

Statistische Inferenz | 02 | Zentraltendenz, Streuung, Standardfehler

Musterlösung

Prof. Dr. Roland Schäfer | Germanistische Linguistik FSU Jena

3. November 2024

Hinweis: Wo nicht anders angegeben, runden Sie die Ergebnisse auf zwei Nachkommastellen.

1 Skalenniveaus

Bestimmen Sie das Skalenniveau von folgenden Messgrößen:

1. Prozentwerte **Verhältnisskala**
2. Wortfrequenz-Rang (häufigstes Wort, ..., seltenstes Wort) **Ordinalskala**
3. Kasus **Nominalskala**
4. Geschwindigkeit **Intervallskala**
5. Akzentsitz (Erstsilbe, Mittelsilbe, Endsilbe) **Nominalskala**
6. Satzlänge, gemessen in Wörtern **Intervallskala**
7. Frequenz eines Wortes im Korpus (absolute Zahl) **Intervallskala**
8. Höhe über NN **Verhältnisskala**
9. DSH-Prüfungsniveau (I – III) **Ordinalskala**
10. Verhältnis Satzlänge in Wörtern zu Wortlänge in Silben in einem Text **Verhältnisskala**
11. Wortklasse (= Wortart) **Nominalskala**
12. Beschleunigung **Verhältnisskala**
13. Textniveau (leicht, mittel, schwer) **Ordinalskala**
14. Frequenz eines Wortes im Korpus pro eine Millionen Wörter **Intervallskala**
15. Textsorte **Nominalskala**

2 Modus und Median

Ermitteln Sie den Modus und wo möglich den Median für folgende Messreihen von Hand (ohne Software):

1. $x = [\text{Nom, Akk, Akk, Akk, Nom, Dat, Gen, Nom, Nom, Akk, Dat, Dat, Akk, Akk}]$
Modus: 6, Median: n. d.
2. $x = [4, 5, 3, 3, 3, 2, 1, 2, 2, 1, 5, 4, 2, 2, 1, 3, 2]$
Modus: 2, Median: 2
3. $x = [4.3, 5.0, 3.0, 3.3, 3.7, 2.3, 1.3, 2.7, 2.0, 1.0, 5.0, 4.3, 2.0, 2.0, 1.3, 3.0, 2.7]$
Modus: 2.0, Median: 2.7

3 Mittel und Streuung

Ermitteln Sie von Hand für die untenstehenden Messreihen das arithmetische Mittel, die Varianz und die Standardabweichung:

1. $x = [2.73, 1.85, 21.24, 17.97, 5.49, 18.90, 12.46, 0.97, 6.45, 7.43]$
 $\bar{x} = 9.55, \text{var}(x) = 57.05, \text{stdev}(x) = 7.55$
2. $x = [1.00, 1.91, 3.12, 4.38, 4.72, 5.29, 3.82, 3.25, 2.04, 0.93]$
 $\bar{x} = 3.05, \text{var}(x) = 2.36, \text{stdev}(x) = 1.54$
3. $x = [1.07, 1.06, 0.94, 1.84, 3.04, 3.22, 4.18, 5.27, 6.27, 6.75]$
 $\bar{x} = 3.36, \text{var}(x) = 4.79, \text{stdev}(x) = 2.19$

4 z-Werte und Standardfehler

Ermitteln Sie für die Messreihen aus Aufgabe 3 die z-Werte für die Messpunkte und die Standardfehler von Hand. Formulieren Sie in eigenen Worten (jeweils ein Satz), was z-Werte und Standardfehler angeben.

1. z-Werte: $[-0.9, -1.02, 1.55, 1.11, -0.54, 1.24, 0.39, -1.14, -0.41, -0.28]$, $SF = 2.39$
2. z-Werte: $[-1.33, -0.74, 0.05, 0.87, 1.09, 1.46, 0.5, 0.13, -0.66, -1.38]$, $SF = 0.49$
3. z-Werte: $[-1.05, -1.05, -1.11, -0.7, -0.15, -0.07, 0.37, 0.87, 1.33, 1.55]$, $SF = 0.69$

Die z-Werte sind die in Standardabweichungen normierten Abweichungen der Messwerte vom Mittelwert der Stichprobe. Der Standardfehler ist die mittlere Abweichung des beobachteten Stichprobenmittels vom wahren Populationsmittel in wiederholten Stichproben der gegebenen Größe (falls die Varianz in der Population der Varianz in der Stichprobe entspricht). (Die „gegebene Größe“ ist die Größe der tatsächlichen Stichprobe.) Vereinfacht gesagt ist er der Erwartungswert, um den Stichproben der vorliegenden Größe vom wahren Mittelwert abweichen.

5 Konfidenzintervalle (Anteilswerte)

5.1 Berechnung des Konfidenzintervalls für Anteilswerte

Berechnen Sie für folgende Anteilswerte (q) die Konfidenzintervalle bei den Stichprobengrößen $n = 10$ und $n = 100$ auf den Konfidenzniveaus $\alpha = 0.9$ und $\alpha = 0.99$ (also je vier Mal den unteren und oberen Wert des Konfidenzintervalls). Die kritischen Werte der Normalverteilung entnehmen Sie bitte der zur Verfügung gestellten Tabelle. Nachkommastellen. Runden Sie auf drei Nachkommastellen.

1. $q = 0.21$
2. $q = 0.49$
3. $q = 0.89$

q	n=10, $\alpha=0.9$	n=100, $\alpha=0.9$	n=10, $\alpha=0.99$	n=100, $\alpha=0.99$
0.21	$[-0.002, 0.422]$	$[0.143, 0.277]$	$[-0.122, 0.542]$	$[0.105, 0.315]$
0.49	$[0.23, 0.75]$	$[0.408, 0.572]$	$[0.083, 0.897]$	$[0.361, 0.619]$
0.89	$[0.727, 1.053]$	$[0.839, 0.941]$	$[0.635, 1.145]$	$[0.809, 0.971]$

5.2 Schlechte Praxis beim Berichten von Konfidenzintervallen

Warum hätte folgende Tabelle ganz ohne Nachrechnen nicht gedruckt werden dürfen?

Measure	<i>M</i>	<i>SD</i>	95% CI	
			Lower	Upper
Age at testing (years)	20.23	2.94	19.59	20.88
Age of onset of L2 learning (years)	5.13	1.78	5.74	5.53

Ingrid Mora-Plaza, Joan C. Mora, Mireia Ortega and Cristina Aliaga-Garcia. Is L2 pronunciation affected by increased task complexity in pronunciation-unfocused speaking tasks? *Studies in Second Language Acquisition*. First View. <https://doi.org/10.1017/S0272263124000470>