



روانسنجی ابزار در مطالعات علوم زیستی (بخش دوم)





مدرس: دکتر ندا گیلانی

گروه آمار و اپیدمیولوژی

دانشگاه علوم پزشکی تبریز دانشکده بهداشت،

آدرس ایمیل: neda.gilani@gmail.com



چه خواهیم آموخت؟

مروری بر بخش یک



تحلیل عاملی اکتشافی



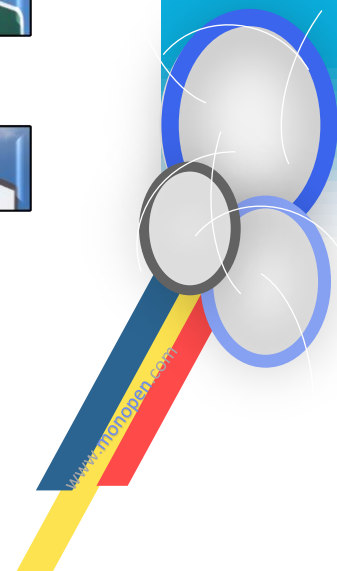
تحلیل عاملی تاییدی



انواع پایایی

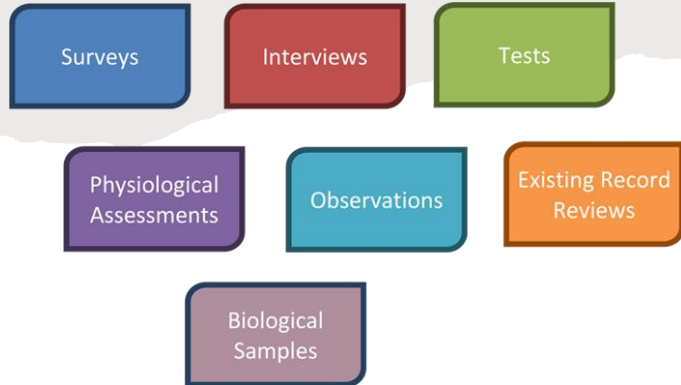


پیش نیازها: آشنایی با مفهوم آزمون های آماری





شیوه‌های جمع‌آوری اطلاعات در پژوهش‌های مختلف



❖ مشاهده (یا معاینه)،

❖ مصاحبه

❖ ارزیابی‌هایی آزمون و پرسشنامه‌ای

اما

قبل از جمع‌آوری اطلاعات هر یک از روش‌های فوق

نیاز است که این سؤال مطرح گردد که

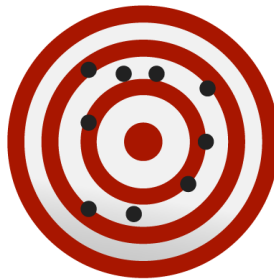
آیا ابزار مورد استفاده روایی و پایایی لازم را دارد یا خیر؟



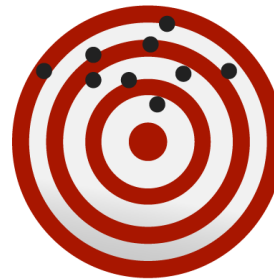
Reliability and Validity



Reliable
Not valid



Low validity
Low reliability



Not reliable
Not valid



Both reliable
and valid

روانشنجی پرسشنامه



اهداف طرح

الف) هدف کلی طرح

روانشنجی پرسشنامه کیفیت زندگی

ب) اهداف اختصاصی طرح

(۱) تعیین روایی صوری و محتوایی پرسشنامه کیفیت زندگی

(۲) تعیین روایی سازه پرسشنامه کیفیت زندگی

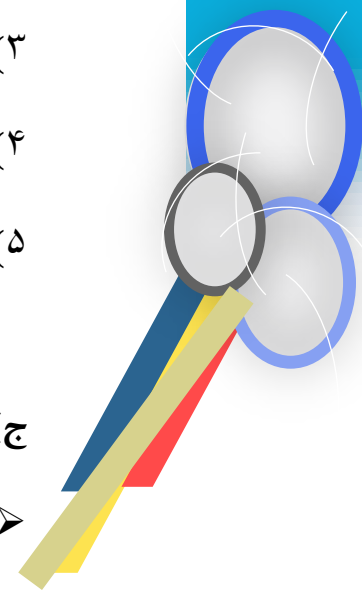
(۳) تعیین روایی ملاکی پرسشنامه کیفیت زندگی

(۴) تعیین پایایی بازآزمایی پرسشنامه کیفیت زندگی

(۵) تعیین پایایی همسانی درونی پرسشنامه کیفیت زندگی

ج) هدف فرعی:

➤ تعیین ارتباط برخی مشخصات فردی-اجتماعی با نمره کیفیت زندگی





روایی (اعتبار) ابزار (Accuracy/Validity)

❖ روایی بیانگر آن است که آیا ابزار قادر به سنجش متغیر یا سازه ای که برای آن ساخته شده است، می باشد یا خیر؟

❖ مطالعه تا چه اندازه ای چیزی را که قصد اندازه گیری آن را داشته است به درستی سنجیده است.

❖ روایی میزان انطباق بین تعریف مفهومی متغیر با تعریف عملیاتی آن است.

❖ روایی آزمون، عبارت است از میزان کارایی آن برای اندازه گیری خصیصه ای که به منظور اندازه گیری آن خصیصه ساخته شده است

❖ روایی نشان دهنده صحت اندازه گیری است

❖ روایی یک موضوع درجه ای است نه یک خصیصه همه یا هیچ

❖ به هر میزان که انواع مختلف روایی یک ابزار مورد سنجش قرار گیرد، شواهد موثرتری جهت اطمینان از صحت ابزار دارد.





انواع روایی ابزار از دید APA، AERA و NCMUE

❖ روایی صوری/ظاهری (Face Validity)

❖ کیفی: دشواری، تناسب و ابهام

❖ کمی: impact score

Impact Score					والدین به سوالات زیر بر اساس طیف لیکرت ۵ گزینه ای (کاملاً موافقم- تا حدودی موافقم- نظری ندارم- تا حدودی مخالفم- کاملاً مخالفم) پاسخ خواهند داد.
اهمیت					
کاملاً مهم	مهم	بطور متوسط مهم	اندکی مهم	اصلاً مهم نیست	
۱	۴	۳	۲	۰	۱- هنگام صحبت کردن با فرزندم احساس راحتی نمی کنم.

$$Impact\ Score = (0.1 \times 5) + (0.4 \times 4) + (0.3 \times 3) + (0.2 \times 2) + (0 \times 1) = 3.4$$

- ❑ در صورتی که امتیاز تاثیر از ۱/۵ بیشتر شود، آیتم برای تحلیل های بعدی مناسب تشخیص داده شده و حفظ خواهد شد.
- ❑ مواردی که امتیاز کمتر از ۱/۵ کسب کرده اند، حذف نخواهند شد و صرفاً مورد بازنگری و اصلاح قرار خواهند گرفت.



انواع روایی ابزار از دید APA، AERA و NCMUE

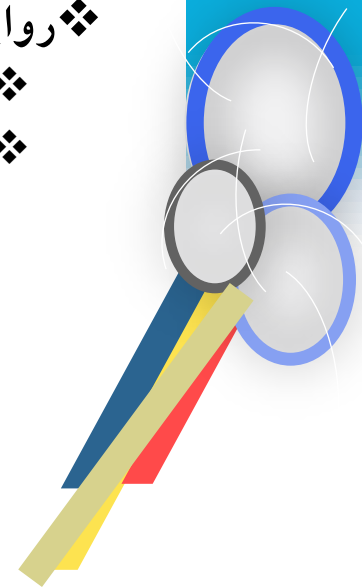
❖ روایی محتوایی (Content Validity)

- ❖ کیفی: ضرورت، اهمیت، واژگان، طیف پاسخدهی، ترتیب سوالات، واژگان، علایم نگارشی و ...
- ❖ کمی: محاسبه دو شاخص نسبت روایی محتوا (Content Validity Ratio (CVR و شاخص روایی محتوا (Content Validity Index (CVI)
- ❖ در CVR ضرورت وجود آیتم از دید متخصصان مورد بررسی قرار می گیرد.
- ❖ در CVI، مرتبط بودن آیتم ها با هدف تحقیق - از دید متخصصان - مد نظر است.

❖ روایی ملاکی / وابسته به معیار (Criterion Validity): پیش بینی رفتار یا توافق بین ارزیابی ها

❖ روایی سازه (Construct Validity)

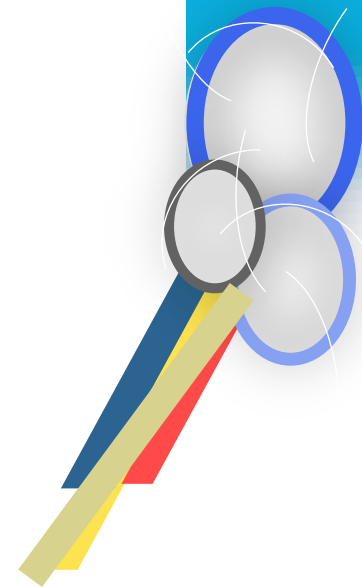
- ❖ روایی همگرا (Convergent)
- ❖ روایی واگرا (Divergent)



نسبت روایی محتوا (CVR)



CVR			والدین به سوالات زیر بر اساس طیف لیکرت ۵ گزینه ای (کاملاً موافقم- تا حدودی موافقم- نظری ندارم- تا حدودی مخالفم- کاملاً مخالفم) پاسخ خواهند داد.
ضرورت			
غیر ضروری	مفید ولی غیر ضروری	ضروری	
۰	۱	۹	۱- هنگام صحبت کردن با فرزندم احساس راحتی نمی کنم.



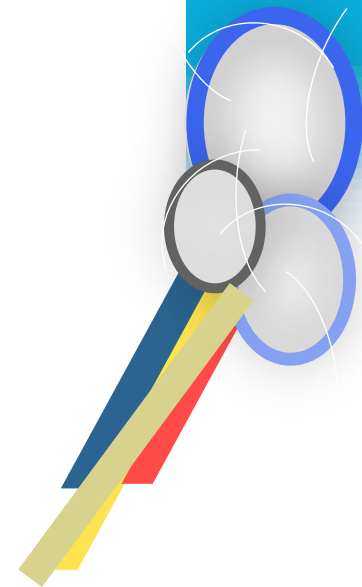


نسبت روایی محتوایی

- بهتر است از تعداد ارزیابان بیشتری (بیشتر از نه نفر) استفاده شود. CVR بر اساس نظر داوران به شرح زیر محاسبه می شود:

$$\text{CVR} = \frac{\text{تعداد کل متخصصین} - \text{تعداد متخصصینی که گزینه ضروری را انتخاب کرده اند}}{2}$$

- مقادیری که به CVR اختصاص می یابد عبارتست از:
- هنگامی که کمتر از نیمی از افراد گزینه "ضروری" را انتخاب می کنند، CVR منفی می شود.
 - هنگامی که نیمی از افراد گزینه "ضروری" و نیمی دیگر گزینه های دیگر را انتخاب کنند، CVR صفر می شود.
 - هنگامی که همه افراد گزینه "ضروری" را انتخاب می کنند، CVR برابر ۱ می شود (که به منظور تبدیل آسان تر به مقدار ۰/۹۹ تعدیل می شود).
 - هنگامی که تعداد افرادی که گزینه "ضروری" را انتخاب می کنند بیشتر از نیم، ولی کل افراد را شامل نمی شوند، CVR بین صفر و ۰/۹۹ بدست می آید.



نسبت روایی محتوایی



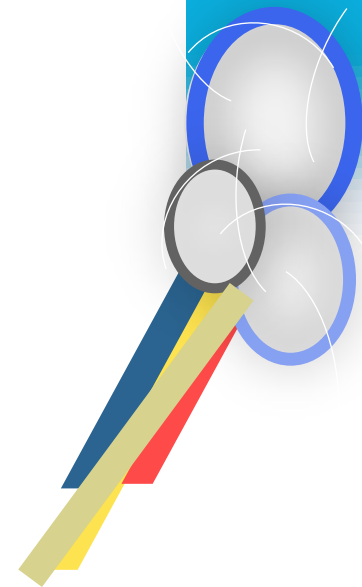
❑ لاوشه جدولی برای تعیین ارزش عددی CVR تهیه کرده است که اصطلاحاً «حداقل ارزش CVR» minimum CVR value نامیده می شود.

❑ نتیجه حاصله پس از مصاحبه باتوجه به تعداد متخصصان با معیار موجود در جدول مقایسه می شود.

❑ در صورتی که عدد حاصله از جدول بزرگتر باشد، حاکی از این است که وجود آیتم مربوطه با سطح معناداری آماری قابل قبولی (۰.۰۵) در این ابزار ضروری و مهم است.

