Grundlagen wissenschaftlichen Arbeitens II

Dr. Christine Groß

Studiengang Wirtschaftsinformatik

DHBW Mannheim

Ablauf der Eye-Tracking Studie im Rahmen der Vorlesung

17.02.2023

Literaturarbeit

Methodenauswahl

Entwickeln/ Erstellen der Stimuli

03.03.2023

Planung des Experiments: Methoden, Ablauf, Auswertung…

Erstellen der Stimuli in der Eye-Tracking Software

Evtl. erste Messdurchläufe

17.03.2023

Durchführung des Experiments

3 Gruppen

Mehrere Durchläufe

31.03.2023

Auswerten des Experiments

Unbenoteter Leistungsnachweis

- Erstellung einer unbenoteten Seminararbeit als Übung (ca. 10 Seiten)
  - Themenbereich: Durchführung eines Experiments
  - Teile der Arbeit
    - Fragestellung
      - Hypothesen
    - Methodenbeschreibung
    - Beschreibung der Durchführung
      - Erstellung der Stimuli
      - Durchführung des Experiments
    - Ergebnisse
    - Diskussion und Ausblick
    - Verbesserungsmöglichkeiten für künftige Studien/ Limitations

Ablauf einer Eye-Tracking Studie I

- Versuchsaufbau im Tobii
  - Wie viele Folien?
  - Welche Aufgaben?
  - Angabe der maximalen Zeitspanne
  - Fehlerfinden mit Space Taste zB?
- Hypothesen!
- Pilotstudie, um mögliche Fehlerquellen beim Studienverlauf zu erkennen
- Formaler Ablauf
  - Einverständniserklärung
  - Fragebogen zur Person
    - Alter
    - Geschlecht
    - Sehhilfe/ Sehbeeinträchtigungen (zB Rot-Grün-Blindheit)
    - Hochschulabschluss

Ablauf einer Eye-Tracking Studie II

- Farb- und Sehtest
- Tutorial: alle sollen denselben Kenntnisstand haben zu Beginn, mit Beispielen- Ablauf der Studie erklären
- Kalibrierung
- Betrachten der Programme (Identifikationsnummer, Zwischenfolie, wie wird Fehler detektiert und notiert, maximale Zeit…)
- Fragebogen zu den Programmen
  - Vorkenntnisse
  - Verständnis
  - Anstrengung
  - Bewertung der Codedarstellungen

Forschungsmethode- und Fragen

Forschungsmethode:

stationäre Eyetracker; Fragebogen; Fehlerdetektion in Programmen mit IDE und ohne

Forschungsfragen

- 1. Ist man mit Hilfe der IDE (Integrated Development Environment) schneller beim Erkennen von Fehlern? (Zeit/(ms))
    - Wenn ja – wie viel schneller?
- 2. Gibt es einen Unterschied zwischen der Anzahl der Fixierungen bis der Fehler gefunden wird mit und ohne IDE? (Gaze Plots/ Blickpunkt-Reihenfolge)
- 3. Ist die Fixierungsdauer am Fehler kürzer mit IDE als ohne? (Heatmap/ ms)
- 4. Sind die Fixierungen mit IDE fokussierter als ohne? (Heatmap = Größe um das rote Feld)
- 5. Haben unterschiedliche Vorkenntnisse/ Konzentration/ Schlaf Auswirkungen auf
    - ...die Geschwindigkeit der Fehlererkennung?
    - ...die Dauer der Fixierung?
    - ...die Anzahl der Fixierungen bis die Fehler entdeckt werden?

Forschungsmethode- und Fragen

Forschungsmethode:

stationäre Eyetracker; Fragebogen; Fehlerdetektion in Programmen mit IDE und ohne

Forschungsfragen

- 1. Ist man mit Hilfe der IDE (Integrated Development Environment) schneller beim Erkennen von Fehlern? (Zeit/(ms))
  - Wenn ja – wie viel schneller?
- 2. Gibt es einen Unterschied zwischen der Anzahl der Fixierungen bis der Fehler gefunden wird mit und ohne IDE? (Gaze Plots/ Blickpunkt-Reihenfolge)
- 3. Ist die Fixierungsdauer am Fehler kürzer mit IDE als ohne? (Heatmap/ ms)
- 4. Sind die Fixierungen mit IDE fokussierter als ohne? (Heatmap = Größe um das rote Feld)
- 5. Haben unterschiedliche Vorkenntnisse/ Konzentration/ Schlaf Auswirkungen auf
  - ...die Geschwindigkeit der Fehlererkennung?
  - ...die Dauer der Fixierung?
  - ...die Anzahl der Fixierungen bis die Fehler entdeckt werden?

