Methoden Verkehrsökonometrie für Master-Studierende

Winter semester 2021/22, Solutions to Tutorial No. 8

Lösungsvorschlag zu Aufgabe 8.1: Elastizitäten

(a) Exogene Variable: Zugangszeiten T_{ni} und Kosten K_{ni}. Beides sind Charakteristika, da sie von den Alternativen (Flughäfen) i abhängen. Sie werden generisch formuliert, also gleiche Zeit- und Kostensensitivitäten für alle Flughäfen.

(b)
- β_1: Globaler Bonus Berlin gegenüber der Referenz Dresden in NE. Der geringfügig positive Wert $\hat{\beta}_1 = 0.195$ der Schätzung bedeutet, dass bei gleichem Preis und gleicher Zugangszeit Berlin gegenüber Dresden um 0.195 NE bevorzugt wird.
- β_2: Globaler Bonus Frankfurt gegenüber der Referenz Dresden in NE. Hier ist die Bevorzugung bei gleicher Anreisezeit und gleichen Kosten 0.581 NE, was ebenfalls plausibel ist: Schließlich ist Frankfurt Knotenpunktsflughafen ("Hub") und bei den Flügen könnte es sich um Anschlussflüge handeln. Diese Eigenschaft ("Hub" oder nicht) wurde jedoch nicht explizit berücksichtigt.
- β_3: Zeitsensitivität. Der negative Wert $\hat{\beta}_3 = -0.0130$ ist konsistent mit der Tatsache, das ein erhöhter Zeitbedarf ein Nachteil ist. Der Zahlenwert (nur 0.013 NE pro Minuten) erscheint sehr gering. Schließlich entsprechen dann erst $1/\hat{\beta}_3 = 77$ Minuten einer Nutzeneinheit. Dies liegt aber daran, dass die Wege insgesamt länger als beim innerstädtischen ÖV sind und die Nutzenunschärfe generell eher relativer (in Prozent) als absoluter Natur (in Minuten) ist.
- β_4: Geldsensitivität. Der negative Wert $\hat{\beta}_3 = -0.0244$ ist konsistent mit der Tatsache, das erhöhte Kosten einen Nachteil darstellen.

(c)
- AC Berlin-Dresden in min: $-\beta_1/\beta_3 = 15.0$
- AC Berlin-Dresden in Euro: $-\beta_1/\beta_4 = 7.98$
- AC Frankfurt-Dresden in min: $-\beta_2/\beta_3 = 44.7$
- AC Frankfurt-Dresden in Euro: $-\beta_2/\beta_4 = 23.80$
- Zeitwert in Euro/h: $60\beta_3/\beta_4 = 32$ Euro/h
- Nutzeneinheit (NE) in Minuten: $1\,\text{NE} = 1/\beta_3 = 77$ Min
- Nutzenunschärfe: $\sigma_\epsilon = -\pi/\sqrt{6}\,1/\beta_3 = 99$ Min.

Der Zeitwert von 32 Euro/h erscheint sehr hoch. Man muss jedoch berücksichtigen, dass der Flugpreis für den Hin- *und* Rückflug gilt, die Zufahrtszeit jedoch für die *einfache* Fahrt. Der tatsächliche Zeitbedarf verdoppelt sich also, so dass der effektive implizite Zeitwert nur 16 Euro/h statt 32 Euro/h beträgt.

(d) Determinist. Nutzenfunktionen:

$$V_{61} = -6.19, \quad V_{62} = -7.47, \quad V_{63} = -8.31.$$

Nenner:

$$\sum_{i=1}^{3} e^{V_{6i}} = 0.00287.$$

Auswahlwahrscheinlichkeiten:

$$\text{Dresden: } P_{61} = 71.6\,\%, \quad \text{Berlin: } P_{62} = 19.8\,\%, \quad \text{Frankfurt: } P_{63} = 8.6\,\%.$$

Im Vergleich die beobachteten relativen Häufigkeiten:

Dresden: $h_{61} = 37/53 = 69.8\,\%$, Berlin: $h_{62} = 12/53 = 22.6\,\%$ Frankfurt: $h_{63} = 4/53 = 7.5\,\%$.