تعداد ارزیابان	حداقل مقدار CVR
5	0.99
6	0.99
7	0.99
8	0.75
9	0.78
10	0.62
11	0.59
12	0.56

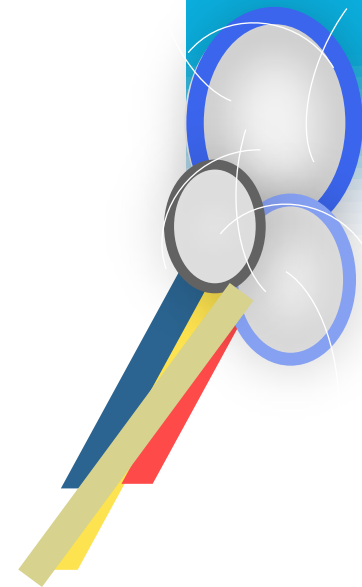
تعداد ارزیابان	حداقل مقدار CVR
13	0.54
14	0.51
15	0.49
20	0.42
25	0.37
30	0.33
35	0.31
40	0.29



شاخص روایی محتوایی



❖ پس از تعیین و محاسبه CVR، میتوان شاخص CVI را با استفاده از روش والتز و باسل (Waltz & Bausell) محاسبه نمود.



شاخص روایی محتوایی (CVI)



❖ برای اولین بار توسط والتز و باسل (Waltz and Bausell, 1983) معرفی شد.

❖ حداقل ۵ نفر و حداکثر ۱۰ نفر متخصص

❖ + غیر مرتبط

❖ + نیاز به بازبینی اساسی

❖ + مرتبط اما نیاز به بازبینی

❖ + کاملاً مرتبط

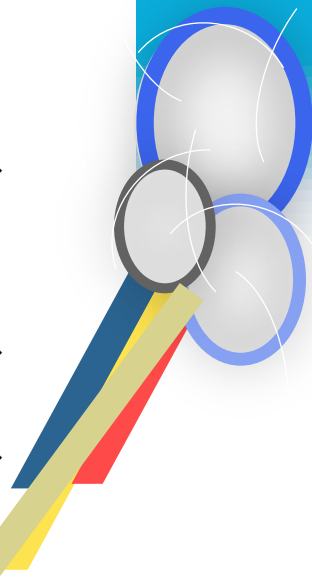
❖ سپس با استفاده از فرمول CVI شاخص روایی محتوا محاسبه میگردد

$$CVI = \frac{\text{تعداد به آیتم نمره 3 و 4 داده اند}}{\text{تعداد کل ارزیابان}}$$

❖ معمولاً ۵ معیار ابهام (Ambiguity)، سادگی (Simplicity)، واضح بودن (Clarity)، اختصاصی بودن (Specificity) و مرتبط بودن (Relevancy) بر اساس طیف لیکرت چهار قسمتی مورد احصا قرار می گیرد.

❖ در سالهای بعد این ۵ معیار به سه معیار وضوح، سادگی و مرتبط بودن تقلیل یافت.

❖ طی سالهای اخیر بر اساس دیدگاه نوین، منحصراً شاخص مرتبط بودن جهت احصا CVI مد نظر قرار می گیرد و دو معیار دیگر به صورت کیفی از منحصراً مورد پرسش و ارزیابی قرار خواهد گرفت.



شاخص روایی محتوایی (CVI)



نمره CVI

❖ بالاتر از ۰/۷۹ مناسب (Adequate) تشخیص داده می شود.

❖ بین ۰/۷۰ تا ۰/۷۹ سوال برانگیز (Questionable) بوده و به اصلاح و بازنگری نیاز دارد.

❖ کمتر از ۰/۷۰ غیر قابل قبول (Unacceptable) بوده و بایستی حذف شود.

CVI								بررسی اعتبار محتوی
سادگی Simplicity				مربوط بودن Relevancy				
کاملا ساده	ساده اما نیاز به بازبینی	نیاز به بازبینی جدی	ساده نیست	کاملا مرتبط	مرتبط اما نیاز به بازبینی	نیاز به بازبینی جدی	مرتبط نیست	
				۵	۲	۱	۰	گویه ها نحوه پاسخ به سوالات به صورت می باشد
								۱- هنگام صحبت کردن با فرزندم احساس راحتی نمی کنم.





روایی ملاکی / معیار / ضابطه

❖ شاخص کارآمدی آزمون در پیش بینی رفتار یک فرد در موقعیت های خاص است.

❖ درجه ای از توافق بین یک تست با یک یا تعداد بیشتری تست معتبر دیگر در همان زمینه (که بهتر است دقیق و ایده آل باشد) است که معمولاً با استفاده از ضریب همبستگی بین آنها سنجیده میشود.

❖ روایی ملاکی بر دو نوع است: هم زمان و پیش بین.

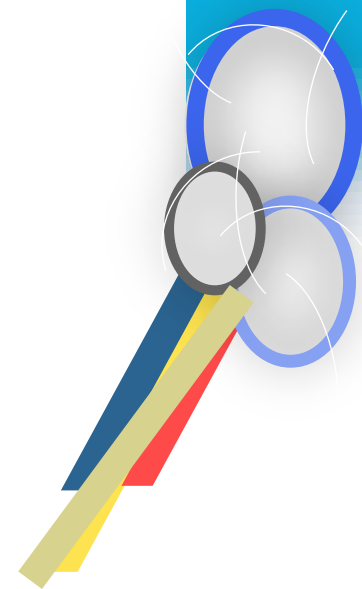
❖ روایی هم زمان

❖ روایی یک معیار برای سنجش یک پدیده خاص در زمان حاضر. زمانی برقرار میشود که یک مقیاس، افرادی را که اساساً با هم تفاوت دارند از یکدیگر تفکیک کند. عبارتی این افراد نمرات متفاوتی در آزمون کسب کنند. (سنجش عملکرد)

❖ روایی هم زمان در مواردی به کار می رود که داده های حاصله از دو اندازه گیری در یک زمان در دسترس باشد.

❖ روایی هم زمان هنگامی برآورد میشود که معیار (استاندارد) در حال حاضر وجود دارد (به عنوان مثال میتوان در یک بررسی از کارمندان میزان حقوق دریافتی را پرسید و به عنوان یک ضابطه از مدارک حسابداری به عنوان ضابطه استفاده نمود و همبستگی مقدار حقوق گزارش شده با مقدار واقعی آن بر اساس مدارک موجود را محاسبه کرد.).

❖ در مواردی محاسبه می شود که هدف جانشین کردن این ابزار اندازه گیری به جای ابزار دیگری باشد.



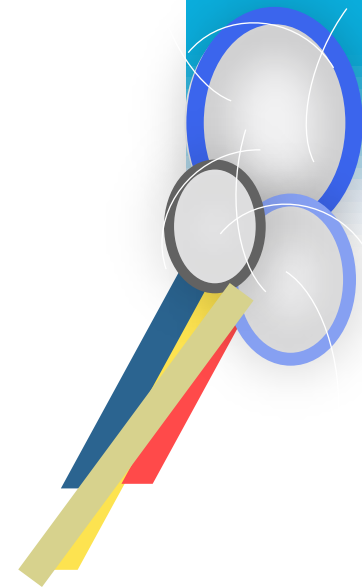


روایی ملاکی / معیار / ضابطه

❖ روایی پیش بین

❖ به استعداد معرف در پیش بینی حوادث آینده مربوط می شود. مثلاً اطلاعات حاصل از اعتبار پیش بین بیشتر در زمینه گزینش افراد و طبقه بندی کارکنان در مساله مورد نظر می باشد.

❖ هنگامی کاربرد دارد که معیار در حال حاضر وجود ندارد و در زمان آینده اتفاق میافتد. هدف آن است تا بدانیم یک مقیاس یا ابزار تا چه اندازه میتواند پیامدی را در آینده پیشبینی کند، به عنوان مثال بر اساس نمره میان ترم دانشجویان قبولی و یا رد شدن آنها را در پایان ترم پیشبینی کنیم.

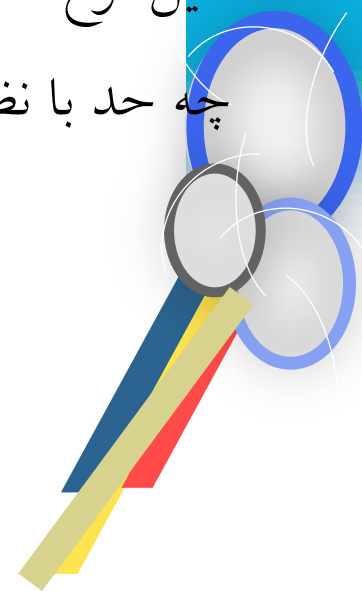


روایی سازه



❖ روایی محتوایی این موضوع را بررسی می کند که آیا محتوای مقیاس مورد نظر قابلیت اندازه گیری هدف را دارد یا خیر؟ ولی روایی سازه ای این موضوع را بررسی می کند که آیا احزای مقیاس مورد نظر توانایی تشکیل مقیاس را دارند یا برخی از آنها نامرتبط اند.

❖ این نوع اعتبار بیان می کند که نتایج بدست آمده از کاربرد یک ابزار اندازه گیری تا چه حد با نظریه هایی که آزمون بر محور آنها تدوین شده است، تناسب دارد.





راه های مختلف ارزیابی روایی سازه

همبستگی

❖ روایی همگرا

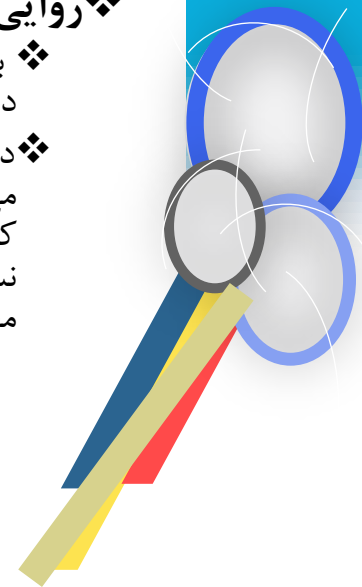
❖ همگرایی (همسو بودن یا همبستگی) آیتم‌های تشکیل دهنده یک مقیاس

❖ همبستگی بین مقیاس مورد بررسی با ابزارهای مشابهی که برای اندازه‌گیری، سازه‌ی مورد نظر (توسط سایر مؤلفین) طراحی شده‌اند. معمولاً بحث در این نحوه‌ی ارزیابی سازه به بحث روایی همزمان در روایی ملاکی مطرح می‌شود. در صورت وجود همبستگی قوی بین سازه‌ی هدف و سازه‌های موجود دیگر، نیازی به تعریف سازه‌ی جدید نیست مگر آن که نسبت به سازه‌های موجود کوتاه‌تر باشد و یا برای آن مفهوم ساده‌تر باشد.

❖ روایی واگرا (تشخیصی/افتراقی)

❖ به این موضوع اشاره دارد که آیتم‌های تشکیل دهنده‌ی دو مقیاس متفاوت نباید به گونه‌ای با یکدیگر همبستگی داشته باشند که بتوان گفت موضوع واحدی را اندازه می‌گیرند.

❖ در صورت وجود یک مقیاس ملاک، در استفاده از این روش، همبستگی مقیاس مورد نظر با مقیاس ملاک محاسبه می‌گردد. سپس همبستگی مقیاس دیگری که غیر مرتبط است با مقیاس ملاک محاسبه می‌شود. در نهایت در صورتی که مقیاس مورد نظر با مقیاس ملاک همبستگی بیشتری داشته باشد، روایی افتراقی آن تأیید می‌شود. ولی این آزمون نسبت به آزمون اول ضعیف‌تر است. همچنین برای دقیق‌تر نمودن روش اخیر، می‌توان همبستگی زیر مقیاس‌های مقیاس مورد بررسی را با مقیاس ملاک یا یک مقیاس مرتبط بررسی نمود.





راه های مختلف ارزیابی روایی سازه

❑ تمایز سنی

❖ (اگر سن یک عامل مهم ویژگی مورد اندازه گیری باشد) مانند آزمون هوش و کسلر، آنگاه نمرات آن آزمون باید با بالا رفتن سن، پیشرفت نشان دهد).

❑ تحلیل خوشه ای

❑ تحلیل عاملی

❖ تحلیل عاملی اکتشافی (Exploratory factor analysis)

❖ تحلیل عاملی تاییدی (Confirmatory factor analysis)

❑ محاسبه همبستگی نمرات آزمون های فرعی با نمره کل آزمون (توجه: خرده آزمون ها نباید بایکدیگر همبستگی زیادی داشته باشند، زیرا در چنین حالتی همه ی آنها یک چیز را اندازه گیری می کنند). Item total correlation بیشتر در مقالات روانشناسی

❑ متوسط واریانس بیان شده (Average Variance Explained: AVE) زمانی که مقدار این شاخص حداقل ۰/۵ باشد می توان روایی همگرایی یک سازه را تأیید نمود.



تحلیل عاملی اکتشافی (EFA)



□ عمدتاً در پرسشنامه

□ هدف بررسی روایی سازه

□ هدف ۱: شناسایی دسته بندی سوالات در چند مفهوم (فاکتور)

□ هدف ۲: پیش بینی فاکتورها و تبیین پیشگویی مقیاس کل (درصد واریانس تبیینی)

□ جنبه اکتشافی

□ متغیر سوالات (وابسته) ترکیب خطی از فاکتورها (متغیر مستقل): سوالات تحت تاثیر کدام فاکتورها هستند. افسردگی به عنوان منشا و سوز وجود دارد که وجود افسردگی نشانه هایی را از خود بجا می گذارد.

