

Lade bitte bis Dienstag 12.12.2017 12:12 den R-Code zu den jeweiligen Aufgaben in moodle hoch.

Bitte Name und die jeweilige Beispielnnummer im Code notieren. Der Code soll so programmiert werden, dass das (richtige) Ergebnis in der Console ausgegeben wird.

Für besonders ausgereifte Lösungen des Beispiels 1 wird ein halber Bonuspunkt vergeben.

1. Schreibe eine Funktion, die für 2 Merkmale eine Kreuztabelle erstellt. Die Funktion soll bei Bedarf die gewünschte Kreuztabelle in der Console ausgeben können und gleichzeitig Werte zurückgeben, nämlich

- die Kreuztabelle selbst als numerische Matrix,
- die Ergebnisse eines χ^2 -Tests: χ^2 -Wert, p-Wert, Carmér's V,
- Warnung, wenn die Zellenhäufigkeiten zu gering sind für die Gültigkeit des χ^2 -Tests.

Für die Erstellung der Kreuztabelle sollen die folgenden Optionen möglich sein, wähle sinnvolle Default-Werte:

- a) Randhäufigkeiten: ja/nein
 - b) Inhalt der Tabelle: Häufigkeiten, relative Häufigkeiten, Zeilenprozent, Spaltenprozent
 - c) Ausgabe mit oder ohne Prozentzeichen
 - d) Anzahl der ausgegebenen Dezimalstellen bei Prozent
 - e) Missing als eigene Kategorie: ja/nein
 - f) Maximale Anzahl an Levels pro Merkmal. Wenn die maximale Anzahl an Levels überschritten wird und wenn alle Levels numerisch sind, soll das Merkmal entsprechend der maximalen Anzahl an Levels recodiert werden. Ansonsten soll die Funktion keine Berechnungen durchführen.
 - g) Maximale Zeichenlänge der Levels. Wenn die Zeichenlänge überschritten wird, soll der Level-Name automatisch gekürzt werden.
2. Teste die Kreuztabellen-Funktion für unterschiedliche Inputparameter auf übersichtliche Art und Weise. Falls Beispiel 1 nicht gelingt, teste die Funktion `chisq.test()`.
 3. Wende die Kreuztabellen-Funktion auf alle Variablenpaare des Schuldatensatzes (siehe Übungsblatt 3) an. Die Teststatistik (χ^2 -Wert) soll in einer entsprechenden Matrix gespeichert werden. Variablenpaare, für die keine Teststatistik berechnet werden kann, sollen in einem eigenen Objekt gespeichert werden. Miss die Laufzeit für die Berechnung. Falls Beispiel 1 nicht gelingt, verwende die Funktion `chisq.test()`.

4. Die Datei `verleihungen.rad` enthält Daten einer Radverleihstation: Geburtsdatum des Entlehners bzw. der Entlehnerin, Verleihstart und Dauer in Minuten. Zur Information: die Daten sind keine hinterhältige Erfindung sondern stammen aus meiner statistischen Praxis.
 - a) Lies die Datei in R ein. Sollte es nicht gelingen, lade `verleihungen.rdata`.
 - b) Wandle die Variable `Verleihung` in ein Datum-Uhrzeit Objekt um.
 - c) Berechne das Ende der Verleihung.
 - d) Wandle das Geburtsdatum in ein Datumsobjekt um. Hier ist bei der Eingabe und Verarbeitung der Daten einiges schief gegangen: Es wurden bei der Eingabe unterschiedliche Datumsformate gewählt und manche Eingaben wurden in einen numerischen Wert umgewandelt. Da die Daten in Mac Excel verarbeitet wurden, sind dies die Tage seit 31.12.1903. Versuche so viele Daten wie möglich zu “retten”.
 - e) Berechne das Alter jeder Person zum Verleihzeitpunkt und überprüfe, ob das Alter plausibel ist.