

Übungsaufgaben – Sitzung 4 – Lösungsskizze

Aufgabe 1

Sie führen eine Umfrage zum Vertrauen in die Europäische Union (EU) durch. Dafür stellen Sie den Teilnehmer*innen die Frage „Wie stark vertrauen Sie der Europäischen Union?“. Die Befragten können dabei Ihr Vertrauen auf einer Skala von 1 bis 100 angeben, wobei höhere Ausprägungen ein stärkeres Vertrauen in die EU bedeuten.

Im Folgenden sehen Sie die Ergebnisse für 10 Teilnehmer*innen:

Teilnehmer*in	A	B	C	D	E	F	G	H	I	J
Vertrauen	67	59	78	92	81	999	8	76	61	73

- a) Welches Skalenniveau hat die Variable „Vertrauen in die EU“ in der oben beschriebenen Form? Welche Mittelwerte dürften Sie auf dieser Grundlage berechnen?
- Skalenniveau: **metrisch**
 - Berechenbare Mittelwerte: **Arithmetisches Mittel, Modus, Median**
- b) Berechnen Sie nun alle erlaubten Mittelwerte und interpretieren Sie Ihre Ergebnisse. Worauf müssen Sie bei der Berechnung achten?

- Fall F hat die Ausprägung 999, d.h. der*die Befragte hat keine Angabe gemacht. Fall F muss aus der Berechnung ausgeschlossen werden. Es verbleiben neun Fälle.
- Arithmetisches Mittel: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{9} (67 + 59 + 78 + 92 + 81 + 8 + 76 + 61 + 73) = \frac{1}{9} * 595 = \mathbf{66,11}$
- Median: Datenmenge muss zunächst geordnet werden. Median ist der Wert, von dem links und rechts jeweils 50% der Werte liegen

G	B	I	A	J	H	C	E	D
8	59	61	67	73	76	78	81	92

- Median: $\tilde{x} = \mathbf{73}$
 - Modus:
 - Häufigster Wert in einer Datenmenge
 - Da jeder Wert nur einmal vorkommt, ist auch jeder Wert ein Modus
 - Interpretation
 - Das Vertrauen in die EU ist über die Stichprobe hinweg relativ hoch (siehe Median).
 - Der Median liegt deutlich über dem arithmetischen Mittelwert. Das arithmetische Mittel ist also nach unten verzerrt. Fall G stellt einen Ausreißer dar. Würde man nur das arithmetische Mittel betrachten, dann würde das Vertrauen unterschätzt, da das arithmetische Mittel ausreißersensitiv ist. Abbildung mit Median sinnvoll, da dieser ausreißerresistent ist.
 - Es gäbe neun Modalwerte. Bei metrisch skalierten Variablen ist der Modalwert oft wenig aussagekräftig.
- c) Welche Angaben bräuchten Sie, um die Verteilung anhand eines Boxplots zu visualisieren? Nennen Sie die Angaben und berechnen Sie diese.
- Boxplot wird auch als 5-Punkte Maß bezeichnet

- Niedrigster Wert, Höchster Wert (Spannweite)
- Unteres Quartil, Oberes Quartil (Interquartilsabstand)
- Median
- Niedrigster Wert: 8
- Höchster Wert: 92
- Spannweite: $92 - 8 = 84$
- Quartile
 - $x_{.25}$ = Links von diesem Wert befinden sich in einer geordneten Datenmenge 25% aller Werte der Verteilung
 - $x_{.75}$ = Links von diesem Wert befinden sich in einer geordneten Datenmenge 75% aller Werte der Verteilung
 - 1. Werte ordnen

G	B	I	A	J	H	C	E	D
8	59	61	67	73	76	78	81	92

- Zur Berechnung der Quartile werden jeweils die beiden Hälften der Datenmenge, die durch den Median separiert werden, zugrunde gelegt. Davon wird jeweils der Median für die Quartile gewählt.
- $x_{.25}$

