UNIVERSITÉ DU LUXEMBOURG

LUXEMBOURG INSTITUTE OF HEALTH
RESEARCH DEDICATED TO LIFE
Multiomics Data Science

BIOSTATISTICS

Lecture 12

Non-parametrical Tests

dr. Petr Nazarov

petr.nazarov@lih.lu

08-05-2020

OUTLINE

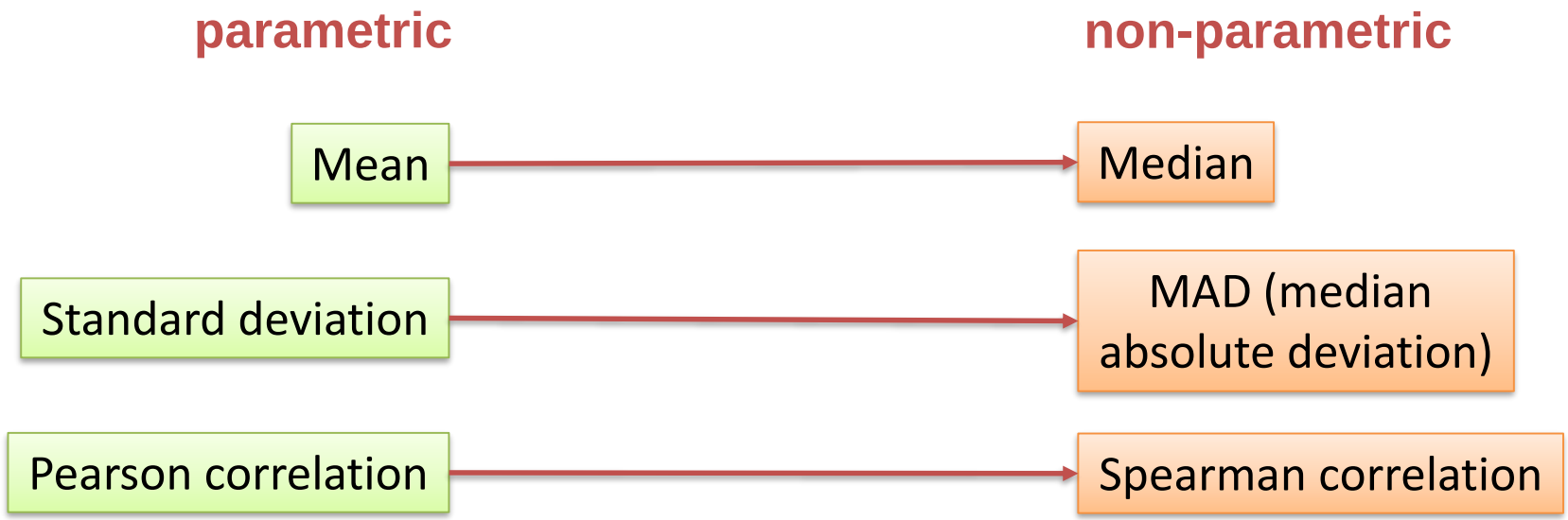
- **Non-parametrical correlation**
  - Kendal`s tau
  - Spearman rank
- **Non-parametical unpaired test**
  - Mann-Whitney-Wilcoxon U-test
- **Non-parametrical paired test**
  - Wilcoxon Signed Rank Test

MEASURES

Non-parametric Measures

Non-parametric measures

A statistical measure which does **not depend on particular data distribution**.
Non-parametric statistics is usually performed on **ranks**.
Non-parametric procedures are more robust to outliers, but less powerful then parametric ones.



MEASURES

Nonparametric Measures of Correlation

Kendal Correlation, τ (Kendall tau rank correlation)

a non-parametric measure of rank correlation: that is, the similarity of the orderings of the data when ranked by each of the quantities.

All combination of data pairs (x_i, y_i) , (x_j, y_j) are checked.

2 pairs are concordant if:

$$(x_i - x_j)(y_i - y_j) > 0$$

2 pairs are discordant if:

$$(x_i - x_j)(y_i - y_j) < 0$$

In case of = 0 pair is not considered.

Let number of corresponding pairs be $n_{concordant}$ and $n_{discordant}$

$$\tau = 2\frac{n_{concordant} - n_{discordant}}{n(n-1)}$$

Spearman's Correlation, ρ (Spearman's rank correlation)

a non-parametric measure of statistical dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function.