□ در تحلیل مولفه های اصلی (PCA) برعکس است. مثلاً وضعیت اقتصادی اجتماعی.

$$\square SES = \alpha_1 Income + \alpha_2 Education + \alpha_3 Occupation + \dots$$

□ سازه، فاکتور، مقیاس، بعد، حوزه، مفهوم، latent variable

□ ایتم، سوال

□ بر خلاف تحلیل هایی مثل تی تست، در این تحلیل می بایست صبور بود و شاید ۲۰ بار اجرا کرد تا به نتیجه مطلوب رسید.

□ داده های مورد استفاده

□ مجموعه داده (1939) Holzinger and Swineford

□ سنجش توانایی ذهنی

□ ۳۰۱ دانش آموز کودکان ۷ و ۹ ساله

□ نسخه اصلی شامل ۲۶ سوال در حالی که نسخه کوتاه شده دارای ۹ سوال



مراحل انجام EFA



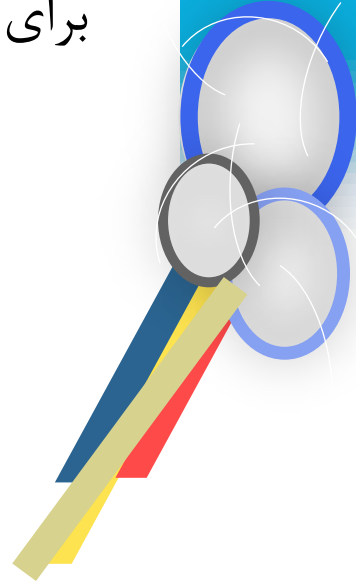
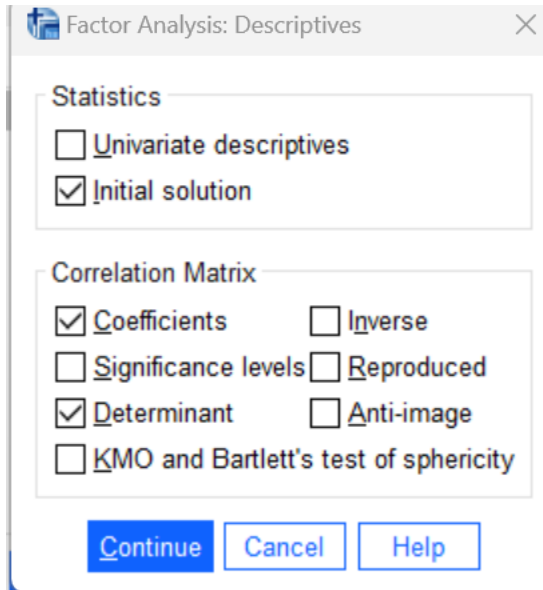
۱. انتخاب آیتم های مناسب برای تحلیل عاملی

■ از منوی descriptive و انتخاب determinant برای آزمون همخطی میان سوالات استفاده میشود.

■ به گونه ای که دترمینان ماتریس R میبایست بزرگتر از ۰.۰۰۰۰۱ باشد. در صورتی که برقرار نباشد در ماتریس R سوالاتی که همبستگی بالای 0.8 دارند را یکی را حذف میکنند.

■ اگر تعداد آیتم هایی که همبستگی بالای ۰/۳ دارند خیلی کم است، اتلاف وقت

برای ادامه تحلیل است (Tabachnick & Fidell (2001)



مراحل انجام EFA

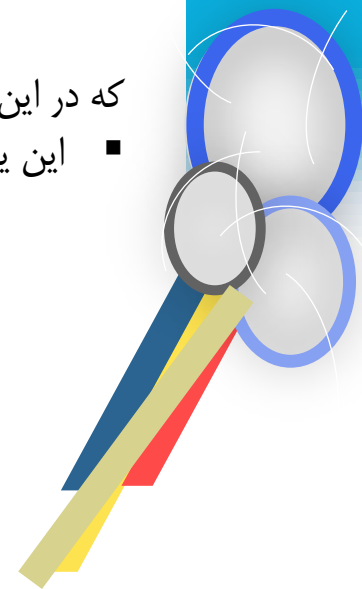


۲. کفایت نمونه گیری

- تشخیص مناسب بودن داده ها برای تحلیل عاملی با شاخص KMO (Kaiser-Meyer-Olkin)
- مقدار آماره ی این آزمون همواره بین ۰ و ۱ تغییر می کند. هر چه بزرگتر بهتر

$$KMO = \frac{\sum \sum r_{ij}^2}{\sum \sum r_{ij}^2 + \sum \sum a_{ij}^2}$$

- که در این رابطه r_{ij} ضریب همبستگی بین سوالات i و j و a_{ij} ضریب همبستگی جزئی بین آنهاست.
- این یک شاخص کلی است.



مراحل انجام EFA



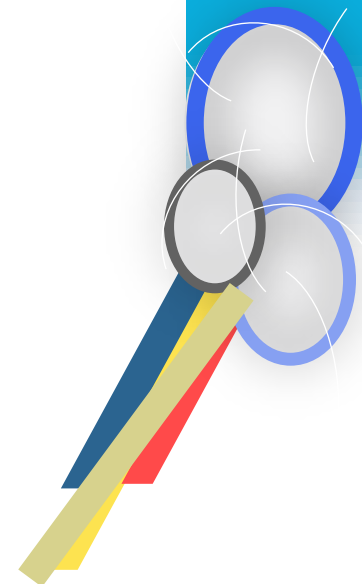
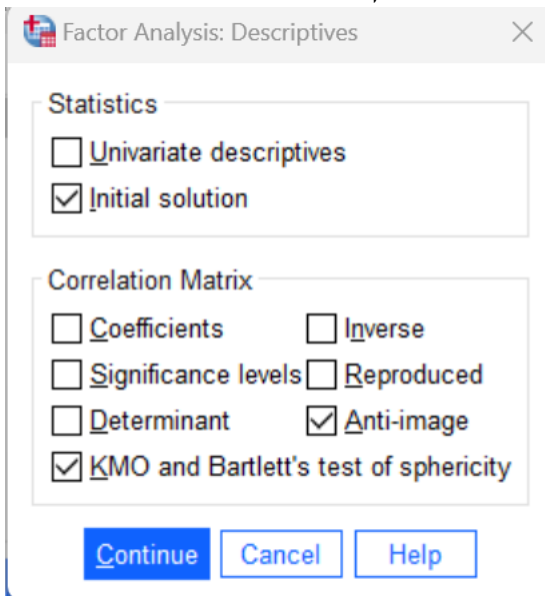
۲. کفایت نمونه گیری

■ ملاک ارزیابی

جدول ۱- قضاوت در مورد ضریب KMO

مقدار KMO	تناسب داده‌ها برای تحلیل عاملی
بزرگتر یا مساوی ۰/۹۰	عالی
۰/۸۰-۰/۸۹	خیلی خوب
۰/۷۰-۰/۷۹	خوب
۰/۶۰-۰/۶۹	متوسط
۰/۵۰-۰/۵۹	ضعیف
کمتر از ۰/۵۰	غیرقابل پذیرش

■ همبستگی روی قطر اصلی ماتریس anti-image برای هر سوال میبایست بالای نیم باشد.



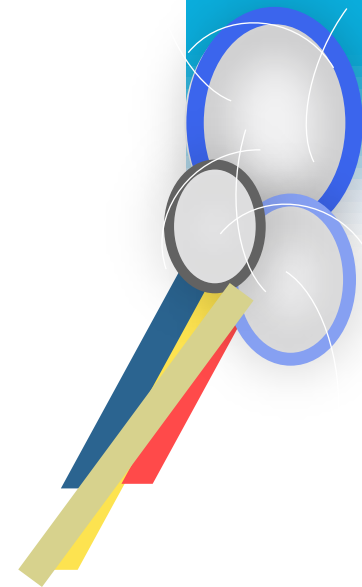
مراحل انجام EFA



۲. کفایت نمونه گیری

■ آزمون بارتلت

- از آنجایی که برای معنا دار بودن یک مدل عاملی لازم است متغیرها همبسته باشند، برای اطمینان از اینکه ماتریس همبستگی هایی که پایه ی تحلیل عاملی قرار می گیرد در جامعه برابر با صفر نیست بایستی از آزمون بارتلت استفاده نمود. آزمون بارتلت این فرضیه را که ماتریس همبستگی های مشاهده شده متعلق به جامعه ای با متغیرهای ناهمبسته است می آزماید. اگر این فرضیه که متغیرها باهم رابطه ندارند رد نشود، کاربرد تحلیل عاملی زیر سؤال خواهد رفت.

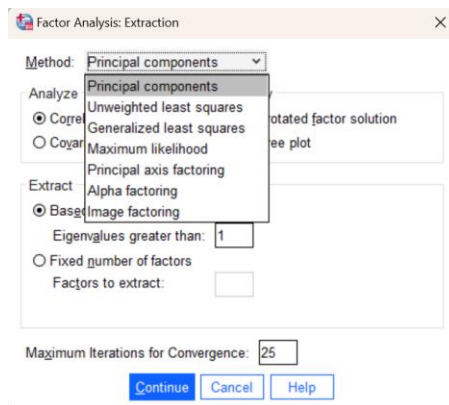


مراحل انجام EFA



۳. تصمیم در خصوص روش استخراج فاکتورها

▪ هفت روش استخراجی در SPSS



- روشهای متداول: **Principal component method** (اغماض خطا در برآوردها و **over estimate** شدن بار عاملی (**loading**) و بیشتر شدن درصد پیش بینی
- **Principal axis factoring**: مناسب ترین روش و مقدار واقعی برآورد ها را بدست می دهد چرا که خطا را نیز در بر می گیرد و مبتنی بر تکرار است.
- در صورتی که در **PAF** با وجود افزایش تکرار (۹۹۹۹) به جواب نرسیدیم از **ML** یا **PCM**

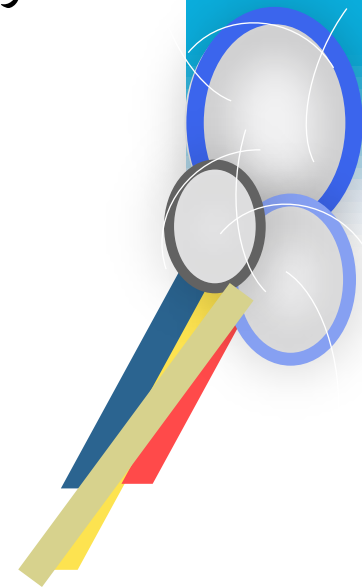


مراحل انجام EFA



۴. تصمیم گیری در خصوص تعداد فاکتورها

- مقادیر ویژه بزرگتر از یک برای ماتریس همبستگی بعنوان ورودی و بزرگتر از میانگین برای ماتریس واریانس - کوواریانس
- واریانس تبیینی مشخص
- نمودار شن ریز
- آزمون معنی داری بیشینه درستمایی درستنمایی

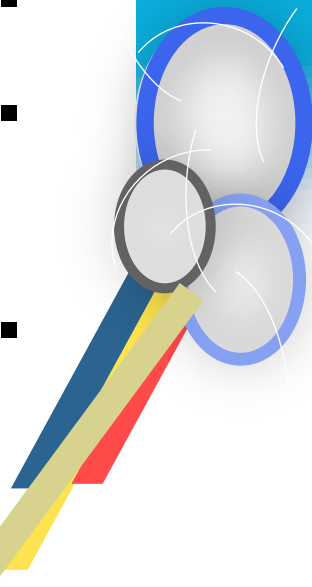


مراحل انجام EFA



۵. تصمیم در خصوص دوران (Rotation)

- یک اصل اساسی
- واضح نمودن ارتباط بین آیتم و عامل (ساده کردن تفسیر ارتباط)
- یکتا (unique) نمودن بار عاملی (loading)
- اشکال دوران
- متعامد (Orthogonal): متعامد (مستقل) شدن عامل ها مانند دوران varimax که مخفف variance maximum است. واریانس بارها را بیشینه می سازد.
- مایل ((oblique (nonorthogonal): ارتباط عامل ها با هم (خروجی همبستگی بالای ۰/۳ و توجیه نظری آن). متداولترین **Pattern matrix Oblimin** به توصیه رنچر پیشنهاد: در ابتدا دوران مایل انتخاب و اگر همبستگی زیر ۰/۳ و یا توجیه وجود نداشت متعامد

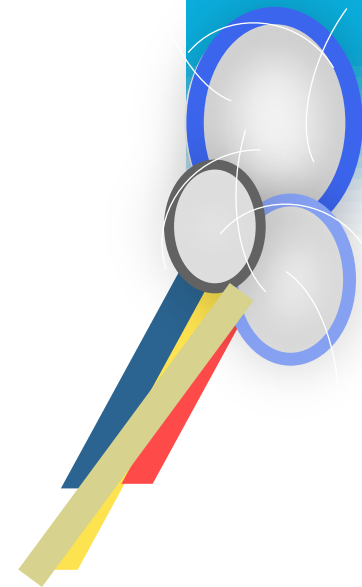


مراحل انجام EFA



۶. نام گذاری فاکتورها

- نام گذاری بر اساس آیت‌های هر فاکتور با نظر متخصصین رشته مربوطه
- در صورت عدم نامگذاری تحلیل عاملی اکتشافی مجدد





مراحل انجام EFA

۷. محاسبه امتیاز (score) برای عامل ها

▪ روش های مختلف محاسبه مثل رگرسیون (امتیازهای وابسته حتی با وجود متعامد بودن)، بارتلت (متعامد شدن امتیازها)، اندرسون روبین (تعمیم یافته بارتلت) میانگین صفر و واریانس یک و

▪ موارد مورد استفاده

▪ توصیفی: مقایسه امتیاز عامل ها برای هر فرد (هر آزمودنی تحت تاثیر کدام فاکتور قرار می گیرد). ویژگی های کدام فاکتور در فرد بیشتر هست.