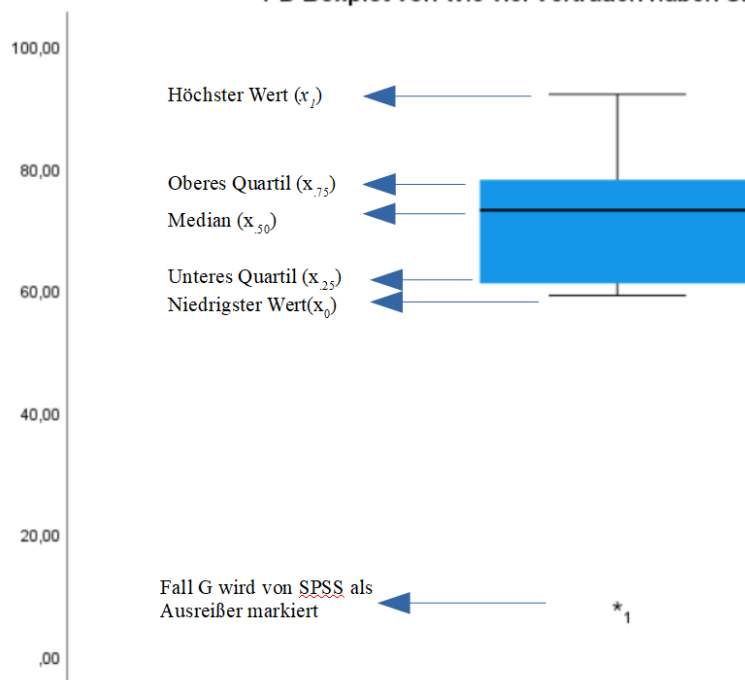
G	B	I	A
8	59	61	67

 - Median dieser Datenhälfte, also $x_{.25}$, liegt zwischen 59 und 61.
 - $x_{.25} = \frac{59+61}{2} = 60$
- $x_{.75}$

H	C	E	D
76	78	81	92

 - Median dieser Datenhälfte, also $x_{.75}$, liegt zwischen 78 und 81.
 - $x_{.75} = \frac{78+81}{2} = 79,5$
 - Interquartilsabstand: $IQR = x_{.75} - x_{.25} = 79,5 - 60 = 19,5$
- So sähe der Boxplot aus (nicht gefordert)

1-D Boxplot von Wie viel Vertrauen haben Sie in die EU



d) Berechnen Sie die Varianz und die Standardabweichung. Welches der beiden Maße lässt sich hier besser interpretieren? Wie interpretieren Sie Ihr Ergebnis?

- $var = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{9} ((67 - 66,11)^2 + (59 - 66,11)^2 + (78 - 66,11)^2 + (92 - 66,11)^2 + (81 - 66,11)^2 + (8 - 66,11)^2 + (76 - 66,11)^2 + (61 - 66,11)^2 + (73 - 66,11)^2) = \frac{1}{9} (0,79 + 50,55 + 141,37 + 670,29 + 221,72 + 3376,77 + 97,81 + 26,11 + 47,47) = \frac{1}{9} * 4632,87 = \mathbf{514,76}$
- Hinweis: Durch die Rundung der Zwischenschritte ergibt sich eine leichte Abweichung. Ohne gerundete Zwischenschritte käme man auf $var = 515,06$.
- $StA = \sqrt{var} = \sqrt{514,76} = \mathbf{22,69}$
- Die Standardabweichung ist einfacher zu interpretieren. Im Durchschnitt streuen die Werte um ca. 23 vom arithmetischen Mittelwert. Es scheint als streuten die Werte relativ weit und die Verteilung sei sehr heterogen. Aber Vorsicht: Die Standardabweichung beruht auf der Varianz, welche wiederum auf dem arithmetischen Mittel aufbaut. Alle diese Maße sind damit ausreißersensitiv. Fall G mit seiner außergewöhnlich niedrigen Abweichung verzerrt alle diese Maße. Zum Vergleich: Schlösse man Fall G aus, läge die Standardabweichung bei 10,49. Die Verteilung erschiene damit deutlich homogener.

e) Sie überlegen nun, die Ergebnisse der Umfrage grafisch darzustellen. Welche Form der Darstellung sollten Sie wählen? Welches Problem könnte aufkommen? Erarbeiten Sie eine Lösung und diskutieren Sie diese kritisch.

