

FAQ Klausuren Analyse




19.7.2024

Klassische Testtheorie (KTT)

Die Klassische Testtheorie analysiert Testergebnisse mithilfe der "klassischen Statistik". Die Annahme dabei ist, dass jede beobachtete Punktzahl eines Individuums aus einem wahren Wert und einem Fehlerwert besteht. Der wahre Wert repräsentiert das tatsächliche Maß der Fähigkeit oder des Merkmals, das der Test zu messen beabsichtigt, während der Fehlerwert zufällige Einflüsse und Messungenauigkeiten widerspiegelt. Die KTT zielt darauf ab, die Zuverlässigkeit und Validität von Tests zu bewerten, indem sie die Konsistenz der Messergebnisse (Reliabilität) und die Genauigkeit, mit der ein Test das misst, was er zu messen beabsichtigt (Validität), untersucht.

Cronbachs Alpha

Cronbachs Alpha ist ein Maß für die interne Konsistenz oder Zuverlässigkeit eines Tests oder Fragebogens, das angibt, wie sehr die einzelnen Items miteinander zusammenhängen. Es wird häufig verwendet, um die Homogenität der Items zu bewerten und sicherzustellen, dass alle Fragen dasselbe messen. Ein hoher Wert von Cronbachs Alpha deutet darauf hin, dass die Items stark miteinander in Beziehung stehen und somit eine hohe interne Konsistenz aufweisen. Werte von Cronbachs Alpha reichen von 0 bis 1, wobei höhere Werte eine bessere Zuverlässigkeit anzeigen.

 CRONBACHS ALPHA -0.04	 CRONBACHS ALPHA 0.63	 CRONBACHS ALPHA 0.82
0.7 < : schlecht	0.7 < x < 0.8 : akzeptabel	> 0.8 : gut

Itemschwierigkeit: Das Verhältnis der Anzahl der Testteilnehmer, die das Item richtig beantwortet haben, zur Gesamtzahl der Testteilnehmer. Ein einfaches Beispiel: Wenn 80 von 100 Teilnehmern ein Item richtig beantworten, beträgt die Itemschwierigkeit 0,8 (oder 80%).

Die Itemschwierigkeit (p_i) wird durch die folgende Formel berechnet:

$$p_i = \frac{R}{N}$$

Dabei ist:

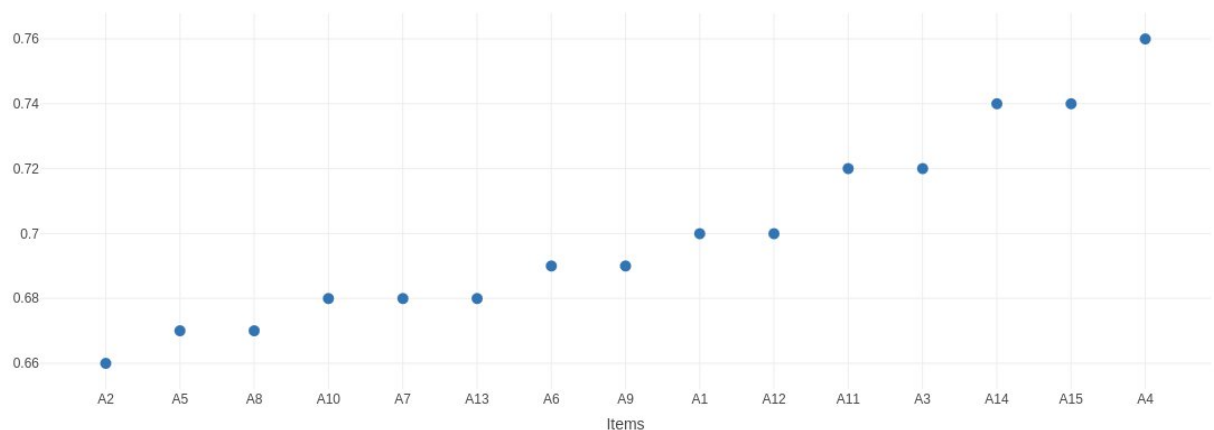
- p_i die Itemschwierigkeit des Items i ,
- R die Anzahl der Testteilnehmer, die das Item richtig beantwortet haben,
- N die Gesamtzahl der Testteilnehmer.

Ein Wert von $p=1$ bedeutet, dass alle Teilnehmer das Item richtig beantwortet haben (leichtes Item), während ein Wert von $p=0$ bedeutet, dass niemand das Item richtig beantwortet hat (schwieriges Item). Optimalerweise sollten Items eine mittlere Schwierigkeit aufweisen (nahe 0,5), um eine gute Differenzierung zwischen den Testteilnehmern zu ermöglichen.

