

# Measures of agreement

---

سید حمید شریف نیا

هیات علمی دانشگاه علوم پزشکی مازندران

# Agreement

Agreement between measurements refers to the degree of concordance between two (or more) sets of measurements. Statistical methods to test agreement are used to assess inter-rater variability or to decide whether one technique for measuring a variable can substitute another.

**Agreement  
Concordance  
Reproducibility**

# Methods Used to Measure Agreement

The statistical methods used to assess agreement vary depending on the type of variable being studied and the number of observers between whom agreement is sought to be assessed.

Type of variable	Number of observers between whom agreement is to be assessed	Method used for assessing agreement
Categorical (nominal)	2	Cohen's kappa
	>2	Fleiss' kappa
Categorical (ordinal)	2	Weighted kappa
	>2	Fleiss' kappa
Continuous	Two or more observers or techniques	Intra-class coefficient
		Bland-Altman plot with limits of agreement

# Bland–Altman plots

When two instruments or techniques are used to measure the same variable on a continuous scale, the Bland–Altman plots can be used to estimate agreement. This plot is a scatter plot of the difference between the two measurements (Y-axis) against the average of the two measurements (X-axis). Thus, it provides a graphical display of bias (mean difference between the two observers or techniques) with 95% limits of agreement.

**Limits of agreement = mean observed difference  $\pm$  1.96  $\times$  standard deviation of observed differences**

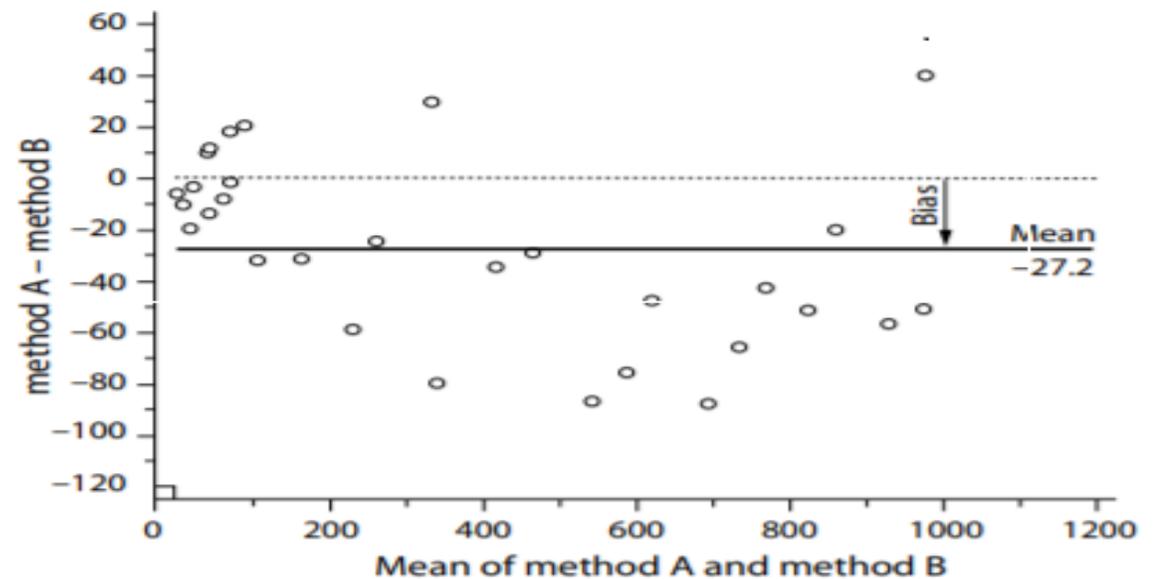
Method A (units)	Method B (units)	Mean (A+B)/2 (units)	(A - B) (units)	(A - B)/Mean (%)	Method A (units)	Method B (units)	Mean (A+B)/2 (units)	(A - B) (units)	(A - B)/Mean (%)
1.0	8.0	4.5	-7.0	-155.6%	1.0	8.0	4.5	-7.0	-155.6%