Data (x_i, y_i) are replaced by their ranks, let's denote them (X_i, Y_i). Then Pearson correlation is measured b/w ranks:

$$\rho_{xy} = \frac{\sum(X_i - m_X)(Y_i - m_Y)}{\sum(X_i - m_X)\sum(Y_i - m_Y)}$$

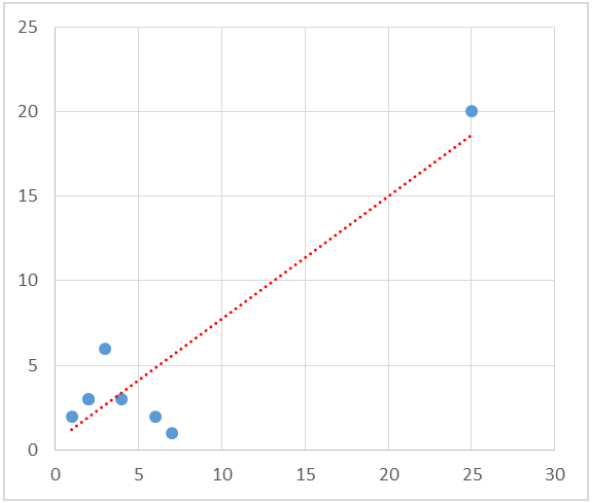
MEASURES

Nonparametric Measures of Correlation

nonparam.xlsx

edu.sablab.net

X	Y
2	3
4	3
3	6
6	2
25	20
2	3
7	1
1	2



Rank(X)	Rank(Y)
2.5	5
5	5
4	7
6	2.5
8	8
2.5	5
7	1
1	2.5

In Excel use:

```
RANK.AVG(x,Data,order)
CORREL(ranks1,ranks2)
```

1. Calculate average rank

2. Calculate Pearson correlation of ranks

Pearson r = 0.91

Spearman ρ = 0.11

NON-PARAMETRIC TESTS

Mann-Whitney-Wilcoxon U-test

Non-paired (simple) t-test → Mann-Whitney test

Number of students on psychology and sociology

Psy	Soc
80	90
95	30
65	65
75	60
60	55
80	70
85	70
90	35
75	75
40	30

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

t-Test: 0.06

In Excel use for rank:
`RANK.AVG(x,TAB,1)`

1. Create ranks for joint data and calculate sum for ranks for each dataset (R_1,R_2)
2. Calculate U1 and U2

$$U_i = n_1 n_2 + \frac{n_i(n_i+1)}{2} - R_i$$

3. Take U = minimum (U1,U2)
4. Calculate z-statistics

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

5. Get p-value from z-stat by normal distribution

In Excel:
`2* NORM.S.DIST(z,TRUE)`

NON-PARAMETRIC TESTS

Mann-Whitney-Wilcoxon U-test

Non-paired (simple) t-test → Mann-Whitney test

Psy	Soc	rank				Psy	Soc
80	90	15.5	18.5		Medians	77.5	62.5
95	30	20	1.5		SumRank	131.5	78.5 ← R1, R2
65	65	8.5	8.5		n=	10	10
75	60	13	6.5		U1, U2	23.5	76.5
60	55	6.5	5		U=	23.5	
80	70	15.5	10.5		z=	-2.00321	
85	70	17	10.5		p-val=	0.045155	
90	35	18.5	3				
75	75	13	13				
40	30	4	1.5				

$$U_i = n_1 n_2 + \frac{n_i(n_i+1)}{2} - R_i$$

$$z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

NON-PARAMETRIC TESTS

Wilcoxon Signed Rank Test

Paired t-test → Wilcoxon signed rank test

before	after
12	412
42	312
31	63
462	632
1	0
25	20
63	124
754	5356
12	83
34	1245

H_0: $\mu_1 = \mu_2$

H_a: $\mu_1 \neq \mu_2$

paired t-Test p-value= **0.165**

Non significant… ???

Procedure for Wilcoxon (Siegel method):

1. Make difference between columns
2. Calculate rank for differences
3. Sum up ranks for positive and negative differences: s+ and s-. Now T = min(s+, s-)
4. Calculate z-statistics and use standard z-test:

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

In Excel use for rank:

`RANK.AVG(diff,DIFF,1)`

NON-PARAMETRIC TESTS

Wilcoxon Signed Rank Test

Paired t-test → Wilcoxon signed rank test

before	after	diff	rank
12	412	400	8
42	312	270	7
31	63	32	3
462	632	170	6
1	0	-1	2
25	20	-5	1
63	124	61	4
754	5356	4602	10
12	83	71	5
34	1245	1211	9

s+ = SUMIF(diff;">0";rank)
s- = SUMIF(diff;"<0";rank)

s+=	52
s- =	3
T =	3
z =	-2.49727
pval(2t)=	0.012515

$$z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

This is an approximate method. To increase power, use W- statistics and special tools, not Excel

QUESTIONS ?

Thank you for your attention