Experimental techniques in research
Groningen, 29-10-2019

Basic statistics

GRIAC Groningen

Maaike de Vries, PhD
m.de.vries04@umcg.nl
Department of Epidemiology
UMCG

Statistics

Aim: statements about population, groups

GRIAC Groningen

Hypotheses

Comparing 2 groups:

H_0: no difference between groups:

X1 mean = X2 mean; X1 mean – X2 mean = 0

H_1: difference between groups:

X1 mean ≠ X2 mean; X1 mean – X2 mean ≠ 0

P-value: Chance to observe H_1 if H_0 is true

GRIAC Groningen

Testing

P < 0.05 Chance to observe H_1 if H_0 is true

< 0.05 (5%)

95% CI 95% chance that the *real* difference between 2 groups falls within the interval

GRIAC Groningen

Testing

Choice for type of test depends upon:

1) Type and distribution of variable

2) (In)dependence of obervations

Testing

Choice for type of test depends upon:

1) **Type and distribution of variable**

2) (In)dependence of obervations

Type of variable

Variable: a quantity with more than 1 value

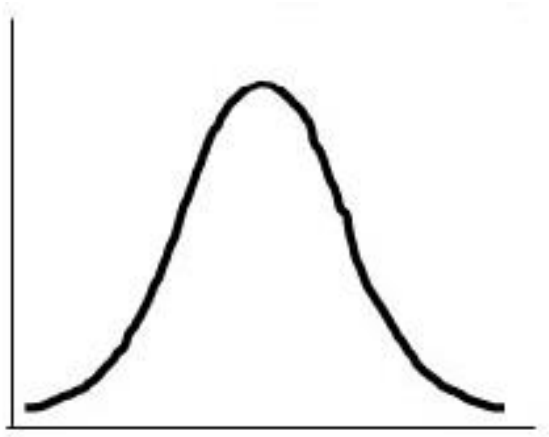
Numerical	- continous:	blood pressure, height
	- discrete:	number of children, number of deaths
Categorical	- ordinal:	better/the same/ worse
	- nominal:	gender, yes/no, blood groups O/A/B/AB place of birth

GRIAC Groningen

Distribution of variable

Numerical variables: Normal (Gaussian)

Normal



GRIAC Groningen

Distribution of variable

Group I: (11, 12, 13, 13, 14, 15)

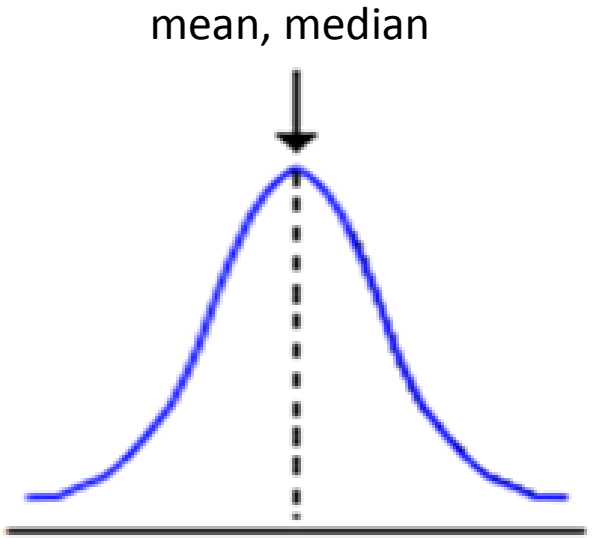
mean = 13

median = 13

Group II: (11, 12, 13, 13, 14, 30)

mean = 15.5

median = 13



Group 1

GRIAC Groningen

Distribution of variable

Mean: Influenced by extreme value, very sensitive to ouliers

Parametric tests are based on mean values

- Data normally distributed?
  - Parametric test
- Data not normally distributed?
  - Non-parametric test

GRIAC Groningen

Non-parametric tests

Advantages:

- Less assumptions (regarding distibution)
- Use of ordinal variables
- Not sensitive for outliers

Disadvantages:

- Less power
- Test only significance wiht p-values, no CI's
- No means, but medians

