salat09.sav

um Wertelabels ergänzt

• Daten der Dozenten ergänzt

• um Alter und BMI erweitert

um fehlende und unzulässige Werte bereinigt

SPSS Kurs, SS 2009, 31

S. Heim/C. Heumann

Daten Aggregieren

Daten > Aggregieren ...

- Ziel:
 - Zusammenfassen von Beobachtungen
 - wiederholte Beobachtungen
 - Gruppenrepräsentant
- Break-Variable zur Unterteilung/Gruppierung
- Aggregierungsfunktionen: Lage-/Streuungsmaße, prozentuale Anteile
- gruppenspezifische Fallzahlen
- Aggregieren in Kombination mit Zusammenfügen von Dateien zum Umbau von Dateien

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 33

Zusammenfügen von Dateien

Datensatz 2009

Daten > Dateien zusammenfügen

- Fälle hinzufügen: Variablen können in unterschiedlicher Reihenfolge vorliegen
- Variablen, die in beiden Dateien vorkommen, bleiben in der Arbeitsdatei unverändert
- bisherige Variablen sind mit (*) gekennzeichnet, hinzukommende Variablen mit (+)
- Verbinden anhand einer Schlüsselvariablen bei ungleicher Datensatzlänge oder bei Referenztabellen;
- Fälle beider Dateien müssen in aufsteigender Reihenfolge der Schlüsselvariablen sortiert sein
- optionaler Herkunftsindikator: 0=Arbeitsdatei, 1=externe Datei

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 34

Zusammenfügen von Dateien: Variablen hinzufügen

- **Fall 1:** Beide Dateien haben gleiche Anzahl an Fällen und die gleichen Fälle. Verwendung einer Schlüsselvariablen optional.
- Fall 2: Externe Datei ist größer als Arbeitsdatei, es sollen aber nur Variablen zu den Fällen der Arbeitsdatei hinzugefügt werden. Externe Datei enthält zu verwendenden Schlüssel.
- **Fall 3:** Externe Datei ist kleiner Arbeitsdatei, es interessieren nur Fälle der externen Datei. Verwendeter Schlüssel ist in Arbeitsdatei.
- Fall 4: In beiden Dateien kommen Beobachtungen vor; Vereinigungsmenge bilden (teilweise fehlende Werte).

Übung 5

- 1. Geben Sie unter Edit > Options ... > File Locations Ihren persönlichen Ordner als standard Datenquelle an.
- 2. Fügen Sie die Umfrage 2009 salat09.sav zur Umfrage 2007 salat07_r.sav hinzu. Behalten Sie nur gemeinsam vorkommende Variablen und ordnen Sie ggfs unterschiedliche Namen paarweise an. Erzeugen Sie eine Variable mit dem Herkunftsjahr, vergeben Sie entsprechende Wertelabels mit Hilfe des Dialogs Data > Define Variable Properties.
- 3(a) Ergänzen Sie fehlende Werte unter Nutzung von Transform > Replace Missing Values Erzeugen Sie dazu eine neue Variable, die fehlenden Werten den Mittelwert aller Beobachtungen zuweist.
- (b) Spezifizieren Sie den zugehörigen Variablentyp über Data > Define Variable Properties.

Übung 5

- 3.(c) Wiederholen Sie den Vorgang unter Beschränkung auf gleiches Herkunftsjahr. Verwenden Sie dazu Transformieren > Fälle auswählen.
- (d) Vergleichen Sie die vorgeschlagenen Ersatzwerte. Welchen Effekt hat die Fallauswahl auf das Ergebnis?
- (e) Erzeugen Sie mit derselben Fallauswahl eine temporäre Datenmatrix. Ermitteln Sie erneut die Ersatzwerte und vergleichen Sie.
- 4. Sortieren Sie nun die Daten der zusammengefügten Umfragen nach Geschlecht und Salatkonsum und ordnen Sie die beiden Spalten nebeneinander an. Lässt sich eine Tendenz hinsichtlich des Salatvorliebe ablesen?
- 5. Teilen Sie die Stichprobe zu gleichen Teilen in Salatliebhaber und Salatverweigerer ein. Verwenden Sie dazu die automatische Trennwerterstellung der visuellen Klassifizierung. Welche Option wählen Sie für die Generierung der Cutpoints? Speichern Sie den Datensatz.

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 37

Übung 5

 Erzeugen Sie eine neue Datei, die die geschlechtsspezifischen Mittelwerte f
ür K
örpergr
ö
ße und -gewicht enth
ält sowie den jeweiligen Prozensatz jener mit mindestens 4-maligen Salatkonsum pro Woche. Verwenden Sie die Aggregierungsfunktion.