Im folgenden dargestellt auf einem Scatterplot:

x-Achse : Item Namen

y-Achse : Item Schwierigkeit



Item-Response Theorie (IRT)

Die Item-Response-Theorie ist ein psychometrisches Modell, das zur Analyse von Testdaten und Fragebögen verwendet wird. Sie bietet eine Methode zur Bewertung der Beziehung zwischen den Eigenschaften der Testteilnehmer und den Eigenschaften der Testitems. Im Gegensatz zur klassischen Testtheorie betrachtet die IRT die Wahrscheinlichkeit, dass eine Person ein bestimmtes Item richtig beantwortet, als Funktion von deren Fähigkeit und den Eigenschaften des Items.

Itemschwierigkeit: Die Itemschwierigkeit wird als der Punkt auf der Fähigkeitsskala definiert, an dem die Wahrscheinlichkeit, das Item richtig zu beantworten, 50 % beträgt. Dies wird durch den Schwellenparameter (b) beschrieben.

Ein häufig verwendetes Modell der IRT ist das Rasch-Modell. Dieses Modell kommt auch hier zum Einsatz. Die Formel zur Berechnung der Wahrscheinlichkeit (p) einer richtigen Antwort auf ein Item i durch eine Person j mit der Fähigkeit Θ_j lautet:

$$P(X_{ij} = 1 \mid \Theta_j, b_i) = \frac{1}{1 + e^{-(\Theta_j - b_i)}}$$

$P(X_{ij} = 1 \mid \Theta_j, b_i)$: die Wahrscheinlichkeit, dass Person j das Item i richtig beantwortet

Θ_j : die Fähigkeit der Person j ,

b_i : die Schwierigkeit des Items i .

In diesem Modell ist die Itemschwierigkeit (b_i) der Punkt, an dem $\Theta_j = b_i$, bedeutet, dass die Fähigkeit der Person gleich der Schwierigkeit des Items ist. Bei diesem Punkt beträgt die Wahrscheinlichkeit einer richtigen Antwort 50%.

Wright-Map

Eine Wright-Map stellt die Fähigkeiten der Testteilnehmer und die Schwierigkeiten der Testitems auf derselben Skala dar. Dies ermöglicht eine anschauliche Visualisierung und den direkten Vergleich der Positionen von Personen und Items in Bezug auf die zugrunde liegende Fähigkeit oder das zu messende Merkmal.

- Auf der Wright-Map sind die Fähigkeiten der Testteilnehmer auf der linken Seite der Achse (Personenachse) und die Schwierigkeiten der Testitems auf der rechten Seite der Achse (Itemachse) dargestellt.
- Jeder Punkt auf der Personenachse repräsentiert die Fähigkeit eines Testteilnehmers, während jeder Punkt auf der Itemachse die Schwierigkeit eines Items repräsentiert.
- Die Skala beschreibt die Wahrscheinlichkeit mit der eine Person ein Item richtig beantwortet.

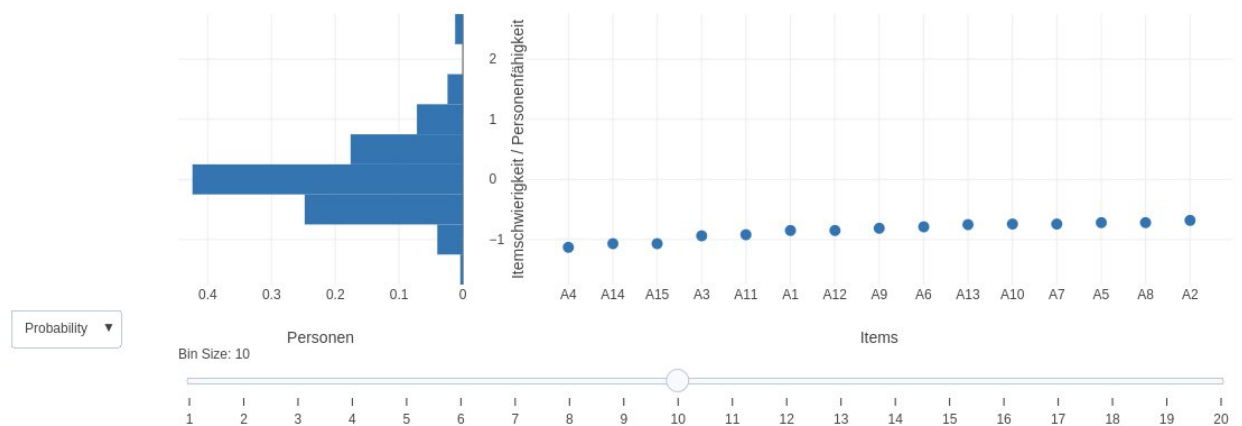
Eine gut gestaltete Wright-Map ist ein wertvolles Werkzeug, um die Passung zwischen Testitems und Testteilnehmern zu beurteilen und hilft dabei, die Qualität und Eignung eines Tests zu verbessern.

