



Hamadan University of Medical Sciences

تعیین روایی و پایایی پرسشنامه با نرم افزار SPSS

Assessment of Questionnaire Validity and Reliability Using SPSS

Sanaz Omid

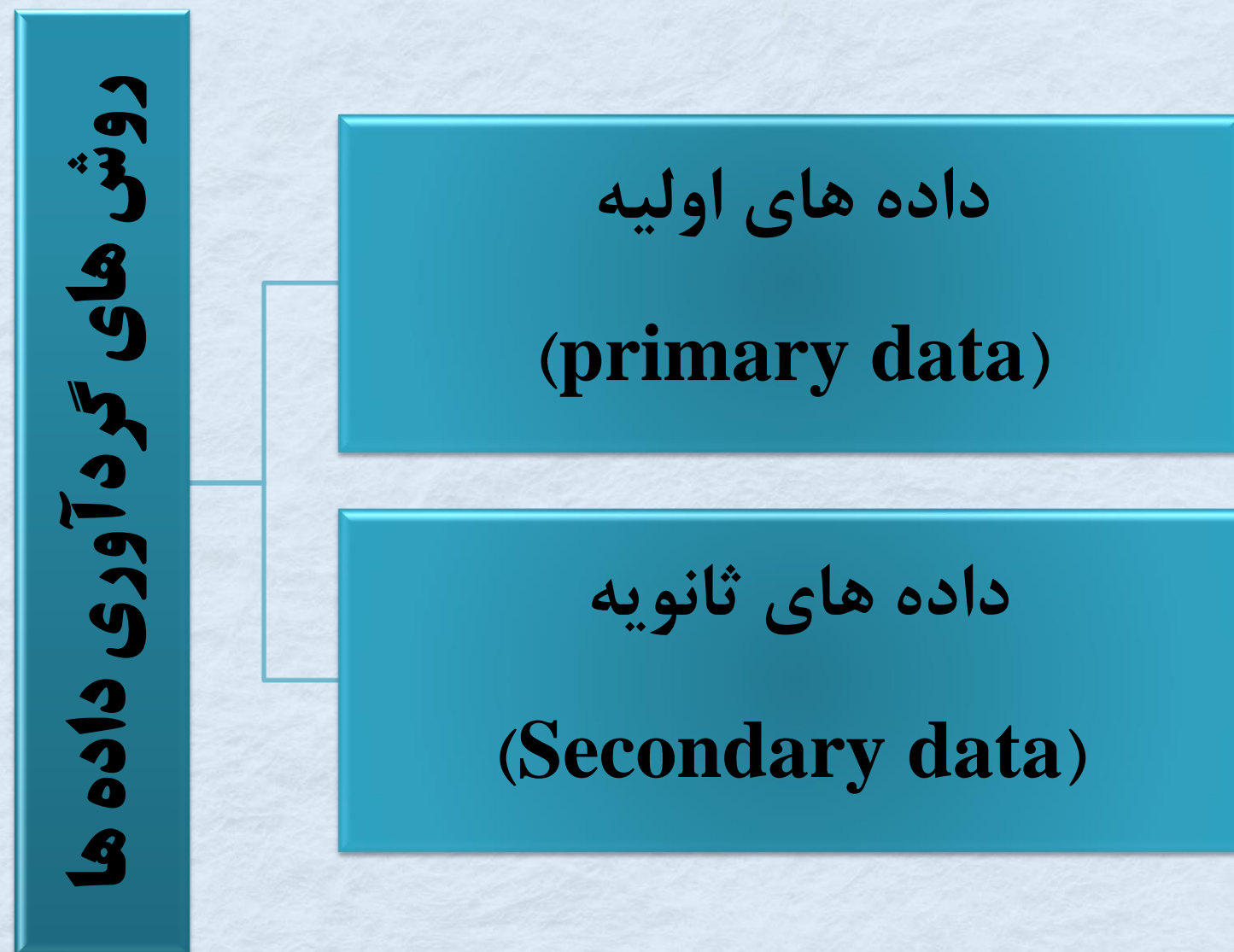
PhD in Epidemiology

Department of Epidemiology, School of Public Health

Sanaz.omidi91@gmail.com

روش‌های جمع‌آوری داده‌ها در پژوهش

□ پیش از آغاز مطالعه، محقق باید نوع داده، منبع داده و روش جمع‌آوری آن‌ها را به‌طور شفاف مشخص کند.



روش‌های جمع‌آوری داده‌ها در پژوهش

داده‌های ثانویه جمع‌آوری: توسط دیگران / سازمان

کم هزینه / زمان

دستری آسان و سریع به اطلاعات

محدودیت در کیفیت یا تناسب با هدف مطالعه داشته باشد

داده‌های بیمارستانی، داده‌های ثبت ملی، مقالات، پایگاه‌های آماری

داده‌های اولیه جمع‌آوری: توسط پژوهشگر

پر هزینه و زمان بر

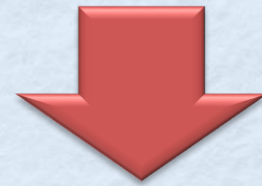
کنترل کامل بر نحوه کیفیت و فرایند جمع‌آوری داده‌ها

طراحی شده مطابق با هدف مطالعه

پرسشنامه، مشاهده، مصاحبه، اندازه‌گیری‌های آزمایشگاهی

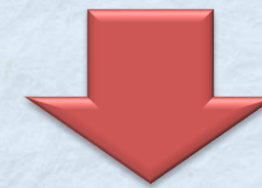
روش‌های جمع‌آوری داده‌ها در پژوهش

روش‌های استخراج داده‌های ثانویه



- ✓ پرونده‌ها و سیستم‌های اطلاعات سلامت
- ✓ پایگاه‌های داده دولتی و آماری
- ✓ داده‌های پژوهشی منتشر شده
- ✓ رجیستری‌ها و سیستم‌های پایش

روش‌های جمع‌آوری داده‌های اولیه



- ✓ مصاحبه
- ✓ پرسشنامه
- ✓ چک لیست
- ✓ مشاهده
- ✓ دلفی
- ✓ بحث گروهی متمرکز

پرسشنامه (Questionnaire)

- "پرسشنامه" یکی از رایج‌ترین ابزارهای جمع‌آوری داده در تحقیقات علمی است که از مجموعه‌ای از سؤال‌های استاندارد تشکیل می‌شود و با هدف استخراج اطلاعات از پاسخ‌دهندگان طراحی می‌گردد.
- ✓ پرسشنامه را می‌توان نوعی مصاحبه‌ی کتبی دانست که در آن تعامل مستقیم حذاقلى وجود دارد.
- ✓ نخستین استفاده رسمی از پرسشنامه به سال ۱۸۳۸ بازمی‌گردد؛ زمانی که انجمن آمار لندن برای اولین بار یک پرسشنامه ساختاریافته را در یک طرح آماری به کار برد.
- ✓ پرسشنامه‌ها ابزاری اقتصادی، سریع و کارآمد برای گردآوری حجم زیادی از اطلاعات از نمونه‌های بزرگ هستند و به‌ویژه در نظرسنجی‌ها، مطالعات جمعیت‌محور و پژوهش‌های اپیدمیولوژیک کاربرد گسترده دارند.

پرسشنامه

تقسیم‌بندی پرسشنامه بر اساس ماهیت تحقیق

- I. پرسشنامه توصیفی: (Descriptive Questionnaire)
- II. پرسشنامه تحلیلی / علت‌یابی: (Analytical / Explanatory Questionnaire)
- III. پرسشنامه ارزیابی نگرش / نظر: (Attitudinal Questionnaire)
- IV. پرسشنامه آزمون یا عملکردی: (Test / Performance Questionnaire)

تقسیم‌بندی پرسشنامه بر اساس نحوه اجرا

- I. حضوری: (Face-to-Face / Paper-based)
- II. تلفنی: (Telephone Questionnaire)
- III. آنلاین / اینترنتی: (Online / Web-based)
- IV. پست الکترونیک یا سنتی: (Email / Postal Questionnaire)

تقسیم‌بندی پرسشنامه بر اساس ماهیت پاسخ‌ها

- I. پرسشنامه با پاسخ‌های بسته (Closed-Ended Questionnaire)
- II. پرسشنامه با پاسخ‌های باز (Open-Ended Questionnaire)
- III. پرسشنامه ترکیبی (Mixed/Open-Closed)



چک لیست و پرسشنامه

| ویژگی | چک لیست (Checklist) | پرسشنامه (Questionnaire) |
|-------------|--|---|
| تعریف | فهرستی از موارد/اقدامات که بررسی یا تأیید می شوند (Objective) | ابزار تحقیقاتی متشکل از سؤالات استاندارد برای جمع آوری اطلاعات از پاسخ دهندگان (Subjective) |
| هدف | اطمینان از انجام یا وجود آیتم های مشخص، ثبت مشاهدات یا رعایت استانداردها | جمع آوری داده های کمی یا کیفی درباره رفتار، نظر یا تجربه افراد |
| ساختار | شامل آیتم های کیفی (بله/خیر)، علامت گذاری یا تأیید انجام شده | شامل سؤال با پاسخ چندگزینه ای، لیکرت یا باز |
| نحوه اجرا | معمولاً توسط ناظر/پژوهشگر تیک زده یا علامت گذاری می شود | پاسخ دهنده به سؤالات پاسخ می دهد |
| کاربرد رایج | کنترل کیفیت، مشاهده، بررسی استانداردها، مطالعات بالینی | نظرسنجی ها، مطالعات اجتماعی، تحقیقات سلامت، اپیدمیولوژی |

