Hochschule RheinMain
Prof. Dr. D. Lehmann

WS 2019/20

Probe-Klausur zur Vorlesung Ökonometrie

Theorie-Teil: Aufgaben 1-3: 30 Punkte

Programmier-Teil: Aufgaben 4-9: 60 Punkte

(die eigentliche Klausur wird nicht so umfangreich, etwa 2+4 Aufgaben mit 20+40 Punkten)

1.Aufgabe (10 Punkte): Es seien $\vec{y} \in \mathbb{R}^n$ und $\vec{x}_0, \vec{x}_1, \cdots, \vec{x}_p \in \mathbb{R}^n$. Wir betrachten das Regressions-Problem

$$\vec{y} \;=\; \sum_{j=0}^{p} \beta_j \, \vec{x}_j \;+\; \vec{\varepsilon}$$

wobei die $\vec{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)$ unabhängige, mit Mittelwert 0 und Standardabweichung σ normalverteilte Zufallszahlen sind.

a) Wie lautet der Kleinste-Quadrate-Schätzer $\hat{\beta}_j$ (oder ordinary least squares, OLS, Schätzer) für die β_j? Geben Sie die allgemeine Formel an, in der die $\vec{x}_j$ und das $\vec{y}$ drin vorkommen.

b) Wie lautet das Minimierungsproblem, aus dem man den Kleinste-Quadrate-Schätzer als Lösung bekommt?

c) Nehmen Sie jetzt an, dass die $\vec{x}_0, \cdots, \vec{x}_p$ ein $(p+1)-$dimensionales Orthonormalsystem sind, also $\vec{x}_j \cdot \vec{x}_k = \delta_{j,k}$ mit $\delta_{j,k} = 1$ für $j = k$ und 0 sonst. Vereinfachen Sie für diesen Fall die Formel für die $\hat{\beta}_j$ soweit wie möglich. Geben Sie ebenfalls eine Formel für den Regression-Fit $\hat{y}$ an und vereinfachen Sie sie soweit wie möglich.

d) Nehmen Sie wie in Teil (c) an, dass die $\vec{x}_0, \cdots, \vec{x}_p$ ein $(p+1)-$dimensionales Orthonormalsystem sind, also $\vec{x}_j \cdot \vec{x}_k = \delta_{j,k}$ mit $\delta_{j,k} = 1$ für $j = k$ und 0 sonst. Wie gross ist die Varianz der $\hat{\beta}_j$? Wie gross ist die Covarianz zwischen einem $\hat{\beta}_j$ und einem $\hat{\beta}_k$? Vereinfachen Sie Ihr Ergebnis soweit wie möglich.

2.Aufgabe (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

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

Geben Sie den OLS-Schätzer (oder Maximum-Likelihood-Schätzer, ist identisch) $\hat{\beta}_1$ an und vereinfachen Sie den Ausdruck soweit wie möglich. Benutzen Sie dazu die konkreten Werte der x_i (die y_i sollen variabel sein, da haben Sie also Buchstaben, keine Zahlen).

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

$$y_i \;=\; \beta_0 \,+\, \beta_1 x_i \,+\, \varepsilon_i$$

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.

3.Aufgabe (10 Punkte): Zufallszahlen $x_1, x_2, \cdots, x_n$ heissen Poisson-verteilt mit Parameter $\lambda > 0$, wenn sie ganzzahlig sind, $x_i \in \mathbb{N}_0 = \{\, 0\,,\, 1\,,\, 2\,,\, \cdots \}$ für alle i, und wenn

$$\mathsf{Prob}[\, x_i = k \,] \;=\; \frac{\lambda^k}{k!}\, e^{-\lambda}\,, \qquad k = 0, 1, 2, \cdots$$

Von der Stichprobe $x_1, x_2, \cdots, x_n$ sei bekannt, dass es mit Parameter $\lambda > 0$ Poisson-verteilte Zufallszahlen sind, aber der Wert von λ sei unbekannt.

a) 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.

b) Ist der Schätzer aus (a) erwartungstreu? Was müssten Sie hier genau zeigen? Versuchen Sie, möglichst weit zu rechnen.