- Variable ist metrisch, deswegen müsste ein Histogramm angefertigt werden.
- Problem: wenige Fälle, viele Ausprägungen, Grafik wäre wenig aussagekräftig. Histogramm sähe so aus (nicht gefordert).



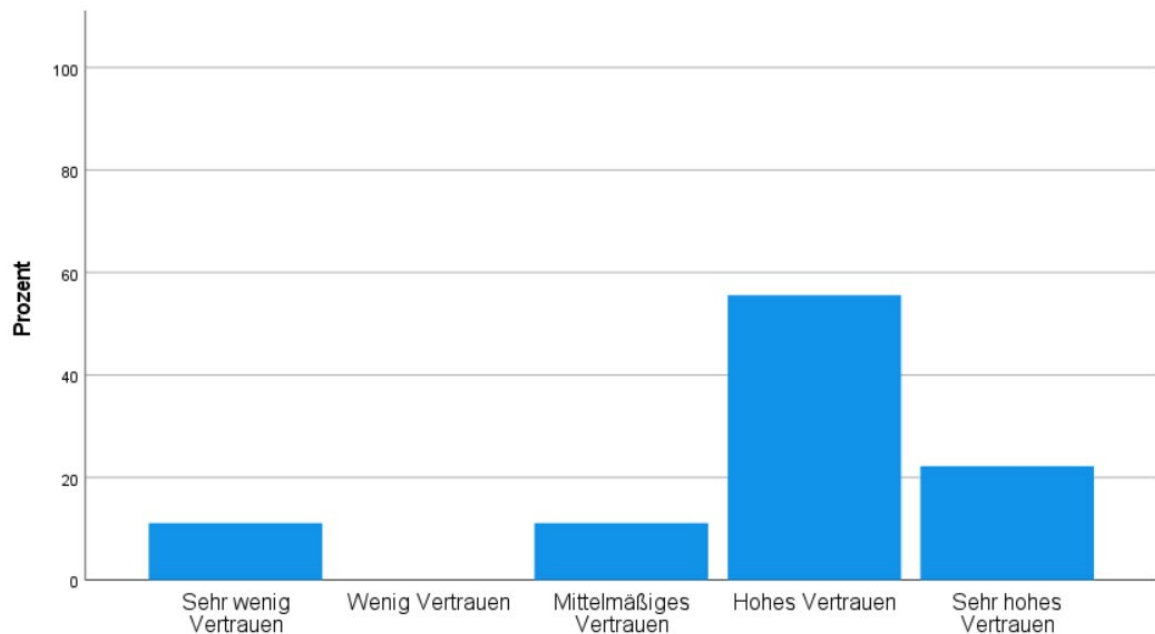
- Mögliche Lösung: Gruppierung. Vorschlag: Gruppierung in fünf Kategorien:

Alte Ausprägung	Neue Ausprägung	
	Beschreibung	Quantitative Zuordnung
1 - 20	Sehr wenig Vertrauen	1
21 - 40	Wenig Vertrauen	2
41 - 60	Mittleres Vertrauen	3
61 - 80	Hohes Vertrauen	4
81 - 100	Sehr hohes Vertrauen	5

- Verteilung sähe nun so aus

Teilnehmer*in	A	B	C	D	E	G	H	I	J
Vertrauen	4	3	4	5	5	1	4	4	4

- Mit der Gruppierung geht ein Abstieg des Skalenniveaus einher und damit auch ein Informationsverlust. Die Variable ist nun nur noch ordinal. Man kann nun nur noch erkennen, dass z.B. Fall D ein höheres Vertrauen in die EU hat als Fall B. Wie viel höher das Vertrauen ist, ist unklar. Dafür kann nun aber ein Säulendiagramm erstellt werden, das aufschlussreicher ist, da es zeigt, dass sich die Werte in der Kategorie „hohes Vertrauen“ ballen.