Deskriptive Statistik

= reine Beschreibung der Daten durch

- Häufigkeitstabellen
  - Zählen = einfachstes statistisches Verfahren
  - Normalskalierte Variablen = einzig mögliche statistische Operation
    - Modalwert = der häufigste vorkommende Wert
    - Mittelwert/ Median nicht möglich, da keine Ordnungsrelation zu Grunde liegt
- Passende Kennwerte
  - Komplette Häufigkeitstabelle
    - zusätzlich zu den *beobachteten* Häufigkeiten,
    - die *prozentualen* Häufigkeiten angeben
  - Kumulierte Häufigkeiten
    - Ab ordinalskalierten Variablen zusätzlich
      - Median
      - Kumulierte Häufigkeiten: bis zur betreffenden Person aufsummierte Häufigkeiten
      - Kumulierte prozentuale Häufigkeiten: auf Basis der Gesamtsumme der Häufigkeiten
  - Klassenbildung: bei intervallskalierten Daten

3.1 Häufigkeitstabellen 31

	Häufigkeit	Prozent	kumulierte Häufigkeit	kumulierte Prozente
mindestens zweimal pro Woche	73	2,6 %	73	2,6 %
einmal pro Woche	360	13,0 %	433	15,7 %
ein- bis dreimal pro Monat	331	12,0 %	764	27,7 %
mehrmals im Jahr	660	23,9 %	1424	51,6 %
seltener	935	33,9 %	2359	85,5 %
nie	402	14,6 %	2761	100,0 %

Tabelle 3.3: Kumulierte Häufigkeiten

Zöfel, 2001 S.31

Lokalisationsparameter - Mittelwert

- Passend für intervallskalierte und normalverteilte Variablen
- Weniger geeignet für nicht normalverteilte oder ordinalskalierte Variablen
- Unsinnig für nominalskalierte Variablen
- Mittelwert = arithmetisches Mittel
  - Von n Werten ist x_i die Summe dieser Werte geteilt durch ihre Anzahl

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Lokalisationsparameter - Median

- Ordinal- bzw intervallskalierte, aber nicht normalverteilte Variablen
- Er ist derjenige Wert, unterhalb und oberhalb dessen jeweils die Hälfte der Messwerte liegen
  - BSP Triglyceridwerte: 489 113 141 120 217 109 675 218 96 225 132
    - 2 Ausreißerwerte (489, 675) -> Median (Ausreißern ggü unempfindlich)
    - Der Größe nach: 96 109 113 120 132 141 217 218 225 489 675 690
      - Median = 141
    - Wenn eine gerade Anzahl besteht : $\frac{141+217}{2} = 179$
- Median ist unempfindlich gegenüber Ausreißern, da es gleichgültig ist, welchen Wert der größte Messwert annimmt, da der Wert des Medians hiervon unberührt bleibt

Dispersionsparameter: Standardabweichung

Lokalisationsparameter = Lage einer Verteilung oder die zentrale Tendenz

Dispersionsparameter = Streumaße = die Breite einer Verteilung

- Einfachstes Streumaß = Spannweite
  - Differenz zwischen größtem und kleinstem Wert
  - basiert lediglich auf Extremwerten
- Standardabweichung
  - Gebräuchlichstes Streumaß
  - Bei intervallskalierten und normalverteilten Variablen
  - Bildung der Summe der quadratischen Abweichungen aller Messwerte vom Mittelwert und diese teilt man durch die um 1 verminderte Fallzahl und zieht daraus die Wurzel
  - Je mehr die einzelnen Werte von

ihrem Mittelwert abweichen, desto

Größer wird die Standardabweichung

Varianz = Quadrat der Standardabweichung

$$\bar{x}_N = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$s_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x}_N)^2}$$

Dispersionsparameter: Quartilabstand

- Median: derjenige Wert, unterhalb und oberhalb dessen 50% der Werte liegen
- 2 weitere ausgezeichnete Punkte der Messwertskala sind
  - 1. Quartil (Q1): unterhalb liegen 25% der Werte
  - 3. Quartil (Q3): unterhalb liegen 75% der Werte
  - Median: 2. Quartil (Q2): unterhalb liegen 50% der Werte
  - Q1, Q2 und Q3 teilen die Messwertskala in 4 Teile mit gleichen Häufigkeiten ein
  - Abstand zwischen Q1 und Q3 beträgt 50% = Maß für Streuung der Werte
  - In der Praxis: halber Quartilsabstand benutzt – dort, wo anstelle des Mittelwertes der Median als Lokalisationsparameter verwendet wird
- Perzentile nach Cole: die so definierten Schnittpunkte können an beliebigen Stellen angesetzt werden – zB unter welchem Meßwert befinden sich 10% der Werte