Testing

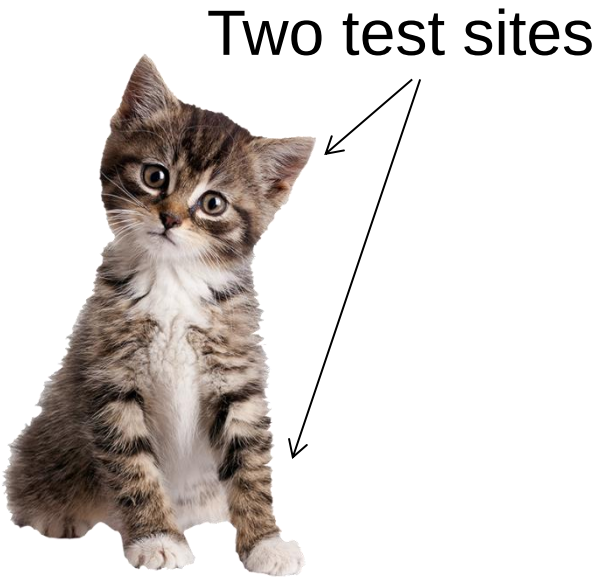
Choice for type of test depends upon:

1) Type and distribution of variable

2) (In)dependence of obervations

(In)dependence of observations

Independent,
two different cats



Dependent (paired)
the same cat

GRIAC
Groningen

(In)dependence of observations

Independent:	males/females
	two treatment groups
Dependent:	repeated measures over time
What about:	fathers and sons
	inhabitants different neighbourhoods
	matched cases

GRIAC Groningen

Dataset

Eur Respir J. 2002 Aug;20(2):383-90.

Perinatal predictors of respiratory symptoms and lung function at a young adult age.

Boezen HM[1], Vonk JM, van Aalderen WM, Brand PL, Gerritsen J, Schouten JP, Boersma ER.

To study the relationship between perinatal factors and the development of respiratory symptoms at young adult age

- Babies born between 1975-1978 in the UMCG
- After ~20 years, a questionnaire was send to mothers and children
  - Random sample of responders was invited for lung function testing

GRIAC Groningen

Type of variable

Variable: a quantity with more than 1 value

Numerical	- continous:	blood pressure, height
	- discrete:	number of children, number of deaths
Categorical	- ordinal:	better/the same/ worse
	- nominal:	gender, yes/no, blood groups O/A/B/AB place of birth

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

	Non-smoking mother	Smoking mother	Total
Total	1388	1701	3089

H_0 : no difference in gestation age < 38 week and maternal smoking

H_1 : difference in gestation age < 38 week and maternal smoking

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

Observed numbers

Gestation < 38 week	Non-smoking mother	Smoking mother	Total
No	1258	1488	2746
Yes	130	213	343
Total	1388	1701	3089

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

Expected numbers

Gestation < 38 week	Non-smoking mother	Smoking mother	Total
No	**1233.9** (2746*1388/3089)	**1512.1** (2746*1701/3089)	2746
Yes	**154.1** (343*1388/3089)	**188.9** (343*1701/3089)	343
Total	1388	1701	3089

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

Expected and Observed numbers

Gestation < 38 week	Non-smoking mother	Smoking mother	Total
No	O = 1258 E = 1233.9	O = 1488 E = 1512.1	2746
Yes	O = 130 E = 154.1	O = 213 E = 188.9	343
Total	1388	1701	3089

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$= \frac{(130-154.1)^2}{154.1} + \frac{(213-188.9)^2}{188.9} + \frac{(1258-1233.9)^2}{1233.9} + \frac{(1701-1512.1)^2}{1512.1}$$

$$= \mathbf{7.713, df=1}$$

(df=(number of rows-1)*(number of columns-1))

GRIAC Groningen

Testing categorical variables

Chi-square test

Gestation age < 38 weeks and maternal smoking

df	0.1	0.05	0.02	0.01	0.005	0.001
1	2.71	3.84	5.41	6.64	7.88	10.83
2	4.61	5.99	7.82	9.21	10.60	13.82
3	6.25	7.82	9.84	11.35	12.84	16.27
4	7.78	9.49	11.67	13.28	14.86	18.47
5	9.24	11.07	13.39	15.09	16.75	20.52
6	10.65	12.59	15.03	16.81	18.55	22.46
7	12.02	14.07	16.62	18.48	20.28	24.32
8	13.36	15.51	18.17	20.09	21.96	26.12
9	14.68	16.92	19.68	21.67	23.59	27.88
10	15.99	18.31	21.16	23.21	25.19	29.59

