

Reliability

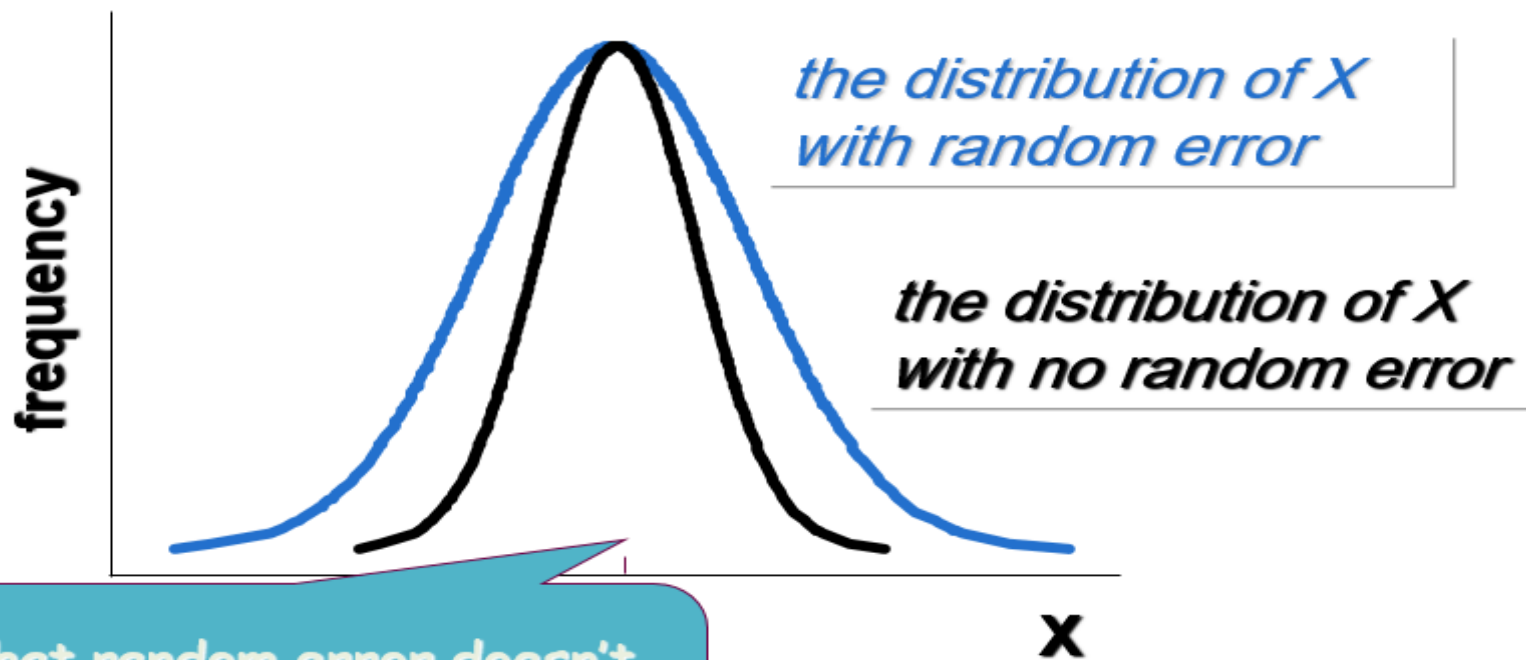
Hamid Sharif Nia

BScN, MScN, PhD in Nursing Education

Associate Professor of Mazandaran University of Medical Science

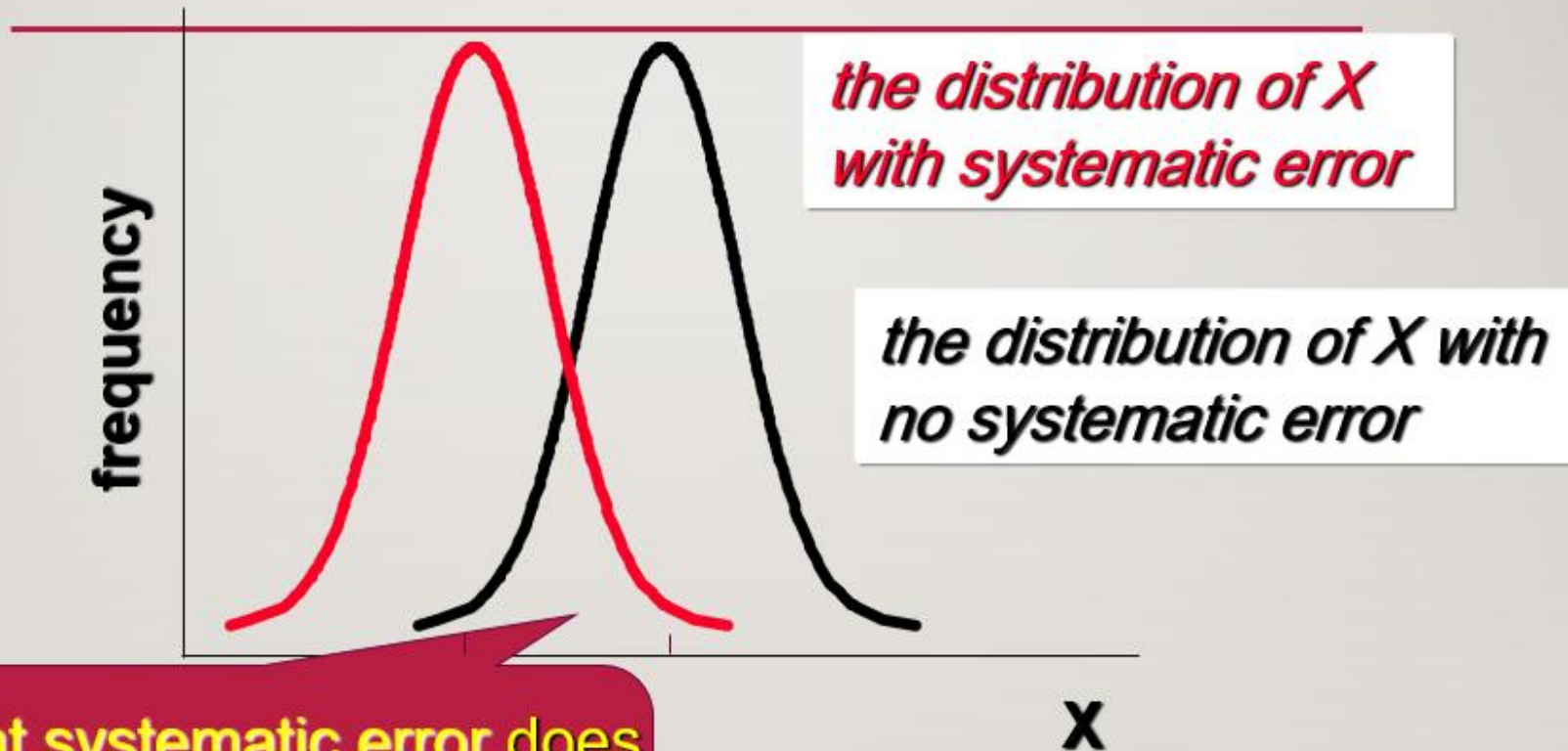
Accuracy and Precision

Random Error



Notice that random error doesn't affect the average, only the **variability** around the average

SYSTEMATIC ERROR



Notice that systematic error does affect the average -- we call this a bias

Random error – measurement not reliable

**Systematic error – measurement biased
(not valid)**

Reliability

It refers to the **consistency** and **reproducibility** of data produced by a given method, technique, or experiment. The form of assessment is said to be reliable if it repeatedly produces stable and similar results under consistent conditions. Consistency is partly ensured if the attribute being measured is stable and does not change suddenly.

However, errors may be introduced by factors such as the **physical and mental state of the examinee**, **inadequate attention**, **distractedness**, **response to visual and sensory stimuli in the environment**, etc. When estimating the reliability of a measure, the examiner must be able to demarcate and differentiate between the errors produced as a result of inefficient measurement and the actual variability of the true score. A true score is that subset of measured data that would recur consistently across various instances of testing in the absence of errors. Hence, the general score produced by a test would be a composite of the true score and the errors of measurement.

واژه های معادل

- قابلیت اعتماد Dependability
- سازگاری Consistency
- ثبات Stability
- قابلیت پیش بینی Predictability
- دقت و صحت Accuracy

الف) ثبات

۱ - پایایی باز آزمون Test-retest

۲ - پایایی به شکل موازی Equivalent form

ب) سازگاری

۳ - کودر ریچاردسون Richardson

۴ - دو نیم کردن Split – halves

۵ - آلفای کرونباخ Cronbach-alpha، تتا و امگا مک دونالد Theta (θ) and Omega (ω)

نکته: پایایی بین مشاهده کنندگان

Inter-rater Reliability

it measures the consistency of the scoring conducted by the evaluators of the test. It is important since not all individuals will perceive and interpret the answers in the same way, hence the deemed accurateness of the answers will vary according to the person evaluating them. This helps in refining and eliminating any errors that may be introduced by the subjectivity of the evaluator. If a majority of the evaluators judge are in agreement with regards to the answers, the test is accepted as being reliable. But if there is no consensus between the judges, it implies that the test is not reliable and has failed to actually test the desired quality. However, the judging of the test should be carried out without the influence of any personal bias.