مصاحبه (Interview)

- مصاحبه یک روش جمع آوری داده اولیه است، مشابه پرسشنامه، اما شفاهی و تعاملی.
- برخلاف چک لیست که بیشتر برای ثبت مشاهدات و کنترل رعایت آیتمها است، مصاحبه با پاسخ دهنده تعامل مستقیم دارد.

| انواع مصاحبه | توضیحات |
|---|---|
| ساختاریافته (Structured) | مشابه پرسشنامه، سوالات مشخص و ثابت دارند |
| نیمه ساختاریافته (Semi-Structured) | ترکیبی از سوالات بسته و باز |
| غیر ساختاریافته / باز (Unstructured / Open) | جریان گفتگو آزاد است و سوالات مطابق پاسخها تنظیم می شود |

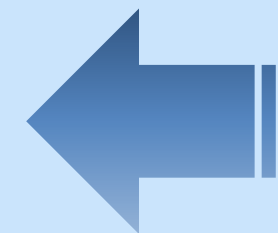
- I. پرسشنامه: داده کتبی از افراد
- II. چک لیست: ثبت یا تأیید آیتمها توسط ناظر
- III. مصاحبه: جمع آوری داده شفاهی و تعاملی

کیفیت اندازه‌گیری

□ کیفیت اندازه‌گیری یعنی دقت و درستی ابزار سنجش؛ اینکه ابزار پژوهش تا چه حد درست، دقیق و قابل اعتماد ویژگی مورد نظر را اندازه‌گیری می‌کند.

اهمیت کیفیت اندازه‌گیری

- ✓ تعیین‌کننده اعتبار نتایج پژوهش
- ✓ پیش‌نیاز تحلیل‌های آماری قابل اعتماد
- ✓ پایه ارزیابی هر پرسشنامه یا ابزار سنجش



کیفیت اندازه‌گیری بر دو ستون اصلی استوار است:

I. **روایی (Validity)**

II. **پایایی (Reliability)**

کیفیت اندازه‌گیری زیربنای هر تحلیل کمی و پژوهش تجربی است. بدون روایی و پایایی مناسب، نتایج پژوهش غیرقابل اعتماد و تکرار ناپذیر خواهند بود.



روایی و پایایی

❖ برای استفاده از هر پرسشنامه در جمع‌آوری داده‌ها، رعایت دو ویژگی اساسی ضروری است:

I. روایی (Validity)

روایی نشان می‌دهد که ابزار تا چه حد واقعاً همان سازه یا ویژگی مورد نظر را اندازه‌گیری می‌کند. ابزار معتبر، سنجشی صحیح و مرتبط با مفهوم هدف را فراهم می‌کند.

II. پایایی (Reliability)

پایایی به میزان ثبات و تکرارپذیری نتایج ابزار اشاره دارد. اگر ابزار در زمان‌های مختلف یا در نمونه‌های مشابه به کار رود و نتایج یکسان یا نزدیک به هم تولید کند، پایاست.

روایی پرسشنامه
Questionnaire Validity

Reliability



Validity

هیچ ابزار نمی‌تواند معتبر باشد مگر اینکه پایا باشد، ولی ابزار پایا ممکن است معتبر نباشد.



Reliable
Not Valid



Low Reliability
Low Validity



Not Reliable
Not Valid



Reliable
Valid

انواع روایی



روایی صوری (Face Validity)

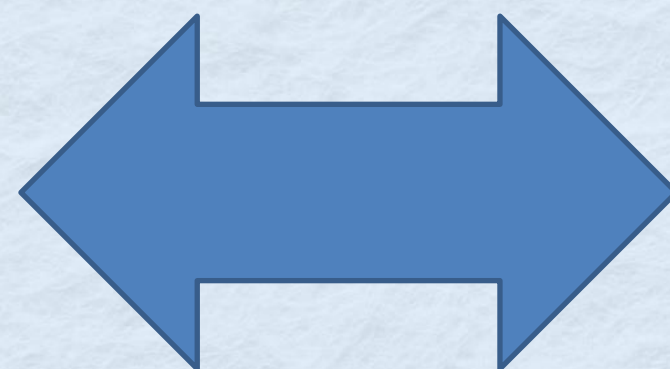
روایی صوری به این بررسی می‌پردازد که آیا ابزار اندازه‌گیری در ظاهر و بر اساس قضاوت افراد، به نظر می‌رسد که متغیر مورد مطالعه را به درستی می‌سنجد یا خیر.

- ✓ اولین و مقدماتی‌ترین روش برای ارزیابی اعتبار ابزار اندازه‌گیری است.
- ✓ پایین‌ترین سطح روایی از میان انواع روایی محسوب می‌شود، زیرا صرفاً بر قضاوت ظاهری مبتنی است.
- ✓ ارزیابی آن بیشتر ذهنی (subjective) است و بر برداشت اولیه از ظاهر، محتوا و وضوح سؤالات تکیه دارد.
- ✓ برخی متخصصان اعتقاد دارند که روایی صوری به تنهایی یک نوع روایی واقعی محسوب نمی‌شود؛ زیرا مبتنی بر داده تجربی یا تحلیل آماری نیست.
- ✓ با این حال، در مراحل اولیه ساخت ابزار اهمیت دارد، چون بر فهم‌پذیری، شفافیت و مناسب بودن ظاهری آیتم‌ها تمرکز دارد.

تعیین روایی صوری: روش کیفی و روش کمی

روایی صوری به روش کیفی (Qualitative Face Validity)

- ✓ بررسی وضوح سؤالات
- ✓ بررسی ساده و قابل فهم بودن
- ✓ بررسی مرتبط بودن با هدف تحقیق
- ✓ اصلاح ابهامات، حذف موارد تکراری، کوتاه‌سازی یا بازنویسی سؤالات
- ✓ تحلیل آماری ندارد — کاملاً بر قضاوت ذهنی مبتنی است.



در این روش، قضاوت انسانی
مبنای کار است. ابزار توسط
چند **متخصص** یا افراد **گروه**
هدف بررسی می‌شود.

پاسخ به این سوالات از نگاه گروه هدف

آیا اجزاء و کلیت
ابزار قابل پذیرش
پاسخ دهندگان می
باشد؟

آیا برداشت افراد غیر
متخصص (گروه هدف)
همان برداشتی است
که مورد نظر محقق
است؟

آیا افرادی که قرار
است به ابزار پاسخ
دهند، با عبارات و
جمله بندی ابزار
موافق هستند؟

آیا ابزار طراحی شده به
صورت ظاهری با هدف
مطالعه مرتبط است؟

- ✓ قضاوت انسانی مبنای اصلی آن است.
- ✓ نظر **متخصصان** برای ارتباط سوالات با سازه مهم است.
- ✓ نظر **گروه هدف** برای وضوح و فهم پذیری سوالات ضروری است.

تعیین روایی صوری: روش کیفی و روش کمی

1

پرسشنامه در یک پایلوت کوچک (۱۰-۱۵ نفر از گروه هدف) اجرا می‌شود.

2

از گروه هدف خواسته می‌شود میزان **اهمیت هر سؤال** را ارزیابی کنند.

3

معمولاً با مقیاس ۵ لیکرتی از «اصلاً مهم نیست» تا «کاملاً مهم است».

$$\text{Impact Score} = \text{Mean Importance} \times \text{Frequency}$$

تفسیر:

✓ اگر نمره تأثیر $< 1/5$ باشد ← آیتم قابل قبول است.

✓ اگر کمتر باشد ← سؤال حذف یا اصلاح می‌شود.

روایی صوری به روش کمی
(Quantitative Face Validity)

در این روش، از **شاخص تأثیر**
آیتم (Item Impact Score)

استفاده می‌شود.

بررسی نمره تأثیر در روایی صوری کمی

❖ برای بررسی نمرات تأثیر، ابتدا از شرکت کنندگان خواسته می شود تا میزان اهمیت هر یک از گویه های پرسشنامه را در یک طیف لیکرتی ۵ قسمتی ارزیابی کنند.

۱- کاملاً مهم است (امتیاز ۵)

۲- تا اندازه ای مهم است (امتیاز ۴)

۳- به طور متوسطی مهم است (امتیاز ۳)

۴- اندکی مهم است (امتیاز ۲)

۵- اصلاً مهم نیست (امتیاز ۱)

✓ منظور از فراوانی بر حسب درصد، تعدادی از افرادی است که به آیتم، امتیاز ۴ و ۵ داده اند.

✓ هدف از اهمیت، میانگین نمرات افراد پاسخ دهنده بر اساس طیف لیکرتی مورد نظر می باشد.