$6.64 < X^2$ of $7.713 < 7.88$ => $0.01 < p < 0.005$ Reject H_0

There is a difference in gestation age < 38 week and maternal smoking

GRIAC Groningen

Type of variable

Variable: a quantity with more than 1 value

Numerical	- continous:	blood pressure, height
	- discrete:	number of children, number of deaths
Categorical	- ordinal:	better/the same/ worse
	- nominal:	gender, yes/no, blood groups O/A/B/AB place of birth

GRIAC Groningen

Which test for which aim?

Non-parametric	Purpose	Parametric
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*

GRIAC Groningen

Which test for which aim?

Non-parametric	Purpose	Parametric
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*

GRIAC Groningen

Testing numerical variables

Two-sample t-test

Birth weight of children and smoking mothers

Group 1 = Non smokers			Group 2 = Heavy smokers		
3.99	**4.08**	**3.54**	**3.18**	**2.84**	**2.90**
3.79	**3.61**	**3.51**	**3.27**	**3.85**	**3.52**
3.60	**3.83**	**2.71**	**3.23**	**2.76**	**3.60**
3.73	**3.31**	**3.26**	**3.75**	**3.59**	**3.63**
3.21	**4.13**	**3.60**	**2.38**	**2.34**	

H_0 : no difference in birth weight between the groups

H_1 : difference in birth weight between the groups

GRIAC Groningen

Testing numerical variables

Two-sample t-test

Birth weight of children and smoking mothers

Group 1 = Non smokers		Group 2 = Heavy smokers	
$X1_{mean}$	**3.5933**	**$X1_{mean}$**	**3.2029**
S1	**0.3707**	**S1**	**0.4927**
N1	**15**	**N1**	**14**

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}.$$

$$t = \frac{(3.5933 - 3.2029)}{0.4337\sqrt{(1/15+1/14)}} = 2.42 \qquad df = 27$$

(df = N1 + N2 – 2)

GRIAC Groningen

Testing numerical variables

Two-sample t-test

Birth weight of children and smoking mothers

The *t* distribution

df	0.1	0.05	0.02	0.01	0.005	0.001
1	6.314	12.706	31.82	63.657	127.321	636.619
5	2.015	2.571	3.365	4.032	4.773	6.869
10	1.812	2.228	2.764	3.169	3.581	4.587
15	1.753	2.131	2.602	2.947	3.286	4.073
20	1.725	2.086	2.528	2.845	3.153	3.85
25	1.708	2.06	2.485	2.787	3.078	3.725
26	1.706	2.056	2.479	2.779	3.067	3.707
27	1.703	2.052	2.473	2.771	3.057	3.69
28	1.701	2.048	2.467	2.763	3.047	3.674
29	1.699	2.045	2.462	2.756	3.038	3.659
30	1.697	2.042	2.457	2.75	3.03	3.646

$2.052 < t \text{ of } 2.42 < 2.473 \Rightarrow 0.01 < p < 0.005$

Reject H_0

There is a difference in birth weight between the groups

Which test for which aim?