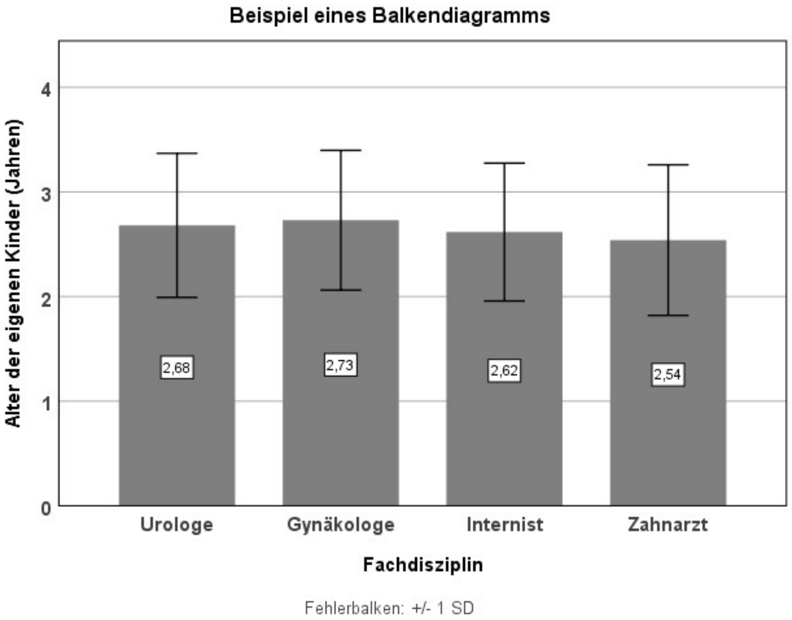
Grafische Darstellung: Balkendiagramm

= Visualisierung statistischer Kennwerte oder Zusammenhänge

Balkendiagramme:

Häufigkeiten, Mittelwerte, Mediane in Abhängigkeit von Gruppierungsvariablen

Fehlerbalken: Streuungsmaß



Kreisdiagramme:

zB bei prozentualen Häufigkeiten von kategorialen Variablen

Statistics

Age

N	Valid	650
	Missing	11
Mean		23.98
Minimum		9
Maximum		69
Sum		15587

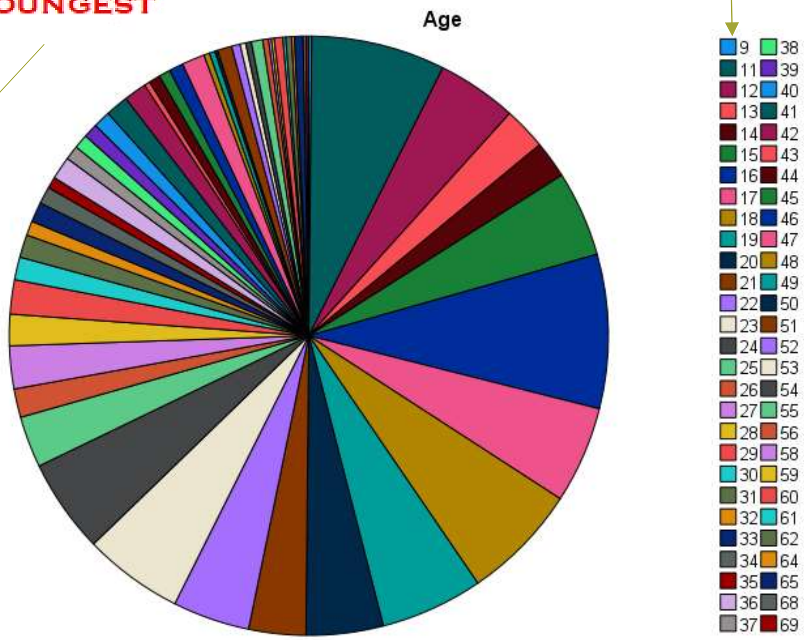
YOUNGEST

OLDEST

AGE

NUMBER OF PARTICIPANTS

YOUNGEST



THE RELATIONSHIP BETWEEN AGE AND LANGUAGE PERCEPTION IS BIASED BY AGE. WHAT CAN I DO NOW?

ANALYSE AT WHICH AGE NO RELATIONSHIP BETWEEN THE VARIABLES CAN BE DETECTED.

Final solution – analyses to be continued

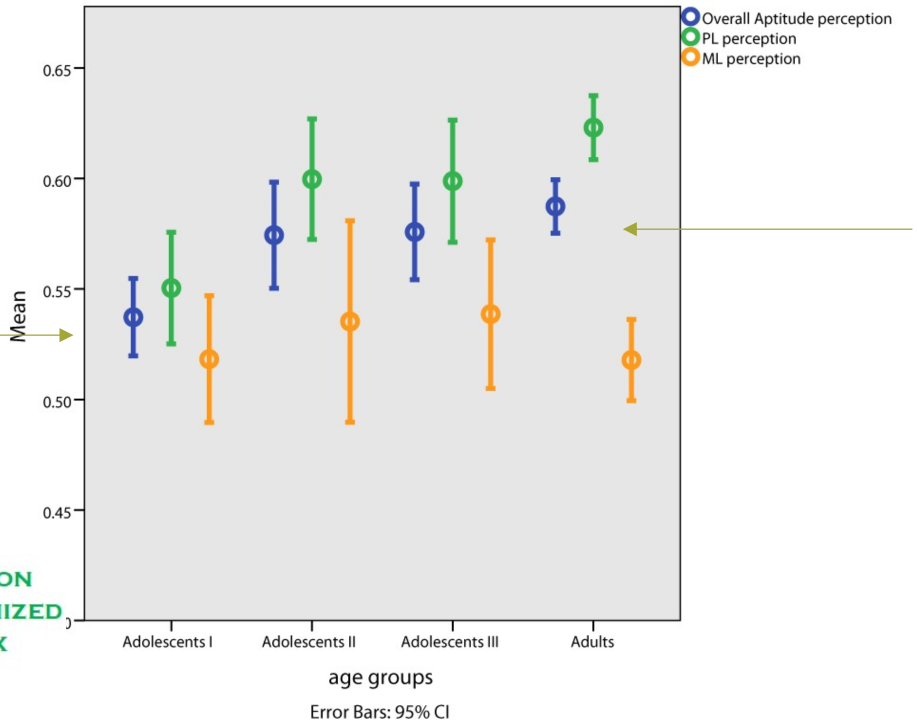
Groups based on mean values have been divided into

adolescents I (aged 11-12),

adolescents II (aged 13-15),

adolescents III (aged 16-17),

adult group (aged 18 – 68).