Item-Achse : Itemschwierigkeit (IRT)

Personen-Achse : Personenfähigkeit

Bin-Size : Einstellen der Bins der Personenfähigkeit

Personen nach Häufigkeit oder Anzahl darstellbar



Gut?	Schlecht?
Klare und detaillierte Darstellung der Verteilung der Fähigkeiten der Testteilnehmer und der Schwierigkeiten der Testitems	Ungleichmäßige Verteilung der Items entlang der Fähigkeitsskala z.B. wenn die meisten Items auf einen bestimmten Schwierigkeitsbereich konzentriert sind, was zu einer schlechten Differenzierung führt.
Ausgewogene Verteilung der Items entlang der Fähigkeitsskala. -> Test enthält Items für verschiedene Schwierigkeitsgrade und ist somit für Testteilnehmer mit unterschiedlichen Fähigkeitsniveaus geeignet	Fähigkeiten der Testteilnehmer und die Schwierigkeiten der Items sind nicht gut aufeinander abgestimmt z.B. wenn viele Items für die Mehrheit der Testteilnehmer zu schwer oder zu leicht sind, ist die Map ebenfalls schlecht.
Items differenzieren gut und sind weder zu leicht noch zu schwer sind, was eine optimale Messung der Fähigkeiten ermöglicht	

Histogramm

Stellt die Verteilung der Itemschwierigkeiten in einem Test oder Fragebogen grafisch dar. Es zeigt, wie oft verschiedene Schwierigkeitsgrade von Testitems vorkommen und ermöglicht es, die allgemeine Verteilung der Schwierigkeitseinschätzungen visuell zu erfassen.

- **Achsen:** Auf der horizontalen Achse (x-Achse) sind die Schwierigkeitsgrade der Items aufgetragen, während auf der vertikalen Achse (y-Achse) die Häufigkeit oder Anzahl der Items, die in einen bestimmten Schwierigkeitsbereich fallen, dargestellt wird.
- **Balken:** Jeder Balken repräsentiert die Anzahl der Items, die in einem bestimmten Schwierigkeitsbereich liegen. Die Höhe eines Balkens zeigt an, wie viele Items in diesem Bereich sind

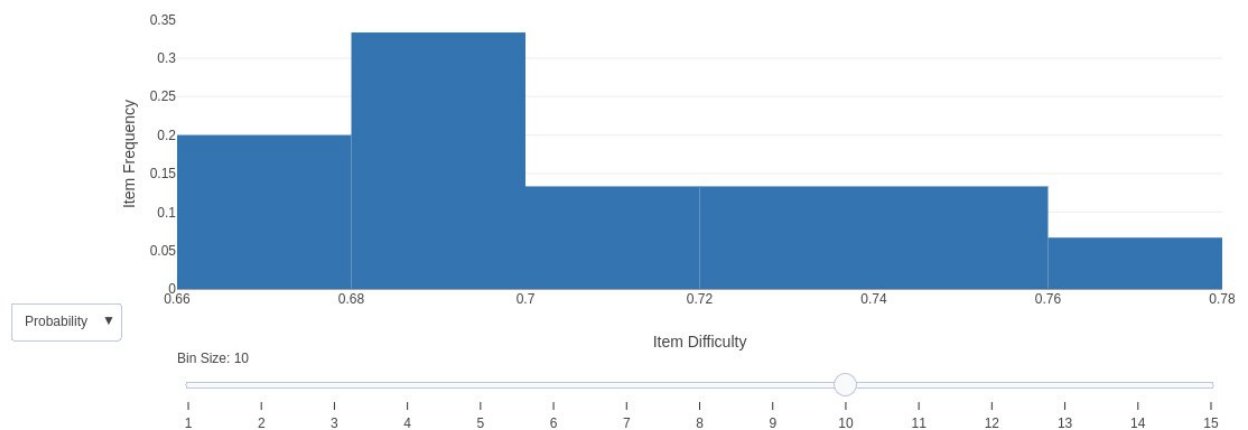
Ein gut gestaltetes Histogramm ermöglicht eine effektive Analyse und hilft dabei, die Qualität des Tests zu bewerten, indem es Einblicke in die Verteilung der Itemschwierigkeiten bietet und mögliche Anpassungen für eine bessere Testabdeckung identifiziert.

