Tutorium Datenauswertung

Klausurvorbereitung

Agenda

- Erinnerungsfotos: letzte Sitzung
- Werbung: Tutor*innen gesucht!
- Infos zur Klausur
- Infos zum Taschenrechner
- Modalität und Schiefe
- Häufigkeitstabellen, Median, Interquartilsabstände
- Regression
- Mittelwertunterschiede
- Korrelationen

Neue Tutor*innen gesucht!

– Voraussichtliche Eckdaten:
  - 2 Tutorien
  - 11 Stunden/Woche
  - 10,00€/Stunde
  - Bezahlung: Oktober bis März
– Aufgaben:
  - Tutorien vorbereiten und geben
  - Den Studis bei Fragen zur Verfügung stehen
  - Methodenaufgaben leiten und korrigieren
  - Klausur korrigieren

→ Alle Infos findet ihr in der PDF, die Frau Jasper herumgeschickt hat!

Infos zur Klausur I

- Datum: 08.07.2019
- Raum: F2
- Anwesenheit bis spätestens 9:55 Uhr
- Ausgabe der Klausuren: 10:00 Uhr
- Offizieller Beginn: 10:15 Uhr
- Bekanntgabe der Ergebnisse bis 20.07.2019

Infos zur Klausur II

- Klausureinsichten:
  - 26.08.2019 vormittags
  - 28.08.2019 nachmittags
- Nachklausur:
  - Montag, 16.09.2019, 10:00 Uhr, E231 (IfK)
  - Bekanntgabe der Ergebnisse: spätestens 23.09.2019
  - Einsicht: Sprechstunde Volker Gehrau
- Gedächtnisprotokolle zu bisherigen Klausuren gibt es bei der Fachschaft

Infos zur Klausur III

– Mitnehmen:
  - Studierendenausweis
  - Füller oder Kugelschreiber (kein Bleistift!)
  - Für Nicht-Muttersprachler: zweisprachiges Wörterbuch
– Die Klausur ist einseitig bedruckt, sodass ihr die Rückseiten als Schmierpapier benutzen könnt
– In der Klausur wird nur eure Matrikelnummer und eure Unterschrift verlangt, aber nicht euer Name

Infos zur Klausur IV

- Falls ihr runden müsst, dann immer auf 2 Nachkommastellen
- Richtige Rechnungen mit falschem Zwischenergebnis geben volle Punkte für die jeweilige Teilaufgabe
- Wurzeln, die nur noch eine einzelne reelle Zahl enthalten (Beispiel: $\sqrt{5}$), gelten bereits als Ergebnis und müssen nicht gezogen werden

Löscht zuletzt gemachte Eingabe

Löscht alle bisherigen Eingaben

Wechselt Vorzeichen

Erzeugt Dezimalzahl

Die Rechenregeln des Taschenrechners sind ähnlich wie die des Windows-Taschenrechners:
⊞ + R + calc + Enter

Taschenrechner

- Quadrieren: $a^2 = a * a$
- Verschachtelung von Formeln ist nicht möglich:

Eingabe 1	Eingabe 2	Eingabe 3	Eingabe 4	Eingabe 5	Eingabe 6	Rechnung	Bewertung
√	20	÷	5	=	[keine]	$\frac{20}{5}$	✕
20	÷	5	√	[keine]	[keine]	$\sqrt{\frac{20}{5}}$	✓

→ Kleinschrittig vorgehen und Gleichungen von innen nach außen lösen!