The **CONCLUSION?**

NO SIGNIFICANT DIFFERENCE IN SPEECH PERCEPTION BETWEEN ADULTS AND ADOLESCENTS ARE RECOGNIZED AT THE AGE OF 13. AGE IS ONLY IMPORTANT IF I MIX GROUPS.

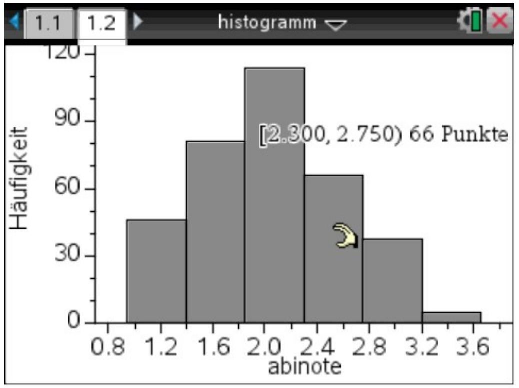
SIGNIFICANT DIFFERENCE FOR THE GROUP BELOW 13.

STATISTICAL METHODS: AS THE DATA SHOWED A NON-NORMAL DISTRIBUTION, A KRUSKAL-WALLIS TEST, FOLLOWED BY MANN-WHITNEY TESTS WERE APPLIED. THIS EXAMPLE IS TAKEN FROM CHRISTINER 2020

Grafische Darstellungen

– Liniendiagramme: zeitliche Verläufe

– Histogramme: Häufigkeitsverteilungen intervallskalierter Variablen

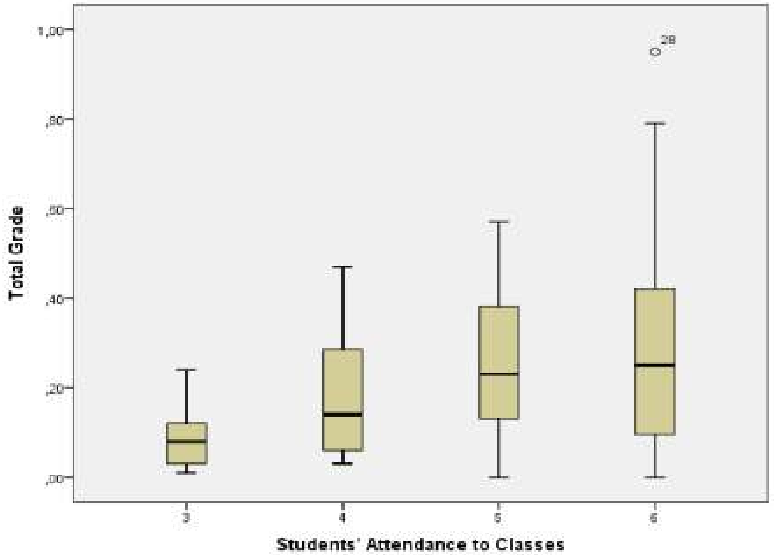


– Boxplots: Median, Quartile von intervallskalierten Variablen

  – Untere und obere Linie markieren den kleinsten bzw größten auftretenden Wert (Minimum und Maximum)

  – Untere Begrenzung der Box: Q1, obere Q3

  – Mittlere Linie: Median



Streudiagramme

Zusammenhang zwischen 2 intervallskalierten Variablen
Form der entstehenden Punkwolke gibt Aufschluss über die Stärke und Form des Zusammenhangs

Age and education

- Please enter your age in years.
- Please indicate your highest level of education.
- Please indicate your mother's highest level of education.
- Please indicate your father's highest level of education.

	Overall Aptitude perception	PL perception	ML perception	Highest level of education	Highest education of mother	Highest education of father	Age
Overall Aptitude perception	1	.82**+	.60**+	.18**+	.20**+	.21**+	.17**+
PL perception	.82**+	1	.15**+	.24**+	.19**+	.20**+	.24**+
ML perception	.60**+	.15**+	1	-.06	.08	.11*	-.07