x-Achse : Itemschwierigkeit

y-Achse : Itemhäufigkeit

Bin-Size : Einstellen der Bins der Itemschwierigkeit

Itemschwierigkeit nach Häufigkeit oder Anzahl darstellbar



Gut?	Schlecht?
<p>Gute Verteilung: Verteilung der Itemschwierigkeiten ist ausgewogen. Idealerweise sollten die Schwierigkeitsgrade der Items eine breite Spanne abdecken, die sowohl leichte als auch schwierige Items umfasst. Dies sorgt dafür, dass der Test für alle Testteilnehmer mit unterschiedlichen Fähigkeitsniveaus geeignet ist.</p>	<p>Ungleichmäßige Verteilung: Verteilung der Itemschwierigkeiten ist stark unausgewogen.</p> <p>Beispielsweise, wenn alle Items in einem engen Schwierigkeitsbereich konzentriert sind, was darauf hindeutet, dass der Test entweder nur sehr einfache oder sehr schwierige Items enthält.</p>
<p>Keine Lücken: Keine großen Lücken in den Schwierigkeitsbereichen zu sehen, was darauf hinweist, dass der Test eine umfassende Abdeckung des gesamten Schwierigkeitsbereichs hat.</p>	<p>Grobe Granularität: Wenn die Schwierigkeitsgrade in zu groben Intervallen gruppiert sind, kann dies wichtige Details über die tatsächliche Verteilung der Itemschwierigkeiten verschleiern.</p>

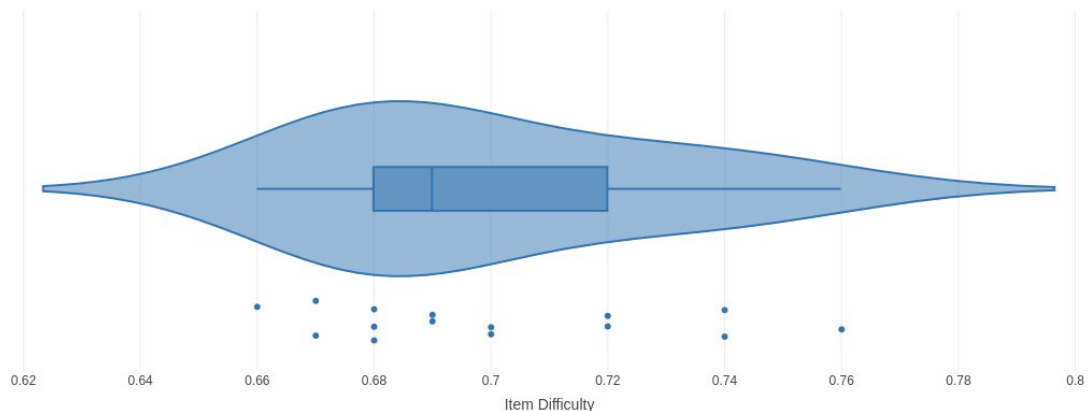
Violin-Plot

Visualisiert die Verteilung der Schwierigkeitseinschätzungen für verschiedene Testitems. Er kombiniert die Merkmale eines Boxplots und eines Dichteplots, indem er die Verteilung der Daten auf beiden Seiten einer vertikalen Achse als Violine darstellt.

- Die Breite der "Violine" an einer bestimmten Stelle zeigt die Dichte der Datenpunkte an dieser Stelle an. Breitere Abschnitte bedeuten, dass mehr Items eine ähnliche Schwierigkeit haben, während schmalere Abschnitte darauf hinweisen, dass weniger Items diese Schwierigkeit teilen
- Die Violinplots bieten zusätzlich zu den üblichen Boxplot-Statistiken (wie Median, Quartile und Ausreißer) eine detaillierte Darstellung der Verteilung der Itemschwierigkeiten, einschließlich möglicher Mehrgipfligkeit oder asymmetrischer Verteilungen.