Taschenrechner – Beispiele

Formel: $\sqrt{\frac{8}{2}+\frac{2}{4}}$

Berechnung:

$$8 \div 2 = 4$$

$$2 \div 4 = 0{,}5$$

$$4 + 0{,}5 = 4{,}5$$

$$4{,}5\surd{=}2{,}12\ldots$$

Formel: $\sqrt{\frac{4+5}{2+3}+\frac{2}{4}}$

Berechnung:

$$4 + 5 = 9$$

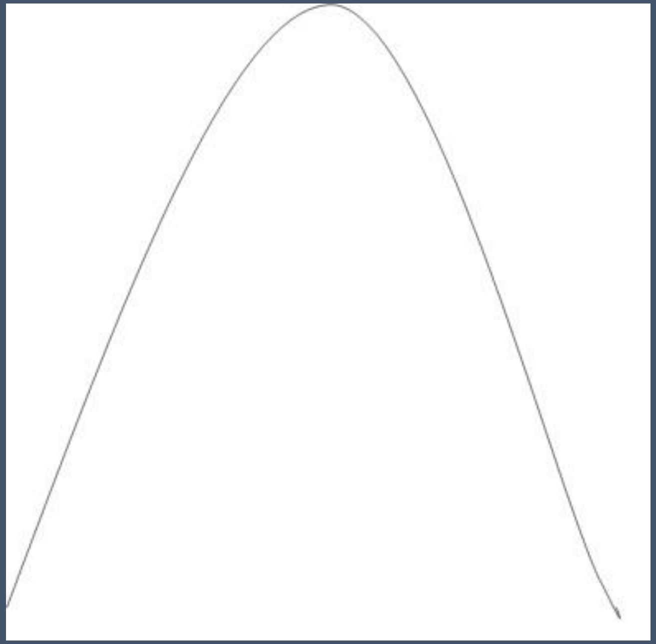
$$2 + 3 = 5$$

$$9 \div 5 = 1{,}8$$

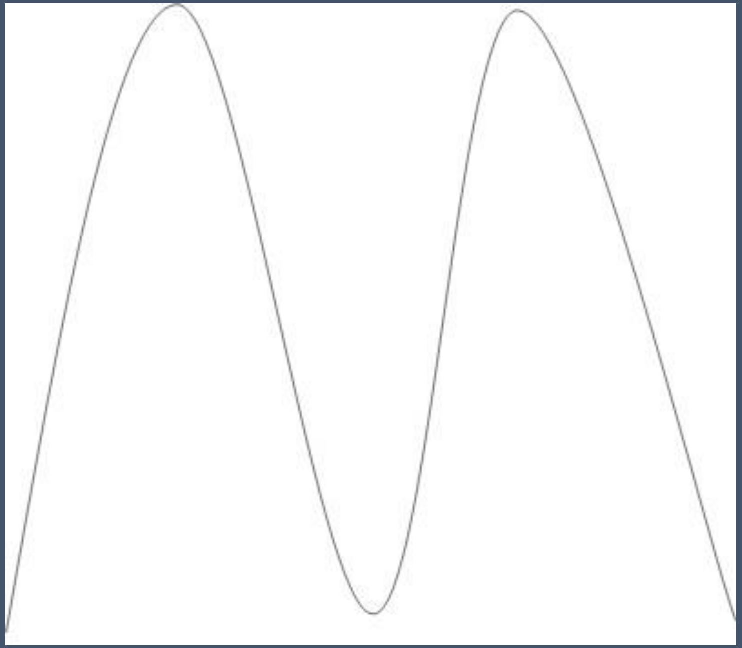
$$2 \div 4 = 0{,}5$$

$$1{,}8 + 0{,}5 = 2{,}3$$

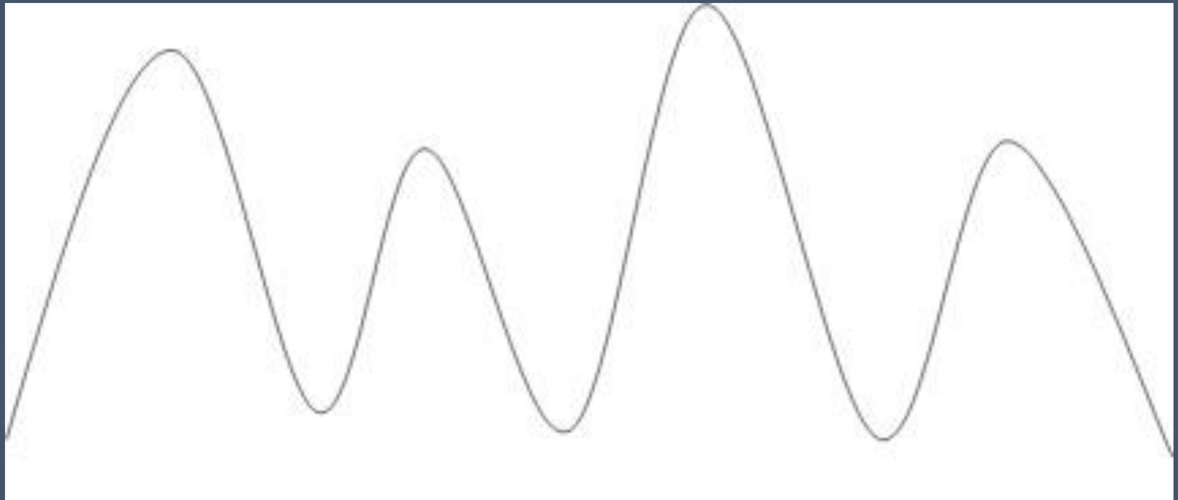
$$2{,}3\surd{=}\ 1{,}51\ \ldots$$



Unimodal

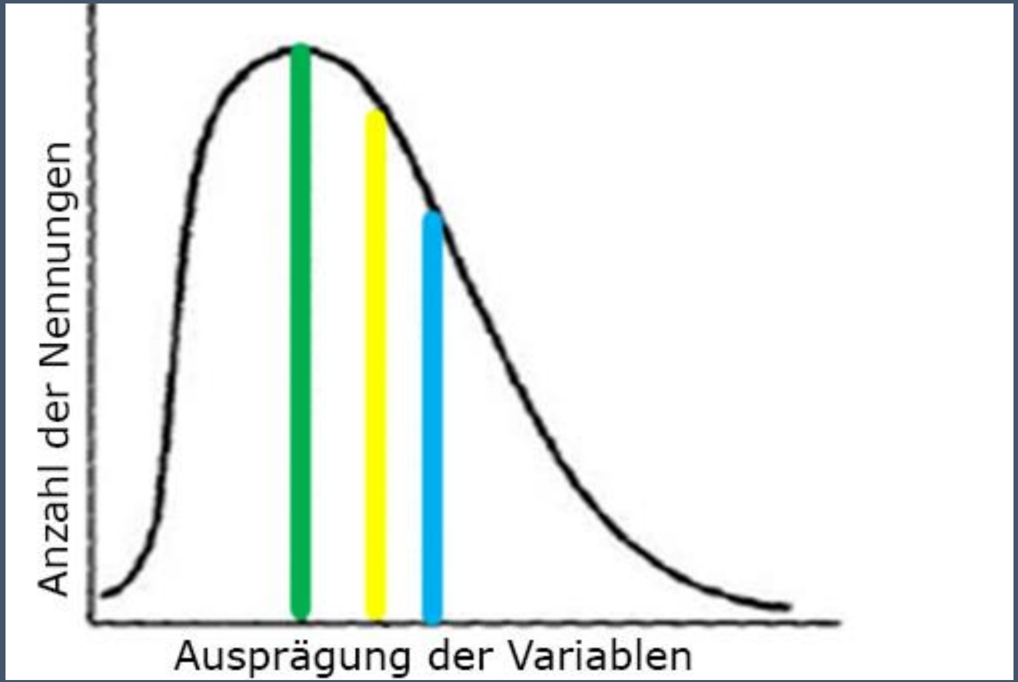


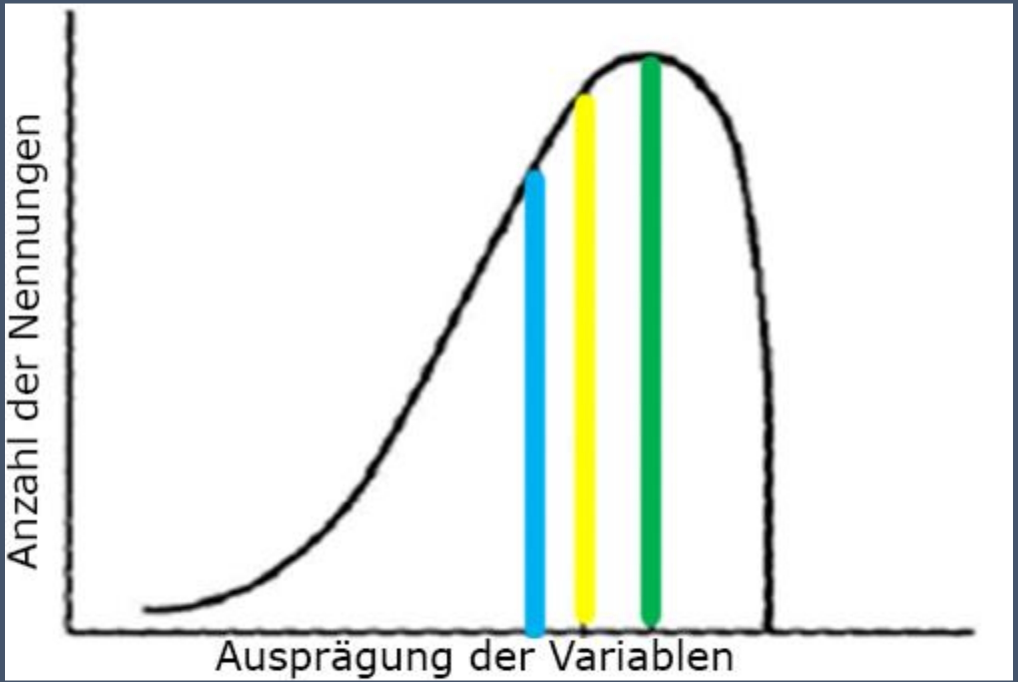
bimodal



multimodal

Schiefe





Rechtsschief: Modus < Median < Mittelwert

Linksschief: Mittelwert < Median < Modus