✓ در صورتیکه امتیاز تأثیر از یک ونیم بیشتر شود، آیتم برای تحلیل های بعدی مناسب تشخیص داده شده و حفظ خواهد شد.

بررسی نمره تأثیر در روایی صوری کمی

| محاسبه امتیاز تأثیر (برای ۱۰ نفر پاسخ دهنده) | | | | | | |
|--|--|-----------------|----------|-------------------|----------------|------------------------|
| شماره | اینم | کاملاً مهم ۵ | مهم ۴ | متوسط مهم ۳ | اندکی مهم ۲ | اصلاً مهم نیست ۱ |
| ۱ | در طول یک ماه گذشته، سیگار کشیده اید؟ | ۲ | ۴ | ۲ | ۱ | ۱ |
| ۲ | در طول یک ماه گذشته با تعارف دوستان سیگار کشیده اید؟ | | | | | |
| ۳ | در طول یک ماه گذشته، از قلیان استفاده کرده اید؟ | | | | | |
| ۴ | در طول یک ماه گذشته، نوشیدنی های الکلی استفاده کرده اید؟ | | | | | |
| ۵ | در طول یک ماه گذشته، از مواد مخدر استفاده کرده اید؟ | | | | | |

$$\text{Mean Importance} = 1+2+3+3+4+4+4+4+5+5=35 / 10=3.5$$

$$\text{Frequency (\%)} = 6/10 \times 100 = 0.6 \%$$

$$\text{Impact Score} = 3.5 \times 0.6 = 2.1 > 1.5 \rightarrow \text{قابل قبول}$$

روایی محتوا (Content Validity)

❖ **روایی محتوا** بیان می‌کند که آیا آیتم‌های یک ابزار اندازه‌گیری (مثل پرسشنامه)، تا چه حد تمام جنبه‌ها و ابعاد آن مفهومی را که قصد سنجش آن را داریم، پوشش می‌دهد و موارد غیرمرتبط در آن دخالت ندارند.

❖ آیا سؤال‌ها دقیقاً همان چیزی را می‌پرسند که باید پرسند؟

❖ ارزیابی روایی محتوا از دو طریق انجام می‌شود:

I. روش کیفی

II. روش کمی

روش ارزیابی کیفی (Qualitative Content Validity)

✓ **کارشناسان حوزه مربوطه (expert judgment)**، آیتم‌ها را بررسی می‌کنند و میزان ارتباط و کفایت آن‌ها را با هدف پژوهش ارزیابی می‌کنند.

□ در این روش، گروهی از متخصصان (۵-۱۰ نفر) موارد زیر را بررسی می‌کنند:

✓ شفافیت سؤالات

✓ تناسب و ارتباط با سازه

✓ کفایت تعداد آیتم‌ها

✓ دستورالعمل مناسب

✓ مقیاس پاسخ‌دهی مناسب

❖ این روش عدد نمی‌دهد؛ بیشتر اصلاح ساختاری پرسشنامه است.

روش ارزیابی کمی (Quantitative Content Validity)

I. نسبت روائی محتوا (CVR :Content Validity Ratio)

- ✓ این شاخص توسط لاوشه lawsche ابداع شد.
- ✓ اطمینان از انتخاب مهم‌ترین و ضروری‌ترین محتوا (ضرورت آیتم)
- ✓ هدف: تعیین اینکه هر آیتم واقعاً ضروری است یا خیر.
- ✓ قضاوت متخصصان بر اساس سه گزینه: ضروری / مفید اما غیرضروری / غیرضروری.
- ✓ مقایسه مقدار CVR با جدول لاوشه برای تأیید یا حذف آیتم‌ها.

II. شاخص روائی محتوا (CVI: Content Validity Index)

- ✓ اطمینان از اینکه آیتم‌های ابزار بهترین طراحی را برای اندازه‌گیری محتوا دارند.
- ✓ تعیین میزان ارتباط، وضوح، سادگی (Simplicity) و شفافیت برای آیتم‌ها.

روش ارزیابی کمی (Quantitative Content Validity)

□ نسبت روایی محتوا (CVR :Content Validity Ratio)

پاسخها بر اساس فرمول زیر محاسبه می گردد:

$$CVR = \frac{n_E - \frac{N}{2}}{\frac{N}{2}}$$

(تعداد افرادی است که گزینه ضروری را نموده اند **منهای** نصف تعداد متخصصان) تقسیم بر نصف تعداد متخصصان

که در این موارد به گزینه **ضروری** نمره ۲ و به دو گزینه دیگر مفید اما ضروری است نمره ۱ و گزینه ضروری نیست نمره صفر

پس طبق فرمول، نسبت روایی محتوا محاسبه می شود و نمره به دست آمده با جدول ارائه شده توسط **لاوشه** از لحاظ قابل قبول بودن مقایسه می شود.

عباراتی که مقدار آنها از ۰/۶۲ بالاتر، حفظ می شوند.

روش ارزیابی کمی (Quantitative Content Validity)

□ نسبت روایی محتوا (CVR :Content Validity Ratio)

○ فرض کنید برای یک آیتم پرسشنامه، ۱۰ متخصص نظر داده‌اند.

هر متخصص باید بگوید آیتم:

1. ضروری است

2. مفید است اما ضروری نیست

3. غیرضروری است

حالا فرض کنیم ۷ نفر از متخصصان گفته‌اند آیتم ضروری است.

بنابراین:

○ $N=10$ تعداد کل متخصصان

○ $Ne=7$ تعداد متخصصانی که آیتم را «ضروری» دانسته‌اند.

$$CVR = (7-5)/5 = 0.4$$

برای ۱۰ متخصص، حداقل مقدار قابل قبول CVR طبق جدول لاوشه = ۰/۶۲ است.

شاخص ما: ۰/۴

نتیجه: این آیتم باید حذف یا اصلاح شود چون ضروری تلقی نشده.

روایی محتوا CVR و CVI برای هر آیتم محاسبه می‌شود، نه برای کل پرسشنامه.



روش ارزیابی کمی (Quantitative Content Validity)

□ شاخص روائی محتوا (CVI: Content Validity Index)

شاخص CVI نشان می‌دهد که آیتم‌های ابزار تا چه حد برای اندازه‌گیری سازه طراحی شده‌اند و مرتبط هستند.

✓ ارزیابی توسط پانل خبرگان انجام می‌شود.

✓ معیار اصلی: اطمینان از اینکه آیتم‌ها به بهترین نحو جهت اندازه‌گیری سازه طراحی شده‌اند.

✓ روش: طیف لیکرت ۴ قسمتی برای هر آیتم و سه معیار زیر.

طیف ۴ لیکرت: (۱) نامرتب / ضعیف (۲) تا حدودی مرتبط (۳) مرتبط (۴) بسیار مرتبط

معیارهای ارزیابی آیتم‌ها (Waltz & Basel)

| معیار | توضیح |
|-------------------|--|
| سادگی و روان بودن | آیتم به صورت ساده و قابل فهم طراحی شده باشد. |
| مربوط بودن | آیتم دقیقاً با سازه مورد نظر مرتبط باشد. |
| شفاف بودن | آیتم واضح، دقیق و بدون ابهام باشد. |



1. هر متخصص برای هر آیتم، هر سه معیار را در طیف ۴ درجه‌ای (۱ تا ۴) امتیاز می‌دهد.
2. برای هر آیتم: تعداد متخصصانی که به آیتم امتیاز ۳ یا ۴ داده‌اند، شمرده می‌شود.
3. میانگین CVIها یا نسبت آیتم‌های قابل قبول.
 - $CVI \geq 0.78$ ← قابل قبول است.
 - $CVI < 0.78$ ← آیتم نیاز به اصلاح یا حذف دارد.

روایی محتوا (Content Validity)

- روایی محتوی به تحلیل محتوی یک آزمون بستگی داشته و **تعیین آن بر اساس قضاوت ذهنی و فردی است.**
- در صورتی که بین افراد مختلف در زمینه روایی آزمون **توافق** وجود داشته باشد، آن آزمون دارای روایی محتوایی است.
- برای سنجش روایی **تنها انجام روایی محتوا کافی نیست.**
- در سنجش روایی محتوا **حداقل از نظر ۱۰ متخصص همان رشته** استفاده شود.

اعتبار ملاکی (Criterion-Related Validity)

- ❖ اعتبار ملاکی نشان می‌دهد که ابزار اندازه‌گیری چقدر با یک معیار خارجی مرتبط است.
- ❖ در روش اعتبار ملاکی، عملکرد فرد در آزمون با یک ملاک مقایسه می‌شود.
- معیار ملاک شامل: همزمان (Concurrent)، پیش بین (Predictive) می‌باشد.

□ اعتبار همزمان (Concurrent Validity)

- ✓ ابزار و معیار ملاک همزمان اندازه‌گیری می‌شوند.
- ✓ به دو یا چند گروه همزمان پرسشنامه داده می‌شود و آزمون می‌کنیم و بررسی می‌کنیم آیا تفاوتی وجود دارد یا خیر.
- ✓ مثال: پرسشنامه افسردگی همزمان با تشخیص روانپزشک.
- ✓ شاخص: ضریب همبستگی پیرسون یا اسپیرمن بین نمره پرسشنامه و معیار ملاک.
- ✓ اگر همبستگی بین نمرات آزمون‌هایی که سازه واحدی را اندازه می‌گیرند بالا باشد، آزمون دارای اعتبار همگرا است.