Aufgabe 2

Öffnen Sie den Datensatz, der die Ergebnisse der Prä-Evaluation der Vorlesung von 2016 enthält, mit SPSS.

- a) Sie interessiert, welche Note die Studierenden ihrem Studium damals gegeben haben. Suchen Sie die Variable „Note für das Studium“.
- a. Auf welchem Skalenniveau ist die Variable gemessen? Und welche möglichen Ausprägungen hat die Variable?
- Skalenniveau: **Ordinal**
 - Ausprägungen: Dafür bei Variable E8 in der Spalte „Werte“ in das blaue Kästchen mit den drei Punkten klicken. Information hier: Ausprägung 1 = „sehr gut“, Ausprägung 5 = „nicht bestanden“.
- b. Lassen Sie sich die absoluten, relativen und kumulierten Häufigkeiten ausgeben.
- *Analysieren > Deskriptive Statistiken > Häufigkeiten*. Dort Variable hinzufügen. Haken bei „Häufigkeitstabellen anzeigen“ setzen. Mit „Ok“ bestätigen.
 - Jetzt werden auch alle möglichen Ausprägungen sichtbar. 1 bis 5 jeweils mit Zwischenschritten.
- c. Wie viele Studierende haben dem Studium die Note 2,0 gegeben? Wie viele Studierende haben ihrem Studium die Note 3,0 oder eine bessere Note gegeben? Woran sehen Sie das?
- 54,2% der befragten Studierenden bewerteten ihr Studium mit 2,0 (Spalte im Output: *Gültige Prozente*).
 - 95,2% der befragten Studierenden bewerteten ihr Studium mit der Note 3,0 oder besser. Abzulesen an der Spalte *Kumulierte Prozente*.
- d. Aufbauend auf Ihrer Überlegung zum Skalenniveau der Variable: Welche Mittelwerte dürfen berechnet werden? Berechnen Sie diese mithilfe von SPSS und interpretieren Sie diese.
- Da ordinal: Modus und Median
 - *Analysieren > Deskriptive Statistiken > Häufigkeiten*. Dort die Variable hinzufügen. Dann rechts auf *Statistiken* klicken. Im erscheinenden Pop-Up-Fenster die Haken bei *Modalwert* und *Median* setzen. Bestätigen. Dann auf *OK* klicken.
 - Modus $h = 2$; Median $\tilde{x} = 2$
- b) Ferner interessiert Sie, wie gut sich die Studierenden, hinsichtlich ihrer eigenen Leistung eingeschätzt haben. Im Datensatz finden Sie eine Variable, mit der gemessen wurde, wo die Studierenden sich relativ zu den anderen einordnen („Ich würde mich unter die Top ?? meines Jahrgangs einordnen“).
- a. Welches Skalenniveau liegt hier vor? Wie weit reicht die Skala?
- Skalenniveau: metrisch
 - *Analysieren > Deskriptive Statistiken > Häufigkeiten*. Dort Variable hinzufügen. Haken bei „Häufigkeitstabellen anzeigen“ setzen. Außerdem unter *Statistiken* die Häkchen bei *Minimum* und *Maximum* setzen. Mit *Ok bestätigen*.
 - Anhand des Outputs kann man u.a. die möglichen Ausprägungen erkennen. Die Variable hat die minimale Ausprägung 10, die maximale Ausprägung 100.