Bezogen auf die Ausprägung der Variablen!

→ Der Modus ist überhaupt nicht ausreißeranfällig. Der Median lässt sich von Ausreißern beeinflussen, aber nicht stark. Der Mittelwert lässt sich sehr stark von Ausreißern beeinflussen!

Häufigkeitstabellen – Basics

Häufigkeitstabellen

– Bestehen aus vier Spalten:

Spalte	**Beantwortung der Frage…**
Ausprägungen	Welche Ausprägungen kann die Variable annehmen?
Anzahl	Wie oft wurde die Ausprägung realisiert (angekreuzt)?
Prozent	Wieviel Prozent der Befragten haben die Ausprägung realisiert?
Kumulierte Prozent	Wieviel Prozent der Befragten haben diese Ausprägung oder eine von den Ausprägungen in den Zeilen davor realisiert?

– Können fehlende Angaben (NAs) enthalten oder nicht

Median und Interquartilsabstände bei gerader Anzahl an Fällen

Fall	1	2	3	4	5	6	7	8	9	10
Ausprägung	30	15	15	30	45	60	75	45	60	60

Ausprägung	Anzahl	Prozent	Kumulierte Prozent
15	2	2/10*100 = 20	20
30	2	20	40
45	2	20	60
60	3	30	90
75	1	10	100

Erste Ausprägung über 25%:
erstes Quartil = 30

Erste Ausprägung über 50%:
Median = 45

Erste Ausprägung über 75%:
drittes Quartil = 60

Mittelwertunterschiede – Übersicht

- Mittelwertunterschiede
  - Zwei verschiedene Studien
    - Vergleich der Intervalle beider Mittelwerte
  - Höchstens zwei Gruppen
    - Verschiedene Personen
      - t-Test für unabhängige Stichproben
    - Dieselben Personen
      - t-Test für abhängige Stichproben
  - Mehr als zwei Gruppen
    - F-Test (einfaktorielle Varianzanalyse)