□ اعتبار پیش‌بینی (Predictive Validity)

- ✓ ابزار پیش‌بینی‌کننده معیار ملاک در آینده است.
- ✓ مثال: تست هوش کودکان و عملکرد تحصیلی ۱ سال بعد.
- ✓ شاخص: ضریب همبستگی یا ROC curve برای متغیر دوتایی.
- ✓ AUC نزدیک به ۱ ← اعتبار بالا

اعتبار ملاکی (Criterion-Related Validity)

روایی همزمان (Concurrent Validity)

پرسشنامه افسردگی متغیر پرسشنامه (DEP_SCORE) تشخیص روانپزشک متغیر ملاک (CLINICAL_SCORE)

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the 'Correlate' option is selected. A sub-menu is displayed, highlighting the 'Bivariate...' option. The data editor window shows two columns: 'ID' and 'DEP_SCORE'. The data rows are as follows:

| ID | DEP_SCORE |
|----|-----------|
| 1 | 15 |
| 2 | 20 |
| 3 | 10 |
| 4 | 18 |
| 5 | 12 |
| 6 | 16 |
| 7 | 18 |
| 8 | 11 |
| 9 | 12 |
| 10 | 15 |
| 11 | 16 |
| 12 | 10 |
| 13 | 9 |
| 14 | 12 |
| 15 | 11 |
| 16 | 18 |

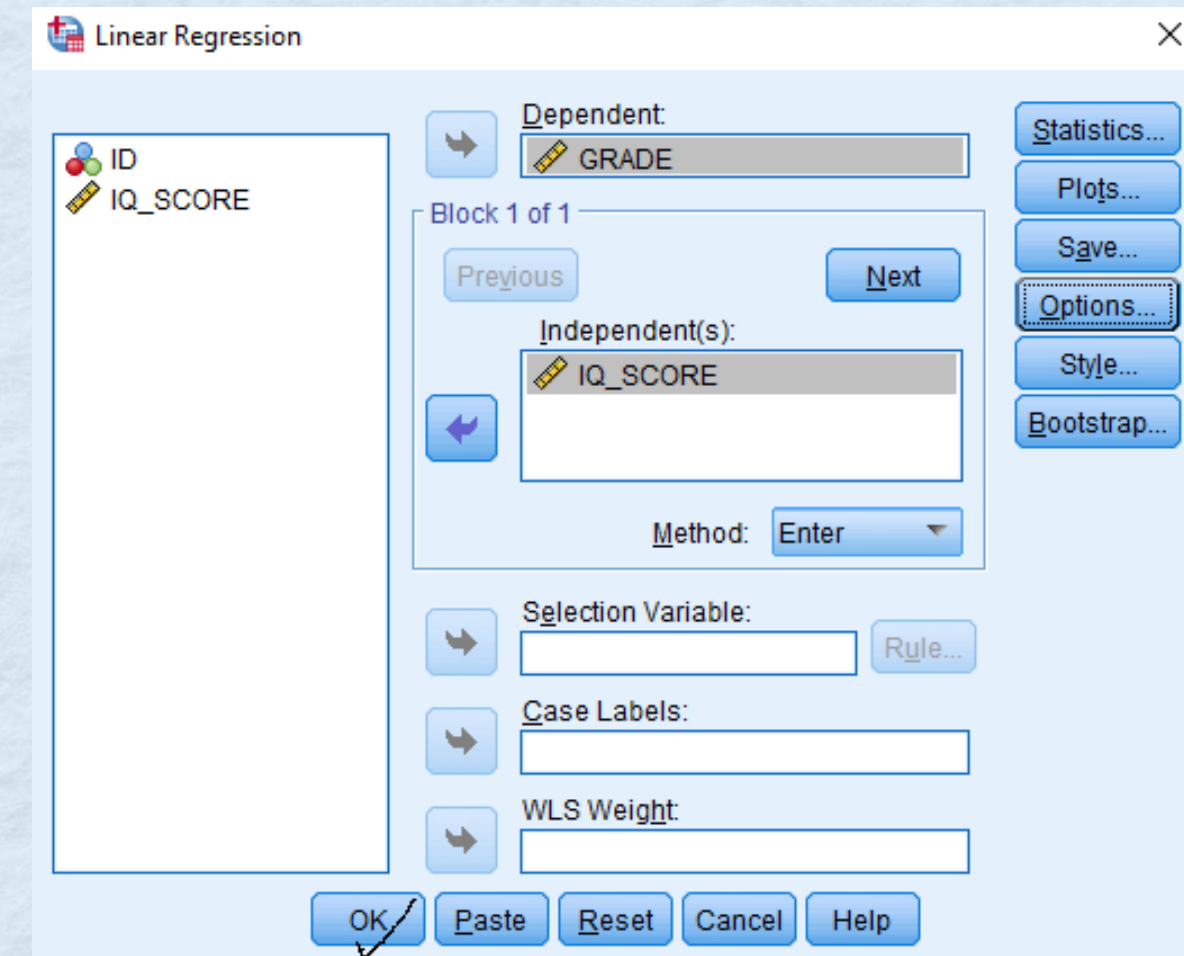
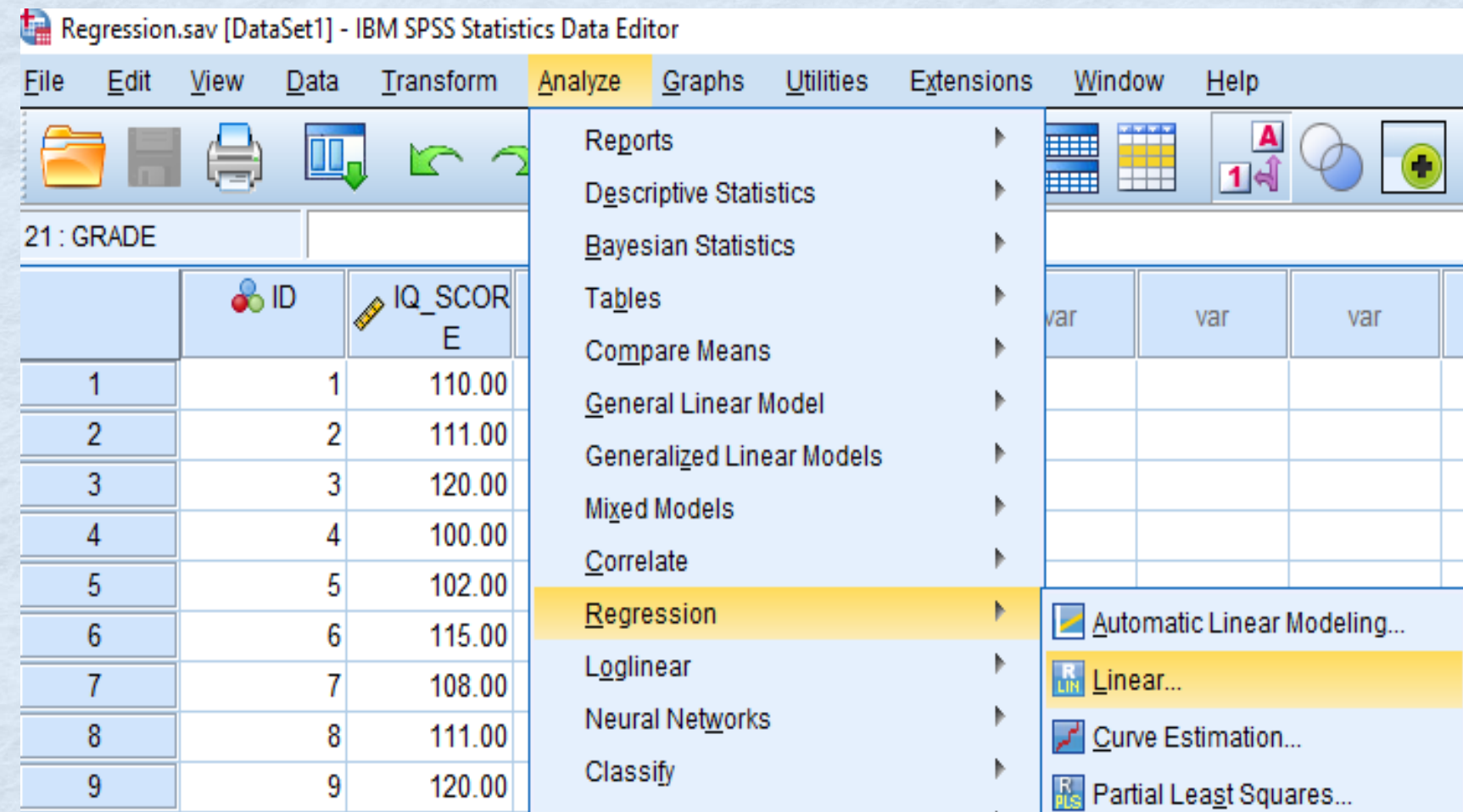
The 'Bivariate Correlations' dialog box is shown. The 'Variables:' list contains 'DEP_SCORE' and 'CLINICAL_SCORE'. The 'Correlation Coefficients' section has 'Pearson' checked. The 'Test of Significance' section has 'Two-tailed' selected. The 'Flag significant correlations' checkbox is also checked. Buttons for 'OK', 'Paste', 'Reset', 'Cancel', and 'Help' are visible at the bottom.