STATISTICAL METHOD
CORRELATIONS

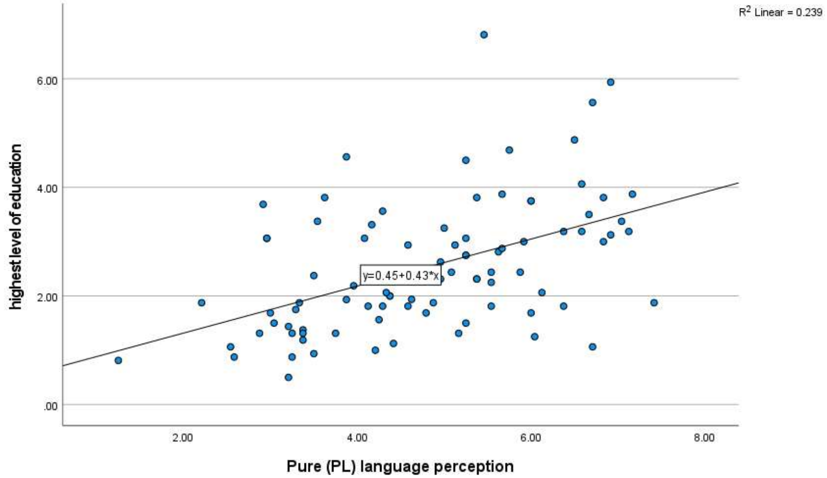
*SIGNIFICANT 0.05

**SIGNIFICANT 0.01

Highest level of education and language perception ability (PL condition)

Correlations

Visual representation

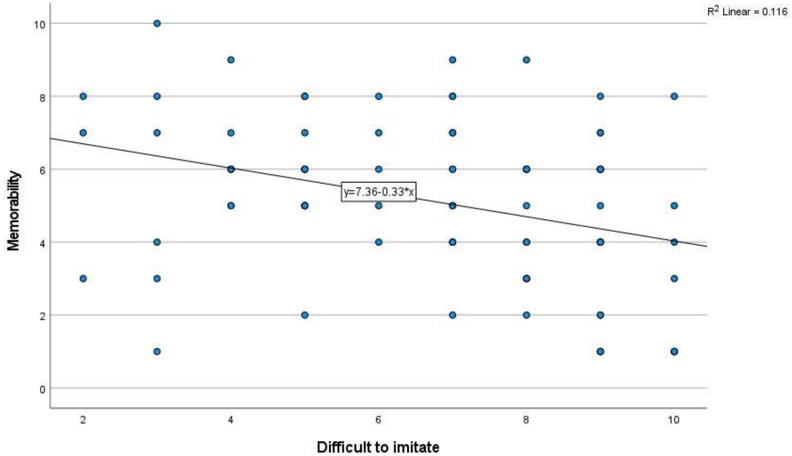


Positive relationship

Could be interpreted as the following: the higher the educational status, the better the ability to perceive unfamiliar language stimuli (positive relationship

Further examples would be

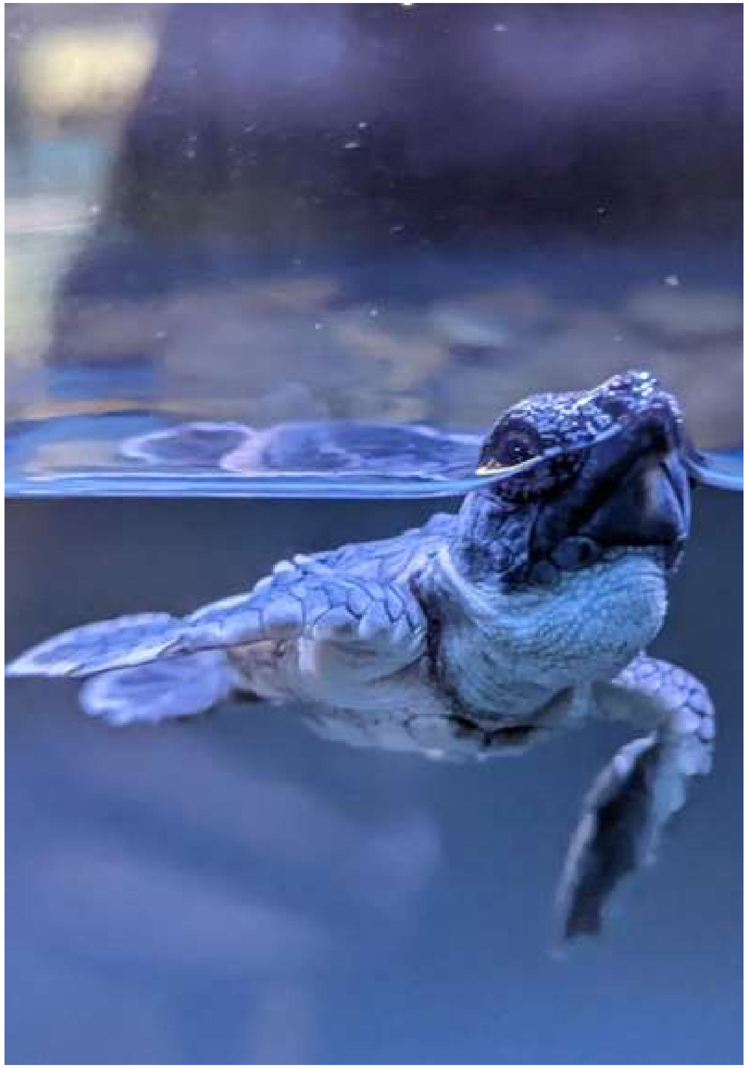
(positive correlation age and grey hair)



Example taken from another research (negative relationship).