▪ ورودی به عنوان متغیر برای تحلیل های جدیدتر مانند انواع رگرسیون ها

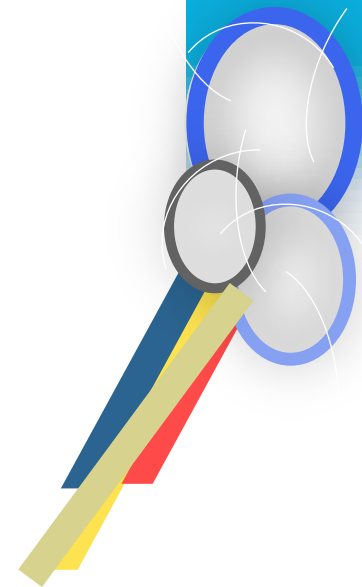


مراحل انجام EFA



۸. اعتبار سنجی (Validation)

- روش هایی مانند cross validation
- مثلاً نمونه ۴۰۰ تایی (۲۰۰ نمونه آیا مدل قبلی را ساپورت می کند؟)
- در صورت تایید می توان داده ها را تلفیق کرد و مدل نهایی را ارائه نمود.



مراحل انجام EFA در SPSS



❑ مبتنی بر کل پرسشنامه (Questionnaire based)

❑ وارد نمودن تمامی آیتم های پرسشنامه

❑ بررسی همزمان همگرایی و افتراق آیتم ها

❑ مبتنی بر سازه (Scale based)

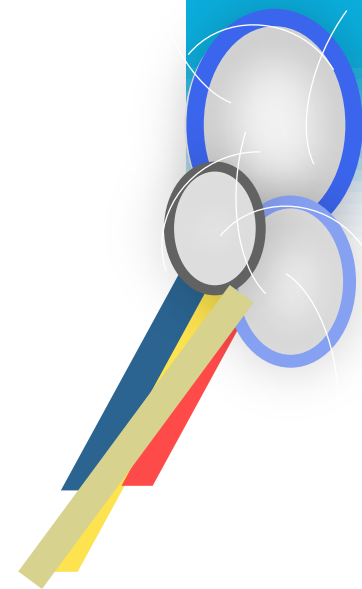
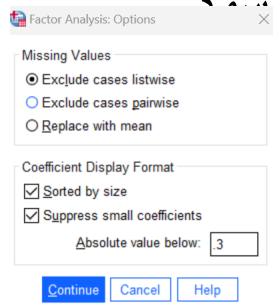
❑ وارد نمودن صرفا آیتم های یک بعد

❑ بررسی همگرایی

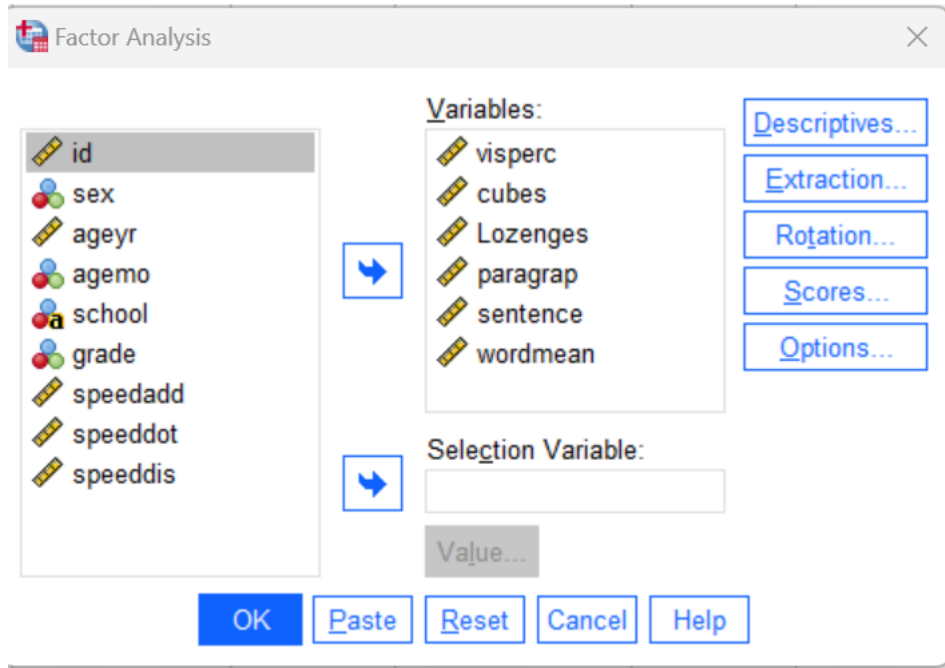
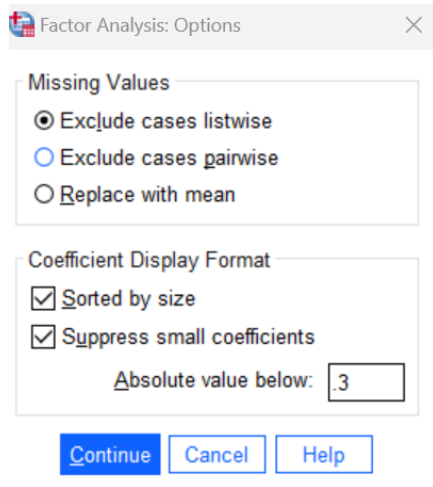
❑ بررسی واگرایی با استفاده از روشهایی همچون تحلیل گروه های شناخته شده

Kwon group analysis مثل تی تست و ...

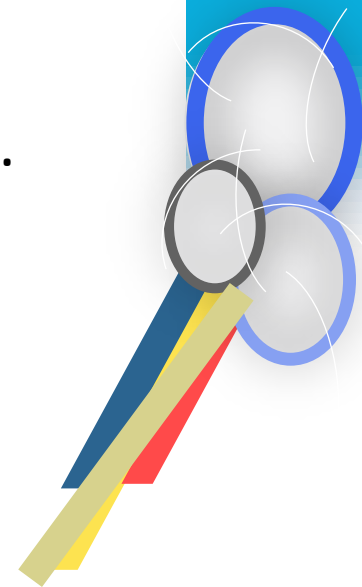
❑ اگر سوالی واریانس صفر داشته باشد کل تحلیل انجام نمی ...



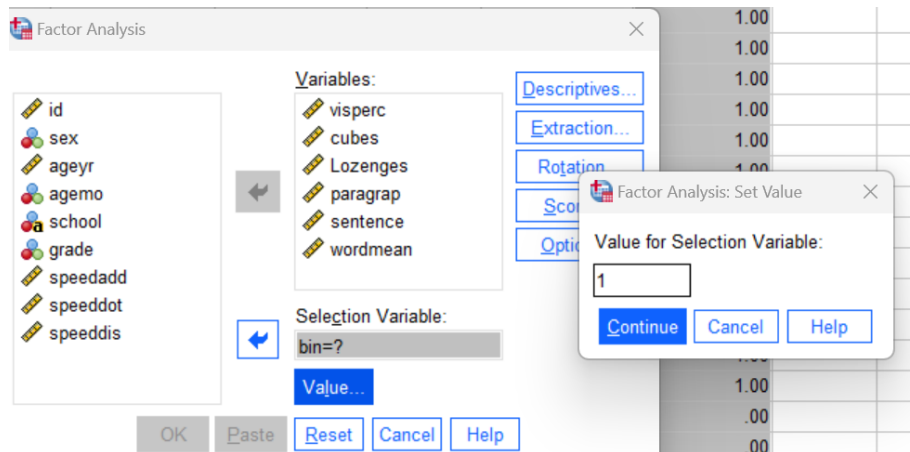
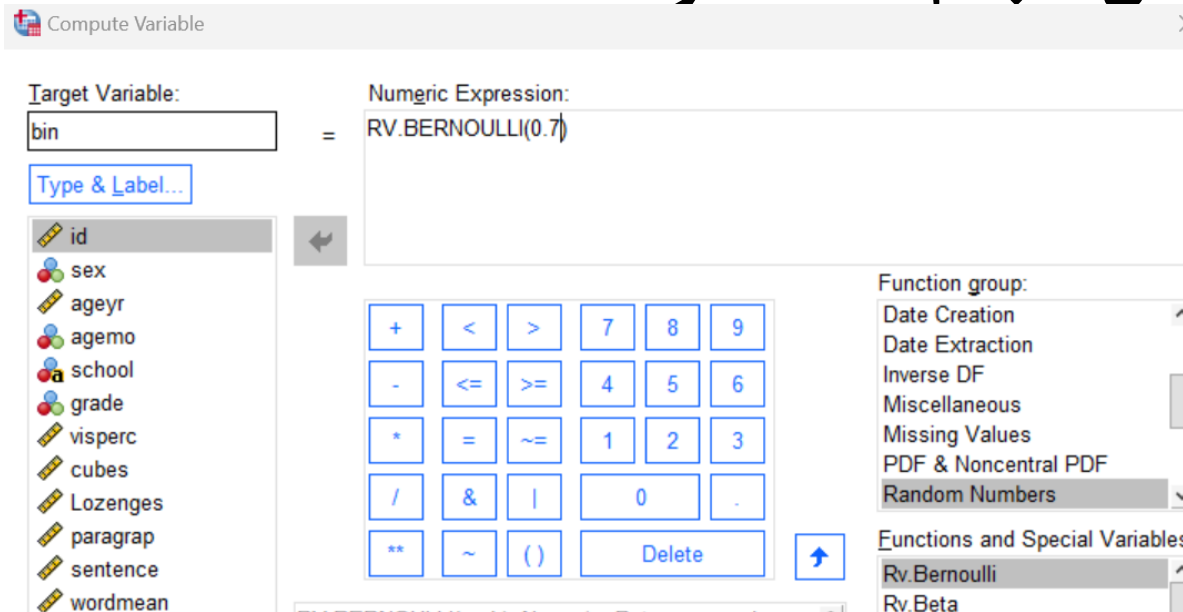
مراحل انجام EFA در SPSS



❑ Transform...Bin....



مراحل انجام EFA در SPSS





روایی افتراقی

۱) روش متوسط واریانس بیان شده (AVE)

این روش، یک روش جایگزین مبتنی بر تحلیل عاملی است که توسط فورنل و ولاکر (۱۹۸۱) ارائه شد. در این روش زمانی بین دو سازه افتراق وجود دارد که متوسط واریانس بیان شده برای یک سازه بزرگتر از واریانس مشترک بین آنها باشد. به عبارت دیگر، مجذور متوسط واریانس بیان شده برای یک سازه باید از مقدار مطلق همبستگی بین این مقیاس و مقیاس حاضر در مطالعه بزرگتر باشد. این همبستگی باید برای مقیاس‌های استاندارد شده به دست آید.

مجموع توان دوم بارهای عاملی مربوط به آیتم‌های سازنده‌ی مقیاس S_1

مجموع مربوط به تمام آیتم‌ها برای شاخص S_2 = [توان دوم بارهای عاملی] - ۱

$$AVE = \frac{S_1}{S_1 + S_2}$$

برای ارائه‌ی این شاخص، از یک ماتریس 2×2 استفاده می‌شود که در آن توان دوم کواریانس بین هر مقیاس با سایر مقیاس‌ها، ارائه می‌شود ولی در این ماتریس عناصر روی قطر اصلی با AVE جایگزین می‌شوند. زمانی روایی افتراقی تأیید می‌شود که هر یک از عناصر روی قطر اصلی این ماتریس نسبت به هر مؤلفه دیگر در سطر یا ستون متناظر بزرگتر باشد.

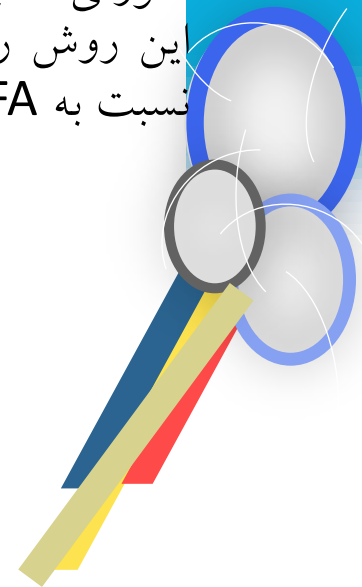
روایی افتراقی



۴) روش تحلیل عاملی تأییدی (Confirmatory Factor Analysis (CFA):

در این روش از تحلیل عاملی تأییدی بدین صورت استفاده می‌شود که اگر به ازای مسیرهای تعریف شده مبتنی بر مبانی نظری برای ارتباط بین آیتم‌ها و سازه‌ی تشکیل دهنده‌ی آن، شاخص‌های برازش مدل مقادیر مناسبی داشته باشند، روایی افتراقی سازه‌های تعریف شده در این تحلیل، تأیید می‌گردد. البته می‌توان با توجه به این موضوع که آیتم‌های تعریف شده در یک مقیاس ضرایب معنی‌دار و قابل ملاحظه‌ای داشته باشند، روایی همگرایی سازه‌ی مورد نظر را نیز تأیید نمود.