| Correlations | | | |
|--------------|---------------------|-----------|--------------------|
| | | DEP_SCORE | CLINICAL_S CORE |
| DEP_SCORE | Pearson Correlation | 1 | .821** |
| | Sig. (2-tailed) | | .000 |
| | N | 20 | 20 |

اعتبار ملاکی (Criterion-Related Validity)

روایی پیش بین (Predictive Validity)

نمره تست هوش کودکان (IQ_SCORE) نمره تحصیلی بعد از یکسال (Grade)



| Model Summary | | | | |
|---------------|-------------------|----------|-------------------|----------------------------|
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .150 ^a | .023 | -.032 | 6.55307 |

a. Predictors: (Constant), IQ_SCORE

| Coefficients ^a | | | | | | | | |
|---------------------------|------------|-----------------------------|------------|---------------------------|-------|------|---------------------------------|-------------|
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. | 95.0% Confidence Interval for B | |
| | | B | Std. Error | Beta | | | Lower Bound | Upper Bound |
| 1 | (Constant) | 97.784 | 22.061 | | 4.432 | .000 | 51.436 | 144.132 |
| | IQ_SCORE | -.128 | .198 | -.150 | -.644 | .527 | -.544 | .289 |

a. Dependent Variable: GRADE

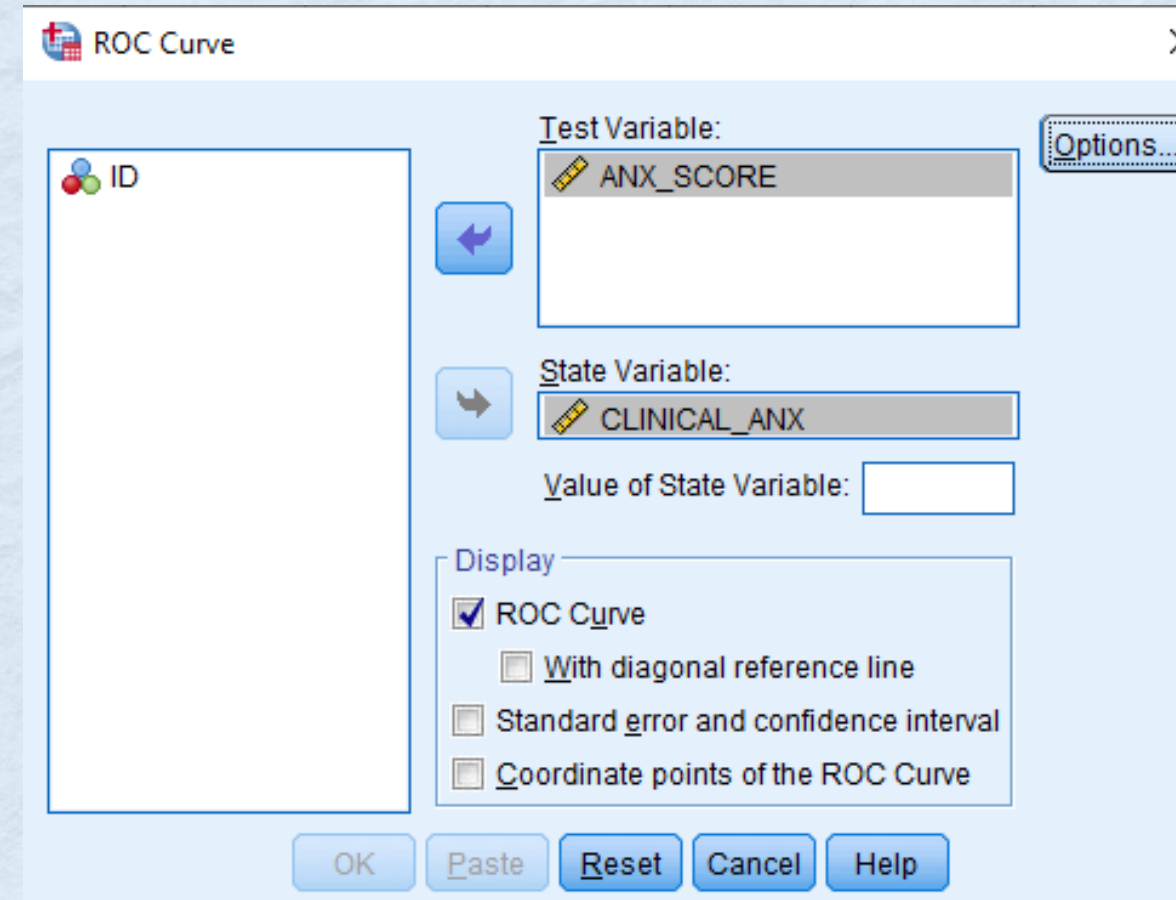
اعتبار ملاکی (Criterion-Related Validity)

روایی پیش بین (Predictive Validity)

تشخیص بالینی اضطراب: (بدون اضطراب (+), با اضطراب (1))

نمره پرسشنامه اضطراب (ANX_SCORE)

| | ID | ANX_SCORE |
|----|----|-----------|
| 1 | 1 | 10.00 |
| 2 | 2 | 18.00 |
| 3 | 3 | 12.00 |
| 4 | 4 | 14.00 |
| 5 | 5 | 10.00 |
| 6 | 6 | 12.00 |
| 7 | 7 | 20.00 |
| 8 | 8 | 19.00 |
| 9 | 9 | 14.00 |
| 10 | 10 | 20.00 |
| 11 | 11 | 10.00 |
| 12 | 12 | 17.00 |
| 13 | 13 | 12.00 |
| 14 | 14 | 13.00 |
| 15 | 15 | 18.00 |
| 16 | 16 | 16.00 |
| 17 | 17 | 15.00 |



| Area Under the Curve | |
|--|-----------|
| Test Result Variable(s): | ANX_SCORE |
| Area | .995 |
| The test result variable(s): ANX_SCORE has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased. | |

AUC نزدیک به ۱ ← ابزار دارای اعتبار پیش‌بینی بالا
برای تشخیص اضطراب

روایی سازه (Construct Validity)

❖ روایی سازه نشان می‌دهد که آیا ابزار ما واقعاً همان سازه‌ای را اندازه‌گیری می‌کند که بر اساس نظریه طراحی شده است و نتایج آن با تئوری‌های مرتبط سازگار است.

روش‌های ارزیابی رایج:

✓ تحلیل عاملی (Factor Analysis)

✓ تحلیل خوشه‌ای (Cluster Analysis)

پایایی پرسشنامه
Questionnaire Reliability

پایایی پرسشنامه

- ❖ **پایایی (Reliability)** به میزان ثبات و تکرارپذیری نتایج حاصل از یک ابزار اندازه‌گیری اشاره دارد. به این معنا که اگر یک پرسشنامه در **زمان‌های مختلف** و در یک **جمعیت ثابت** مورد استفاده قرار گیرد، نتایج به‌دست‌آمده تفاوت قابل‌توجهی با یکدیگر نخواهند داشت.
- ❖ اهمیت پایایی از آن جهت است که نشان می‌دهد ابزار اندازه‌گیری در **شرایط یکسان** تا چه اندازه نتایج مشابه و قابل اعتماد تولید می‌کند.
- ❖ برای سنجش پایایی از شاخصی به نام **ضریب پایایی (Reliability Coefficient)** استفاده می‌شود. دامنه این ضریب از **صفر تا یک** متغیر است؛ به طوری که مقدار **صفر** بیانگر عدم پایایی و مقدار **یک** نشان‌دهنده پایایی کامل ابزار است. با این حال، پایایی کامل به ندرت در عمل مشاهده می‌شود و در صورت گزارش چنین مقادیری، معمولاً نسبت به نتایج به‌دست‌آمده با دیده تردید نگریسته می‌شود.

