

Übungen zu Wirtschaftsstatistik - Blatt 1

Abgabe am 03. 05. vor Beginn der Übung

Hinweis: Einen Link zum Download der Statistik-Software **R** sowie eine kurze Einführung sind auf der Homepage der Vorlesung zu finden.

Aufgabe 1 (4 Punkte)

Nach dem zentralen Grenzwertsatz ist bekannt, dass für $X_n \sim \text{Bin}(n, p)$ die Zufallsvariable $Y_n = (X_n - np) / \sqrt{np(1-p)}$ für $n \rightarrow \infty$ in Verteilung gegen die Standardnormalverteilung konvergiert. Simulieren Sie für $p = 0.7$ und $n = 100, 1000, 10000$ und 100000 jeweils 10000 Realisierungen von Y_n und plotten Sie die Histogramme der relativen Häufigkeiten der simulierten Daten jeweils gemeinsam mit der Dichte der Standardnormalverteilung und einer geeigneten Beschriftung in ein einzelnes Grafikfenster. Was ist zu beobachten?

Aufgabe 2 (2+2+2+2 Punkte)

Betrachten Sie folgenden Datensatz „aflspend.dat“, der auf der Homepage bereitsteht:

Beschreibung: Die Ausgaben von AFL-Football-Vereinen (australische Liga) und ihre Erfolge von 2003 bis 2007.

Spaltenname	Bedeutung
Club	Name des Vereins
Spent	Gesamtausgaben (in Millionen australische Dollar)
Wins	Anzahl der Siege von 2003 bis 2007
GrandFinals	Anzahl der „Grand Final“-Teilnahmen von 2003 bis 2007
Place3to4	Häufigkeit der Platzierungen 3 oder 4
Place5to8	Häufigkeit der Platzierungen 5 bis 8
Top8	Häufigkeit der Platzierungen in der Top 8

- Lesen Sie mit dem Befehl `read.table()` den Datensatz in **R** ein und geben Sie ihn mit `print()` aus.
- Berechnen Sie mit den Befehlen `mean()` und `var()` das Stichprobenmittel und die Stichprobenvarianz für die (Einträge in den) Spalten **Spent** und **Wins**.
- Berechnen Sie auch das Stichprobenmittel und die Stichprobenvarianz für den Quotienten **Wins / Spent**. Wie kann man dieses Stichprobenmittel interpretieren?

Hinweis: Sind x und y zwei gleich große Vektoren, so teilt der Befehl x/y sie komponentenweise.

- Betrachten Sie die beiden Spalten **Spent** und **Wins**. Angenommen, man kennt die australische Football-Szene nicht und muss deshalb davon ausgehen, dass die Vereine alle gleich sind (bzgl. dieser Werte), wobei es natürlich zufallsbedingte Abweichungen gibt. Macht es dann bei den Einträgen in den Spalten Sinn, davon auszugehen, dass sie unabhängig sind? Diskutieren Sie die Annahme kurz jeweils für die zwei Spalten.

Aufgabe 3 (1+2+3 Punkte)

Simulieren Sie 200 Realisierungen von $\text{Exp}(2)$ -verteilten Zufallsvariablen.

- (a) Erstellen Sie ein Histogramm der relativen Häufigkeiten der simulierten Daten und plotten Sie in dieses Histogramm die Dichte der $\text{Exp}(2)$ -Verteilung.
- (b) Plotten Sie die empirische Verteilungsfunktion der simulierten Daten mit der in R enthaltenen Funktion `ecdf`.
- (c) Schreiben Sie eine eigene Funktion (ohne den Befehl `ecdf` zu verwenden), die die empirische Verteilungsfunktion bestimmt. Plotten Sie diese Funktion zusammen mit der theoretischen Verteilungsfunktion in ein Schaubild. *Mögliche Vorgehensweise:* Wenn `v` der Vektor mit den Stichprobenwerten ist, können Sie mit `v[v<x]` den Teilvektor erzeugen, der die Werte enthält, die kleiner sind als `x` (analog kleiner gleich) sind. Die Länge eines Vektors erhalten Sie mit `length()`. Sortieren kann man ihn mit `sort()`. Mit `lines(c(x1, x2), c(y1, y2))` können Sie eine Linie zwischen (x_1, y_1) und (x_2, y_2) erzeugen. Alternative: mit `function()` eine Funktion definieren und plotten lassen. Achtung: Die Funktion muss vektorwertig sein.