5.0	16.0	10.5	-11.0	-104.8%	5.0	16.0	10.5	-11.0	-104.8%
10.0	30.0	20.0	-20.0	-100.0%	10.0	30.0	20.0	-20.0	-100.0%
20.0	24.0	22.0	-4.0	-18.2%	20.0	24.0	22.0	-4.0	-18.2%
50.0	39.0	44.5	11.0	24.7%	50.0	39.0	44.5	11.0	24.7%
40.0	54.0	47.0	-14.0	-29.8%	40.0	54.0	47.0	-14.0	-29.8%
50.0	40.0	45.0	10.0	22.2%	50.0	40.0	45.0	10.0	22.2%
60.0	68.0	64.0	-8.0	-12.5%	60.0	68.0	64.0	-8.0	-12.5%
70.0	72.0	71.0	-2.0	-2.8%	70.0	72.0	71.0	-2.0	-2.8%
80.0	62.0	71.0	18.0	25.4%	80.0	62.0	71.0	18.0	25.4%
90.0	122.0	106.0	-32.0	-30.2%	90.0	122.0	106.0	-32.0	-30.2%
100.0	80.0	90.0	20.0	22.2%	100.0	80.0	90.0	20.0	22.2%
150.0	181.0	165.5	-31.0	-18.7%	150.0	181.0	165.5	-31.0	-18.7%
200.0	259.0	229.5	-59.0	-25.7%	200.0	259.0	229.5	-59.0	-25.7%
250.0	275.0	262.5	-25.0	-9.5%	250.0	275.0	262.5	-25.0	-9.5%
300.0	380.0	340.0	-80.0	-23.5%	300.0	380.0	340.0	-80.0	-23.5%
350.0	320.0	335.0	30.0	9.0%	350.0	320.0	335.0	30.0	9.0%
400.0	434.0	417.0	-34.0	-8.2%	400.0	434.0	417.0	-34.0	-8.2%
450.0	479.0	464.5	-29.0	-6.2%	450.0	479.0	464.5	-29.0	-6.2%
500.0	587.0	543.5	-87.0	-16.0%	500.0	587.0	543.5	-87.0	-16.0%
550.0	626.0	588.0	-76.0	-12.9%	550.0	626.0	588.0	-76.0	-12.9%
600.0	648.0	624.0	-48.0	-7.7%	600.0	648.0	624.0	-48.0	-7.7%
650.0	738.0	694.0	-88.0	-12.7%	650.0	738.0	694.0	-88.0	-12.7%
700.0	766.0	733.0	-66.0	-9.0%	700.0	766.0	733.0	-66.0	-9.0%
750.0	793.0	771.5	-43.0	-5.6%	750.0	793.0	771.5	-43.0	-5.6%
800.0	851.0	825.5	-51.0	-6.2%	800.0	851.0	825.5	-51.0	-6.2%
850.0	871.0	860.5	-21.0	-2.4%	850.0	871.0	860.5	-21.0	-2.4%
900.0	957.0	928.5	-57.0	-6.1%	900.0	957.0	928.5	-57.0	-6.1%
950.0	1001.0	975.5	-51.0	-5.2%	950.0	1001.0	975.5	-51.0	-5.2%
1000.0	960.0	980.0	40.0	4.1%	1000.0	960.0	980.0	40.0	4.1%
mean ( $\bar{d}$ )								-27.17	-17.40%
standard deviation (s)								34.81	-12.64%

