Frage:

Warum haben bei den Klassenkenngrößen im Skript (Kapitel 2, Seite 3) die Klassenbreite und Klassenmitte dieselbe Notation d_j?

Antwort:

Im Buch zur Vorlesung, "Deskriptive Statistik", S.25, wird die Klassenbreite mit d_j und die Klassenmitte mit a_j bezeichnet. Die selbe Notation wird in der Formelsammlung verwendet.
Im Skript wird fälschlicherweise die Klassenmitte mit d_j bezeichnet, richtig wäre a_j

Frage:

Was ist der Unterschied zwischen der Gruppierung und Klassierung von Daten?

Antwort:

Gruppierte Daten sind Daten, bei denen einige Ausprägungen mehrfach vorkommen. Hier kann es vorteilhaft sein diese Struktur auszunutzen, vgl Formel für arithmetisches Mittel bei gruppierten Daten.
Klassierte Daten sind Daten, die ursprüngliche stetig waren und nachträglich in gewisse Klasse eingeteilt wurden. Damit gibt es auch bei klassierten Daten natürlich Ausprägungen die mehrfach vorkommen, so dass klassierte Daten als Spezialfall von gruppierten Daten angesehen werden können.

Frage:

Was ist der Unterschied zwischen S_{xx} und s_x^2?

Antwort:

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{1}$$

Definiert die (empirische) Varianz des Merkmals.
Dies entspricht der **mittleren** quadratischen Abweichung vom arithmetischen Mittel.
Die Varianz ist in der Statistik das gebräuchliche Maß für die Streuung eines Merkmals.
vgl. Skript Kapitel 4 und im Buch zur Vorlesung, "Deskriptive Statistik", Kapitel 3.2.3, sowie Kapitel 4.4.

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{2}$$

S_{xx} ist die Summe der quadrierten Abweichungen der Beobachtungen vom arithmetischem Mittel $\bar{x}$. Dieses Maß wird of verwendet, wenn sich der zusätzliche Term $\frac{1}{n}$ aus der Varianz sowieso rauskürzen würde.

Frage:

Müsste bei der Berechnung der empirischen Varianz nicht noch um einen Freiheitsgrad bereinigt und durch n-1 geteilt werden?

Antwort:

Dies ist eine andere Definition der Varianz, die sog. korrigierte Stichprobenvarianz. Diese spielt aber erst im Zusammenhang mit der Definition von Zufallsvariablen eine Rolle und wird somit erst im nächsten Semester im Zuge der Induktiven Statistik relevant. Bei genügend großem Stichprobenumfang spielt die Bereinigung $\frac{1}{n-1}$ im Gegensatz zu $\frac{1}{n}$ sowieso nur eine geringe Rolle.

Frage:

Warum können Kontingenztafeln keine metrischen Merkmale abbilden?

Antwort:

Kontingenztafeln sollten nur bei Merkmalen mit eher wenigen diskreten Ausprägungen verwendet werden, was zumeist auf nominale oder ordinale Merkmale zutrifft. Es sind aber auch Kontingenztabellen für diskrete metrische Merkmale denkbar, beispielsweise für die beiden Variablen Anzahl Personen im Haushalt und Anzahl der Autos.

Frage:

Zur Verknüpfung von Laypeyres- und Paasche-Indizes und dem dadurch enstehenden Umsatzindex: Wie aussagekräftig ist dieser, denn wird er nicht durch die Inflation und die i.d.R. höheren Preise der Berichtsperiode verfälscht?
Und gibt es einen tieferen Sinn in der Verknüpfung außer dem Umsatzindex?

Antwort:

Die Interpretation des Umsatzindexes ist sicherlich komplexer als die eines Preisindexes oder eines Mengenindexes, da die Veränderung des Werts des Warenkorbs sowohl durch Preis- als auch durch Mengenveränderungen bewirkt worden sein könnten. Das Problem der Inflation kann jedoch auch durch die Verwendung inflationsbereinigter Preise behoben werden. Die Verknüpfung der Indizes soll nur darauf hinweisen, dass in beiden Fällen der Umsatzindex entsteht und der Umsatzindex somit auf zweierlei Arten hergeleitet/interpretiert werden könnte.

Frage:

Ist die empirische Verteilungsfunktion eine nicht-stetige Funktion?

Antwort:

Nein, die empirische Verteilungsfunktion ist eine rechtsseitig stetige Funktion.