Further examples would be

(negative correlation age and no hair)



Übung:
Erstellen sie die deskriptive Statistik für ihren Datensatz mit SPSS

Analytische Statistik - Grundlagen

- Auch schließende Statistik; Interferenzstatistik
- Befasst sich mit dem Problem, wie aufgrund von Ergebnissen, die anhand einer vergleichsweise kleinen Zahl von Personen (oder Objekten) gewonnen wurde, allgemeingültige Aussagen hergeleitet werden können
- Die analytische Statistik versucht, von den Verhältnissen der Stichprobe auf die Verhältnisse in der Grundgesamtheit zu schließen
  - Grundgesamtheit = alle untersuchbaren Personen (Objekte) = diejenige Grundgesamtheit, für die die Stichprobe repräsentativ ist (möglichst ähnliche Verhältnisse aufweist)
- Zwei Problemkreise beim Schluss von Verhältnissen der Stichprobe auf die betreffende Grundgesamtheit
  - Schluss von Kennwerten der Stichprobe auf die entsprechenden Parameter der Grundgesamtheit (z.B. Mittelwerte der Körpergrößen aller erwachsenen Männer)
  - Überprüfung von Hypothesen (Überprüfung Kennwerte mehrerer SP darauf ob sie zur gleichen GS gehören = Prüfstatistik)

Analytische Statistik – objektive Testverfahren

- Überprüfung, ob ein Mittelwertsunterschied zufällig zustande gekommen ist oder nicht = *signifikant* unterschiedlich
- Prüfung auf Signifikanz bei Mittelwerten möglich, aber auch bei Standardabweichungen, Prozentwerten, Häufigkeitsverteilungen, Korrelations- und Regressionskoeffizienten (unterscheiden sie sich signifikant von 0)
- Hypothesen
  - H0 = Nullhypothese: Die beiden Stichprobenmittelwerte gehören zu der gleichen Grundgesamtheit
  - H1 = Alternativhypothese: Die beiden StichprobenMW gehören zu verschiedenen Grundgesamtheiten

→ Ob die Nullhypothese beibehalten wird oder zugunsten der Alternativhypothese verworfen wird, ist anhand der betreffenden Prüfstatistik zu überprüfen – je nach Testsituation gibt es hier unterschiedliche Tests

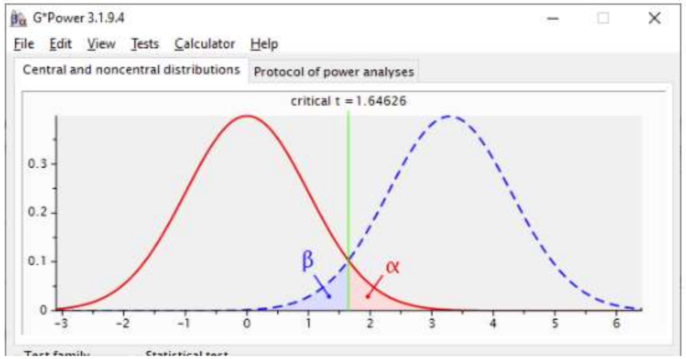
- Z:b: t-Test nach Student zum Vergleich zweier Mittelwerte
  - Prüfgröße *t* – anhand derer Freiheitsgrade bestimmen *df* – *t*-Verteilung, die wie die Normalverteilung eine eingipfelige Verteilung, deren Gestalt von der Anzahl der df abhängt – bei hohen Freiheitsgraden nähert sie sich der Normalverteilung an
  - Mit Hilfe der Verteilung kann zur Prüfgröße *t* und zur Anzahl *df* die sogenannte Irrtumswahrscheinlichkeit *p* bestimmt werden

  → Wenn die Nullhypothese verworfen wird und die Alternativhypothese angenommen wird, ist die Irrtumswahrscheinlichkeit 0,03

  Klassische Signifikanzwerte: p$\leq$ 0,05 signifikant $*$ p $\leq$ 0,01 sehr signifikant $**$ p $\leq$ 0,001 höchst signifikant $**$
- Weitere Prüfverteilungen: Standardnormalverteilung (Gauß), F-Verteilung (Fischer), x^2 – Verteilung (Pearson)

Fehler erster und zweiter Art

Hat man Nullhypothese und Alternativhypothese formuliert, so kann man beim Überprüfen dieser Hypothesen mit einem passenden statistischen Test offenbar zwei Fehler machen:



- Fehler 1. Art oder $\alpha - \text{Fehler}$
  - Die Nullhypothese wird verworfen, obwohl sie richtig ist.
  - = Irrtumswahrscheinlichkeit *p*
- Fehler 2. Art oder $\beta - \text{Fehler}$
  - Die Nullhypothese wird behalten, obwohl sie falsch ist.
  - Die Wahrscheinlichkeit einen solchen Fehler zu begehen ist nicht berechenbar.
  - Lediglich: Die Gefahr, einem $\beta - \text{Fehler}$ zu erliegen ist desto kleiner, je deutlicher die berechnete Irrtumswahrscheinlichkeit *p* die Signifikanzgrenze übersteigt