---

Limits of agreement = mean observed difference  $\pm 1.96 \times$  standard deviation of observed differences

$$d - 1.96s = -27.2 - (1.96 \times 34.8) = -95.4$$

$$d + 1.96s = -27.2 + (1.96 \times 34.8) = 41.1$$



# فرضیات

الف) استقلال نمرات

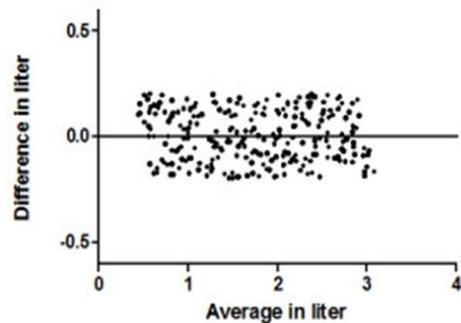
ب) کفایت حجم نمونه

ج) توزیع طبیعی داده‌ها

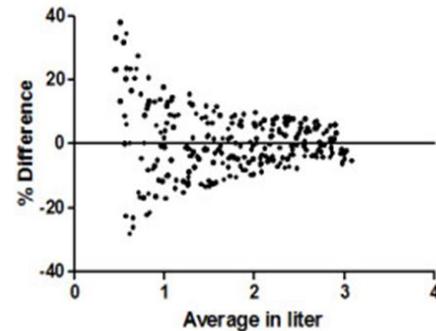
د) همسانی پراکندگی (Homoscedasticity)

## Theoretical types of scatter

Example Homoscedastic scatter



Example of heteroscedastic scatter



# Example

محققى در يك مطالعه تصميم دارد به بررسى ميزان توافق بين اندازه گيرى دما بدن به كمك يك ترمومتر پيشانى به عنوان يك وسيله تازه در بيمارستان در مقايسه با ترمومتر پرده تمپان به عنوان معيار طلايى و استاندارد در جمعيت بزرگسال پردازد. پس از تخمين كفايت حجم نمونه و جمع آورى داده، به كمك نرم افزار MedCalc آناليز ادامه مى يابد...



Arial 10

- Data
- Notes
- Variables
- Filters
- Saved tests and graphs

negative200		A	B	C	D
A1		forehead_C	L.Ear.C		
1	36.8	36.80	36.39		
2		36.10	35.78		
3		36.40	36.00		
4		36.50	36.00		
5		36.10	35.61		
6		37.20	36.56		
7		36.90	36.17		
8		36.70	36.28		
9		37.10	35.89		
10		36.40	35.78		
11		36.80	36.17		
12		36.80	35.56		
13		36.20	36.00		
14		36.20	36.00		
15		36.00	35.61		
16		36.00	35.61		
17		37.10	36.22		
18		36.00	36.00		
19		36.30	35.78		
20		36.60	35.94		
21		36.90	36.28		
22		36.90	36.11		
23		36.30	35.89		
24		36.70	36.06		
25		36.50	35.89		



Arial 10

- Data
- Notes
- Variables
- Filters
- Saved tests and graphs

- Summary statistics...
- Outlier detection...
- Distribution plot >
- Correlation >
- Regression >
- T-tests >
- Rank sum tests >
- Variance ratio test (F-test)...
- ANOVA >
- Crosstabs >
- Survival analysis >
- Meta-analysis >
- Serial measurements... >
- Reference intervals >
- Method comparison & evaluation >**
- Agreement & responsiveness >
- ROC curves >
- Create tables >

- Bland-Altman plot...
- Bland-Altman plot with multiple measurements per subject...
- Comparison of multiple methods...
- Mountain plot...
- Deming regression...
- Passing & Bablok regression...
- Coefficient of variation from duplicate measurements...

Bland-Altman plot



**First method:**

L.Ear.C

**Second method:**

forehead\_C

**Filter:**

|

**Graph**

Plot against (X-axis): Mean of both methods

- Plot differences
- Plot differences as %
- Plot ratios

**Options**

Maximum allowed difference between methods:

- Draw line of equality (difference=0)
- 95% CI of mean difference
- 95% CI of limits of agreement
- Draw regression line of differences
  - 95% Confidence Interval