6 : Observer2

	Observer1	Observer2	var	var	var	var	var	var	var	var	var	var	var
1	1.00	1.00											
2	1.00	1.00											
3	.00	1.00											
4	.00	1.00											
5	1.00	.00											
6	.00	1.00											
7	1.00	1.00											
8	1.00	1.00											
9	1.00	.00											
10	1.00	1.00											
11	.00	1.00											
12	1.00	1.00											
13	1.00	1.00											
14	.00	1.00											
15	.00	1.00											
16	1.00	.00											
17	1.00	1.00											
18	1.00	1.00											
19	1.00	1.00											
20	.00	1.00											
21	1.00	.00											
22	1.00	1.00											
23	.00	1.00											
24	1.00	1.00											
25	.00	1.00											
26	1.00	.00											
27	.00	1.00											
28	1.00	1.00											

Crosstabs

Row(s):

Column(s):

Layer 1 of 1

Display clustered bar charts
 Suppress tables

Display layer variables in table layers

Crosstabs: Statistics

Chi-square Correlations

Nominal
 Contingency coefficient
 Phi and Cramer's V
 Lambda
 Uncertainty coefficient

Ordinal
 Gamma
 Somers' d
 Kendall's tau-b
 Kendall's tau-c

Nominal by Interval
 Eta

Kappa
 Risk
 McNemar

Cochran's and Mantel-Haenszel statistics
 Test common odds ratio equals:

Crosstabs

Observer1 ^ Observer2 Crosstabulation

Count

		Observer2	Total
		No	Yes
Observer1	No	11	11
	Yes	8	19
Total		19	30

>۷۵
۷۵-۴۰
۴۰.>

Symmetric Measures

		Value	Asymptotic Standardized Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.648	.143	3.557	.000
N of Valid Cases		30			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

چگونه باید نتیجه پایایی کوهن را گزارش کرد؟

کاپای کوهن جهت ارزیابی توافق بین ... و ... بر روی نمونه صورت گرفت. توافق عالی / خوب / ضعیف بین

دو n

ارزیاب وجود داشت؛

$K = \dots (95\% \text{ CI, } \dots \text{ to } \dots), p <$

	Q6	Q7	Q8	Q9
5.00	5.00	5.00	5.00	
5.00	5.00	5.00	5.00	
4.00	5.00	4.00		
5.00	4.00	3.00		
4.00	3.00	2.00		
4.00	3.00	4.00		
1.00	5.00	3.00		
4.00	5.00	4.00		
5.00	5.00	4.00		
5.00	4.00	5.00		
5.00	5.00	4.00		
3.00	4.00	3.00	2.00	
4.00	5.00	4.00	2.00	
5.00	3.00	5.00	2.00	
4.00	5.00	4.00	3.00	
5.00	5.00	4.00	3.00	
4.00	5.00	3.00	3.00	
4.00	4.00	4.00	3.00	
4.00	5.00	4.00	3.00	

Inter-rater agreement (kappa) [?] [X]

Data for observer A:

Data for observer B:

Filter:

Options

Weighted Kappa

Linear weights

Quadratic weights

[?] [OK] [Cancel]

- Reliability Analysis...
- Weighted Kappa**
- Fleiss Kappa
- Multidimensional Unfolding (PREFSCAL)...
- Multidimensional Scaling (PROXSCAL)...
- Multidimensional Scaling (ALSCAL)...

Correlations

Correlations

		DAS1	DAS2
DAS1	Pearson Correlation	1	1.000^{**}
	Sig. (2-tailed)		.000
	N	55	55
DAS2	Pearson Correlation	1.000^{**}	1
	Sig. (2-tailed)	.000	
	N	55	55

******. Correlation is significant at the 0.01 level (2-tailed).

پایایی فرم های موازی

پایایی اشکال موازی متعلق به نسخه های مختلف یک ابزار است که اثر Recall را از بین می برد.

معایب:

تعداد گویه ها باید برابر باشد.