β = Wahrscheinlichkeit, dass ein Unterschied nicht erkannt wird

	H_0 ist falsch	H_0 ist richtig
Test lehnt H_0 ab	richtige Entscheidung	Fehler erster Art (α)
Test lehnt H_0 nicht ab	Fehler zweiter Art (β)	richtige Entscheidung

Teststärke = 1 - β

Wahrscheinlichkeit dafür, dass ein bestehender Unterschied auch aufgezeigt wird

Streubereiche und Konfidenzintervalle

- Der Schluss von Kennwerten einer Stichprobe auf die Parameter der zugehörigen Grundgesamtheit erfolgt bei intervallskalierten und normalverteilten Variablen über Streubereiche und Konfidenzintervalle
  - Streubereiche: Sagen einen Bereich voraus, in dem die einzelnen Messwerte liegen
    - Wie viel Prozent der Werte liegen in einem bestimmten Intervall?
    - ein um den MW symmetrisches Intervall
    - Intervallbreite vorgeben (idR in ganzzahligen Einheiten der Standardabweichung angegeben) und dann die Prozentzahl der im Intervall enthaltenen Werte ermitteln
    - Oder diese Prozentzahl ist vorgegeben und die Breite des Intervalls wird bestimmt
  - Konfidenzintervalle: geben an zwischen welchen Grenzen sich mit vorgegebener Wahrscheinlichkeit Mittelwert und Standardabweichung in der Grundgesamtheit bewegen
    - Wie kann man vom Kennwert (Mittelwert, Standardabweichung, prozentuale Häufigkeiten) einer Stichprobe auf den entsprechenden Parameter der Grundgesamtheit schließen?

Normalverteilung

- Je nachdem, ob Normalverteilung der Werte vorliegen oder nicht, sind gegebenenfalls verschiedene analytische Tests durchzuführen
  - Vergleich zweier Stichproben
    - Normalverteilung: klassischer t-Test nach Student
    - Nicht Normalverteilung: U-Test nach Mann und Whitney
- Vor der Anwendung eines statistischen Tests, der eine Normalverteilung der Werte zur Voraussetzung hat, ist diese zunächst zu überprüfen, zB mit
  - Chiquadrat-Test
    - Werte der zu überprüfenden Variablen werden in Klassen eingeteilt und dann die beobachteten Klassenhäufigkeiten mit den unter Normalverteilung zu erwartenden verglichen
    - Nur für recht große Fallzahlen
    - Wenn nicht signifikant, kann Normalverteilung angenommen werden
  - Kolmogrov-Smirnov-Test
    - Für kleinere Fallzahlen
    - Nicht anfällig für Ausreißerwerte: aber bei Ausreißerwerte besser Methoden verwenden, die keine Normalverteilung voraussetzen
    - Wenn nicht signifikant, kann Normalverteilung angenommen werden

Beziehung zwischen 2 Variablen I

- Je nach Skalenniveau der beteiligten Variablen und je nachdem ob bei intervallskalierten Variablen, Normalverteilung vorliegt oder nicht, gelangen unterschiedliche Tests zur Anwendung
  - **Test auf signifikante Unterschiede**
    - Vergleich verschiedener Stichproben hinsichtlich ihrer Mittelwerte oder Mediane
      - Unabhängige oder abhängige (eine Variable zu mehreren Zeitpunkten) Stichproben
      - Vergleich von zwei oder mehr als zwei Stichproben
      - Intervallskalierte und normalverteilte Werte, ordinalskalierte und nicht normalverteilte intervallskalierte Werte?
        - t-Test nach Student (unabhängige Stichproben, normalverteilt)
        - t-Test für abhängige Stichproben
        - U-Test Mann und Whitney: bei nicht Normalverteilung
        - Einfaktorielle Varianzanalyse: Vergleich von mehr als zwei unabhängigen Stichproben; normalverteilt; Varianzhomogenität vorausgesetzt
          - Post-hoc Test: bei signifikantem Ergebnis einer Varianzanalyse: welche Gruppen sind für die Signifikanz verantwortlich
        - H-Test nach Kruskal und Wallis bei nicht Normalverteilung und mehr als 2 Stichproben

Beziehung zwischen 2 Variablen II

- Korrelationen: Analyse des Zusammenhangs zwischen 2 Variablen
  - Mindestens ordinalskalierte aber dichotome Variablen (nicht mehr als 2 Kategorien)
    - je größer die eine Variable, desto größer die andere
    - je größer die eine Variable, desto kleiner die andere
  - Produkt-Moment-Korrelation, Maßkorrelationskoeffizient oder Korrelationskoeffizient nach Pearson: klassische Korrelation zwischen 2 intervallskalierten und normalverteilten Variablen

- **Endliche Varianz (und Kovarianz)**: bei Erhöhung des Stichprobenumfangs darf sich die Variabilität nicht immer weiter erhöhen, sondern sollte sich stabilisieren. Bei Variablen, die bivariat normalverteilt sind, ist diese Voraussetzung automatisch gegeben. Der Korrelationskoeffizient ist damit auch gleichzeitig der *Maximum-Likelyhood Schätzer* des Korrelationskoeffizienten in der Grundgesamtheit (asymptotisch *erwartungstreu* und *effizient*).