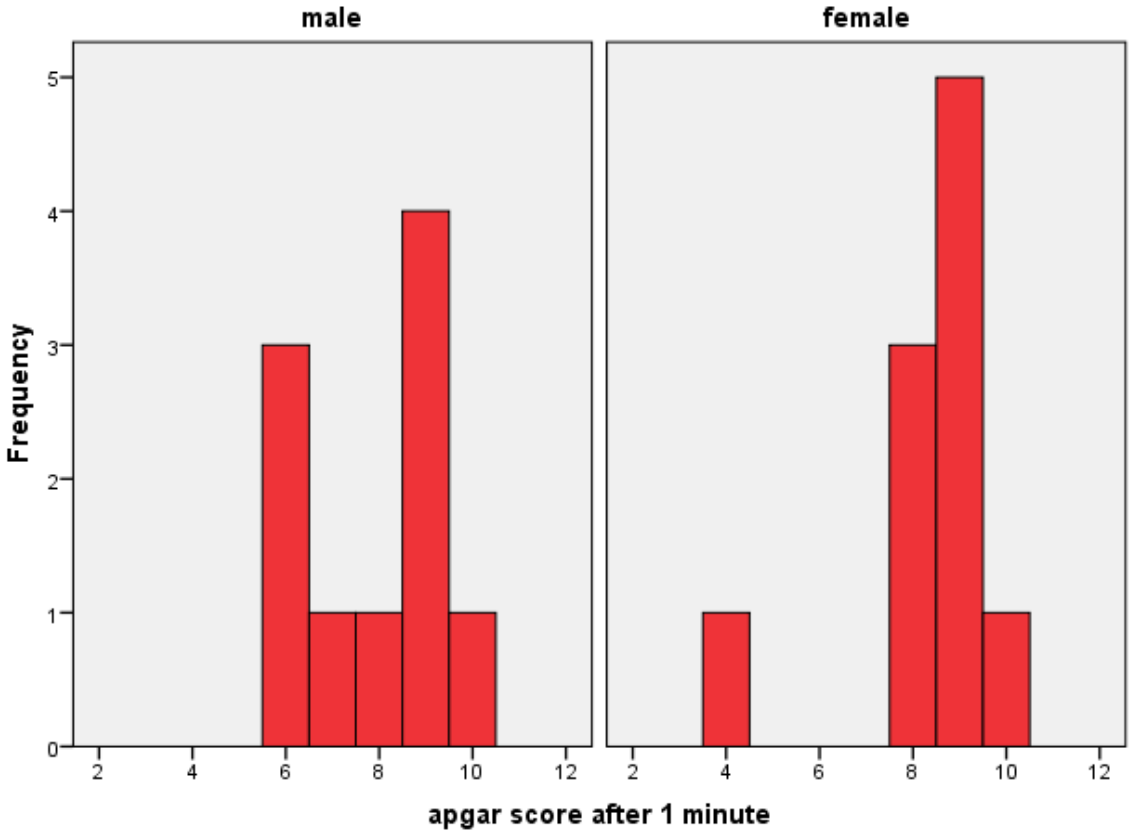
Non-parametric	Purpose	Parametric
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*

GRIAC Groningen

Testing numerical variables

Mann-Whitney U test

APGAR score and gender of the children



GRIAC Groningen

Testing numerical variables

Mann-Whitney U test

H_0: no difference in APGAR score between boys and girls

H_1: difference in APGAR score between boys and girls

1. Rank observations of both groups together from low to high
2. Assign rank-scores (in case of equal rank give mean rank)
3. Sum the ranks for each group
4. Compare these with critical ranges

a.k.a. Wilcoxon rank sum test

gender	apgar1m
1	7
1	8
1	6
1	6
1	10
1	6
1	9
1	9
1	9
1	9
2	9
2	8
2	8
2	9
2	9
2	8
2	9
2	4
2	9
2	10

GRIAC Groningen

Testing numerical variables

1. Rank observations from low to high

Gender	APGAR 1m
2	4
1	6
1	6
1	6
1	7
1	8
2	8
2	8
2	8
1	9
1	9
1	9
1	9
2	9
2	9
2	9
2	9
2	9
1	10
2	10

GRIAC Groningen

Testing numerical variables

2. Assign rank scores (equal rank = mean rank)

Gender	APGARI 1m		Rank
2	4		1
1	6		
1	6		
1	6		
1	7		
1	8		
2	8		
2	8		
2	8		
1	9		
1	9		
1	9		
1	9		
2	9		
2	9		
2	9		
2	9		
2	9		
1	10		
2	10		

GRIAC Groningen

Testing numerical variables

3. Sum the ranks for each group

Gender	APGARI 1m		Rank		Males		Females
2	4		1				1
1	6		3		3		
1	6		3		3		
1	6		3		3		
1	7		5		5		
1	8		7.5		7.5		
2	8		7.5				7.5
2	8		7.5				7.5
2	8		7.5				7.5
1	9		14		14		
1	9		14		14		
1	9		14		14		
1	9		14		14		
2	9		14				14
2	9		14				14
2	9		14				14
2	9		14				14
2	9		14				14
1	10		19.5		19.5		
2	10		19.5				19.5
					SUM		**SUM**
					97		**113**