انواع پایایی

| روش ارزیابی معمول | تعریف | انواع پایایی |
|--|---|---|
| همبستگی نتایج در دو زمان مختلف Pearson/Spearman Correlation, ICC | میزان ثبات نتایج ابزار در طول زمان را می‌سنجد. در این روش، پرسشنامه در دو زمان مختلف بر روی یک گروه یکسان اجرا شده و میزان همبستگی بین نمرات دو بار اجرا محاسبه می‌شود. | پایایی بازآزمون (Test-Retest Reliability) |
| همبستگی آیتم‌های سازه‌ها Cronbach's Alpha, McDonald's Omega | میزان همخوانی و هماهنگی سؤالات یک ابزار را در سنجش یک سازه واحد بررسی می‌کند. | پایایی همسانی درونی (موازی) (Internal Consistency) |
| Spearman-Brown Coefficient | در این روش، ابزار به دو نیمه معادل تقسیم شده و همبستگی بین نمرات دو نیمه محاسبه می‌شود. آیا نیمه‌های مختلف پرسشنامه نتایج مشابهی دارند؟ | پایایی تصنیف (دو نیمه) (Split-Half Reliability) |
| Cohen's Kappa, ICC | میزان توافق بین ارزیابان مختلف را هنگام سنجش یک متغیر بررسی می‌کند. کاربردی در ابزارهای مشاهده‌ای و چک‌لیست‌ها | پایایی ناظر-ناظر (Inter-Rater Reliability) |

پایایی باز آزمایی (Test-Retest Reliability)

- در این روش برای سنجش پایایی، سوالات آزمون در دو نوبت و تحت شرایط مشابه به یک گروه واحد داده می شود و نمرات حاصل با هم مقایسه می شوند.
- در این روش، **یک پرسشنامه برای همان افراد در دو نوبت زمانی متفاوت** اجرا می شود و در نهایت ضریب همبستگی بین نمرات حاصل از اجرای دو آزمون برآورده شده تا میزان شباهت امتیازات مشخص شود و به عنوان ضریب پایایی به کار رود.
- اگر ابزار پایا باشد، انتظار داریم افراد در دو نوبت نمرات مشابهی کسب کنند.
- روش آزمون - باز آزمون فرض می کند که متغیرها یا مفاهیم مورد اندازه گیری و همچنین خصوصیات آزمون شوندگان در طول دوره تغییر نخواهند کرد.

فاصله زمانی پیشنهادی:

- معمولاً ۲ تا ۴ هفته
- نه آن قدر کوتاه که پاسخها یادشان بماند.
- نه آن قدر طولانی که واقعاً سازه تغییر کند.

چه زمانی از باز آزمایی استفاده می کنیم؟

- وقتی سازه نسبتاً پایدار است (مانند: اضطراب، نگرش، دانش، باورها)
- نه برای سازه های ناپایدار (مانند: درد لحظه ای، خلق روزانه)

پایایی باز آزمایی (Test–Retest Reliability)

□ شاخص آماری مورد استفاده بسته به نوع داده دارد.

| شاخص | نوع داده |
|------------------------------------|----------|
| ضریب همبستگی پیرسون (r Pearson) | پیوسته |
| Spearman | رتبه‌ای |
| Kappa | طبقه‌ای |

| مقدار r | تفسیر |
|---------------|-------|
| < 0.50 | ضعیف |
| $0.50 - 0.69$ | متوسط |
| $0.70 - 0.79$ | خوب |
| ≥ 0.80 | عالی |

پایایی باز آزمایی (Test-Retest Reliability)

□ مراحل تحلیل در SPSS

Analyze → Correlate → Bivariate

The screenshot shows the SPSS interface with the 'Analyze' menu open. The 'Correlate' option is selected, and the 'Bivariate...' sub-menu is visible. The data editor shows two variables: 'ID' and 'T1'.

| ID | T1 |
|----|-------|
| 1 | 18.00 |
| 2 | 22.00 |
| 3 | 15.00 |
| 4 | 30.00 |
| 5 | 25.00 |
| 6 | |
| 7 | |
| 8 | |
| 9 | |

Descriptive Statistics

| | Mean | Std. Deviation | N |
|----|---------|----------------|---|
| T1 | 22.0000 | 5.87367 | 5 |
| T2 | 22.2000 | 5.26308 | 5 |

Correlations

| | | T1 | T2 |
|----|---------------------|--------|--------|
| T1 | Pearson Correlation | 1 | .987** |
| | Sig. (2-tailed) | | .002 |
| | N | 5 | 5 |
| T2 | Pearson Correlation | .987** | 1 |
| | Sig. (2-tailed) | .002 | |
| | N | 5 | 5 |

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

ضریب همبستگی پیرسون بین نمرات آزمون و بازآزمایی برابر با $r = 0.91$ و از نظر آماری معنادار است ($p < 0.05$)، که نشان‌دهنده پایایی بسیار بالا ابزار می‌باشد.

پایایی باز آزمایی (Test-Retest Reliability)

Analyze → Descriptive Statistics → Crosstabs

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, and the path 'Descriptive Statistics' > 'Crosstabs...' is highlighted. A purple arrow points to the 'Crosstabs: Statistics' dialog box, where the 'Kappa' checkbox is checked. Other options like Chi-square, Correlations, and Nominal/Ordinal statistics are also visible.

| Symmetric Measures | | | | | |
|----------------------|-------|-------|--|----------------------------|--------------------------|
| | | Value | Asymptotic Standard Error ^a | Approximate T ^b | Approximate Significance |
| Measure of Agreement | Kappa | .667 | .287 | 1.732 | .083 |
| N of Valid Cases | | 6 | | | |

a. Not assuming the null hypothesis.
b. Using the asymptotic standard error assuming the null hypothesis.

پایایی باز آزمایی (Test-Retest Reliability)

مقدار **Kappa = 0.667** نشان‌دهنده توافق خوب بین آزمون و باز آزمایی است.

ضریب کاپای کوهن برای ارزیابی پایایی باز آزمایی برابر با ۰/۶۶۷ بود که نشان‌دهنده **توافق خوب** بین دو نوبت اندازه‌گیری است. با توجه به حجم نمونه محدود، آزمون معناداری به سطح بزرگتر از ۰/۰۵ است، که قابل انتظار بوده و از اعتبار تفسیر اندازه اثر نمی‌کاهد. □ مقدار **Kappa** مهم‌تر از **p-value** است.

بسیاری از مقالات اصلاً **p-value** را گزارش نمی‌کنند.

✓ بنابراین غیرمعنادار بودن، به معنی نبود پایایی نیست.

| مقدار Kappa | تفسیر |
|-------------|-------|
| < 0.20 | ضعیف |
| 0.21 – 0.40 | کم |
| 0.41 – 0.60 | متوسط |
| 0.61 – 0.80 | خوب |
| ≥ 0.81 | عالی |

پایایی همسانی درونی (Internal Consistency Reliability)

پایایی همسانی درونی به میزان انسجام و هماهنگی آیتم‌های یک ابزار در سنجش یک سازه واحد اشاره دارد. در این روش، ابزار اندازه‌گیری تنها یک بار اجرا می‌شود و بررسی می‌گردد که آیا آیتم‌های مختلف پرسشنامه به‌طور همسو یک مفهوم مشخص را اندازه‌گیری می‌کنند یا خیر.

این نوع پایایی زمانی استفاده می‌شود که:

- ✓ ابزار فقط **یک بار** اجرا شده باشد
- ✓ هدف، بررسی **انسجام درونی** آیتم‌ها باشد.
- شاخص‌های سنجش:

I. آلفای کرونباخ (Cronbach's Alpha)

II. KR-20 (کودر-ریچاردسون) است.

در مجموع، ضریب پایایی آلفای کرونباخ زمانی کاربرد دارد که سؤالات پرسشنامه به‌صورت **دو حالتی** (صحیح-غلط) طراحی نشده باشند و برای سنجش ثبات یا همسانی درونی آیتم‌ها به‌کار می‌رود.

محاسبه آلفای کرونباخ (SPSS)

یک پرسشنامه ۵ سؤالی شدت درد پس از عمل جراحی داریم (لیکرت ۱ = بدون درد تا ۵ = درد شدید)
آیتم‌ها:

✓ Q1 شدت درد در استراحت

✓ Q2 شدت درد هنگام حرکت

✓ Q3 نیاز به مسکن

✓ Q4 اختلال خواب به علت درد

✓ Q5 محدودیت فعالیت روزانه

محاسبه آلفای کرونباخ (SPSS)

Analyze → Scale → Reliability Analysis

Alpha.sav [DataSet2] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

6 : Q3 3.00

| | ID | Q1 |
|----|-------|------|
| 1 | 1.00 | 4.00 |
| 2 | 2.00 | 3.00 |
| 3 | 3.00 | 5.00 |
| 4 | 4.00 | 2.00 |
| 5 | 5.00 | 4.00 |
| 6 | 6.00 | 3.00 |
| 7 | 7.00 | 5.00 |
| 8 | 8.00 | 2.00 |
| 9 | 9.00 | 4.00 |
| 10 | 10.00 | 3.00 |
| 11 | | |
| 12 | | |
| 13 | | |
| 14 | | |
| 15 | | |
| 16 | | |

Reports
Descriptive Statistics
Bayesian Statistics
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

Q5 var var var
4.00
3.00
5.00
2.00
4.00
3.00
5.00
2.00
4.00
3.00

Reliability Analysis...
Multidimensional Unfolding (PREFSCAL)...
Multidimensional Scaling (PROXSCAL)...
Multidimensional Scaling (ALSCAL)...

Reliability Analysis

Items:
Q1
Q2
Q3
Q4
Q5

Ratings:

Model: Alpha
Scale label: Alpha

Statistics...