- b. Welche Mittelwerte dürften berechnet werden? Berechnen Sie mit Hilfe von SPSS die Mittelwerte und geben das Ergebnis an. Sofern erlaubt, berechnen Sie auch die Quartile, die Varianz und Standardabweichung. Wie interpretieren Sie Ihr Ergebnis?
- Da metrisch: Alle Mittelwerte berechenbar: *Analysieren > Deskriptive Statistiken > Häufigkeiten*. Dort die Variable hinzufügen. Dann rechts auf *Statistiken* klicken. Im erscheinenden Pop-Up-Fenster die Haken bei *Modalwert, Median, Mittelwert, Quartile, Varianz und Standardabweichung* setzen. Bestätigen. Dann auf *OK* klicken.
 - Arithmetisches Mittel: $\bar{x} = 45,2$; Median: $\tilde{x} = 40$; Modus: $h = 50$;
 - Median und arithmetisches Mittel liegen nah beieinander. Bedeutende Ausreißer sind eher unwahrscheinlich
 - $x_{.25} = 30$; $x_{.75} = 50$; Daraus folgt: IQR = 20; Quartile liegen recht nah beieinander. Wahrscheinlich eher homogene Verteilung.
 - $Var = 439,04$; $StA = 20,96$. Standardabweichung relativ zum Mittelwert nicht so groß. Auch Indiz für eher homogene Verteilung.
- c. Fertigen Sie dann mit der Software ein Boxplot an. Welche Schlüsse ziehen Sie aus der Grafik?
- *Grafik > Diagrammerstellung*. Die erscheinende Meldung mit *OK bestätigen*. Links unten *Boxplot* auswählen. Dann rechts *1D-Boxplot* anklicken und mit gedrückter Maustaste nach oben ziehen. Oben die Variable aus der Liste links in das Kästchen *Y-Achse* ziehen. Mit *OK* bestätigen.
 - Wenige Ausreißer oben. Die „Box“ ist eher eng. Auch hier sieht es wieder nach einer eher homogenen Verteilung aus.
- d. Welche Form der grafischen Darstellung wäre noch sinnvoll? Begründen Sie und erstellen Sie eine entsprechende Grafik.
- Da Skalenniveau metrisch: Histogramm.
 - *Grafik > Diagrammerstellung*. Die erscheinende Meldung mit *OK bestätigen*. Links unten *Histogramm* auswählen. Dann rechts *Einfaches Histogramm* anklicken und mit gedrückter Maustaste nach oben ziehen. Oben die Variable aus der Liste links in das Kästchen *X-Achse* ziehen. Mit *OK* bestätigen.
- c) ZUSATZ: Sie finden, die Messung der Variable zu feingliedrig. Gruppieren Sie neu, begründen und diskutieren ihre Gruppierung.

a. *Mögliche Gruppierung*

Alte Ausprägung	Neue Ausprägung	
	Beschreibung: Ich zähle mich zu	Quantitative Zuordnung
Bis 25	... den Leistungstärksten	1
26 - 50	... den Leistungsstarken	2
51 - 75	... dem Mittelfeld	3
76 - 100	... den Leistungsschwächeren	4

- Dafür: *Transformieren > Umcodieren in andere Variable*. Variable E4 hinzufügen. *Alte und Neue Werte* klicken. Im erscheinenden Pop-Up nun die Neucodierung vornehmen. Dafür zunächst *Bereich KLEINSTER bis Wert* anwählen und darunter 25

eingeben. Rechts bei *Neuer Wert* > *Wert 1* eingeben und dann *Hinzufügen*. Ähnlich für die weiteren Kategorien vorgehen. Um die folgenden beiden Kategorien hinzufügen, *Bereich* anwählen und jeweils 26 und 50 sowie 51 und 75 eingeben, dann jeweils rechts bei *Neuer Wert* > *Wert 2* und *3* eingeben und jeweils hinzufügen. Für die letzte Kategorie *Bereich*, *Wert bis GRÖSSTER* wählen und 76 angeben. Rechts dann 4 als neuen Wert eingeben, dann *Hinzufügen*. Unten nun *Weiter* klicken. Rechts nun einen *Namen* für die neue Variable eingeben. Zum Beispiel: E4gr. Bei *Beschriftung* nun spezifizieren. Zum Beispiel: „E4 in vier Kategorien gruppiert“. Dann *Ändern* und *OK klicken*. In der Variablenansicht sollte die neue Variable E4gr nun auftauchen. Unter *Werte* und dem blauen Kästchen mit den drei Punkten können die Ausprägungen neu beschriftet werden.