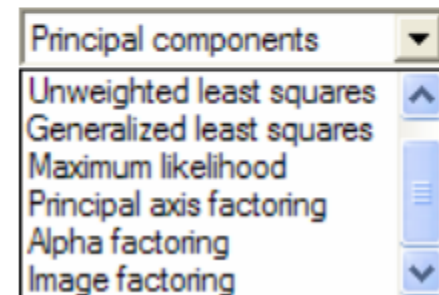
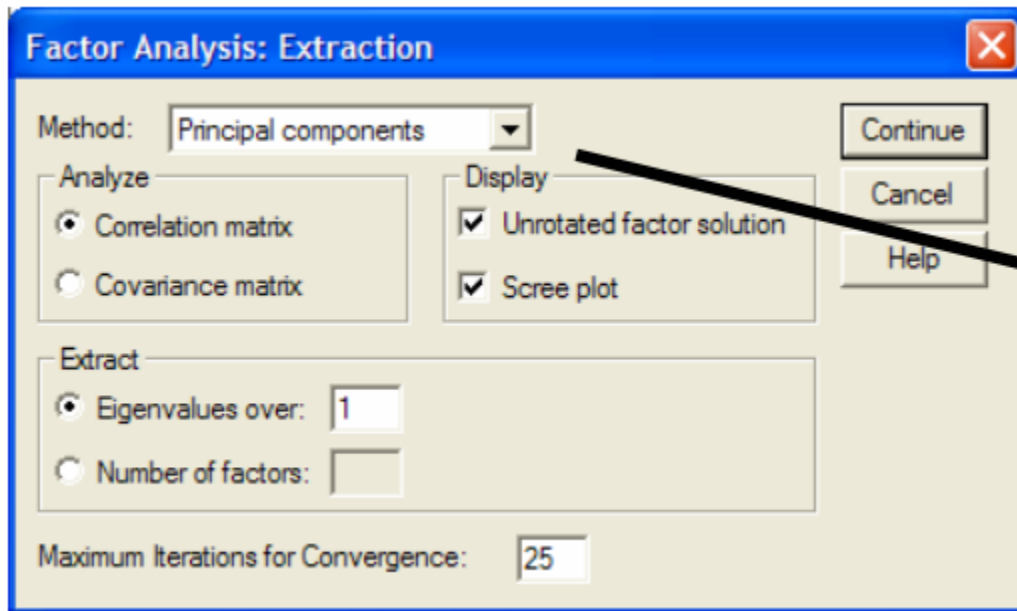
استراتژی دیگری که در این روش می‌توان استفاده کرد این است که ابتدا همبستگی بین مقیاس‌ها (سازه‌ها) را در تحلیل آزاد گذاشت تا مدل آن را برآورد کند، سپس در اجرای بعدی، همبستگی بین مقیاس‌ها را روی عدد یک ثابت نمود و سپس این دو مدل را با شاخص کای دو، مقایسه نمود. در صورتی که بین این دو مدل تفاوت معنی‌داری وجود نداشته باشد، روایی افتراقی مقیاس‌ها تأیید نمی‌شود. این روش را باید به ازای هر دو مقیاس حاضر در مطالعه به صورت زوج گونه انجام داد. این روش نسبت به CFA معمولی و همچنین نسبت به روش AVE دقیق‌تر عمل می‌کند.



تحلیل عاملی در spss



روش انجام تحلیل عاملی

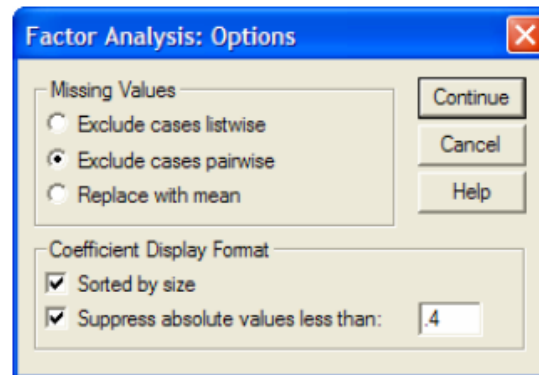


تحلیل عاملی در spss



منوی option

Click on **Options...** in the main dialog box. By default SPSS will list variables in the order in which they are entered into the data editor. Although this format is often convenient, when interpreting factors it can be useful to list variables by size. By selecting *Sorted by size*, SPSS will order the variables by their factor loadings. There is also the option to *Suppress absolute values less than* a specified value (by default 0.1). This option ensures that factor loadings within ± 0.1 are not displayed in the output. This option is useful for assisting in interpretation; however, it can be helpful to increase the default value of 0.1 to either 0.4 or a value reflecting the expected value of a significant factor loading given the sample size (see Field section 15.3.6.2.). For this example set the value at 0.4.





تحلیل عاملی در spss

حجم نمونه

یک سوال بسیار با اهمیت در تحلیل عاملی تعیین حداقل حجم نمونه است. (کلاین، ۱۹۹۰) تعیین حداقل حجم نمونه لازم برای گردآوری داده‌های مربوط به مدل‌یابی معادلات ساختاری بسیار با اهمیت است. (مک‌کیتی، ۲۰۰۴) با وجود آنکه در مورد حجم نمونه لازم برای تحلیل عاملی و مدل‌های ساختاری توافق کلی وجود ندارد (شریبر، ۲۰۰۶)، اما به زعم بسیاری از پژوهشگران حداقل حجم نمونه لازم ۲۰۰ نمونه می‌باشد. (هولتر، ۱۹۸۳؛ گارور و منتزر، ۱۹۹۹؛ سیوو و همکاران، ۲۰۰۶؛ هو، ۲۰۰۸).

کلاین نیز معتقد است در تحلیل عاملی اکتشافی برای هر متغیر ۱۰ یا ۲۰ نمونه لازم است اما حداقل حجم نمونه ۲۰۰ قابل دفاع است. اما در تحلیل عاملی تأییدی حداقل حجم نمونه براساس عامل‌ها تعیین می‌شود نه متغیرها. اگر از مدل‌یابی معادلات ساختاری استفاده شود حدود ۲۰ نمونه برای هر عامل (متغیر پنهان) لازم است. (جکسون، ۲۰۰۳) حجم نمونه توصیه شده برای تحلیل عاملی تأییدی حدود ۲۰۰ نمونه برای ده عامل توصیه شده است. (شه و گلداشتاین، ۲۰۰۶، کلاین، ۲۰۱۰)

نکته) برای اعتبار سنجی پرسشنامه. انجام تحلیل عاملی روی آن حداقل حجم نمونه مورد نیاز ۲۰۰ مورد است (کلانتری ۱۳۹۱، کلاین ۲۰۱۰). تحلیل اکتشافی نیازمند نمونه‌هایی با حجم بسیار زیاد می‌باشد.

البته (Field (2005) بیان میدارد که حجم نمونه ۳۰۰ بخوبی کفایت می نماید (Field, A.P. Discovering statistics using spss).

اگر حجم نمونه کم باشد از PLS و معادلات ساختاری بر مبنای واریانس استفاده کرد.

تحلیل عاملی در spss



روش استخراج عامل ها

علاوه بر انتخاب مدل تحلیل، محقق باید مشخص کند که عامل ها چگونه باید استخراج شوند. دو روش کلی برای این کار وجود دارد، روش عامل های متعامد و روش عامل های متمایل.

انتخاب بین چرخش متعامد یا متمایل به نظر محقق بستگی دارد. اینکه او معتقد است عوامل میبایست مستقل از هم باشند یا خیر. اگر معتقد است که مستقل هستند از چرخشهای متعامد استفاده میکنیم. برای تحلیلهای اکتشافی متعامد را بکار میبریم چون معمولاً محقق فرضیه ای در خصوص همبستگی میان عاملها ندارد.

برای دستیابی به یک ساختار ساده، اگر راه حل متعامد انتخاب شود، روش **Varimax** و در صورت انتخاب راه حل متمایل، چرخش **Direct Oblimin** مناسبتر است.

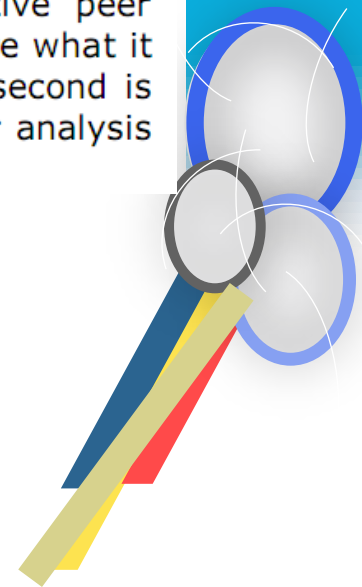


تحليل عاملی در spss



Interpretation

The next step is to look at the content of questions that load onto the same factor to try to identify common themes. If the mathematical factor produced by the analysis represents some real-world construct then common themes among highly loading questions can help us identify what the construct might be. The questions that load highly on factor 1 seem to all relate to using computers or SPSS. Therefore we might label this factor *fear of computers*. The questions that load highly on factor 2 all seem to relate to different aspects of statistics; therefore, we might label this factor *fear of statistics*. The three questions that load highly on factor 3 all seem to relate to mathematics; therefore, we might label this factor *fear of mathematics*. Finally, the questions that load highly on factor 4 all contain some component of social evaluation from friends; therefore, we might label this factor *peer evaluation*. This analysis seems to reveal that the initial questionnaire, in reality, is composed of four sub-scales: fear of computers, fear of statistics, fear of maths, and fear of negative peer evaluation. There are two possibilities here. The first is that the SAQ failed to measure what it set out to (namely SPSS anxiety) but does measure some related constructs. The second is that these four constructs are sub-components of SPSS anxiety; however, the factor analysis does not indicate which of these possibilities is true.



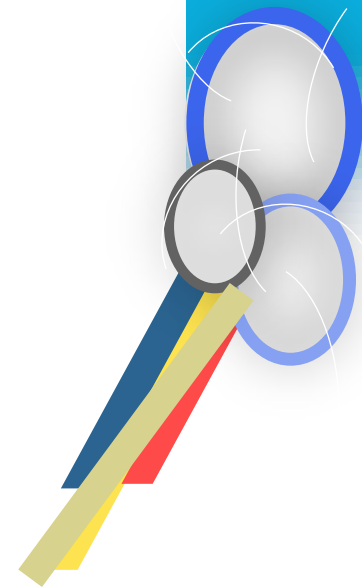
آلتمن بلاند نمودار



سنجش روایی

Tukey Mean-Difference Plot

plot of the average of the two measures (on the x-axis) versus the y-axis). the difference between the two measures (on



پایایی



- ❑ سنجه ای دارای پایایی نباشد نمی تواند مقدار درست و واقعی یک پدیده را نشان دهد. بنابراین در اعتبار سنجی ابزارها، سنجش روایی باید پس از تایید پایایی صورت گیرد.
- ❑ از بعد روش شناختی، روایی با حقیقت در ارتباط است و پایایی با قطعیت مرتبط می باشد.
- ❑ از آنجائیکه راهی برای مشاهده یا محاسبه امتیاز **true** وجود ندارد $(Reliability = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{\sigma_X^2})$ در نتیجه روشهای متفاوتی برای برآورد پایایی یک آزمون پیشنهاد شده است.
- ❑ ضریب پایایی عددی بین صفر تا یک می باشد که صفر نشان دهنده عدم وجود پایایی است و یک پایایی صد در صد را نشان می دهد.
- ❑ روش مورد استفاده برای سنجش پایایی یک ابزار (آزمون یا پرسشنامه) به دلیل عملکرد و طبیعت ابزار و چگونگی استفاده از آن در زمینه مورد استفاده متفاوت می باشد.



انواع سنجش پایایی

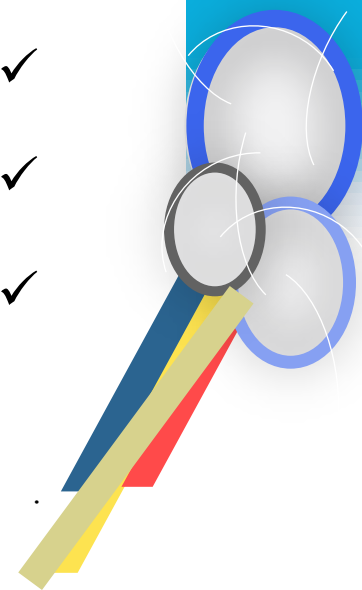


سنجش توافق درونی (Internal Consistency) در پایایی

- ✓ با استفاده از آلفا کرونباخ (رفرنسهای جدید: امگا مک دونالد)
- ✓ پایایی دو نیمه آزمون (Split half): ضریب اسپیرمن - براون
- ✓ پایایی کودر ریچاردسون

سنجش ثبات (Stability) در پایایی

- ✓ با استفاده از آزمون-بازآزمون (Test-Retest): ضریب اسپیرمن - براون
- ✓ فرمهای یکسان (موازی) و فرمهای متعادل (Equivalent Forms)
- ✓ فرمهای دقیقاً یکسان (Parallel & strictly parallel forms)،
- ✓ پایایی بین مشاهده کنندگان، مصاحبه گران یا ارزیابان (پایایی درون رده ای - Intraclass)



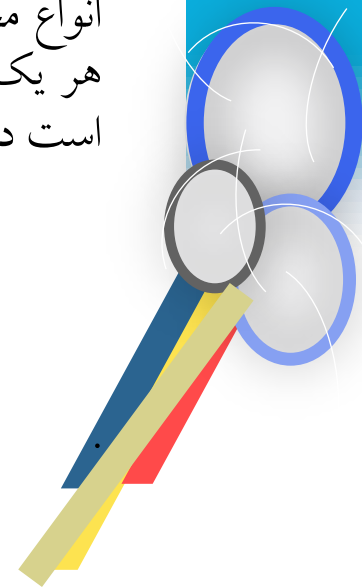
انواع سنجش پایایی



پایایی بین مشاهده کنندگان، مصاحبه گران یا ارزیابان (پایایی درون رده ای - Intraclass)

برآوردی از میزان توافق نمره‌های میان دو یا تعداد بیشتری از داورهایی است که یک مقیاس را ارزیابی نموده‌اند که در مورد داده‌های اسمی، رتبه‌ای و کمی به ترتیب با شاخص‌های کاپای کوهن، کاپای وزنی و همبستگی درون-رده‌ای (ICC)، Intraclass Correlation Coefficient ارزیابی می‌گردد.

