Hochschule RheinMain
Prof. Dr. D. Lehmann WS 2016/17

12. Übungsblatt zur Vorlesung Ökonometrie
(Probe-Klausur)

Theorie-Teil: Aufgaben 1+2, 50 Minuten (wenn Sie eher fertig sind, können Sie abgeben und schon mit dem Programmier-Teil beginnen)
Programmier-Teil: Aufgaben 3-6, 100 Minuten

Aufgabe 1 (10 Punkte): Es sei

$$\vec{x} \;=\; (-5,-4,-3,-2,-1,0,1,2,3,4,5)$$

und $\vec{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_{11})$ seien mit Mittelwert 0 und Standardabweichung σ normalverteilte Zufallszahlen.

a) Betrachten Sie das Regressionsproblem

$$\vec{y} \;=\; \beta_1\,\vec{x} \,+\, \vec{\varepsilon} \tag{1}$$

Geben Sie den OLS-Schätzer $\hat{\beta}_1$ an und vereinfachen Sie den Ausdruck soweit wie möglich.

b) Betrachten Sie das Regressionsproblem ($i = 1, 2, ..., 11$)

$$y_i \;=\; \beta_0 \,+\, \beta_1\,x_i \,+\, \varepsilon_i \tag{2}$$

Geben Sie die OLS-Schätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ an und vereinfachen Sie die Ausdrücke soweit wie möglich.

c) Geben Sie Formeln für die Varianzen von $\hat{\beta}_0$ und $\hat{\beta}_1$ aus Teil (b) an und vereinfachen Sie die Ausdrücke soweit wie möglich.

Aufgabe 2 (10 Punkte): Zufallszahlen $x_1, x_2, \cdots, x_n$ heissen Poisson-verteilt mit Parameter $\lambda > 0$, wenn $x_i \in \mathbb{N}$ für alle i und

$$\mathsf{Prob}[x_i = k] \;=\; \frac{\lambda^k}{k!}\,e^{-\lambda}$$

Von der Stichprobe $x_1, x_2, \cdots, x_n$ sei bekannt, dass es mit Parameter λ Poisson-verteilte Zufallszahlen sind, aber der Wert von λ sei unbekannt. Schätzen Sie λ mit Hilfe der Maximum-Likelihood-Methode, d.h. geben Sie den Maximum-Likelihood-Schätzer $\hat{\lambda}_{\mathrm{ML}} = \hat{\lambda}_{\mathrm{ML}}(x_1, ..., x_n)$ für λ an.

Aufgabe 3 (12 Punkte): Laden Sie sich von der Vorlesungs-homepage das file `co2-levels_MaunaLoaObsHawaii.csv` herunter. Es enthält Zahlen zur CO_2-Konzentration in der Atmosphäre für den Zeitraum 1958 - 2015.

a) Importieren Sie die Daten nach R und speichern Sie sie in dem Dataframe `co2levels`. Eliminieren Sie sämtliche Zeilen, die `NA`'s enthalten. Wieviele Observations bleiben dann noch übrig?

b) Speichern Sie die Daten der Spalte `co2 level` in dem Vektor `yfull` und die Daten der Spalte `trend comp` in dem Vektor `ytrend`. Speichern Sie weiterhin die Daten der Spalte `decimal date` in den Vektor `zeit`. Plotten Sie dann `yfull` und `ytrend` als Funktion von `zeit`. Dabei soll der Plot jeweils aus einer durchgezogenen Linie bestehen, also keine einzelnen Punkte.

c) Fitten Sie folgende Modelle (mit $t_0 = 1958.208 = \texttt{zeit}[1]$)

$$\begin{aligned} \texttt{ytrend}(t) &= C_{t_0}\, e^{r(t-t_0)} && (3)\\ \texttt{ytrend}(t) &= a_0 + a_1(t-t_0) && (4)\\ \texttt{ytrend}(t) &= b_0 + b_1(t-t_0) + b_2(t-t_0)^2 && (5) \end{aligned}$$

an die Daten, d.h., berechnen Sie die Koeffizienten $C_{t_0}, r, a_0, a_1, b_0, b_1$ und b_2.

d) Stellen Sie den Regression-Fit für alle drei Modelle, in rot, grün und blau, zusammen mit den Original-Daten in einem Diagramm dar.

e) Geben Sie für die Koeffizienten a_1 und b_1 jeweils ein 90%-Vertrauensintervall an.