Paste Reset Cancel Help

محاسبه آلفای کرونباخ (SPSS)

Reliability Analysis: Statistics

Descriptives for

- Item
- Scale
- Scale if item deleted

Inter-Item

- Correlations
- Covariances

Summaries

- Means
- Variances
- Covariances
- Correlations

ANOVA Table

- None
- F test
- Friedman chi-square
- Cochran chi-square

Interrater Agreement: Fleiss' Kappa

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

Asymptotic significance level (%): 95

Missing

- Exclude both user-missing and system missing values

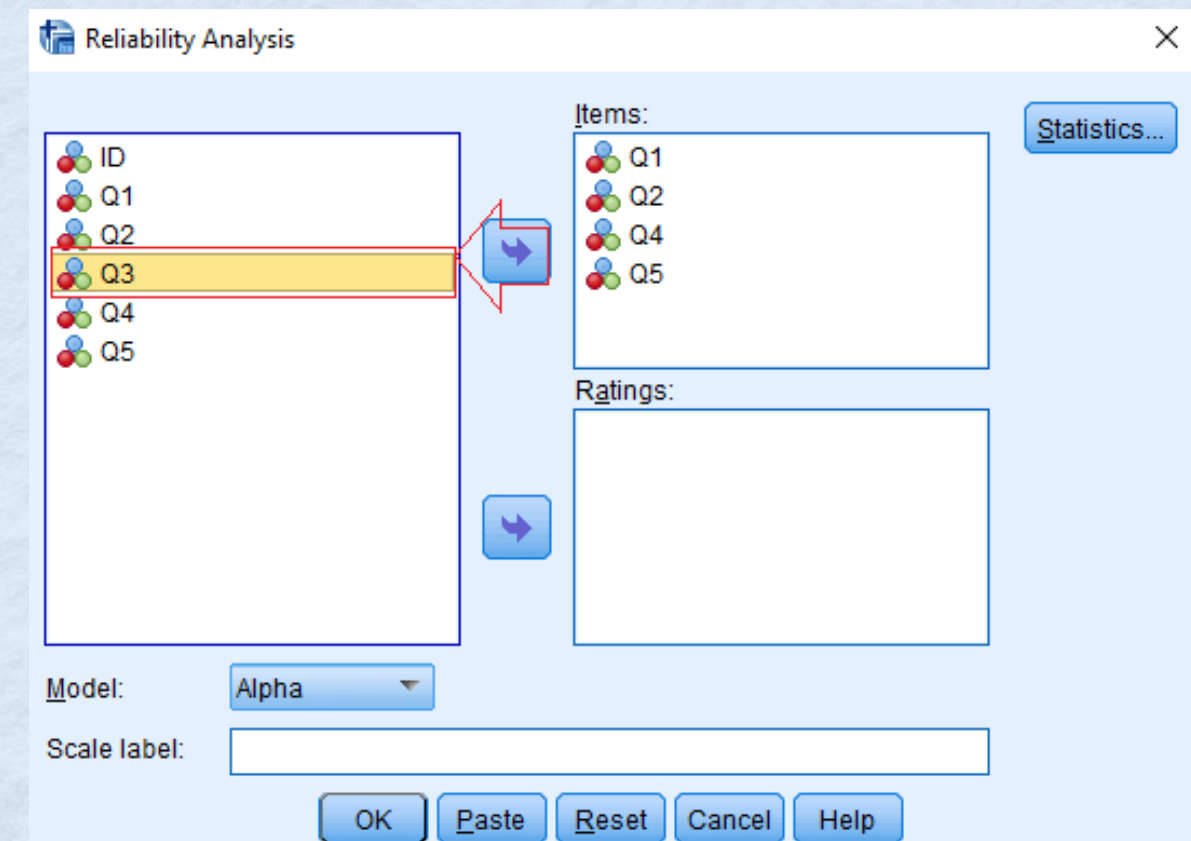
Statistics...

| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .964 | .964 | 5 |

- ضریب آلفای کرونباخ شاخصی است برای سنجش پایایی درونی یک پرسشنامه؛ یعنی نشان می‌دهد آیتم‌ها تا چه حد با هم هم‌راستا و سازگار هستند.
- مقدار 0.96 : **بسیار بالا** است و نشان می‌دهد که آیتم‌های پرسشنامه تقریباً به‌طور کامل با یکدیگر هم‌خوانی دارند.
- پرسشنامه از نظر **پایایی عالی** است.
- نیازی به حذف هیچ آیتمی نیست، چون همه آیتم‌ها همسو هستند.

محاسبه آلفای کرونباخ (SPSS)

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Squared Multiple Correlation | Cronbach's Alpha if Item Deleted |
|----|----------------------------|--------------------------------|----------------------------------|------------------------------|----------------------------------|
| Q1 | 14.3000 | 15.789 | .945 | . | .948 |
| Q2 | 14.3000 | 16.233 | .881 | . | .958 |
| Q3 | 14.0000 | 16.889 | .838 | . | .965 |
| Q4 | 14.3000 | 16.233 | .881 | . | .958 |
| Q5 | 14.3000 | 15.789 | .945 | . | .948 |



| Cronbach's Alpha | Cronbach's Alpha Based on Standardized Items | N of Items |
|------------------|--|------------|
| .965 | .965 | 4 |

➤ پرسشنامه از نظر پایایی عالی است.

➤ نیازی به حذف هیچ آیتمی نیست، چون همه آیتمها همسو هستند.

کودر-ریچاردسون (SPSS)

- ❖ نشان می‌دهد هر آیت‌م پرسشنامه چقدر با کل پرسشنامه همبستگی دارد.
- ❖ اگر همبستگی پایین باشد (معمولاً کمتر از ۰/۳)، آن آیت‌م ممکن است به هم‌خوانی کل پرسشنامه کمک نکند و قابل حذف است
- ❖ در SPSS، کودر-ریچاردسون عملاً همان **Corrected Item-Total Correlation** است که در تحلیل پایایی با آلفای کرونباخ قابل دسترسی است.

Analyze → Scale → Reliability Analysis

پایایی تنصیفی (Split-Half Reliability)

- ❖ این روش یکی از روش‌های سنجش پایایی درونی (Internal Consistency) پرسشنامه است. پرسشنامه به دو نیمه مساوی تقسیم می‌شود (مثلاً آیتم‌های زوج و فرد، یا نصف اول و نصف دوم).
- ❖ سپس همبستگی بین نمره مجموع دو نیمه محاسبه می‌شود.
- ❖ هدف: بررسی این که آیا نیمه‌های مختلف پرسشنامه نتایج مشابهی می‌دهند.
- ❖ چون تقسیم پرسشنامه باعث کاهش تعداد آیتم‌ها در هر نیمه می‌شود، معمولاً از فرمول اسپیرمن-براون (Spearman-Brown) برای اصلاح همبستگی استفاده می‌کنیم تا پایایی کل پرسشنامه تخمین زده شود.
- ❖ روش Split-Half سریع و ساده برای ارزیابی پایایی داخلی معمولاً همراه با آلفای کرونباخ استفاده می‌شود. اگر همبستگی بین دو نیمه پایین باشد ← نشان‌دهنده آیتم‌های ناسازگار

پایایی تنصیفی (Split-Half Reliability)

Analyze → Scale → Reliability Analysis

Split-Half ✓

ضریب Spearman–Brown گزارش می‌شود. ✓

| Reliability Statistics | | | |
|--------------------------------|------------------|------------|----------------|
| Cronbach's Alpha | Part 1 | Value | .954 |
| | | N of Items | 3 ^a |
| | Part 2 | Value | .945 |
| | | N of Items | 2 ^b |
| | Total N of Items | | 5 |
| Correlation Between Forms | | | .880 |
| Spearman-Brown Coefficient | Equal Length | | .936 |
| | Unequal Length | | .938 |
| Guttman Split-Half Coefficient | | | .895 |

a. The items are: Q1, Q2, Q4.
b. The items are: Q5, Q3.

• همه شاخص‌ها بالای ۰/۹ ← پرسشنامه پایایی بسیار خوبی دارد

• همبستگی بین آیت‌ها و نیمه‌ها نشان می‌دهد تمام آیت‌ها با هم همسو هستند.