انواع مختلف پایایی ارائه شده فوق لزوماً از یکدیگر مجزا نیستند و نتایج یکسانی نیز ارائه نمی‌دهند. هر یک از شاخص‌های بیان شده، مشابه با ضریب همبستگی هستند نکته قابل توجه این که ممکن است در یک مطالعه، بیش از یک نوع از شاخص‌های فوق استفاده گردد.





سنجش توافق درونی (آلفا کرونباخ)

□ کرونباخ در سال ۱۹۵۱ ضریب آلفا را برای حل مشکل تعیین پایایی آزمونه‌های چندسوالی ابداع کرد.

□ مقدار صفر این شاخص نشان دهنده‌ی این است که امتیاز واقعی مقیاس مورد نظر اندازه‌گیری نشده و نتیجه‌ی حاصل از محاسبه ناشی از خطای اندازه‌گیری است و در واقع عدم پایایی مقیاس مورد نظر را نشان می‌دهد. مقدار عددی یک این شاخص (در نقطه مقابل صفر) پایایی کامل مقیاس مورد نظر را نشان می‌دهد به عبارت دیگر، افراد برداشت کاملاً یکسانی از این مقیاس داشته‌اند و در اندازه‌گیری مقیاس خطایی وجود ندارد.

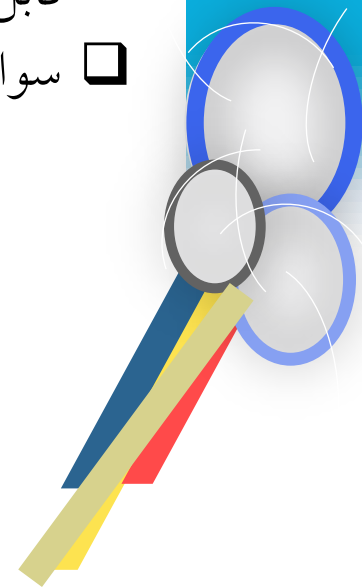
□ این روش معمولترین ضریب پایایی ثبات داخلی است و معرف میزان تناسب گروهی از آیتمهایی است که یک سازه را می‌سنجند. آلفا باید حداقل برابر ۰.۷ باشد تا یک سوال در یک ابزار بماند و بسیاری از محققان نقطه ۰.۸ را برای باقی ماندن یک سوال در ابزار ضروری می‌دانند.





سنجش توافق درونی (آلفا کرونباخ)

- ❑ روش آلفای کرونباخ فقط به انجام یکبار آزمون نیاز دارد تا برآوردی از پایایی آزمون را فراهم نماید.
- ❑ در کل این ضریب هنگامی مفید است که سوالات یک ابزار به صورت صحیح-غلط امتیاز بندی نشده باشند و جهت سنجش ثبات درونی سوالات در مقیاس لیکرت بکار میرود.
- ❑ مقدار آلفای بیش از ۰.۹ را عالی، ۰.۸-۰.۹ را خوب، ۰.۷-۰.۸ را قابل قبول، ۰.۶-۰.۷ را قابل بحث، ۰.۵-۰.۶ را ضعیف و کمتر از ۰.۵ را غیرقابل قبول معرفی کردند.
- ❑ سوالات باید همجهت باشند.





سنجش توافق درونی (آلفا کرونباخ)

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum s_i^2}{s_T^2} \right)$$
$$\alpha = \frac{N\bar{C}}{\bar{v} + (N-1)\bar{C}}$$

که در آن $\bar{C}$ برابر با میانگین کوواریانس داخلی آیتم‌ها inter-item و $\bar{v}$ برابر با متوسط واریانس است. اگر سوالات به صورت دو حالتی (درست=۱، غلط=۰) در نظر گرفته شوند، ضریب آلفا از رابطه زیر قابل محاسبه است؛ KR20 (Kuder-Richardson)

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum pq}{s_T^2} \right)$$

که در آن p تعداد پاسخ‌های درست و q تعداد پاسخ‌های غلط است.

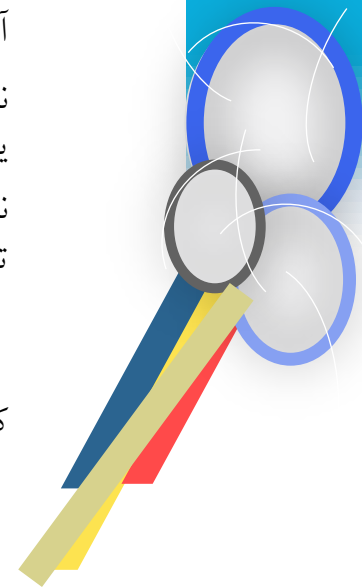
آلفای کرونباخ برای پرسشنامه‌های غیر استاندارد استفاده می‌شود.

نکته مهم این است که سؤالات (۱: کاملاً مخالفم به ۵: کاملاً موافقم) برای تمامی سؤالاتی که آلفا را می‌سنجند یکسان باشد.

نکته مهم: با توجه به کم برآورد پایایی توسط آلفای کرونباخ توصیه می‌شود پایایی سؤالات ترتیبی با استفاده از تتای ترتیبی سنجیده شود. چرا که در آلفای کرونباخ فرض بر این است که پاسخهای هر سوال پیوسته هستند.

$$\theta = \frac{k}{k-1} \left(1 - \frac{1}{\lambda_{max}} \right)$$

که در آن λ_{max} برابر با بیشترین مقدار ویژه در تحلیل مولفه‌های اصلی است.





سنجش توافق درونی (آلفا کرونباخ)

نکته) گاهی اوقات پیش می آید که در سنجش پایایی از طریق آلفای کرونباخ، مقدار آلفای کل منفی در می آید. بهتر است قبل از اقدام به تجدید نظر در طراحی مقیاس، این چند مورد را واریسی شود.

- مقادیر پرت و دورافتاده در متغیرها وجود داشته باشد که بعضا بخاطر اشتباه در ورود داده هاست. مثلا به جای کد ۲ ممکن است ۲۲ خورده باشد. لذا ابتدا باید از تمامی متغیرها جدول فراوانی تهیه نمود.
- ممکن است مقادیر اشتباه (Missig value) را تعریف نکرده باشید. مثلا کدهای ۹۹ را در میزینگ و لیو منظور نکرده باشید. بعضا پیش می آید دانشجویان تعریف مقادیر اشتباه در قسمت برچسبها (Value) را حمل بر تعریف Missig value می کنند که اشتباه است.
- ممکن است بعضی آیتها به صورت مثبت و بعضی دیگر به صورت منفی باشد (مخصوصا در مقیاس لیکرت و مقیاس های مشابه). در این حالت حتما ابتدا باید یک دسته از این متغیرها را با سایر متغیرها هم جهت نمود.
- ممکن است متغیرها هیچ تناسبی با یکدیگر نداشته باشند. یعنی عقلا لزومی نداشته باشد که بین متغیرها همبستگی وجود داشته باشد.

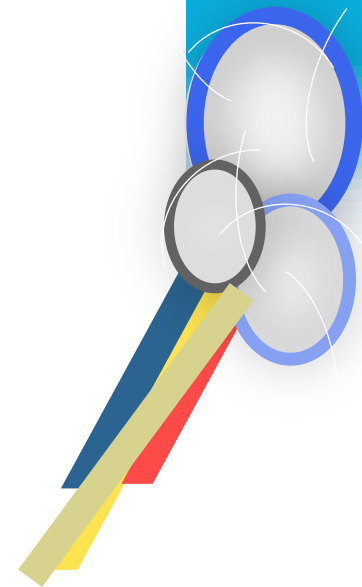




سنجش توافق درونی (ضریب امگا)

Hayes, A. F., & Coutts, J. J. (2020a). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But.... *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>

آموزش نصب و اجرا (Cronbach's alpha جایگزین ضریب آلفا کرونباخ) (McDonald's omega ضریب امگا مک دونالد)
(analysisacademy.com)



سنجش توافق درونی (دو نیمه آزمون)



□ در این روش یک آزمون به دو نیمه تقسیم شده و رابطه همبستگی بین دو نیمه آزمون سنجیده میشود. این آزمون معمولاً برای متغیرهای دو حالتی که کد صفر به پاسخهای غلط و کد یک به پاسخهای صحیح داده میشود به کار می رود.

□ در این روش دو نیمه آزمون باید در محتوی و سختی سوالات با هم مشابه باشند و از آنجا که تعداد سوالات ابزار به دو قسمت تقسیم شده می بایست روش ضریب همبستگی به کار رود که با استفاده از ضریب اسپیرمن-براون یا گاتمن محاسبه میشود.

□ فرمول محاسباتی ضریب اسپیرمن-براون به صورت زیر است:

$$r_{SB} = \frac{k \times r_{ij}}{[1 + (k - 1) \times r_{ij}]}$$

که در آن r_{SB} ، ضریب همبستگی دونیمه‌ای اسپیرمن براون، r_{ij} ضریب همبستگی پیرسون بین فرم‌های i و j و k برابر تعداد کل نمونه تقسیم بر تعداد نمونه در هر فرم است که معمولاً برابر عدد ۲ در نظر گرفته می‌شود.



سنجش توافق درونی (دو نیمه آزمون)

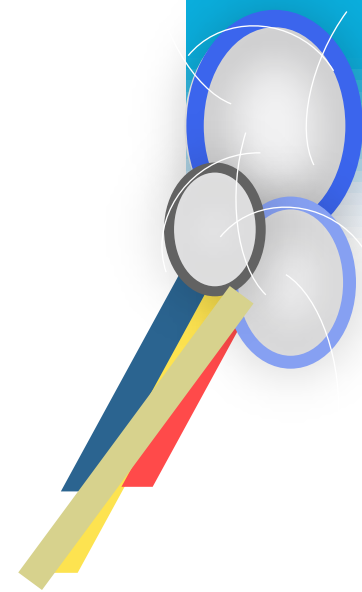


Reliability Statistics

Cronbach's Alpha	Part 1	Value	.586
		N of Items	12 ^a
	Part 2	Value	.598
		N of Items	11 ^b
	Total N of Items		23
Correlation Between Forms		.649	
Spearman-Brown Coefficient	Equal Length	.787	
	Unequal Length	.787	
Guttman Split-Half Coefficient		.787	

- a. The items are: Statistics makes me cry, My friends will think I'm stupid for not being able to cope with SPSS, Standard deviations excite me, I dream that Pearson is attacking me with correlation coefficients, I don't understand statistics, I have little experience of computers, All computers hate me, I have never been good at mathematics, My friends are better at statistics than me, Computers are useful only for playing games, I did badly at mathematics at school, People try to tell you that SPSS makes statistics easier to understand but it doesn't.
- b. The items are: I worry that I will cause irreparable damage because of my incompetence with computers, Computers have minds of their own and deliberately go wrong whenever I use them, Computers are out to get me, I weep openly at the mention of central tendency, I slip into a coma whenever I see an equation, SPSS always crashes

□ نکته) اگر تمامی ضرایب همبستگی حاصل از دو نیمه های ممکن محاسبه شود میانگین این ضرایب برابر با آلفای کرونباخ خواهد بود.



سنجش توافق درونی (ضریب گاتمن)

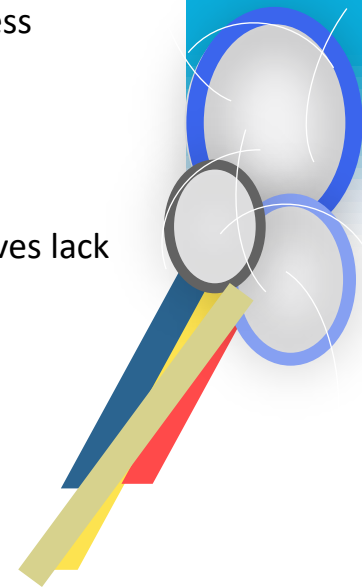


ضریب گاتمن: اصلاح شده‌ای از ضریب اسپیرمن براون است که در شش شکل مختلف L1 تا L6 برای اهداف مختلف ارائه می‌شود. نکته قابل توجه در این ضریب و ضریب گاتمن برای محاسبه پایایی با روش دو نیمه کردن در این است:

(نکته) مقادیر شدت تحت تأثیر نحوه دو نیمه کردن آیتم هاست. یکی از پیشنهادها، اختصاص تصادفی مجموعه‌ی آیتم‌ها به دو گروه است به گونه‌ای که انتظار می‌رود واریانس بین دو مجموعه یکسان باشد.