(e) Mikroskopische Preis-Eigenelastizitäten:

$$\epsilon_{nii}^{(\text{mic,K})} = \frac{K_{ni}}{P_{ni}} \frac{\partial P_{ni}}{\partial K_{ni}} = \beta_4 K_{ni} \left(1 - P_{ni}\right).$$

Für Person 6:

$$\begin{aligned} \epsilon_{611}^{(\text{mic,K})} &= \beta_4 K_{61} \left(1 - P_{61}\right) = -1.4, \\ \epsilon_{622}^{(\text{mic,K})} &= \beta_4 K_{62} \left(1 - P_{62}\right) = -4.9, \\ \epsilon_{633}^{(\text{mic,K})} &= \beta_4 K_{63} \left(1 - P_{63}\right) = -6.7 \end{aligned}$$

Mikroskopische Anreisezeit-Eigenelastizitäten analog:

$$\begin{aligned} \epsilon_{611}^{(\text{mic,T})} &= \beta_3 T_{61} \left(1 - P_{61}\right) = -0.37, \\ \epsilon_{622}^{(\text{mic,T})} &= \beta_3 T_{62} \left(1 - P_{62}\right) = -1.3, \\ \epsilon_{633}^{(\text{mic,T})} &= \beta_3 T_{63} \left(1 - P_{63}\right) = -1.4 \end{aligned}$$

(f) Mikroskopische Preis-Kreuzelastizitäten ϵ_{nij}=prozentualer relativer Anstieg der Wahrscheinlichkeit, dass Person n Flughafen i wählt, wenn Flughafen $j \neq i$ den Preis um 1 % anhebt: $\epsilon_{nij} = -\hat{\beta}_4 K_{nj} P_{nj}$, also z.B. für Person $n = 6$ und preissteigernder Flughafen Frankfurt ($j = 3$):

$$\epsilon_{613}^{(\text{mic,K})} = \epsilon_{623}^{(\text{mic,K})} = -\beta_4 K_{63} P_{63} = 0.63$$

Die Kreuzelastizität ist positiv: Erhöht der Flughafen Frankfurt seinen Preis um 1 %, *steigert* sich die Nachfrage in Dresden und in Berlin um jeweils 0.63 %. Die relative Steigerung ist aufgrund der IIA-Eigenschaft der Logitmodelle für beide Flughäfen dieselbe.

Natürlich gilt auch allgemein als Folge einer relativen Preiserhöhung, beispielsweise des Flughafens $j=3$, folgende Summenbedingung:

$$\begin{aligned}\sum_i P_{ni}\epsilon_{ni3} &= -\beta_4 P_{n1}K_{n3}P_{n3}) - \beta_4 P_{n2}K_{n3}P_{n3}) + \beta_4 P_{n3}K_{n3}(1-P_{n3}) \\ &= \beta_4 K_{n3}P_{n3}(-P_{n1}-P_{n2}+1-P_{n3}) \\ &= 0 \qquad\qquad [\text{da} \sum_i P_{ni}=1].\end{aligned}$$

Dies spiegelt wider, dass es sich hier um eine *Verschiebungselastizität*, also ein Nullsummenspiel handelt: Was jemand gewinnt, muss einem anderen weggenommen werden.

(g) Makroskopische Preis-Eigenelastizität:

$$\epsilon_{ii}^{(\text{mac,K})} = \frac{K_i}{N_i}\frac{\partial N_i}{\partial K_i}. \tag{1}$$

- $N_i = \sum_{n=1}^N P_{ni}$ die vom Modell geschätzte Gesamtnachfrage nach Flügen am Flughafen i (Gesamtzahl an potenziellen Passagieren: $N \geq N_i$),
- $K_i = \frac{1}{N}\sum_n K_{ni}$ der Mittelwert des Preises über alle *potenzielle* Kunden am Flughafen i (oder auch die Summe aller Preise, das kürzt sich bei der Elastizität weg).