$CI_{\bar{x}} = \bar{x} \pm 2 * SE_{\bar{x}}$

$CI_{\bar{x}_1-\bar{x}_2} = \bar{x}_1 - \bar{x}_2 \pm 2 * SE_{\bar{x}_1-\bar{x}_2}$

$CI_{x-y} = \bar{x}_{x-y} \pm 2 * SE_{x-y}$

[nur mit R]

Regressionskoeffizienten

Koeffizient	Zeichen	Interpretation
Regressionskoeffizient	b	Wenn die UV um eine $Einheit_{UV}$ steigt, steigt die AV um b $Einheiten_{AV}$
Betakoeffizient	β	Wenn die UV um eine $Standardabweichung_{UV}$ steigt, steigt die AV um β $Standardabweichungen_{AV}$

UV: Alkoholpegel (in ‰)
AV: Lautstärke (in dB)

b = 20

Wenn der Alkoholpegel um ein Promille steigt, steigt (nach dem Regressionsmodell) die Lautstärke um 20 Dezibel.

β = 0,7

Wenn der Alkoholpegel um eine Standardabweichung steigt, steigt (nach dem Regressionsmodell) die Lautstärke um 0,7 Standardabweichungen.

Mittelwertunterschiede – Aufgabe

Aufgabe: In einem Experiment wurde die Wirkung von Psilocybin im Rahmen von psychotherapeutischen Sitzungen erforscht, indem die Versuchspersonen nach dem Erfolg der Therapie (1-10) gefragt wurden. Nennen Sie die dem Experiment zugehörige Hypothese und überprüfen Sie diese anhand der untenstehenden Daten.

Fall	1	2	3	4	5	6	7	8	9	10
Versuchsgruppe	2	9	4	1	3	5	7	3	9	2
Kontrollgruppe	9	1	4	6	8	1	9	10	1	10

$\bar{x}_1 = 4{,}5, \quad s^2{}_1 \approx 8{,}5$

$\bar{x}_2 = 5{,}9, \quad s^2{}_2 \approx 14{,}77$

$$SE_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s^2{}_1}{n_1} + \frac{s^2{}_2}{n_2}}$$

$$CI_{\bar{x}_1-\bar{x}_2} = \bar{x}_1 - \bar{x}_2 \pm 1{,}96 * SE_{\bar{x}_1-\bar{x}_2}$$

Mittelwertunterschiede: Lösung

H_1: Die psychotherapeutische Behandlung mit Psilocybin ist erfolgreicher als die psychotherapeutische Behandlung ohne Psilocybin.

H_0: Die psychotherapeutische Behandlung mit Psilocybin ist nicht erfolgreicher als die psychotherapeutische Behandlung ohne Psilocybin.

$\bar{x}_1 = 4{,}5, \qquad s^2{}_1 \approx 8{,}5$

$\bar{x}_2 = 5{,}9, \qquad s^2{}_2 \approx 14{,}77$

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{8{,}5}{10} + \frac{14{,}77}{10}} \approx 1{,}53$$

$$CI_{\bar{x}_1 - \bar{x}_2} = 4{,}5 - 5{,}9 \pm 2 * 1{,}53 \rightarrow [-4{,}45; 1{,}65]$$

Das Konfidenzintervall der Mittelwertdifferenzen enthält den Wert 0. Die Mittelwerte zwischen Gruppe 1 und Gruppe 2 unterscheiden sich also nicht signifikant. Damit ist die Forschungshypothese falsifiziert!