GRIAC Groningen

Testing numerical variables

4. Compare with critical ranges (Wilcoxon rank sum test)

Use T-statistics of Mann-Whitney-Wilcoxon test

- Take the sum ranks in the smaller group
  (either group can be taken if the sample size is equal)

The Mann-Whitney test (Wilcoxon two sample test)

N1	N2	0.1	0.05	0.02	0.01	0.001
5	10	26-54	23-57	21-59	19-61	15-65
6	10	35-67	32-70	29-73	27-75	23-79
7	10	45-81	42-84	39-87	37-89	31-95
8	10	56-96	53-99	49-103	47-105	41-111
9	10	69-111	65-115	61-119	58-122	52-128
10	10	82-128	78-132	74-136	71-139	63-147

t = 97 or 113 => p > 0.1 => Retain H_0

There is no difference in APGAR score between boys and girls

GRIAC Groningen

Which test for which aim?

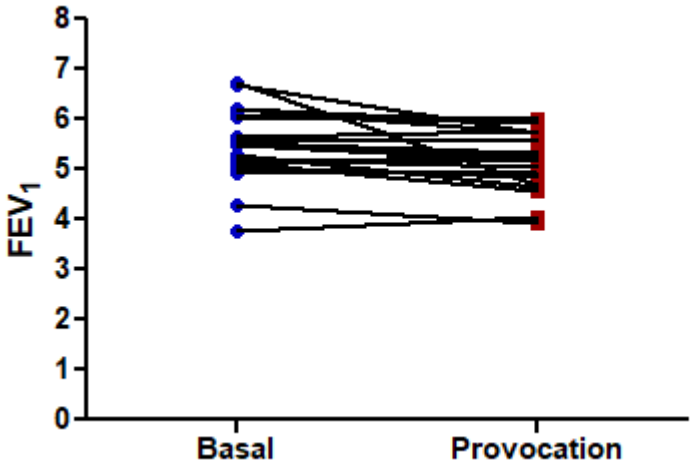
Non-parametric	Purpose	Parametric
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*

GRIAC Groningen

Testing numerical variables

Paired t-test

Lung function levels (FEV_1) after provocation test



H_0 : no difference in lung function levels after provocation test

H_1 : difference in lung function levels after provocation test

GRIAC Groningen

Testing numerical variables

Paired t-test

ID	FEV_basal	FEV_prov
1	5.50	5.18
2	5.07	5.34
3	5.30	4.65
4	5.53	5.32
5	6.10	5.92
6	3.76	4.03
7	6.71	4.68
8	5.50	5.25
9	6.68	5.73
10	6.22	5.72
11	6.05	6.01
12	5.15	5.19
13	5.21	5.03
14	5.65	5.77
15	5.61	5.55
16	4.30	3.92
17	6.22	5.94
18	5.13	4.56
19	4.93	4.91
20	5.06	4.84

Calculate individual difference between basal and provocation:

mean difference = $X_{meandiff} = 0.31$

Calculate standard deviation (sd) of the differences:

sd = 0.49 => se = 0.49/√20 = 0.11

$t = X_{meandiff}/se$ => 0.31 / 0.11

t = 2.818 df = n-1 = 19

Compare with critical ranges

GRIAC Groningen

Testing numerical variables

Paired t-test

The *t* distribution

df	0.1	0.05	0.02	0.01	0.001
1	6.314	12.706	31.82	63.657	636.619
2	2.92	4.303	6.965	9.925	31.599
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.61
5	2.015	2.571	3.365	4.032	6.869
10	1.812	2.228	2.764	3.169	4.587
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.537	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850

$2.537 < t \text{ of } 2.818 < 2.861 \Rightarrow 0.02 < p < 0.01$

Reject H_0

There is a difference in lung function after provocation

GRIAC Groningen

Which test for which aim?