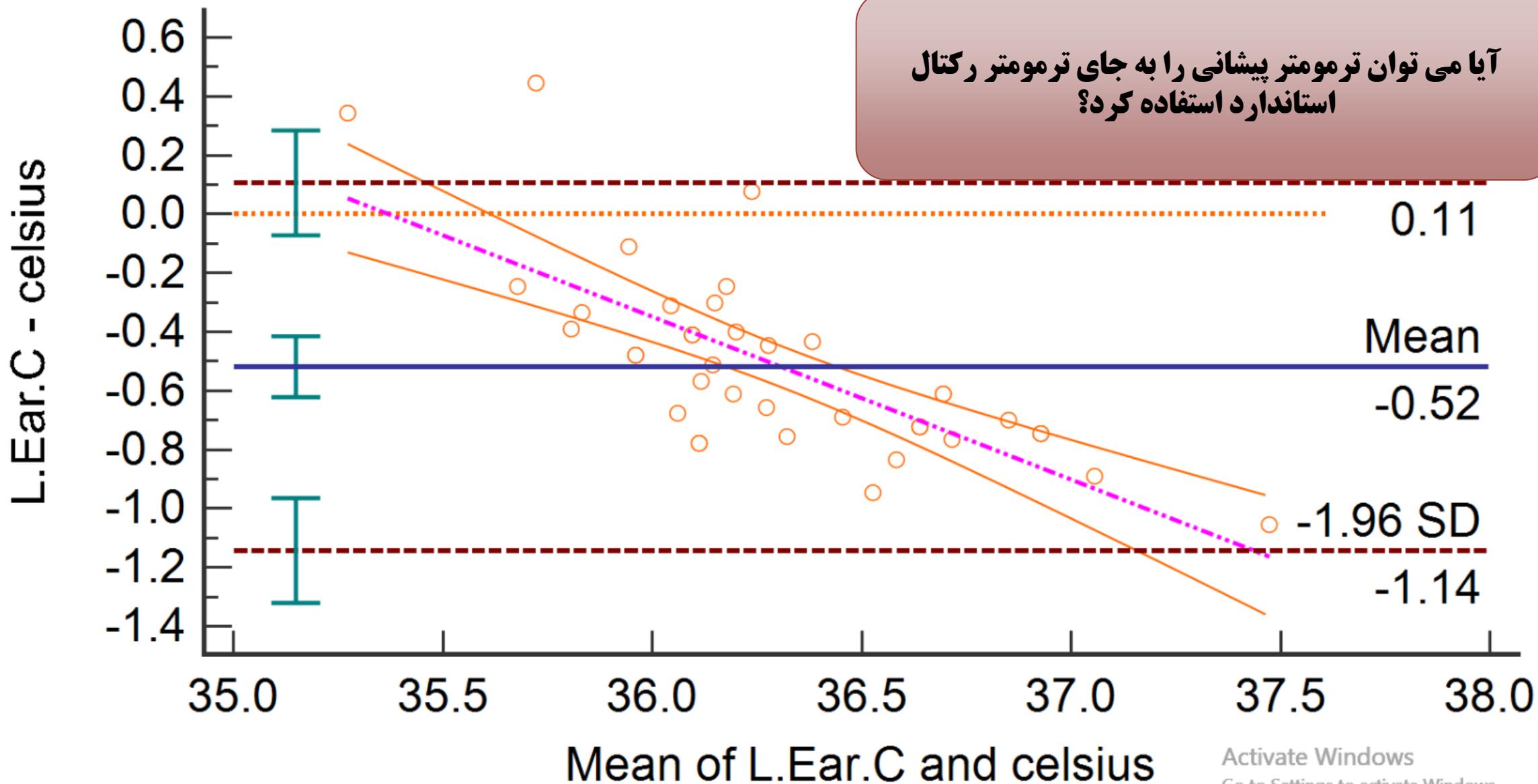
Subgroups...