Ein guter Violinplot hilft dabei, Muster in den Itemschwierigkeiten zu erkennen, wie z.B. Cluster von ähnlich schwierigen Items oder Unterschiede in der Schwierigkeit über verschiedene Itemgruppen hinweg. Ein schlechter Violin-Plot hingegen kann irreführend sein und wichtige Details der Verteilung übersehen lassen.

x-Achse : Itemschwierigkeit



ViolinPlot mit Boxplot

Gut?	Schlecht?
Eine klare Visualisierung der Dichte der Datenpunkte	Zu wenig Datenpunkte, was zu einer ungenauen Dichteabschätzung führt
Eine leicht erkennbare zentrale Tendenz und Variabilität der Itemschwierigkeiten	Übermäßige Glättung der Dichtefunktion, was wichtige Details der Verteilung verschleiert
Keine übermäßigen Verzerrungen oder Missinterpretationen der Daten	Fehlende oder missverständliche Beschriftungen und Achsenskalierungen, die die Interpretation erschweren.

Faktorenanalyse

Die Faktorenanalyse wird verwendet, um die zugrunde liegenden Dimensionen (Faktoren) eines Datensatzes zu identifizieren, indem sie die Beziehungen zwischen den Variablen untersucht. Im Kontext der Itemschwierigkeiten zielt die Faktorenanalyse darauf ab, herauszufinden, ob es gemeinsame Dimensionen oder Gruppen gibt, die die Schwierigkeit verschiedener Items in einem Test erklären. Dies kann helfen, die Struktur eines Tests zu verstehen und zu beurteilen, ob die Items zusammen ein konsistentes Konstrukt messen.

Eine gute Faktorenanalyse zeigt eine klare und interpretierbare Struktur, hohe Faktorladungen und hohe Kommunalität. Schlechte Faktorenanalysen führen zu unklaren oder schwer interpretierbaren Faktoren und niedrigen Ladungen.

Was sind Faktoren?

Faktoren sind latente Variablen oder Dimensionen, die die beobachteten Variablen (z.B. Itemschwierigkeiten) erklären. Sie repräsentieren zugrunde liegende Konstrukte oder Muster in den Daten.

Was ist eine MR (Minimal Residual)?

Minimal Residual bezieht sich auf die Methode zur Schätzung der Faktorladung in der Faktorenanalyse. Die Methode minimiert die Residuen (Differenzen zwischen den beobachteten und den durch das Modell geschätzten Korrelationsmatrizen) zur Bestimmung der besten Passung.

Was ist R^2 ?

ist ein Maß für den Anteil der Varianz einer Variablen, der durch die Faktoren erklärt wird. Es gibt an, wie gut die Faktoren die Variabilität der Itemschwierigkeiten erklären. Ein hohes R^2 bedeutet, dass ein großer Anteil der Varianz durch die identifizierten Faktoren erklärt wird, was auf eine gute Modellanpassung hinweist.

Was ist die Rotation?

Rotation ist ein Verfahren zur Vereinfachung und Verbesserung der Interpretierbarkeit der Faktoren. Es gibt zwei Hauptarten der Rotation: orthogonale Rotation (z.B. Varimax) und oblique Rotation (z.B. Promax). Orthogonale Rotation führt zu unkorrelierten Faktoren, während oblique Rotation korrelierte Faktoren zulässt.

Gut?	Schlecht?
<p>Klarheit der Faktorenstruktur: Klare und interpretierbare Struktur von Faktoren. Dies bedeutet, dass die Items gut gruppiert sind, basierend auf ihren Schwierigkeitsgraden, und dass diese Gruppen (Faktoren) logisch und theoretisch sinnvoll sind.</p>	<p>Unklare Faktorenstruktur: Unklaren oder schwer interpretierbaren Faktoren. Dies kann passieren, wenn Items auf mehreren Faktoren hohe Ladungen haben oder wenn keine klare Gruppierung der Items erkennbar ist.</p>
<p>Hohe Faktorladungen: Items sollten hohe Ladungen auf nur wenigen Faktoren haben, was bedeutet, dass sie stark mit den Faktoren korrelieren, auf denen sie platziert sind. Dies zeigt eine gute Unterscheidung zwischen den Faktoren und den Items.</p>	<p>Niedrige Faktorladungen: Wenn Items niedrige Ladungen auf den identifizierten Faktoren haben, deutet dies darauf hin, dass die Items nicht gut zu den Faktoren passen und möglicherweise nicht gut das zugrunde liegende Konstrukt messen.</p>
<p>Gute Kommunalität: Items sollten eine hohe Kommunalität aufweisen, d.h. der Großteil der Varianz eines Items sollte durch die identifizierten Faktoren erklärt werden.</p>	<p>Hohe Kreuzladungen: Wenn Items hohe Ladungen auf mehreren Faktoren haben, kann dies zu Verwirrung führen und die Interpretation der Faktoren erschweren.</p>