درجه سختی سوال یکسان باشد

از نظر هموزنیستی باید یکسان باشد

میانگین و انحراف معیار باید نزدیک به هم باشد










روش کودر ریچاردسون

آزمون کودر ریچاردسون در آزمون‌هایی که پاسخ آن‌ها بصورت کدهای صفر و یک برای پاسخ‌های غلط و صحیح هست، مورد استفاده قرار می‌گیرد. کودر و ریچاردسون دو فرمول KR20 و KR21 را ارائه کردند.

- Reports ▶
- Descriptive Statistics ▶
- Custom Tables ▶
- Compare Means ▶
- General Linear Model ▶
- Generalized Linear Models ▶
- Mixed Models ▶
- Correlate ▶
- Regression ▶
- Loglinear ▶
- Neural Networks ▶
- Classify ▶
- Dimension Reduction ▶
- Scale ▶**
- Nonparametric Tests ▶
- Forecasting ▶
- Survival ▶
- Multiple Response ▶
- Missing Value Analysis...
- Multiple Imputation ▶
- Complex Samples ▶
- Simulation...
- Quality Control ▶

 CDS.Total




- Items:
-  I have dropped many of i...
 -  My concentration is as go...
 -  I can't be bothered doing...
 -  I get pleasure from life at ...
 -  I am concerned about th...
 -  I may not recover comple...
 -  My sleep is restless and ...
 -  I am not the person I use...
 -  I wake up in the early hou...

Statistics...

Reliability Statistics

Cronbach's Alpha	Part 1	Value	.941
		N of Items	11
	Part 2	Value	.905
		N of Items	10
	Total N of Items		21
Correlation Between Forms			.855
Spearman-Brown Coefficient	Equal Length		.922
	Unequal Length		.922
Guttman Split-Half Coefficient			.913

Reliability Statistics



Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.958	.958	21

چهار نکته که موقع محاسبه ضریب آلفا باید توجه داشت:

- ۱- مقیاس اندازه گیری انعکاسی باشد
- ۲- کدگذاری گویه های یک جهت باشند
- ۳- ضریب آلفا تک بعدی است. برای هر بعد جداگانه حساب شود
- ۴- در صورت مقیاس های متفاوت اندازه گیری، از ضریب آلفای استاندارد استفاده

کنید

روش‌های اصلاح پایایی

برای اصلاح پایایی، باید سعی نمود قبل از حذف سوالات، روش‌های دیگر آزمایش شود، شاید لازم باشد مفاهیم سوالات و گویه‌ها به صورت شفاف‌تر و دقیق‌تر بیان شود. گاهی اوقات، مفاهیم و سوالات تاریخی در پرسشنامه وجود دارد که بهتر است با افزایش چند سوال در کنار آنها، مفهوم شفاف‌تر شود. در این صورت همبستگی بین سوالات افزایش یافته و پایایی ارتقاء خواهد یافت. راهکار دیگر، افزودن دستورالعمل پاسخ-گویی جهت راهنمایی پاسخگویان می‌باشد. در صورتی که این موارد قبل از توزیع پرسشنامه در حد وسیع رعایت شده باشد، می‌توان راه کار آخر را به شرط زیر به کار بست، با حذف سوالاتی که بیشترین ارتقاء را در ضریب پایایی خواهند داشت، پایایی را افزایش می‌دهیم. البته حذف سوالات، باید به صورت یک به یک و بر اساس بزرگترین افزایش در ضریب α باشد. فرایند حذف تا جایی ادامه پیدا می‌کند که ضریب آلفا به آستانه مد نظر برسد.

نقطه برش های متداول برای تفسیر آلفا

بلند و آلمن (۱۹۹۷) معتقد هستند که مقیاس هایی که برای ارزیابی های بالینی به کار می روند بایستی در مقایسه

با مقیاس های مورد استفاده در پژوهش های غیر بالینی دارای پایایی بیشتری باشند. بر این اساس آنها ضریب

آلفای بین ۰/۷ و ۰/۸ را برای مقاصد پژوهشی پیشنهاد می کنند ولی در مقاصد بالینی مقدار آلفای مقیاس های

مورد بررسی باید بزرگتر از ۰/۹ باشد.



