9. Übung am 19. Juni 2023

Übungsaufgabe 46 (Schadenssummen)**.** Der Datensatz Insurance.RData enthält für 100.000 Sturmschäden einer Versicherung die ausbezahlte Summe, wobei selbige maximal 5.000 Euro (Deckungssumme im Vertrag) beträgt obwohl die echte Schadenshöhe höher sein kann[vii]. Angenommen, wir wissen, dass die echte Schadenshöhe S logarithmisch normalverteilt mit Parameter μ, σ^2 ist. Überlegen Sie sich zwei verschiedene Zugänge, um in diesem Fall μ und σ^2 und damit $F_S(10000)$ und $F_S(100000)$ zu schätzen, wobei F_S wie gewohnt die Verteilungsfunktion von S bezeichnet.

Übungsaufgabe 47. Das R-Snippet R-Codes_permtest.R implementiert einen auf den ersten Blick etwas ungewöhnlichen Hypothesentest. Finden Sie heraus, was der R-Code macht, und berechnen Sie approximativ die power-Funktion des Tests. Vergleichen Sie die Powerfunktion mit jener des entsprechenden t-tests. Adaptieren Sie weiters den Code so, dass im Falle zweier Exponentialverteilung auf Gleichheit der Parameter getestet wird, und berechnen Sie approximativ den Fehler erster Art.

Übungsaufgabe 48. Wir haben in der UV bisher großteils klassische Hypothesentests kennengelernt. Eine in der Praxis oft auftretende Situation ist, ausgehend von Stichproben $(X_1, Y_1), \ldots (X_n, Y_n)$ auf Unabhängigkeit von X und Y zu testen. Für diese Situation gibt es sowohl klassische als auch moderne Tests, einer davon ist im schon erwähnten R-package 'qad' implementiert. Wählen Sie verschiedene Verteilungen von (X, Y)[viii], generieren Sie jeweils R Mal Stichproben und lesen Sie aus dem produzierten output den berechneten p.value (q(x1,x2)) aus. Überprüfen Sie damit, ob Abhängigkeiten auch für keine Stichproben gut entdeckt werden können.

Übungsaufgabe 49. Wir nehmen an, dass die Zufallsvariablen X, Y, ε die Gleichheit

$$Y = 2X + 1 + \varepsilon$$

erfüllen, wobei $X \sim \mathcal{U}(-3, 3), \varepsilon \sim \mathcal{N}(0, 0.5)$ gilt, X und ε unabhängig sind, und Y dann via $Y = 2X + 1 + \varepsilon$ berechnet wird. Erzeugen Sie in R eine Stichprobe $(X_1, Y_1), \ldots, (X_n, Y_n)$ der Größe $n = 100$ von (X, Y)[ix], berechnen Sie dann in R jene Werte für $a, b \in \mathbb{R}$, die

$$\sum_{i=1}^{n} (Y_i - aX_i - b)^2$$

minimieren (Sie können dafür bspweise die schon bekannte Funktion *nls* oder irgendeinen anderen Minimierer verwenden), und nennen Sie die Werte $\hat{a}_n$ und $\hat{b}_n$. Was ist für großes n zu beobachten?

Übungsaufgabe 50. Lesen Sie Abschnitt 6.1 und leiten Sie die Schätzer $\hat{\theta}_1, \hat{\theta}_0$ gemäß Gleichung (6.5) her.

Übungsaufgabe 51. Beweisen Sie die in Bemerkung 6.5 getätigte Aussage.

[vii] diese Fälle sind im Datensatz mit 'censored=1' markiert, d.h. in diesen Fällen ist nur bekannt, dass die echte Schadenshöhe mehr als 5.000 Euro betrug

[viii] zumindest ein Mal unabhängig und ein Mal abhängig

[ix] d.h., erzeugen Sie $X_1, \ldots, X_n$ sowie $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ und setzen Sie $Y_i = 2X_i + 1 + \varepsilon_i$