Aufgabe 4 (10 Punkte): Betrachten Sie das Regressionsproblem aus Aufgabe 1b mit $\beta_0 = 4$ und $\beta_1 = -1$ und $\sigma = 2$. Generieren Sie $N = 10000$ Datenvektoren $\vec{y}$ mit der Spezifikation (2) und führen Sie für jedes dieser $\vec{y}$'s eine lineare Regression für das Modell (2) durch. Zeigen Sie dann, dass die Grösse

$$[n-(p+1)]\frac{\hat{\sigma}^2}{\sigma^2} \;=\; (11-2)\frac{\hat{\sigma}^2}{2^2} \;=:\; \xi$$

$\chi^2_{11-2} = \chi^2_9$-verteilt ist, indem Sie ein Histogramm der ξ's erzeugen und dann in dieses Histogramm einen Plot der Dichte der χ^2_9-Verteilung, in rot, addieren.

Aufgabe 5 (10 Punkte): Bestimmen Sie die Koeffizienten $\{c_k\}_{k=0}^{20}$ der Fourierreihen-Entwicklung

$$F_\ell(x) \;:=\; \frac{\sin\left[(\ell+\frac{1}{2})x\right]}{\sin(\frac{x}{2})} \;=\; c_0 + \sum_{k=1}^{\infty} 2c_k\,\cos(kx)$$

für den Fall $\ell = 10$ numerisch mit Hilfe einer linearen Regression. Gehen Sie dazu folgendermassen vor:

a) Legen Sie die Variablen $n = 1000$, $m = 20$ und den Vektor

```
x = seq( from = -pi , to = pi , length = n )
```

in R an, wir diskretisieren also das Intervall $[-\pi, \pi]$ mit $n = 1000$ Punkten.

b) Legen Sie die Matrix $X = \big(\cos(x), \cos(2x), \cdots, \cos(mx)\big)$ der Regressoren an, $m = 20$.

c) Legen Sie den Vektor $y = F_{\ell=10}(x)$ an und führen Sie schliesslich die lineare Regression durch.

d) Geben Sie explizit die Grössen $\{\, c_k\}_{k=0}^{20}$ an und plotten Sie sie als Funktion von k.

e) Plotten Sie die Funktion $F_{10}(x)$ und den Regression-Fit über dem Intervall $[-\pi, \pi]$ in einem Diagramm, den Regression-Fit in rot.

Aufgabe 6 (8 Punkte): In den Übungen haben wir viel mit Zufallszahlen zu tun gehabt.

a) Aus normal-verteilten Zufallszahlen kann man durch geeignete Kombination t_n-verteilte, $F_{k,\ell}$-verteilte und χ_n^2-verteilte Zufallszahlen generieren, wie geht das genau? (Hier müssen Sie nichts programmieren.)

b) Erzeugen Sie $N = 10000$ t_4-verteilte Zufallszahlen in R und stellen Sie sie in einem Histogramm dar. Wählen Sie die Skalierung des Histogramms so, dass die empirische Verteilung der Zufallszahlen mit der theoretischen Dichte der t_4-Verteilung vergleichbar ist. Stellen Sie beides in einem Diagramm dar, die theoretische Dichte in rot. Benutzen Sie den `breaks`-Parameter, um die Anzahl der dargestellten Rechtecke im Histogramm zu erhöhen.

c) Nehmen wir an, Sie wüssten von den Zufallszahlen aus (b) nur, dass sie t_n-verteilt sind, kennen aber den Wert von n nicht. Berechnen Sie die log-Likelihood-Funktion zum Schätzen von n und plotten Sie sie als Funktion von $n \in \{1, 2, 3, \cdots, 19, 20\}$.