OK

Cancel

آیا می توان ترمومتر پیشانی را به جای ترمومتر رکتال  
استاندارد استفاده کرد؟



# قابلیت تکرارپذیری (Repeatability) قابلیت تجدیدپذیری (Reproducibility)

در کارهای آزمایشگاهی برای اینکه اطمینان کافی از دقت (Precision) و صحت (Accuracy) مناسب بودن عواملی که می‌توانند به‌طور تصادفی بر پراکندگی نتایج اندازه‌گیری‌های در طول زمان و همچنین اختلاف قابل قبول نتایج حاصل از استفاده از روش‌ها، تجهیزات، نفرات، مکان‌ها و همچنین زمان‌های کاری مختلف تاثیرگذار باشند نیاز به کنترل تکرارپذیری ( $r$ ) و تجدیدپذیری (R) است.

## دقت چیست؟

اما دقت اندازه‌گیری به نزدیک بودن نتایج اندازه‌گیری‌ها به یکدیگر تعریف می‌شود به‌طوری‌که هرچه اندازه‌گیری‌های مختلف به یکدیگر نزدیک‌تر باشند دقت بالاتر است. در حقیقت دقت توصیف خطای تصادفی است.

## صحت چیست؟

صحت یک اندازه‌گیری به معنای نزدیک بودن نتایج به مقدار واقعی است. اگرچه مقدار واقعی نیز میزانی از یک متغیر است که بدون هرگونه خطا و عدم قطعیتی وجود دارد اما ما می‌دانیم که اندازه‌گیری عاری از خطا نخواهد بود و خطای مخصوص به خود را دارد. زمانی یک وسیله اندازه‌گیری صحت دارد که میانگین اندازه‌گیری‌های آن در زمان‌های متفاوت به مقدار واقعی نزدیک باشد. هرچه این اختلاف کمتر باشد صحت اندازه‌گیری بیشتر است. به عبارت دیگر صحت یک اندازه‌گیری توصیف خطای سیستماتیک است.

# خطا

**خطای تصادفی:** بدون الگوی خاص، غیر قابل پیش بینی و معمولا غیر قابل اجتناب

**خطای سیستماتیک:** خطایی ثابت و تکرارپذیر، قابل پیش بینی، اجتناب پذیر

**Random Error.....Precision.....Reliability**

**Systematic Error.....Accuracy.....Validity**

# تکرارپذیری

**تکرارپذیری** مقدار تخمینی است که پیش‌بینی می‌شود با یک احتمال مثلاً ۹۵٪، قدرمطلق اختلاف بین دو نتیجه آزمون انفرادی حاصل از اندازه‌گیری بر روی یک نمونه ثابت و همچنین با روش اندازه‌گیری یکسان در شرایط یکسان (اپراتور یا اندازه‌گیر، دستگاه یکسان، آزمایشگاه یکسان و فاصله زمانی کوتاه بین اندازه‌گیری‌ها) از آن کوچک‌تر باشد.

$$CR = 1.96 \times \sqrt{\left( \frac{\sum \sqrt{d^2 - d^1}}{n} \right)}$$

# تجدید پذیری

تجدیدپذیری بر خلاف تکرارپذیری به تغییر اندازه‌گیری‌های مختلف در شرایط متفاوت (اپراتور یا اندازه‌گیر، دستگاه یکسان، آزمایشگاه یکسان و فاصله زمانی کوتاه بین اندازه‌گیری‌ها) اشاره دارد.

## Bland-Altman plot

Method A L.Ear.C  
Method B forehead\_C  
celsius

**Differences**

Sample size	39
Arithmetic mean	-0.5182
95% CI	-0.6215 to -0.4149
P (H <sub>0</sub> : Mean=0)	<0.0001
Lower limit	-1.1427
95% CI	-1.3208 to -0.9647
Upper limit	0.1063
95% CI	-0.07177 to 0.2843

Regression Equation  $y = 19.5454 + -0.5526 x$

Parameter	Coefficient	SE	t	P	95% CI
Intercept	19.5454	2.9489	6.6281	<0.0001	13.5704 to 25.5204
Slope	-0.5526	0.08121	-6.8043	<0.0001	-0.7172 to -0.3880
Coefficient of Repeatability	1.1882				
95% CI	0.9733 to 1.5256				