Gilt für den Prozentsatz 'echt größer 4' oder 'größer gleich 4'? Leiten Sie eine Hypothese über das Ernährungsbewusstein unter Geschlechtern ab. S. Heim/C. Heumann

SPSS Kurs, SS 2009, 38

Datenexploration

- Analysieren > Deskriptive Statistiken > Häufigkeiten erzeugt Statistiken, Graphiken und Häufigkeitstabelle
 - Variable Geburtsdatum: Minimum, Maximum, Modalwert, Median
 - Variable Salatkonsum: Balkendiagramm
- Analysieren > Deskriptive Statistiken > Deskriptive liefert Statistiken und erzeugt ggfs standardisierte Variablen
- Analysieren > Deskriptive Statistiken > Explorative Datenanalyse... liefert Statistiken und Graphiken (Boxplots), Tests auf NV und vieles mehr

Übung 6

• Analysieren > Nichtparametrische Tests > Chi-Quadrat... " H_0 : Daten folgen einer Gleichverteilung" \rightarrow Erwartete Werte: Alle Kategorien gleich

Tests auf Verteilung

- Analysieren > Nichtparametrische Tests > K-S bei einer Sichprobe "H₀: Daten folgen einer XX-Verteilung", wobei XX für Normal-, Gleich-, Poisson- und Exponentialverteilung stehen kann.
- Analysieren > Deskriptive Statistiken > ...
 ... Explorative Datenanalyse testet mit Kolmogorov-Smirnow oder Shapiro-Wilks auf NV
- Analysieren > Deskriptive Statistiken > Häufigkeiten...
 - Diagramme: Histogramme mit Normalverteilungskurve
 - $\pm \sigma$ entspricht den Wendepunkten in der Dichte
 - NV(0,1): $\bar{x} \pm \{1/2/3\}\hat{\sigma}$ entspricht $\{68.26/95.44/99.72\}\%$

- 1. Laden Sie die Datendatei Eier_oKat.sav, die für 1000 Hühnereier der Hühnerrasse "Loheimer Braun" verschiedene Kennzeichen enthält.
- 2. Beschreiben Sie den Datensatz, indem Sie jede Variable geeignet visualisieren und sinnvolle Statistiken berechnen. Sind die metrischen Variablen normal-verteilt? Wie lässt sich dies bereits an Schiefe und Wölbung ablesen? Welche graphischen Darstellungen können Sie zur Überprüfung auf NV generieren? Lassen sich Ausreißerwerte identifizieren?
- 3. Erstellen Sie einen neuen Datensatz Eier_Fallausschluss.sav, der die Daten für braunfarbige Eier enthält, die schwerer als 48 g und leichter als 77 g sind mit einer Breite > 41 mm und < 47 mm und einer Höhe > 49 mm und < 63 mm.

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 41

Übung 6

- 4. Erweitern Sie den reduzierten Datensatz um die Information der Hühnerhaltung (Eier_Haltung.sav). Inspizieren Sie die Daten vorher und bringen Sie sie in die benötigte Reihenfolge.
- Kategorisieren Sie die Höhe eines Hühnereis in 'klein' (kleiner 50 mm), 'mittel' und 'groß' (mindestens 60 mm) durch (a) Umkodierung in eine andere Variable und (b) visuelles Klassieren. Vergleichen Sie die Ergebnisse rechnerisch.
- 6. Geben Sie das durchschnittliche Gewicht für nach Haltungsart und Höhenkategorie gruppierte Daten an, zunächst durch Aggregieren. Fügen Sie dann diejenigen Werte ein, die Sie über Analysieren > Mittelwerte... erhalten und vergleichen Sie.

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 42

Übung 6

- 7. Klassieren Sie die Variable Gewicht in eine neue Variable in (möglichst) gleichmäßig besetzte Kategorien 'leicht', 'mittel' und 'schwer'. Überprüfen Sie die angestrebte Gleichverteilung graphisch und per Test. Warum sind die Daten nicht exakt gleichverteilt?
- 8. Prüfen Sie die metrischen Variablen auf Normalverteilung. Verwenden Sie Ihnen dabei bekannte graphische Verfahren und Tests.
- 9. Speichern Sie den Datensatz als Eier_Uebung6.sav.

zente

Pivot–Tabelle

Nominalniveau

Viewer: Pivot–Tabelle

- klare, schlüssige und aussagekräftige Präsentation der Ergebnisse
- kategoriale Variablen definieren Aufbau (Zeilen/Spalten/Schichten)
- häufig auch Strukturierung durch berechnete Zellenstatistiken (Mittelwert, Median, #Beobachtete, #Erwartete, Prozente etc.)
- geschachtelte versus gestapelte Tabellen
- spezifischen SPSS Pivot Table Editor öffnen:
 - Doppelklick auf Tabelle
- Tabelle im Viewer auswählen und über Menüleiste
 Edit > Edit Content > In Separate Window
- Format > TableLooks Tabellenvorlagen zur schnellen Layoutänderung z. B. zum akademischen Stil

