

Lösung zu Kapitel 4: Beispiel 1

Lottozahlen in Österreich 2007 und 2008

In Österreich ist Lotto sehr beliebt. Sechs Zahlen plus Zusatzzahl werden von einer Maschine aus der Menge der Zahlen von 1 bis 45 gezogen. Natürlich erwartet man, dass jede Zahl die gleiche Wahrscheinlichkeit hat, gezogen zu werden. Im Datenfile sind die Ziehungshäufigkeiten für alle 45 Lottozahlen des österreichischen Lottos für die Jahre 2007 und 2008 gegeben. Kann man aus den Daten schließen, dass der Mechanismus, mit dem die Lottozahlen ermittelt werden, fair ist, d.h., dass für alle Zahlen die Auswahlwahrscheinlichkeit gleich ist?

Daten: **lotto0708.sav**

Variablen: **lottozahl**, **h2007**, **h2008**

Die Nullhypothese lautet somit: Die Auswahlwahrscheinlichkeit ist für jede einzelne Lottozahl gleich hoch.

Einen ersten Überblick über die Verteilung der Daten kann man sich mit Hilfe eines gestapelten Balkendiagrammes schaffen.

PASW

Diagramme

Diagrammerstellung

in Galerie auswählen von Balken

Doppelklick auf Galeriediagramm

Variable **lottozahl** in das Feld **X-Achse?** ziehen

Variable **h2007** in das Feld **Anzahl** auf der **Y-Achse** ziehen

Variable **h2008** in das Feld **Anzahl** auf der **Y-Achse** ziehen

Zusammenfassungsgruppe erstellen mit bestätigen

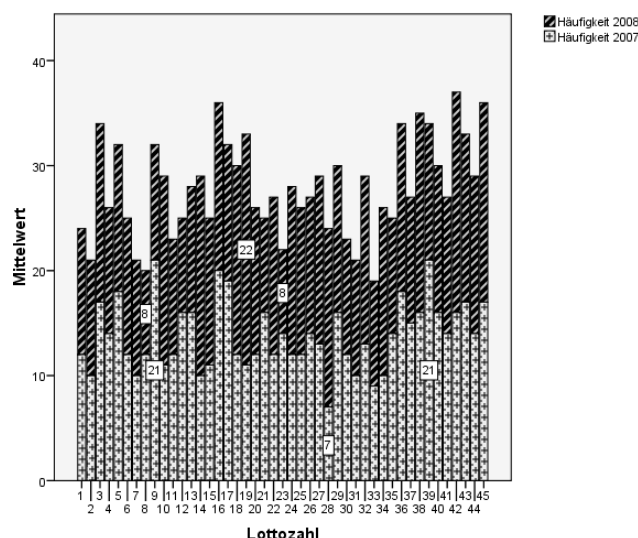


Abbildung 1: gestapeltes Balkendiagramm mit den Variablen **h2007** und **h2008**

Anhand des Balkendiagrammes lässt sich erkennen, dass es sehr wohl Unterschiede in der Ziehungshäufigkeit bei den einzelnen Zahlen gibt. Nun soll überprüft werden, ob diese dem Zufall zuzuschreiben sind oder ob doch eine signifikante Abweichung von der Gleichverteilung gegeben ist. D.h. es soll die kategoriale Variable `lottozahl` auf Gleichverteilung getestet werden. Diese Frage kann mit einem Chi-Quadrat-Test beantwortet werden.

Hinweis: An dieser Stelle sei darauf hingewiesen, dass es sich bei den vorliegenden Daten (eigentlich) um eine Vollerhebung der Lottoziehungen für die beiden Jahre handelt. Will man jedoch von den Ziehungshäufigkeiten der Jahre 2007 und 2008 auf die Fairnis des Mechanismus über einen unbestimmten Zeitraum hinweg (also für immer und ewig) schließen, stellen die Daten eine Stichprobe dar.

Da es sich bei den Variablen `h2007` und `h2008` um Häufigkeitsvariablen handelt, müssen diese jeweils zuerst gewichtet werden.

PASW

Daten

Fälle gewichten ...

☒ Fälle gewichten mit *markieren der Variable (h2007)* ☐ Häufigkeitsvariable:

Erst jetzt kann die Verteilung der Ziehungshäufigkeiten der Lottozahlen aus dem Jahr 2007 einer Analyse unterzogen werden.

PASW

Analysieren

Nichtparametrische Tests

Chi-Quadrat ...

markieren der Variable (lottozahl) ☐ Testvariable(n):

Statistik für Test	
	Lottozahl
Chi-Quadrat	33,837 ^a
df	44
Asymptotische Signifikanz	,866

a. Bei 0 Zellen (0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 13,9.

Abbildung 2: Chi-Quadrat-Test auf Gleichverteilung mit der Variable `h2007`

Der berechnete p-Wert für das Jahr 2007 (0,866) weist auf keine wesentlichen Unterschiede bei den Ziehungshäufigkeiten der einzelnen Zahlen hin.

Im nächsten Schritt muss man die Gewichtung für die Variable `h2007` **Zurücksetzen** und die Fälle mit `h2008` gewichten. Unterzieht man nun die Variable `lottozahl` erneut einem Chi-Quadrat-Test auf Gleichverteilung, werden die Ziehungshäufigkeiten aus dem Jahr 2008 zur Berechnung herangezogen.

Statistik für Test	
	Lottozahl
Chi-Quadrat	35,000 ^a
df	44
Asymptotische Signifikanz	,832

a. Bei 0 Zellen (0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 14,0.

Abbildung 3: Chi-Quadrat-Test auf Gleichverteilung mit der Variable h2008

Auch für die Ziehungen aus dem Jahr 2008 weist der p-Wert (0,832) auf einen fairen Mechanismus bei der Ermittlung der Lottozahlen hin. Das bedeutet, dass aufgrund der annähernd gleichverteilten Häufigkeiten bei den Lottozahlen-Ziehungen dieser beiden Jahre die Nullhypothese bei zu behalten ist.

Zum Schluss sollte nicht darauf vergessen werden, die Gewichtung der Variable h2008 wieder aufzuheben!

PASW

Daten

Fälle gewichten ...

Zurücksetzen

OK

Hinweis: Als Alternative zur gesonderten Analyse der Daten aus den beiden Jahren kann auch in einem ersten Schritt die Summe der Häufigkeiten aus den beiden Jahren berechnet werden, um anschließend die aggregierten Ziehungshäufigkeiten mit Hilfe des Chi-Quadrat-Tests auf eine Gleichverteilung hin zu überprüfen. Dies wäre vor allem dann sinnvoll, wenn die Verteilungen in den einzelnen Jahren signifikante Unterschiede aufweisen und man prüfen möchte, ob diese im Laufe der Zeit verschwinden oder ob sie sich verstärken.

Für dieses Beispiel erhält man nach Aggregation der Ziehungshäufigkeiten der Jahre 2007 und 2008 einen p-Wert in Höhe von 0,854. D.h. es ist mit einer Wahrscheinlichkeit von 85,4 % anzunehmen, dass die Auswahlmethode "fair" ist. Es wird also jede Lottozahl mit gleich hoher Wahrscheinlichkeit gezogen und die Nullhypothese ins bei zu behalten.