# آزمون های دیگر؟

Paired T test

Correlation

Linear regression

Correlation	
Variable Y	forehead_C
Variable X	L.Ear.C
Sample size	39
Correlation coefficient r	0.8678
Significance level	P<0.0001
95% Confidence interval for r	0.7605 to 0.9200

## Paired samples t-test

Mean difference	-0.5182
Standard deviation of differences	0.3186
Standard error of mean difference	0.05102
95% CI of difference	-0.6215 to -0.4149
Test statistic t	-10.157
Degrees of Freedom (DF)	38
Two-tailed probability	P < 0.0001

## Regression Equation

$$y = 17.3828 + 0.5105 x$$

Parameter	Coefficient	Std. Error	95% CI	t	P
Intercept	17.3828	1.7575	13.8218 to 20.9438	9.8907	<0.0001
Slope	0.5105	0.04806	0.4131 to 0.6078	10.6219	<0.0001

## Analysis of Variance

# Passing and Bablok Regression

Passing and Bablok regression analysis is a statistical procedure that allows valuable estimation of analytical methods agreement and possible systematic bias between them. It is robust, non-parametric, non sensitive to distribution of errors and data outliers.

$$y = a + bx$$

# پیش فرض های رگرسیون پاسینگ و بابلوک

- ۱- متغیرهای  $X$  و  $Y$  همبستگی مثبت و بالایی با یکدیگر باید داشته باشند.
- ۲- ارتباط بین متغیرهای  $X$  و  $Y$  باید خطی باشد.
- ۳- هیچ فرض خاصی بین متغیرهای  $X$  و  $Y$  مانند توزیع واریانس ها در نظر گرفته نمی شود.

# عرض از مبدا (Intercept) و شیب خط (Slope)

**شیب خط:** نشان می‌دهد که میزان حساسیت متغیر وابسته به متغیر مستقل چقدر است. به این معنی که با افزایش یک واحد به مقدار متغیر مستقل چه میزان متغیر وابسته تغییر خواهد کرد..... **(خطای سیستماتیک)**

**عرض از مبدا:** بیانگر مقداری از متغیر وابسته است که به ازاء مقدار متغیر مستقل برابر با صفر محاسبه می‌شود. به عبارت دیگر می‌توان مقدار ثابت یا عرض از مبدا را مقدار متوسط متغیر وابسته به ازاء حذف متغیر مستقل در نظر گرفت..... **(خطای نسبی)**

# فرض مورد تست رگرسیون پاسینگ و بابلک

فواصل اطمینان ۹۵٪ نشان می‌دهد آیا ارزش‌های تخمین زده شده از صفر و یک به ترتیب برای عرض از مبدا و شیب خط رگرسیونی متفاوت است یا نه؟ که اگر متفاوت باشد فقط ناشی از شانس است. به عبارت دیگر اگر فاصله اطمینان ۹۵٪ برای عرض از مبدا شامل عدد صفر شود می‌توان نتیجه گرفت که هیچ اختلاف معنی‌داری بین ارزش عرض از مبدا بدست آمده و ارزش صفر وجود ندارد و بنابراین اختلاف ثابتی بین دو روش اندازه‌گیری وجود ندارد.

همچنین اگر فاصله اطمینان ۹۵٪ شیب خط شامل عدد یک شود می‌توان نتیجه گرفت که هیچ اختلاف معنی‌داری بین ارزش‌های بدست آمده از شیب خط و ارزش عددی یک وجود ندارد و بنابراین هیچ اختلاف نسبتی بین دو روش اندازه‌گیری وجود ندارد و می‌توان روش دوم را به جای روش اول (روش استاندارد یا روش طلایی) استفاده کرد ( $x=y$ ).