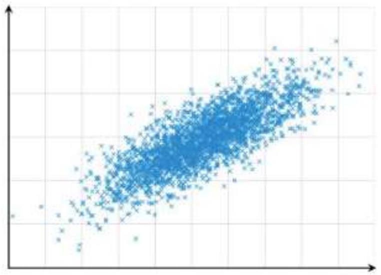
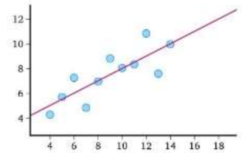
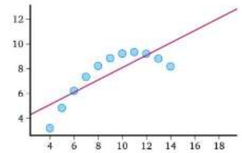


Abbildung 5: Endliche Varianz[2]

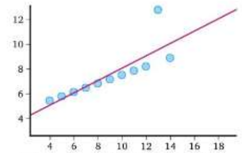
- **Linearität**: die Korrelation ist ein Maß für **lineare Abhängigkeit**. Abweichungen der Daten von dieser Linearitätsannahme führen zu einer mehr oder weniger starken Verzerrung des Korrelationskoeffizienten, wie in den nachfolgenden Beispielen gezeigt wird:



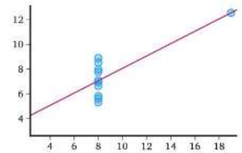
Die Variablen im ersten Diagramm scheinen normalverteilt zu sein. Ein Diagramm wie dieses würde man erwarten, wenn man davon ausgeht, dass beide Variablen (etwa) normalverteilt sind.



Während ganz klar auch hier ein Zusammenhang zwischen beiden Variablen besteht, ist dieser nicht linear. Hier sagt uns die Produkt-Moment-Korrelation zwar das ein Zusammenhang besteht, aber nur so weit, wie er sich linear abbilden lasst.



In diesem Diagramm besteht eine perfekte lineare Beziehung zwischen beiden Variablen, bis auf einen einzigen Ausreißer. Dies zeigt wie anfällig der Korrelationskoeffizient für das Vorhandensein von Ausreißern sein kann.



In diesem Beispiel wird gezeigt, wie ein Ausreißer den Korrelationskoeffizienten künstlich erhöhen kann, auch wenn die Beziehung zwischen beiden Variablen nicht linear ist.

Abbildung 6: Linearität und Korrelation

https://wgruber.github.io/Modellbildung2/korrelationen.html

Beziehung zwischen 2 Variablen III

- Rangkorrelation nach Spearman
  - zwischen 2 mind. ordinalskalierten Variablen
  - Oder 2 intervallskalierte und nicht normalverteilte Variablen
- Rangkorrelation nach Kendall
  - Zwischen 2 mindestens ordinalskalierten Variablen
- Intraclass Correlation Coefficient ICC
  - Dann zu verwenden, wenn die Übereinstimmung von 2 Variablen nicht nur bezüglich ihrer Richtung bestimmt werden soll, sondern auch bezüglich des mittleren Niveaus (Mittelwert der geschätzten Werte) der beiden Variablen
  - Diese Berechnung ist auf mehr als 2 Variablen ausdehnbar
- Vierfelderkorrelation
  - Zwischen 2 dichotomen Variablen (zB Zusammenhang zwischen Alkohol- und Nikotinkonsum in einer Stichprobe)
- Punktbiserale Korrelation
  - Zwischen einer dichotomen und einer intervallskalierten, normalverteilten Variablen
- Partielle Korrelation
  - Wenn ein formaler, nicht aber ein kausaler Zusammenhang besteht – Zusammenhang wird von einer anderen Variable mitbestimmt, die in gleicher Höhe sowohl mit der einen als auch der anderen Variable korreliert
  - Partielle Korrelation bietet die Möglichkeit, solche Störvariablen, die eine Scheinkorrelation erzeugen, auszuschließen

Beziehung zwischen 2 Variablen IV

- Regression
  - Korrelation = Stärke des Zusammenhangs zwischen 2 Variablen
  - Regression = Zusammenhang soll formelmäßig erfasst werden
    - Man versucht Formeln zu finden, nach denen aus der Kenntnis des Wertes der einen Variablen den zu erwartenden Wert der anderen (abhängigen) Variablen bestimmen kann
    - Unterscheidung zwischen linearen und nicht-linearen Zusammenhängen
      - Lineare Regression
        - Linearer Zusammenhang: Parameter a und b der Geradengleichung: $y=b*x+a$
        - Regressionsgerade ist die Gerade, für welche die Summe der Quadrate der Abweichungen aller Punkte von dieser Geraden ein Minimum wird – die abhängige Variable wird auf der y-Ordinate aufgetragen
        - Parameter b : Regressionskoeffizient – gibt Tangens des Steigungswinkels an
        - Parameter a: Ordinatenabschnitt; gibt den Punkt an, an dem die Regressionsgerade die y-Achse schneidet
        - In die Formeln für b und a gehen ausschließlich die Größen ein, die bereits bei der Berechnung des Produkt-Moment-Korrelationskoeffizienten anfallen

Starten Sie nun die Statistische Auswertung der Ergebnisse der Eye-Tracking Studie