Aufgabe 4 (10 Punkte): Bestimmen Sie die Koeffizienten $\{c_k\}_{k=0}^{10}$ der Potenzreihenentwicklung

$$\log[1 + x] \;=\; \sum_{k=0}^{10} c_k\, x^k \,+\, \text{error}$$

numerisch mit Hilfe einer linearen Regression. Gehen Sie dazu folgendermassen vor:

a) Legen Sie die Variablen $n = 1000$, $m = 10$ und den Vektor

```
x = seq( from = -0.5 , to = 0.5 , length = n )
```

in R an, wir diskretisieren also das Intervall $[-0.5, 0.5]$ mit $n = 1000$ Punkten.

b) Legen Sie die Matrix X der Regressoren $x, x^2, ..., x^m$ an.

c) Legen Sie den Vektor $y = \log[1 + x]$ an und führen Sie schliesslich die lineare Regression durch.

d) Plotten Sie die Grössen $\{\, c_k\}_{k=0}^{10}$ und $\{k\, c_k\}_{k=0}^{10}$ als Funktion von $k = 0, 1, ..., 10$.

e) Plotten Sie die Funktion $\log[1+x]$ und den Regression-Fit über dem Intervall $[-0.5, 0.5]$ in einem Diagramm, den Regression-Fit in rot.

Aufgabe 5 (10 Punkte):

a) Aus normal-verteilten Zufallszahlen kann man durch geeignete Kombination t_n-verteilte, $F_{k,\ell}$-verteilte und χ^2_n-verteilte Zufallszahlen generieren, wie geht das genau? (Hier müssen Sie nichts programmieren; Sie können Ihre Antwort auf ein Extra-Blatt schreiben oder als Text in den Code-Editor schreiben.)

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 $n \in \{1, 2, 3, \cdots, 19, 20\}$. Das Maximum sollte dann also in der Nähe von $n = 4$ liegen.

Aufgabe 6 (15 Punkte): Laden Sie sich von der Vorlesungshomepage die Datei `co2levels-MaunaLoa.csv` herunter. Sie 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`. Informieren Sie sich über den `na.omit()`-Befehl und eliminieren Sie dann 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 $t_0 = 1958.208 = \texttt{zeit}[1]$)

$$\texttt{ytrend}(t) = C_{t_0}\, e^{r(t-t_0)} \tag{1}$$
$$\texttt{ytrend}(t) = a_0 + a_1(t - t_0) \tag{2}$$
$$\texttt{ytrend}(t) = b_0 + b_1(t - t_0) + b_2(t - t_0)^2 \tag{3}$$

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 7 (10 Punkte): Betrachten Sie das Regressionsproblem aus Aufgabe 2b,

$$y_i \;=\; \beta_0 + \beta_1 x_i + \varepsilon_i \tag{4}$$

mit

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

und

$$\begin{aligned} \beta_0 &= 4 \\ \beta_1 &= -1 \end{aligned}$$

und $\sigma = 2$. Generieren Sie $N = 10000$ Datenvektoren $\vec{y}$ mit der Spezifikation (4) und führen Sie für jedes dieser $\vec{y}$'s eine lineare Regression für das Modell (4) durch. Zeigen Sie dann, dass die Grösse (für jedes Regressions-Resultat können Sie ein ξ berechnen)

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

mit

$$\hat{\sigma}^2 \;=\; \frac{(\vec{y}-\hat{y})^2}{n-(p+1)} \;=\; \frac{(\vec{y}-X\hat{\beta})^2}{n-(p+1)}$$

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

Aufgabe 8 (10 Punkte):

a) Erzeugen Sie $N = 10000$ mit Parameter $\lambda = 2$ exponential-verteilte Zufallszahlen.

b) Wieviele Zahlen würde man theoretisch in dem Intervall $[0, \frac{1}{2}]$ erwarten? Geben Sie eine konkrete Zahl an. Wieviele Zufallszahlen befinden sich tatsächlich in dem Intervall $[0, \frac{1}{2}]$?

c) Es gibt die R-Funktionen

```
dexp()
pexp()
qexp()
```

Wenn wir die Dichte der Exponential-Verteilung mit $p_\lambda(x)$ bezeichnen, was wird in diesen R-Funktionen jeweils berechnet? Das Resultat lässt sich also jeweils durch $p_\lambda(x)$ oder eine geeignete Transformation davon ausdrücken. (Hier müssen Sie nichts programmieren)

d) Wenn Sie die $N = 10000$ Zufallszahlen mit dem `sort()`-Befehl sortieren und dann plotten, stimmt das Bild bis auf eine Skalierung mit einem Plot der `qexp()`-Funktion überein. Finden Sie diese Skalierung und erzeugen Sie dann den entsprechenden Plot dazu, in dem die sortierten Zufallszahlen zusammen mit der `qexp()`-Funktion dargestellt sind, die `qexp()`-Funktion etwa in rot.

Aufgabe 9 (5 Punkte): Wie lautet die Dichte-Funktion $p_{\mu,\sigma}(x)$ der Normalverteilung mit Mittelwert μ und Standardabweichung σ? Erzeugen Sie $N = 10000$ mit Parametern $(\mu, \sigma) = (20, 4)$ normalverteilte Zufallszahlen und stellen Sie sie in einem Histogramm dar.