Guttman's lower bounds (lambda 1-6) are a set of six coefficients, L1 to L6, generated when in SPSS one selects "Guttman" under the Model button.:

1. L1: An intermediate coefficient used in computing the other lambdas.
2. L2: More complex than Cronbach's alpha and preferred by some researchers, though less common.
3. L3: Equivalent to Cronbach's alpha.
4. L4: Guttman split-half reliability.
5. L5: Recommended when a single item highly covaries with other items, which themselves lack high covariances with each other.
6. L6: Recommended when inter-item correlations are low in relation to squared multiple correlations



سنجش توافق درونی (ضریب گاتمن)



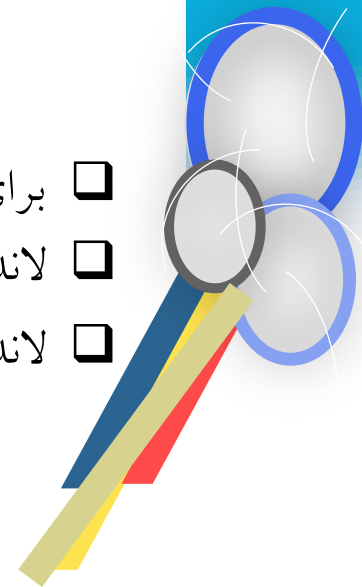
Reliability Statistics

Lambda	1	.721
	2	.801
	3	.754
	4	.787
	5	.766
	6	.819
N of Items		23

❑ برای تفسیر پایایی با استفاده از روش گاتمن از حد پایینتر استفاده میکنند.

❑ لاندای ۳ برابر آلفای با کرونباخ

❑ لاندای ۶ برابر با ضریب پایایی به روش تنصیف

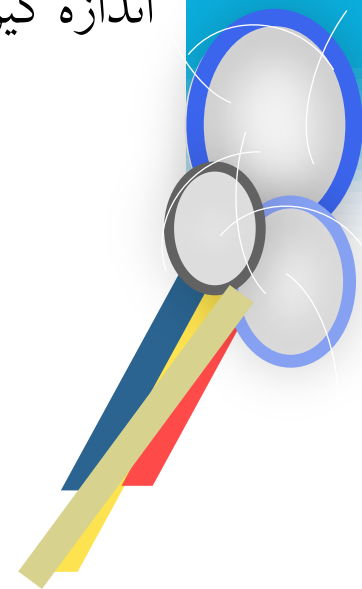




سنجش ثبات (آزمون-باز آزمون/ثبات در طول زمان)

در این روش جهت سنجش پایایی، سوالات آزمون در دو نوبت به یک گروه واحد، تحت شرایط مشابه داده می شود و نمرات حاصل با هم مقایسه می شوند. در نهایت ضریب همبستگی بین نمرات حاصل از اجرای دو بار آزمون برآورد شده تا چگونگی شباهت امتیازات مشخص گردد.

معمولترین روش یافتن همبستگی بین نمرات در این روش محاسبه ضریب همبستگی اسپرمن براون می باشد. روش آزمون-باز آزمون فرض میکند که متغیرها یا مفاهیم مورد اندازه گیری و همچنین خصوصیات آزمون شوندگان در طول دوره تغییر نخواهند کرد.





سنجش ثبات (آزمون-باز آزمون/ثبات در طول زمان)

❑ دلیل برخی معایب آن از قبیل آشنایی آزمون شوندگان با محتوی سوالات، مشکل در تعیین فاصله زمانی مناسب برای اجرای دو نوبت آزمون و عدم علاقه آزمون شوندگان به اجرای نوبت دوم توسط بسیاری از محققان بی ارزش دانسته شده است.

❑ محققین سه نوع اثر مختلف شامل اثر حافظه (memory effect)، اثر خستگی (Fatigue effect) و اثر ناشی از تغییرات واقعی (Genuine change effect) در مقدار پایایی آزمون بازآزمون موثر خواهند بود.

❑ اثر حافظه که به دلیل یادآوری سوالات در آزمون است سبب ایجاد نوعی خطای همبستگی (غیر تصادفی) شده و باعث بالا رفتن همبستگی نمرات آزمون و باز آزمون می گردد که پایایی را بالا می برد.

❑ اثر خستگی نوعی خطای تصادفی در بازآزمون شده و پایایی را کاهش می دهد.

❑ خطای ناشی از تغییرات واقعی نیز معمولاً سبب کاهش پایایی می گردد.



سنجش ثبات (آزمون-باز آزمون/ثبات در طول زمان)



□ فاصله زمانی مناسب برای انجام دو آزمون به ثبات متغیرهایی که عامل تحت بررسی را تحت تاثیر قرار می-دهد بستگی دارد. چنانچه متغیر مورد اندازه گیری فشار خون باشد زمان کوتاه خواهد بود اما جهت سنجش نگرش در مورد یک پدیده روانشناختی به فاصله زمانی طولانی-تری (دو هفته یا یک ماه) نیاز است.

□ این روش برای مقیاس‌هایی قابل استفاده است که در طول زمان پایدار باشند، به عنوان مثال مقیاس‌ها (سازه‌ها) بی نظیر هوش، اضطراب یا افسردگی این ویژگی را دارند ولی سازه‌ای مثل درد از لحظه‌ای به لحظه دیگر تغییر می‌کند و در این مورد نمی‌توان از این روش برای ارزیابی پایایی این مقیاس‌ها استفاده کرد. البته از این روش برای ارزیابی پایایی میانگین شدت درد که میانگین شدت را در ۲۴ ساعت گذشته یا هفته گذشته ارزیابی می‌کند، قابل استفاده است.

□ از لحاظ آماری، پایایی آزمون- بازآزمون، نسخه‌ای از پایایی با روش دو نیمه کردن است و برای ارزیابی آن از ضریب اسپیرمن - براون استفاده می‌شود.



سنجش ثبات (آزمون-باز آزمون/ثبات در طول زمان)



نگارش

❖ یکی دیگر از روشهای بررسی پایایی ابزار، ارزیابی ثبات آن در طول زمان است. آزمون کلاسیک ثبات، روش آزمون باز آزمون و به معنای تکرار اندازه گیری در زمان دیگر توسط همان ابزار و بر روی همان افراد است (Furr and Bacharach 2018).

❖ فاصله زمانی بین دو آزمون است که بهتر است بین دو هفته الی یک ماه باشد (DiIorio 2006).

❖ در مطالعه حاضر از روش آزمون باز آزمون برای تعیین ثبات ابزار استفاده خواهد شد. با در نظر گرفتن حداقل پایایی برابر 0.6، توان برابر 0.95 و خطای نوع اول برابر 0.05 با استفاده از پکیج calculate ICC sample size در نرم افزار R، (Rathbone, Shaw et al. 2015) حداقل حجم نمونه برابر ۳۰ بدست آمد. در این مطالعه جهت ارزیابی دقیقتر از ۵۰ نفر از بیماران (افزایش حجم نمونه برای ارزیابی روایی همگرایی در دو بازه زمانی است) درخواست می شود که پرسشنامه را تکمیل کنند و مجدد دو هفته بعد تکمیل پرسشنامه در شرایط یکسان انجام خواهد شد و میزان همبستگی بین نمرات از طریق ICC (Intraclass Correlation Coefficient) سنجیده خواهد شد.

1. Furr R, Bacharach V. Chapter 9: estimating and evaluating convergent and discriminant validity evidence. Psychometrics: an introduction Thousand Oaks, CA: SAGE Publications Inc. 2018.
2. DiIorio CK. Measurement in health behavior: methods for research and evaluation: John Wiley & Sons; 2006.
3. Rathbone A, Shaw S, Kumbhare D, Rathbone MA. Package 'ICC. Sample. Size'. CRAN-R Project. 2015.





سنجش ثبات (فرمهای موازی و متعادل)

❖ این روش بر این فرض است که نمرات واقعی گویه ها دارای واریانس یکسان هستند ولی لزوماً میانگین آنها یکسان نیست.

❖ این روش به آزمون برابری میانگین و واریانس گویه های ابزار اندازه گیری میپردازد.

❖ در این روش دو فرم معادل یا موازی (دارای درجه دشواری یکسان) طراحی و به طور همزمان بر روی یک نمونه واحد از آزمودنیها اجرا می گردد و سپس به محاسبه همبستگی بین نتایج دو آزمون پرداخته میشود. چنانچه مقدار معنی دار آزمون کای اسکوئر کوچکتر از ۰.۰۵ و مقدار پایایی بیشتر از ۰.۷ باشد میتوان گفت ابزار پایاست.



سنجش ثبات (فرمهای موازی و متعادل)



Test for Model Goodness of Fit

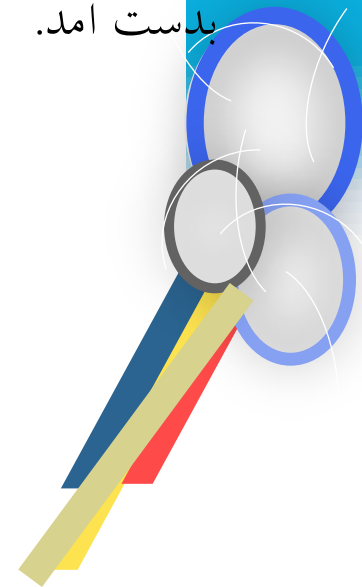
Chi-Square	Value	16835.814
	df	274
	Sig	.000
Log of Determinant of	Unconstrained Matrix	-8.363
	Constrained Matrix	-1.792

Under the parallel model assumption

این میزان پایایی دقیقاً برابر با میزانی است که در دو روش الفای کرونباخ و لانداى سوم گاتمن بدست آمد.

Reliability Statistics

Common Variance	.986
True Variance	.116
Error Variance	.870
Common Inter-Item Correlation	.118
Reliability of Scale	.754
Reliability of Scale (Unbiased)	.754

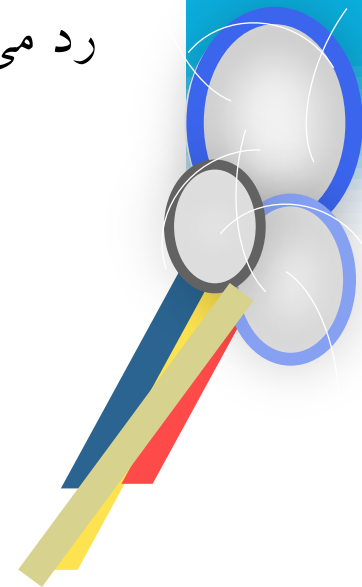




سنجش ثبات (فرمهای دقیقاً موازی)

❖ این روش نیز همانند روش فرمهای موازی است، منتهی با این تفاوت که در این روش فرض بر این است که نمرات واقعی گویه ها میانگین و واریانس یکسان دارند.

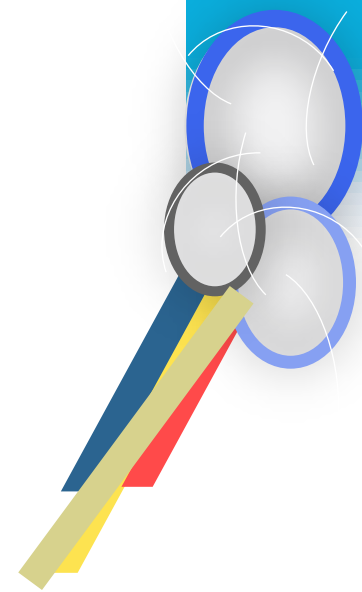
❖ در این روش نیز همانند روش فرمهای موازی چنانچه معنی داری نتیجه آزمون از ۰.۰۵ کوچکتر باشد در آن صورت فرض صفر مبنی بر موازی بودن مدل را رد می کنیم.





سنجش ثبات (فرمها)

❖ به دلیلی مشکل در طراحی و ساخت دو فرم یا نمونه سوال همسان و مساوی در مورد یک مفهوم، عدم علاقه آزمون شوندگان برای پر کردن فرم دوم و یادگیری ناشی از فرم اول و تاثیر آن در تکمیل فرم دوم چندان کاربردی ندارد.





سنجش ثبات (فرمها)

❖ در این روش دو مقیاس الف و ب برای ارزیابی هدف واحدی وجود دارد و هر دو فرم برای مجموعه واحدی از آزمونی‌ها اجرا می‌شود و میزان همبستگی (اسپیرمن - براون) بین این دو فرم میزانی از پایایی مقیاس هدف را نشان می‌دهد. ضریب حاصل، میزانی از ثبات موقت و همسانی مقیاس مورد نظر را نشان می‌دهد و برای اجرای فرم‌های الف و ب نیز باید یک فاصله زمانی رعایت شود. بنابراین این روش ارزیابی پایایی، ترکیبی از دو روش دو نیمه کردن و آزمون - بازآزمون را شامل می‌شود. اگر دو فرم بلافاصله پس از یکدیگر و بدون فاصله زمانی اجرا شود، میزانی از همسانی دو فرم را نشان می‌دهد و ثبات زمانی مقیاس را ارزیابی نمی‌کند.





