-17.453
⊙ parameter	double [1]	18
p.value	double [1]	9.973664e-13
conf.int	double [2]	-223 -175
⊙ estimate	double [2]	201 400
⊙ null.value	double [1]	0
stderr	double [1]	11.42956
alternative	character [1]	'two.sided'
method	character [1]	' Two Sample t-test'
data.name	character [1]	'mice2$before and mice2$after'

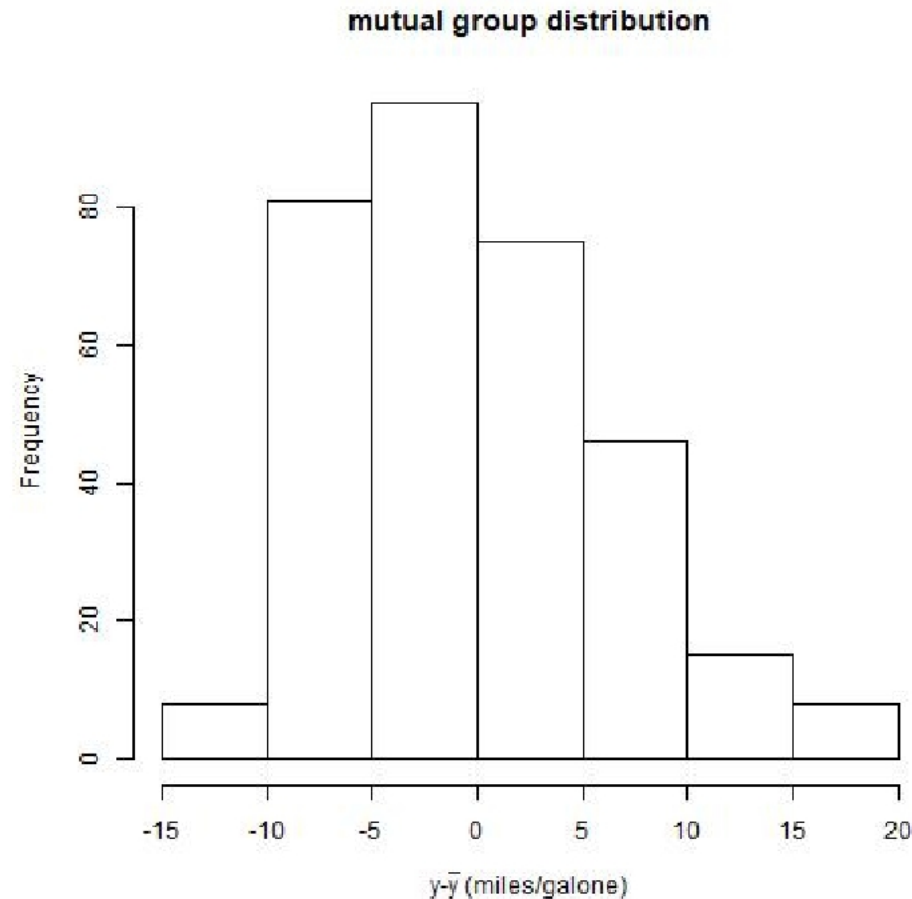# Comparison of un-paired sample means

Assumptions:
- Mutual normality of groups
- Equality of variances

Test mutual normality of groups

Construct mutual sample: subtract group mean from values:

$$\forall i \ in \ Y_1, \ y'_{1i} = y_{1i} - \bar{Y}_1$$
$$\forall i \ in \ Y_2, \ y'_{2i} = y_{2i} - \bar{Y}_2$$

mutual group distribution

Test normality of mutual sample

$$Y' = Y - \bar{Y} \sim N$$

Perform **Shapiro–Wilk test** on Y'

$$H_0: Y' \sim N$$
$$H_A: Y' \neq N$$

demand $P_v > 0.05$

🔽 shapiro_h_unpaired	list [4] (S3: htest)	List of length 4
▶ statistic	double [1]	0.9629943
p.value	double [1]	2.158936e-07
method	character [1]	'Shapiro-Wilk normality test'
data.name	character [1]	'data_flat$vals_shifted'

If $Y' \sim N$ ➔ perform Levene test
Test equality of group variances

	Df	F value	Pr(>F)
group	1	0.1624512	0.6871739
	326	NA	NA

$$H_0: \sigma_1 = \sigma_2$$
$$H_A: \sigma_1 \neq \sigma_2$$

If $P_v > 0.05$ variances are not different

Comparison of un-paired sample means

Normal Distribution

⊙ ttest_h	list [10] (S3: htest)	List of length 10
▶ statistic	double [1]	-12.62059
▶ parameter	double [1]	326
p.value	double [1]	5.272935e-30
conf.int	double [2]	-11.95 -8.73
▶ estimate	double [2]	20.1 30.5
▶ null.value	double [1]	0
stderr	double [1]	0.8190135
alternative	character [1]	'two.sided'
method	character [1]	' Two Sample t-test'
data.name	character [1]	'values by ind'

Un-normal Distribution

Wilcoxon signed-rank test

⊙ wilcoxon_h	list [7] (S3: htest)	List of length 7
▶ statistic	double [1]	2521.5
parameter	NULL	Pairlist of length 0
p.value	double [1]	2.070329e-23
▶ null.value	double [1]	0
alternative	character [1]	'two.sided'