Descriptives for

- I**tem
- S**cale
- S**cale if item deleted

Inter-Item

- C**orrelations
- C**ovariances

Summaries

- M**eans
- V**ariances
- C**ovariances
- C**orrelations

ANOVA Table

- N**one
- F** test
- F**riedman chi-square
- C**ochran chi-square

Interrater Agreement: Fleiss' Kappa

- Display agreement on individual categories
- Ignore string cases
- String category labels are displayed in uppercase

Asymptotic significance level (%):

Missing

- E**xclude **b**oth user-missing and system missing values
- U**ser-missing values are treated as valid

 Hotelling's T-square **T**ukey's test of additivity **I**ntraclass correlation coefficientModel: Type: Confidence interval: %Test value:

چند نکته مهم در خصوص ضریب آلفای کرونباخ

- آلفای کرونباخ بر اساس همبستگی و واریانس کار می‌کند: هر قدر همبستگی مثبت بین سؤالات بیشتر شود، میزان آلفای کرونباخ بیشتر خواهد شد و بالعکس هر قدر واریانس میانگین سؤالات بیشتر شود آلفای کرونباخ کاهش پیدا خواهد کرد.
- از آنجایی که در فرمول محاسبه آلفای کرونباخ ضریب تعداد سؤالات موجود است، افزایش تعداد سؤالات تأثیر مثبت یا منفی (بسته به نوع همبستگی بین سؤالات) بر میزان آلفای کرونباخ خواهد گذاشت.
- افزایش حجم نمونه نیز باعث کاهش واریانس میانگین سؤالات و در نتیجه باعث افزایش آلفای کرونباخ می‌شود. بنابراین در مواقعی که حجم نمونه شما کم است، رسیدن به آلفای کرونباخ خوب کار آسانی نیست
- ضریب آلفای کرونباخ از ضرایب محافظه کارانه و سخت گیرانه شناخته می‌شود. زیرا ضریب پایایی محاسبه شده از سایر شیوه‌ها (نظیر دونیمه کردن) مقادیر بیشتری از آلفا را بدست می‌دهد.

ضریب تتا و ضریب امگا مک دونالد

$$\theta = \frac{N}{N-1} \left(1 - \frac{1}{\lambda} \right)$$

$$\Omega = 1 - \frac{N - \sum h_i^2}{N + 2r} = 1$$

ت گویه‌ها، I= مجموع بارهای عاملی گویه‌ها در هر عامل

Myocardial Infarction						
Factors	Factors name	Items	Loading	h^2	% of Variance	Eigenvalues
1	death anxiety and despair	Q13. The possibility of sudden death worries me	.938	.783	45.027	6.768
		Q14. There is only misery in the future for me	.831	.749		
		Q25. I feel frustrated	.819	.721		
		Q18. Things which I regret about my life are bothering me	.804	.496		
		Q5. I am concerned about the uncertainty of my health	.793	.665		
		Q8. I am not the person I used to be	.741	.644		
		Q21. I become tearful more easily than before...	.708	.437		
		Q26. I am concerned about my capacity for sexual activity	.625	.388		
		Q17. My problems are not yet over	.614	.406		
		Q6. I may not recover completely	.583	.591		
2	life satisfaction	Q12. I feel in good spirits	.790	.474	11.765	3.102
		Q4. I get pleasure from life at present	.789	.620		
		Q15. My mind is as fast and alert as always	.757	.533		
		Q2. My concentration is as good as it ever was	.702	.699		
		Q19. I gain just as much pleasure from my leisure activities as I used to	.675	.435		
		Q20. My memory is as good as it always was	.578	.446		

Abbreviation: h^2 : Communalities

Table 3: Construct Validity and Reliability Results (Fornell Larcker Criterion Table)

	α	θ	Ω	CR	AVE	MSV	ASV	Factor 1	Factor 2
Factor 1	.928	.962	.953	.900	.568	.223	.223	.753	-
Factor 2	.865	.843	.879	.829	.620	.223	.223	.472	.787

Abbreviations: α : Cronbach's alpha coefficients, θ : Theta Coefficient, Ω : McDonald Omega Coefficient, CR: Construct reliability, AVE: Average variance extracted, MSV: maximum shared squared variance, ASV: Average shared squared variance



Intraclass Correlation Coefficient

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.920 ^a	.786	.971	23.876	15	15	.000
Average Measures	.958 ^c	.880	.985	23.876	15	15	.000

Missing

Exclude both user-missing and system missing values

User-missing values are treated as valid

Hotelling's T-square

Tukey's test of additivity

Intraclass correlation coefficient

Model: **Two-Way Mixed**

Type: **Consistency**

Confidence interval: **95** %

Test va **Consistency**

Absolute Agreement

Continue **Cancel** **Help**

I appreciate your concern