سنجش ثبات (توافق میان ارزیابان)

در این روش همگنی اندازه‌گیری‌ها که توسط یک ابزار روی مجموعه واحدی از آزمودنی‌ها و توسط دو یا چند ارزیاب بررسی می‌شود، آزمون می‌گردد، تا میزان توافق ارزیاب‌ها روی ابزار مورد نظر بررسی گردد.

برای ارزیابی پایایی با این روش برای

- داده‌های اسمی از ضریب کاپای کوهن
- داده‌های ترتیبی از ضریب کاپای وزنی
- داده‌های کمی همبستگی درون رده‌ای (ICC) استفاده می‌شود.





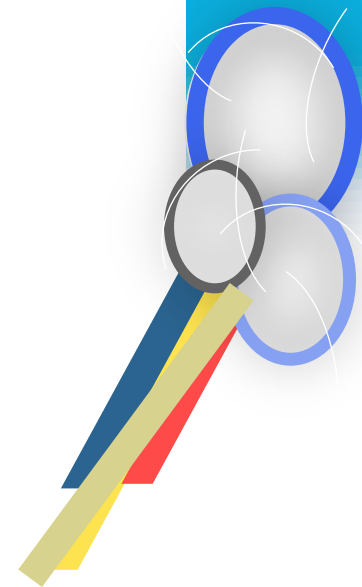
سنجش ثبات (توافق میان ارزیابان (inter rater reliability)

کاپای کوهن

این شاخص میزان توافق دو ارزیاب را روی یک صفت کیفی نشان می دهد. بر این اساس حداقل مقدار قابل قبول ضریب کاپا بیش از ۰.۶ می باشد و مقادیر بالاتر از ۰.۸ در توافق دو داور یا ارزیاب ایده آل می باشد.

جدول ۱- قدرت مقادیر مختلف ضریب کاپا در تعیین میزان توافق
بین مشاهده کنندگان یا داوران

قدرت توافق	مقدار آماره کاپا
ضعیف	کمتر از صفر
کم	۰-۰/۲
پایین تر از متوسط	۰/۲۱-۰/۴
متوسط	۰/۴۱-۰/۶۰
خوب	۰/۶۱-۰/۸۰
عالی	۰/۸۱-۱





سنجش ثبات (توافق میان ارزیابان)

در بررسی توافق بین دو ارزیاب شاخص کاپای کوهن معرفی شد. در برخی از موقعیت‌های عملی هدف بررسی توافق میان بیش از دو ارزیاب است. در این مورد نیاز است شاخصی میزان توافق بین ارزیاب‌ها را بررسی نماید. برای این منظور ارزیاب‌ها توافق خود را در مورد هدف‌های مورد نظر به صورت کیفی اعلام می‌نمایند و برای بررسی آن می‌توان از ضریب W کندال یا فلایز کاپا استفاده کرد (از آزمونهای ناپارامتری related samples).

ضریب کندال بین ۱- و ۱ قرار می‌گیرد؛ عدد ۱ توافق کامل در جهت مثبت و عدد ۱- توافق کامل در جهت منفی را نشان می‌دهد و عدد صفر، بین این دو عدد، عدم توافق رتبه‌ها (تصادفی بودن آنها) را نشان می‌دهد.





- Fleiss JL, Levin B, Paik MC (2003) Statistical methods for rates and proportions, 3rd ed.

Value of K	Strength of agreement
< 0.20	Poor
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Good
0.81 - 1.00	Very good

McBride (2005) suggests the following descriptive scale for values of the concordance correlation coefficient (for continuous variables):

Value of ρ_c	Strength of agreement
< 0.90	Poor
0.90 - 0.95	Moderate
0.95 - 0.99	Substantial
>0.99	Almost perfect





سنجش ثبات (توافق میان ارزیابان)

همبستگی درون رده ای (ICC)

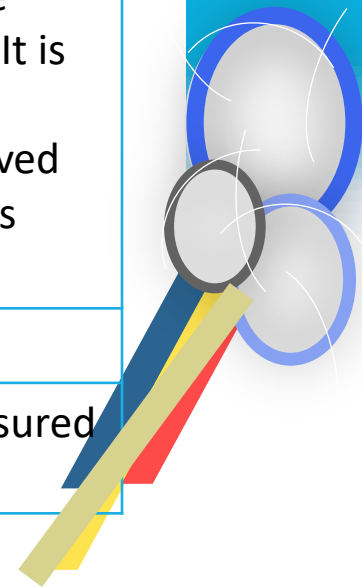
- ❑ از این شاخص برای ارزیابی میزان توافق دو یا تعداد بیشتری از ارزیاب‌ها در اندازه‌گیری یک متغیر کمی استفاده می‌شود.
- ❑ برای ارزیابی پایایی اندازه‌گیری‌های تکراری روی یک هدف واحد نیز می‌توان از این شاخص استفاده کرد.
- ❑ این شاخص به عنوان نسبتی از واریانس‌های درون ارزیاب‌ها یا تکرار اندازه‌گیری‌ها به واریانس کل تعبیر می‌شود.
- ❑ (۱) برای ارزیابی توافق، این شاخص به کار می‌رود و نمی‌توان از ضریب همبستگی پیرسون که به نوعی روند تغییرات دو متغیر را نسبت به هم می‌سنجد، برای این منظور استفاده کرد.
- ❑ در برخی از منابع، از این شاخص برای ارزیابی توافق چند ارزیاب یا اندازه‌گیری‌های تکراری یک صفت رتبه‌ای (طیف‌های لیکرت) نیز استفاده می‌شود که در یک حالت خاص مقدار آن با مقدار آلفای کرونباخ برابر است.





- There are 10 forms of ICCs.
- The best practice of reporting ICC should include software information, “model,” “type,” and “definition” selections.

Types	Definitions
Interrater reliability	It reflects the variation between 2 or more raters who measure the same group of subjects.
Test-retest reliability	It reflects the variation in measurements taken by an instrument on the same subject under the same conditions. It is generally indicative of reliability in situations when raters are not involved or rater effect is neglectable, such as self-report survey instrument.
Intrarater reliability	It reflects the variation of data measured by 1 rater across 2 or more trials.





McGraw and Wong (1996) defined 10 forms of ICC based on the “Model” (1-way random effects, 2-way random effects, or 2-way fixed effects), the “Type” (single rater/measurement or the mean of k raters/measurements), and the “Definition” of relationship considered to be important (consistency or absolute agreement).

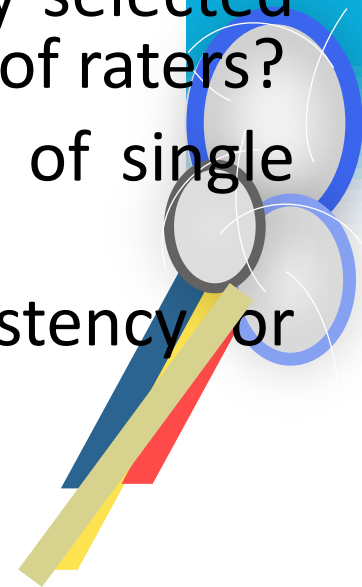
McGraw and Wong (1996) Convention ^a	Shrout and Fleiss (1979) Convention ^b	Formulas for Calculating ICC ^c
One-way random effects, absolute agreement, single rater/measurement	ICC (1,1)	$\frac{MS_R - MS_W}{MS_R + (k+1)MS_W}$
Two-way random effects, consistency, single rater/measurement	–	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$
Two-way random effects, absolute agreement, single rater/measurement	ICC (2,1)	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$
Two-way mixed effects, consistency, single rater/measurement	ICC (3,1)	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E}$
Two-way mixed effects, absolute agreement, single rater/measurement	–	$\frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$
One-way random effects, absolute agreement, multiple raters/measurements	ICC (1,k)	$\frac{MS_R - MS_W}{MS_R}$
Two-way random effects, consistency, multiple raters/measurements	–	$\frac{MS_R - MS_E}{MS_R}$
Two-way random effects, absolute agreement, multiple raters/measurements	ICC (2,k)	$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$
Two-way mixed effects, consistency, multiple raters/measurements	ICC (3,k)	$\frac{MS_R - MS_E}{MS_R}$
Two-way mixed effects, absolute agreement, multiple raters/measurements	–	$\frac{MS_R - MS_E}{MS_R + \frac{MS_C - MS_E}{n}}$

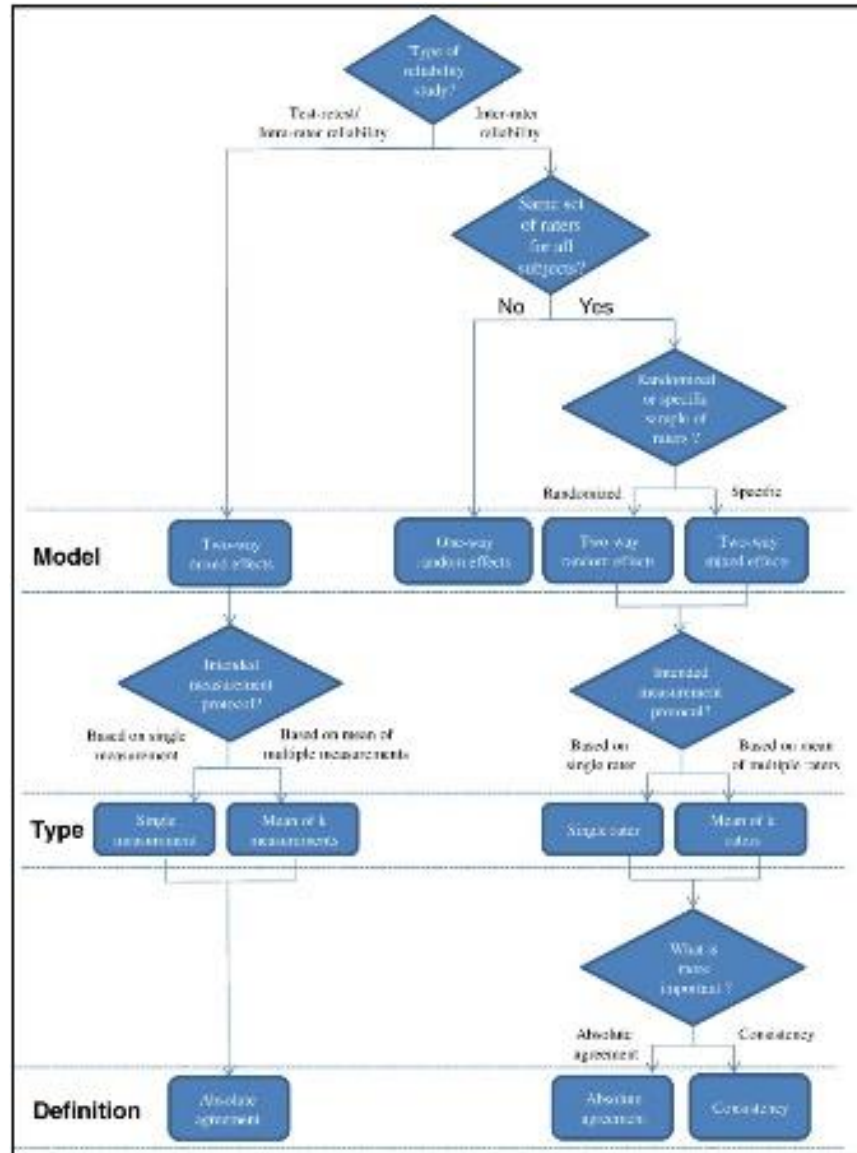


Selection of the correct ICC

Selection of the correct ICC form for interrater reliability study can be guided by 4 questions:

- Model:
 - (1) Do we have the same set of raters for all subjects?
 - (2) Do we have a sample of raters randomly selected from a larger population or a specific sample of raters?
- Type:(3) Are we interested in the reliability of single rater or the mean value of multiple raters?
- Definition:(4) Do we concern about consistency or agreement?







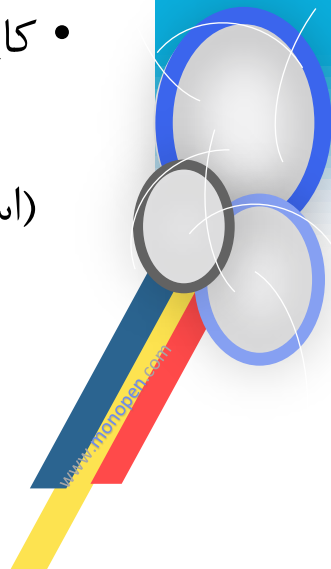
ضریب کاپا

- کاپای کوهن: بررسی توافق بین دو ارزیاب و ارزیابی دو حالتی و همه آزمودنی ها توسط

دو ارزیاب بررسی شوند

- کاپای فلایز (Fleiss' Kappa): توافق بیش از دو ارزیاب برای متغیرهای کیفی

(اسمی/ترتیبی) شرط همه آزمودنی ها توسط ارزیابها بررسی شود وجود ندارد.





42



<https://survey.porsline.ir/s/vHpGqivA>