MedCalc - negative.sav

File Edit View Format Tools **Statistics** Graphs Tests Sample size Window Help



Arial 10

- Data
- Notes
- Variables
- Filters
- Saved tests and graphs

- Summary statistics...
- Outlier detection...
- Distribution plot >
- Correlation >
- Regression >
- T-tests >
- Rank sum tests >
- Variance ratio test (F-test)...
- ANOVA >
- Crosstabs >
- Survival analysis >
- Meta-analysis >
- Serial measurements...
- Reference intervals >
- Method comparison & evaluation >**
- Agreement & responsiveness >
- ROC curves >
- Create tables >



	C	D	E	F	G
C					
39					
78					
00					
00					
31					
56					
17					
28					

14	36.20	36.
15	36.00	35.
16	36.00	35.
17	37.10	36.

- Bland-Altman plot...
- Bland-Altman plot with multiple measurements per subject...
- Comparison of multiple methods...
- Mountain plot...
- Deming regression...
- Passing & Bablok regression...**
- Coefficient of variation from duplicate measurements...

## Regression Equation

$$y = 15.555556 + 0.555556 x$$

### Systematic differences

Intercept A	<u>15.5556</u>
95% CI	15.5556 to 18.1007

### Proportional differences

Slope B	<u>0.5556</u>
95% CI	0.4861 to 0.5556

### Random differences

Residual Standard Deviation (RSD)	0.1772
± 1.96 RSD Interval	-0.3472 to 0.3472

### Linear model validity

Cusum test for linearity	No significant deviation from linearity (P=0.23)
--------------------------	---

# cumulative sum linearity test (CUSUM)

آیا باقیمانده‌ها به‌طور تصادفی در بالا و پایین خط رگرسیون پراکنده شده‌اند؟

Variable Y	L.Ear.C
Variable X	forehead_C celsius

پیشنهاد...

Sample size	200
Kendall's Tau	0.554
Significance level	P<0.0001
95% Confidence Interval for Tau <sup>a</sup>	0.480 to 0.621

<sup>a</sup> BC<sub>a</sub> bootstrap confidence interval (500 iterations; random number seed: 978).

# حجم نمونه برای استفاده از رگرسیون پاسینگ و بابلک

---

?

# Lin's concordance correlation coefficient (CCC)

The concordance correlation coefficient measures the agreement between two variables, e.g., to evaluate reproducibility or for inter-rater reliability.

ضریب همبستگی لین ترکیبی از صحت و دقت ( $pc = pC_b$ ) را به صورت همزمان ارزیابی و مشخص می‌کند

MedCalc - negative200.sav - [negative200]

File Edit View Format Tools **Statistics** Graphs Tests Sample size



Arial 10

- Data
- Notes
- Variables
- Filters
- Saved tests and graphs

- Summary statistics...
- Outlier detection...
- Distribution plot
- Correlation
- Regression
- T-tests
- Rank sum tests
- Variance ratio test (F-test)...
- ANOVA
- Crosstabs
- Survival analysis
- Meta-analysis
- Serial measurements...
- Reference intervals
- Method comparison & evaluation
- Agreement & responsiveness**
- ROC curves
- Create tables

Variable Y	forehead_C celsius
Variable X	L.Ear.C

Sample size	200
Concordance correlation coefficient	0.2259
95% Confidence interval	0.1798 to 0.2710
Pearson $\rho$ (precision)	0.7083
Bias correction factor $C_b$ (accuracy)	0.3189

Concordance correlation coefficient

Variable Y: forehead\_C

Variable X: L.Ear.C

Filter:

Options

- Logarithmic transformation Y
- Logarithmic transformation X

OK Cancel

< 0.90: poor  
0.90 to 0.95: moderate  
0.95 to 0.99: substantial  
> 0.99 almost perfect

16	36.00	35.61
17	37.10	36.22
18	36.00	36.00

- Intraclass correlation coefficient...
- Concordance correlation coefficient...**
- Inter-rater agreement (kappa)...
- Cronbach's alpha...
- Responsiveness...

---

**از توجه شما سپاسگزارم**