S. Heim/C. Heumann SPSS Kurs, SS 2009, 48

Kreuztabelle – CROSSTABS

• Statistiken: z. B. χ^2 -Unabhängigkeitstest, Kontingenzkoeffizient bei

• Zellen: z. B. beobachtete/erwartete Häufigkeiten, zeilenweise Pro-

Analysieren > Deskriptive Statistiken > Kreuztabelle

Gruppiertes Balkendiagramm generierbar

Beispiel: Haltung
 – versus Gewichtskategorien

Viewer: Pivot-Tabelle

- Edit > Toolbar zur Modifikation von Schriftart/-größe
- Pivot > Pivoting Trays Drag & Drop zum Umsortieren der Variablen in Zeilen, Spalten und Schichten
- Reihenfolge der Icons im Pivoting Trays Fenster repräsentiert Reihenfolge in der Tabelle
- schneller Schichtwechsel über Pfeile des Pivot-Symbols
- Crtl + Alt + Klick auf Kategorielabel (z.B. 'weiblich'), View > Hide um einzelne Kategorien auszublenden
- Label markieren, Bearbeiten > Auswählen > Datenzellen und Format > Zelleneigenschaften: Horizontale/vertikale Ausrichtung, Schattierung, Werteformat
- Umwandlung in interaktive Graphik möglich

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 49

Übung 7: Kontingenztafel

- Öffnen Sie den in Übung 5 abgespeicherten Datensatz.
- Erstellen Sie eine Kontingenztafel f
 ür Geschlecht versus Salatliebhaber/verweigerer und schichten Sie diese gem
 äß Herkunftsjahr. Nehmen Sie neben den beobachteten und erwarteten Werten auch die zeilenweise Prozente in die Tafel auf.
- Explorieren Sie die Darstellungsweisen durch Pivoting. Blenden Sie die erwarteten Werte aus.
- Gibt es einen signifikanten Zusammenhang im Jahr 2007 oder 2009?

Übung 8: Klinische Studie

- **Diagnose:** temporaler Arteriitis (TA) mit und ohne neuroophthalmologischen Komplikationen (NOC)
- **Methode:** Dopplersonographie der periorbitalen Arterien (DSPA) liefert in der Gruppe der Patienten mit TA und mit NOC 10 pathologische und 15 unauffällige Befunde und bei Patienten mit TA ohne NOC 5 pathologische und 57 unauffällige Befunde.
- **Frage:** Weisen Patienten mit zusätzlichen NOC eher Schädigungen der PA auf als Patienten ohne NOC?

Kontingenztafel mit Sekundärdaten

- direkte Eingabe der Sekundärdaten:
 - dichotome Variablen Gruppe und Befund
 - Variable mit Gewichten der Kontingenztafelfelder
- Daten > Fälle gewichten (Kontrollfeld links unten)
- Analysieren > Deskriptive Statistiken > Kreuztabellen...
- Faustregel von mindestens 5 Fällen pro Zelle "verletzt"
 → exakter Test nach Fisher

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 52

Korrelationsanalyse

Analysieren > Korrelation > Bivariat...

- eine Korrelationsmatrix für alle metrischen Variablen
- parametrisch: Produkt–Moment–Korrelation nach Bravais & Pearson
 - misst linearen Zusammenhang
 - NV–Annahme

$$\rho = \frac{\operatorname{Cov}(x, y)}{s_x s_y}$$

d. h. für $\rho = 0$ könnte immer noch ein anderer als linearer Zusammenhang bestehen und für $\rho \to 1$ könnte zusätzlich ein anderer Zusammenhang bestehen.

- nicht-parametrisch: Korrelation nach Spearman
 misst monotonen Zusammenhang
- Pivot: Umstrukturierung in reine Korrelationsmatrix

S. Heim/C. Heumann

SPSS Kurs, SS 2009, 53

Korrelationsanalyse

Analysieren > Korrelation > Partiell...

- nur parametrische Variante verfügbar (NV-Annahme!)
- Kontrollvariable, um deren Einfluss der Zusammenhang zweier anderer Merkmale bereinigt wird
- Beispiel: "Breite und Höhe gleichgewichtiger Eier sind (partiell linear) negativ korreliert", d. h. "H₀ : ρ_{Breite*Hoehe|Gewicht} ≥ 0" → einseitige Testung

Übung 9: Korrelationsanalyse

- 1. Führen Sie eine Korrelationsanalyse der metrischen Variablen in Eier_Uebung6.sav durch. Erstellen Sie ein Streudiagramm Gewicht versus Breite. Lässt sich eine Gerade, die den linearen funktionalen Zusammenhang zweier Merkmale widerspiegelt, hinzufügen? Sind Gewicht und Breite von einander unabhängig? Welcher Zusammenhang besteht zwischen Bestimmtheitsmaß und Korrelationskoeffizient?
- 2. Generieren Sie eine Streudiagrammmatrix für Gewicht, Breite und Höhe. Fügen Sie Regressionslinien hinzu. Ist der Zshang zwischen Breite und Höhe plausible?
- 3. Generieren Sie ein gemäß Gewichtskategorie eingefärbtes Streudiagramm für Breite versus Höhe. Wie beurteilen Sie jetzt die gruppenspezifische Korrelation?
- 4. Überprüfen Sie sodann anhand der Betrachtung partieller Korrelationen, ob sich Breite und Höhe eines Hühnereis umgekehrt proportional zueinander verhalten.