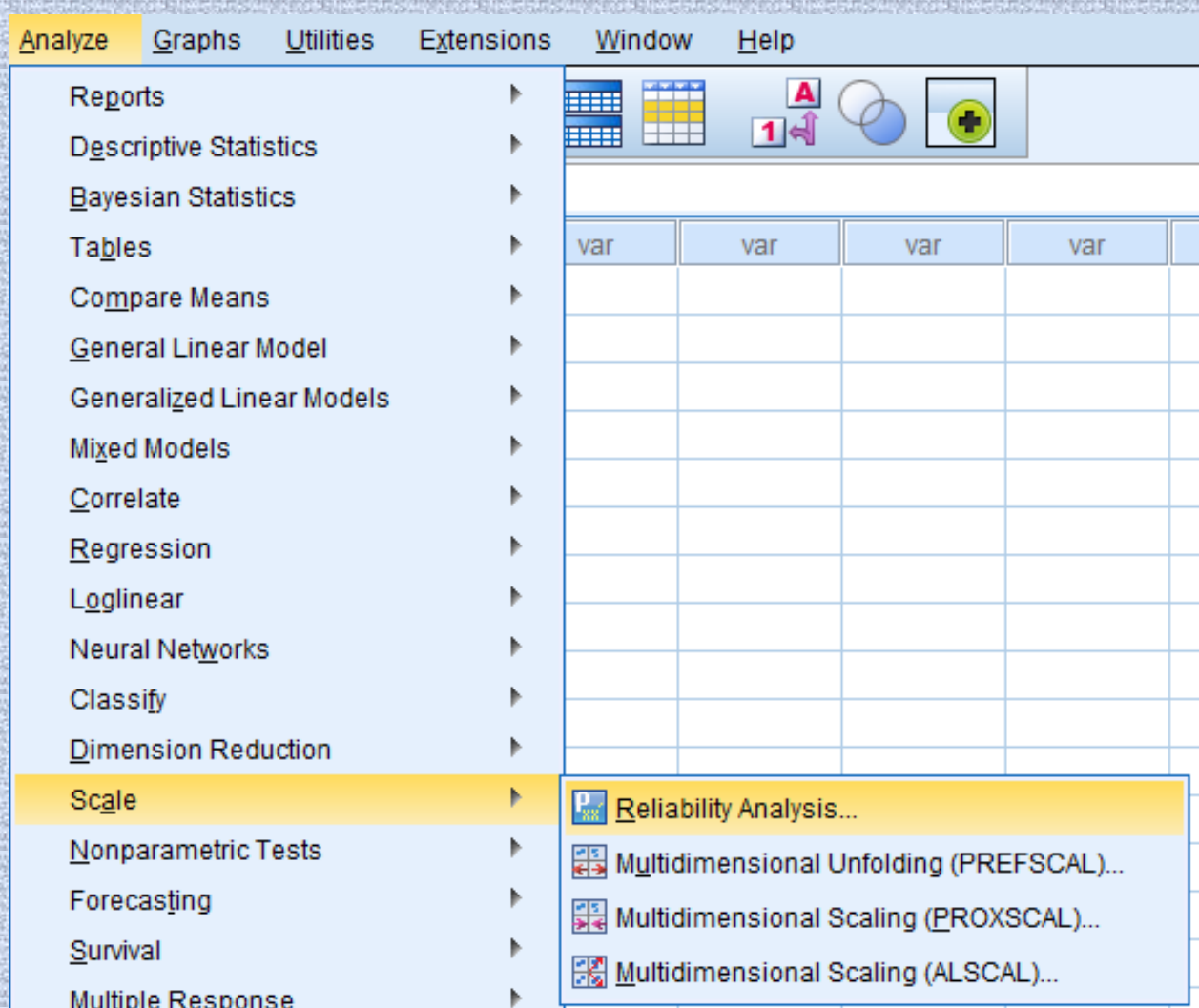
پایایی بین ارزیابان (Inter-Rater Reliability)

- ❖ این پایایی زمانی استفاده می‌شود که چند ارزیاب یک پدیده یا پرسشنامه را ارزیابی می‌کنند و هدف سنجش میزان توافق بین آنهاست.
- IRR نشان می‌دهد که نتایج ثبت شده توسط چند ارزیاب مستقل چقدر با یکدیگر همخوانی دارند.
- کاربرد: در مطالعات بالینی، روان‌شناسی، اپیدمیولوژی و هر جایی که قضاوت انسانی وارد شود.
- شاخص‌های ارزیابی IRR
 - ضریب کاپا (Cohen's Kappa) داده‌های طبقه‌ای
 - ضریب همبستگی درون کلاسی (Intraclass Correlation Coefficient) برای داده‌های پیوسته یا نمره‌های عددی

| شاخص | نوع داده | بازه | تفسیر |
|------|----------|------|-----------------|
| ICC | پیوسته | 0-1 | بالای ۰/۷۵ عالی |

پایایی بین ارزیابان (Inter-Rater Reliability)

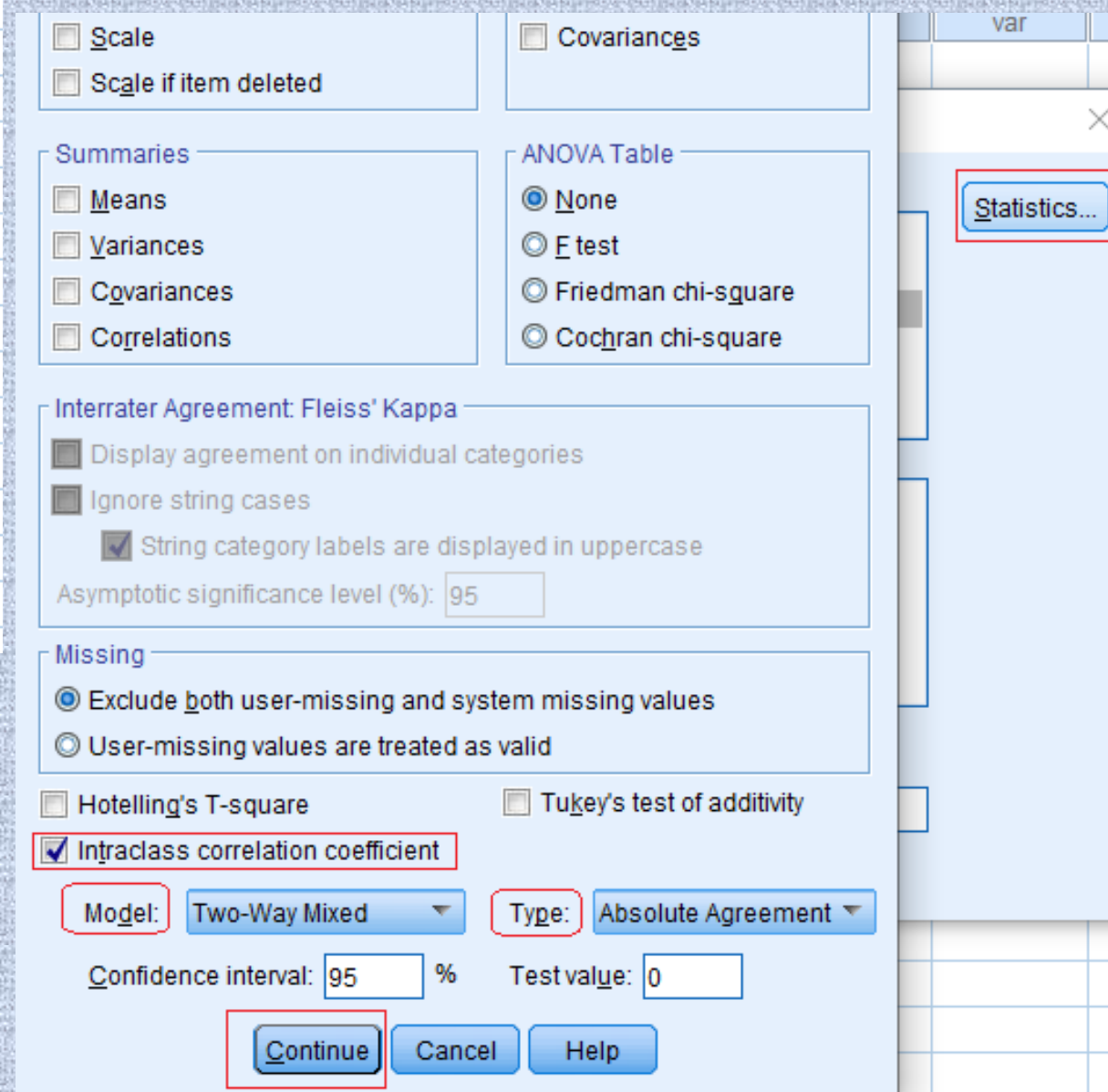
Analyze → Scale → Reliability Analysis



□ مدل

I. Two-Way Mixed ارزیاب‌ها ثابت هستند.

II. One-Way Random نمونه‌ای از ارزیاب‌های تصادفی



پایایی بین ارزیابان (Inter-Rater Reliability)

| Model | Type | هدف مطالعه |
|--------------------------------|--------------------|--|
| One-Way Random | Consistency | ارزیابها تصادفی و می خواهیم تعمیم بدهیم |
| Two-Way Mixed | Consistency | ارزیابها ثابت، پایایی فقط بین همانها مهم است |
| Two-Way Mixed / One-Way Random | Absolute Agreement | دنبال تطابق دقیق نمره ها هستیم |

پایایی بین ارزیابان (Inter-Rater Reliability)

| 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 | .878 ^a | .537 | .985 | 22.667 | 4 | 8 | .000 |
| Average Measures | .956 ^c | .777 | .995 | 22.667 | 4 | 8 | .000 |

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.


b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

| مقدار | تفسیر | شاخص |
|-------|--|------------------|
| 0.878 | پایایی یک ارزیاب منفرد خوب تا عالی است | Single Measures |
| 0.956 | پایایی میانگین چند ارزیاب عالی است | Average Measures |

□ Single Measures: اگر فقط یک ارزیاب داده‌ها را ثبت کند، پایایی بین ارزیابان چقدر است.

□ Average Measures: وقتی میانگین نمرات چند ارزیاب را می‌گیریم، پایایی حتی بالاتر می‌رود.



از توجه شما سپاسگزارم