Statistische Inferenz | 02 | Zentraltendenz, Streuung, Standardfehler

Prof. Dr. Roland Schäfer | Germanistische Linguistik FSU Jena

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

2 Modus und Median

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

1. $x_1 = [\text{Nom, Akk, Akk, Akk, Nom, Dat, Gen, Nom, Nom, Akk, Dat, Dat, Akk, Akk}]$

2. $x_2 = [4, 5, 3, 3, 3, 2, 1, 2, 2, 1, 5, 4, 2, 2, 1, 3, 2]$

3. $x_3 = [4.3, 5.0, 3.0, 3.3, 3.7, 2.3, 1.3, 2.7, 2.0, 1.0, 5.0, 4.3, 2.0, 2.0, 1.3, 3.0, 2.7]$

3 Mittel und Streuung

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

1. $x_4 = [2.73, 1.85, 21.24, 17.97, 5.49, 18.90, 12.46, 0.97, 6.45, 7.43]$
2. $x_5 = [1.00, 1.91, 3.12, 4.38, 4.72, 5.29, 3.82, 3.25, 2.04, 0.93]$
3. $x_6 = [1.07, 1.06, 0.94, 1.84, 3.04, 3.22, 4.18, 5.27, 6.27, 6.75]$

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.

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 und $\alpha = 0.99$ (also je vier Mal den unteren und den oberen Wert des Konfidenzintervalls). Die kritischen Werte der Normalverteilung entnehmen Sie bitte der zur Verfügung gestellten Tabelle. Runden Sie auf drei Nachkommastellen.

1. $q_1 = 0.21$
2. $q_2 = 0.49$
3. $q_3 = 0.89$

5.2 Konfidenzintervalle für Mittelwerte

5.2.1 Fehler finden

Warum hätte folgende Tabelle ganz nicht gedruckt werden dürfen? Der Fehler ist ohne nachzurechnen erkennbar?

			95% CI	
Measure	*M*	*SD*	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

5.2.2 Transfer: Stichprobengröße

Aus den Zahlen für *Age at testing* können wir den Standardfehler und die Stichprobengröße rekonstruieren **auch ohne in den Originalartikel zu schauen**. Finden Sie zuerst den Standardfehler und dann die Stichprobengröße. Mit der Stichprobengröße können Sie dann das korrekte KI für *Age of onset* berechnen.

6 Transfer: Anteilswerte und Mittelwerte

In dieser Transferaufgabe geht es darum, zu zeigen, wie sich der Standardfehler für Anteilswerte aus dem Standardfehler für Mittelwerte ableiten lässt. Der allgemeine Standardfehler für Normalverteilungen wird berechnet mit:

$$SE = \frac{s}{\sqrt{n}} = \frac{\sqrt{s^2}}{\sqrt{n}} = \sqrt{\frac{s^2}{n}} \tag{1}$$

In der Wurzel der dritten Variante steht also die Varianz (hier der Einfachheit halber die nicht korrigierte Version mit n statt $n-1$ im Nenner). Der Standardfehler für Anteilswerte — also für Stichproben von n Einsen und Nullen (also letztlich Zähldaten aus einem Bernoulli-Experiment) wie $x_7 = [0, 1, 1, 0, 0, 1, 1, 1, 1, 0]$ — wird berechnet mit:

$$SE = \sqrt{\frac{q(1-q)}{n}} \tag{2}$$

Was bei dem SE für Mittelwerte die Varianz ist, erscheint hier als $p(1-p)$. Um zu zeigen, dass dies tatsächlich die Varianz für Bernoulli-Stichproben ist, können wir die Varianz zunächst genauso berechnen wie für Mittelwerte, wenn wir den Anteilswert von Einsen q als Mittelwert ansetzen.

$$q = \bar{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n} \tag{3}$$

Die Varianz für eine solche Stichprobe (oder Grundgesamtheit) ist dann wie zu erwarten gegeben:

$$s^2(x) = \frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n} \tag{4}$$

Zeigen Sie für solche Fälle, dass:

$$\sqrt{\frac{s^2}{n}} = \sqrt{\frac{q(1-q)}{n}} \tag{5}$$

Es ist also hinreichend, zu zeigen, dass:

$$\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}{n} = q(1-q) \tag{6}$$

Berechnen Sie anschließend zur Kontrolle auf beiden Wegen das Ergebnis für $x_7 = [0, 1, 1, 0, 0, 1, 1, 1, 1, 0]$.