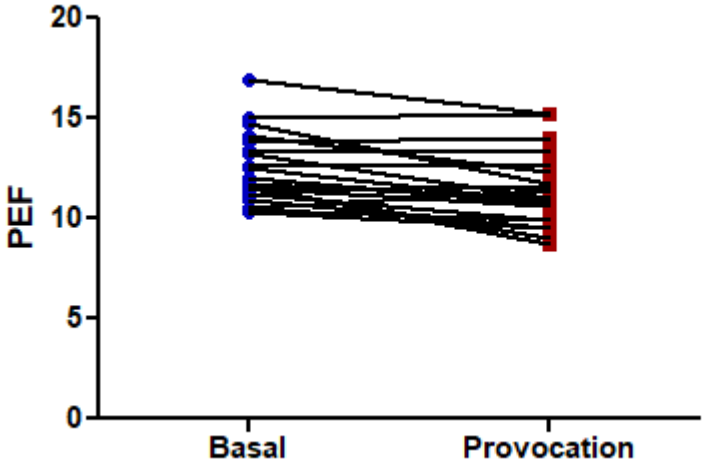
Non-parametric	Purpose	Parametric
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*

GRIAC Groningen

Testing numerical variables

Wilcoxon signed rank test

Lung function levels (PEF) after provocation test



H_0 : no difference in lung function levels after provocation test

H_1 : difference in lung function levels after provocation test

GRIAC Groningen

Testing numerical variables

Wilcoxon signed rank test

ID	PEF_basal	PEF_prov
1	12.51	10.80
2	13.83	14.02
3	11.52	8.98
4	11.45	11.62
5	13.25	11.05
6	11.93	10.67
7	11.65	8.69
8	14.12	12.27
9	10.37	9.90
10	13.43	13.35
11	14.97	15.20
12	10.48	11.46
13	10.87	9.50
14	14.71	11.67
15	16.94	15.25
16	10.30	9.52
17	12.61	12.73
18	11.26	9.95
19	11.55	10.58
20	11.75	11.04

1. Calculate difference per individual (exclude “zero” differences)

2. Rank the differences in increasing sequence in size
   (ignore the sign)
   Give mean value in case of equal rank

3. Sum up positive ranks (T+) and the negative ranks (T-)

4. Compare with critical values

Testing numerical variables

Wilcoxon signed rank test

ID	PEF_bas	PEF_prov		Sign	Abs diff	Rank		Neg	Pos
10	13.43	13.35		+	0.08	1			1
17	12.61	12.73		-	0.12	2		2	
4	11.44	11.62		-	0.18	3.5		3.5	
2	13.83	14.02		-	0.18	3.5		3.5	
11	14.97	15.20		-	0.23	5		5	
9	10.37	9.90		+	0.47	6			6
20	11.75	11.04		+	0.71	7			7
16	10.30	9.52		+	0.79	8			8
19	11.55	10.58		+	0.97	9			9
12	10.48	11.46		-	0.98	10		10	
6	11.98	10.67		+	1.31	11.5			11.5
18	11.26	9.95		-	1.31	11.5		11.5	
13	10.87	9.50		+	1.38	13			13
15	16.94	15.25		+	1.70	14			14
1	12.51	10.80		+	1.71	15			15
8	14.12	12.27		+	1.85	16			16
5	13.25	11.05		+	2.20	17			17
3	11.52	8.98		+	2.54	18			18
7	11.65	8.69		+	2.96	19			19
14	14.71	11.67		+	3.04	20			20
								SUM	**SUM**
								35.5	**174.5**

GRIAC Groningen

Testing numerical variables

Wilcoxon signed rank test

Wilcoxon one sample (or matched pairs) test

N	0.1	0.05	0.02	0.01	0.001
15	30-90	25-95	19-101	15-105	6-114
16	35-101	29-107	23-113	19-117	9-127
17	41-112	34-119	28-125	23-130	11-142
18	47-124	40-131	32-139	27-144	14-157
19	53-137	46-144	37-153	32-158	18-172
20	60-150	52-158	43-167	37-173	21-189

T = 35.5 or 174.5 => $0.01 < p < 0.001$

Reject H_0

There is a difference in lung function after provocation

GRIAC Groningen

Which test for which aim?

Non-parametric	Purpose	Parametric	
Mann-Whitney U test	Test differences between 2 groups (unpaired observations)	*Two sample t-test (unpaired t-test)*	Numerical variables
Kruskall Wallis	Test differences between >2 groups (unpaired observations)	*One-way analysis of variance*	
Wilcoxon signed rank test	Test differences within 1 group (paired observations)	*Paired t-test*	
	Test differences between 2 categorical variables	*Chi-square (or Fisher's exact)*	

GRIAC Groningen