Pearson's r – Aufgabe

<u>Aufgabe</u>: Stellen Sie in Bezug auf die unten stehenden Variablen eine Hypothese auf und überprüfen Sie diese anhand des Korrelationskoeffizienten r.

$$cov\,(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) * (y_i - \bar{y})}{n-1}, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \mathrm{r} = \frac{cov\,(x,y)}{s_x * s_y}, \qquad s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Fall i	1	2	3	4	5
Temperatur x_i	30	20	0	40	15
Wasserkonsum y_i	2	2	1	3	1

Pearson's r – Lösung 0: Hypothesen

Fall i	1	2	3	4	5
Temperatur x_i	30	20	0	40	15
Wasserkonsum y_i	2	2	1	3	1

Pearson's r kann nur Zusammenhangshypothesen prüfen. Deshalb müssen wir auch eine Zusammenhangshypothese formulieren:

H_1: Es gibt einen Zusammenhang zwischen der Temperatur (in Grad Celsius) und dem Wasserkonsum (in Litern).

H_0: Es gibt keinen Zusammenhang zwischen der Temperatur (in Grad Celsius) und dem Wasserkonsum (in Litern).

Pearson's R – Lösung I: Mittelwert und Standardabweichung (x)

Fall i	1	2	3	4	5
Temperatur x_i	30	20	0	40	15

$$\bar{x} = \frac{30 + 20 + 0 + 40 + 15}{5} = 21$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{920}{5-1}} \approx 15{,}17$$

Fall i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	30 – 21 = 9	9² = 81
2	30 – 21 = -1	-1² = 1
3	0 – 21 = -21	-21² = 441
4	40 – 21 = 19	19² = 361
5	15 – 21 = -6	-6² = 36
		$\sum = 920$

Pearson's R – Lösung II: Mittelwert und Standardabweichung (y)

Fall i	1	2	3	4	5
Wasserkonsum y_i	2	2	1	3	1

$$\bar{y} = \frac{2+2+1+3+1}{5} = 1{,}8$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{2{,}8}{5-1}} \approx 0{,}84$$

Fall i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	2 – 1,8 = 0,2	0,2² = 0,04
2	2 – 1,8 = 0,2	0,2² = 0,04
3	1 – 1,8 = -0,8	-0,8² = 0,64
4	3 – 1,8 = 1,2	1,2² = 1,44
5	1 – 1,8 = -0,8	-0,8² = 0,64
		$\sum = 2{,}8$

Pearson's r – Lösung III: Kovarianz (x,y)

Fall i	1	2	3	4	5
Temperatur x_i	30	20	0	40	15
Wasserkonsum y_i	2	2	1	3	1

Fall i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) * (y_i - \bar{y})$
1	30 – 21 = 9	2 – 1,8 = 0,2	9 * 0,2 = -1,8
2	30 – 21 = -1	2 – 1,8 = 0,2	-1 * 0,2 = -0,2
3	0 – 21 = -21	1 – 1,8 = -0,8	-21 * (-0,8) = 16,8
4	40 – 21 = 19	3 – 1,8 = 1,2	19 * 1,2 = 22,8
5	15 – 21 = -6	1 – 1,8 = -0,8	-6 * (-0,8) = 4,8
			$\sum = 46$

$$cov\,(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) * (y_i - \bar{y})}{n-1} = \frac{46}{5-1} = 11,5$$

$x_i - \bar{x}$ und $y_i - \bar{y}$ haben wir bereits für die Varianzen berechnet – einfach übernehmen!

Pearson's r – Lösung IV: Standardisierung, Interpretation

$cov\ (x, y) = 11{,}5, \qquad s_x \approx 15{,}17, \qquad s_y \approx 0{,}84$

$$\text{r} = \frac{cov\ (x,y)}{s_x * s_y} = \frac{11{,}5}{15{,}17 * 0{,}84} \approx 0{,}90$$

→ Es gibt einen <u>extrem starken positiven Zusammenhang</u> zwischen den Variablen Temperatur und Wasserkonsum! Die Hypothese ist daher nicht falsifiziert.

FRAGEN?

Danke für das tolle Semester mit euch!