Also allgemein

$$\epsilon_{ii}^{(\text{mac,K})} = \frac{K_i}{N_i}\sum_n \frac{\partial P_{ni}}{\partial K_i} \tag{2}$$

Fall 1: Alle Preise werden um einen festen *Betrag* verändert

Dann gilt die mittlere Preisänderung $\mathrm{d}K_i$ auch für jeden einzelnen potenziellen Kunden n: $\mathrm{d}K_{ni} = \mathrm{d}K_i = \text{const.}$ bzw.

$$\frac{\partial P_{ni}}{\partial K_i} = \frac{\partial P_{ni}}{\partial K_{ni}} = \frac{P_{ni}}{K_{ni}}\epsilon_{nii}$$

und damit

$$\epsilon_{ii}^{(\text{mac,abs,K})} = \frac{K_i}{N_i}\sum_n \frac{P_{ni}}{K_{ni}}\epsilon_{nii} \tag{3}$$

Fall 2: Alle Preise werden um einen festen *relativen Betrag* verändert

Dann gilt für jeden potenziellen Kunden n eine Preisänderung gemäß $\mathrm{d}K_{ni}/K_{ni} = \mathrm{d}K_i/K_i$, also

$$\frac{\partial P_{ni}}{\partial K_i} = \frac{K_{ni}}{K_i}\frac{\partial P_{ni}}{\partial K_{ni}} = \frac{P_{ni}}{K_i}\epsilon_{nii}$$

und damit

$$\epsilon_{ii}^{(\text{mac,rel,K})} = \frac{1}{N_i}\sum_n P_{ni}\epsilon_{nii} \tag{4}$$

Dies kann man auch als gewichtetes Mittel der Mikro-Elastizitäten schreiben:

$$\epsilon_{ii}^{(\text{mac,rel,K})} = \sum_n w_{ni}\epsilon_{nii}, \quad w_{ni} = \frac{P_{ni}}{N_i} = \frac{P_{ni}}{\sum_j P_{nj}} \tag{5}$$

die Wichtung beträgt dabei den erwarteten Anteil, den Kunde n zur Gesamtnachfrage N_i dieses Flughafens beiträgt.

Während die Mikro-Elastizitäten um so negativer sind, je kleiner die Auswahlwahrscheinlichkeit ist, mittelt sich dies bei den Makro-Elastizitäten weg. Die Makro-Elastizitäten sind die für die Betreiber eigentlich relevanten Größen, da sie die relativen Kunden-Reaktionen auf absolute bzw. relative Preisänderungen zeigen.

Hier würde bei einer Preiserhöhung am jeweiligen Flughafen um 1 %

- Dresden 2.6 %,
- Berlin 1.4 %,
- und Frankfurt 1.3 %

der Kunden verlieren. Ist die jeweilige makroskopische Preiselastizität größer als -1, könnte der Flughafen durch eine Preiserhöhung mehr Einnahmen erzielen, zumindest, wenn sich die Nachfrageänderungen gleichmäßig auf alle Preissegmente bezieht (worüber keine der beiden Makro-Elastizitäten Auskunft gibt!). Hier sollten jedoch (zumindest, wenn die jeweiligen anderen Flughäfen passiv bleiben!) alle FLughäfen ihren preis erniedrigen, um nicht nur mehr Kunden zu haben, sondern auch mehr Einnahmen zu erzielen.

(h) Für gleiche modellierte Auswahlwahrscheinlichkeiten in Dresden und Berlin,

$$\frac{P_1}{P_2} = e^{V_1 - V_2} = 1 \quad \Rightarrow \quad V_1 = V_2$$

kann man eine "Äquipotentialkurve" bezüglich der Flugkosten in Dresden und der Anreisezeit nach Dresden formulieren:

$$\ln\left(\frac{P_1}{P_2}\right) = V_1 - V_2 = -\beta_1 + \beta_3(T_1 - T_2) + \beta_4(K_1 - K_2)$$

und nach Auflösen bezüglich der Anreisezeit T_1 nach Dresden:

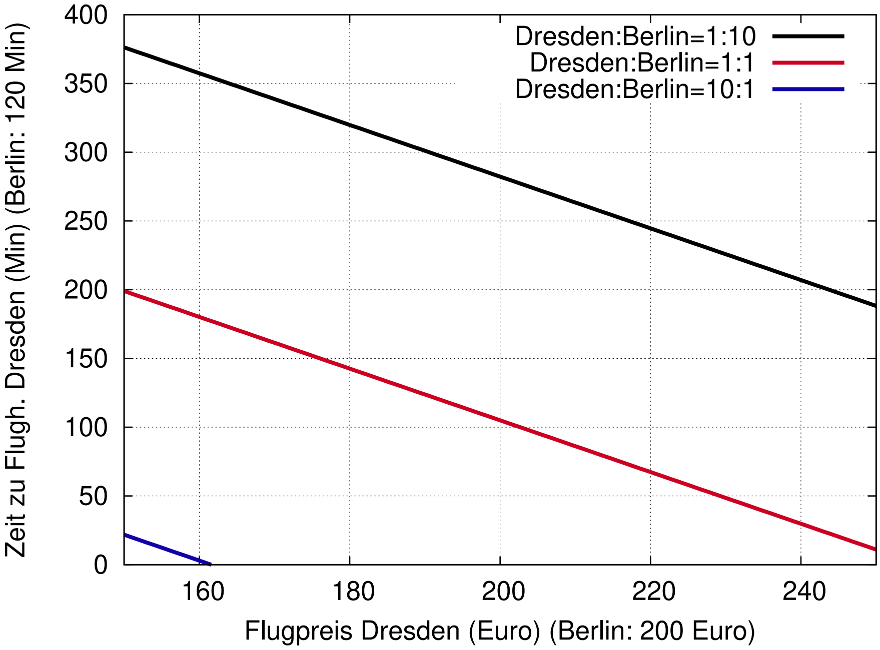
$$T_1(K_1) = \frac{\ln\left(\frac{P_1}{P_2}\right) + \beta_1 + \beta_3 T_2 - \beta_4(K_1 - K_2)}{\beta_3}$$

Bei gleichen Kosten und $P_1 = P_2$ ergibt sich die Anreisezeit

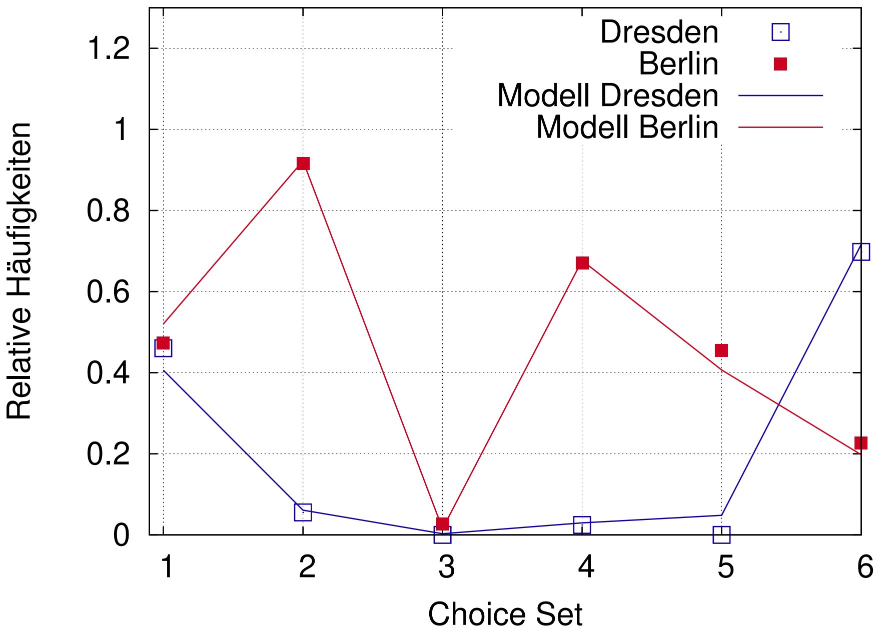
$$T_1 = T_2 + \frac{\beta_1}{\beta_3} = T_2 - 15\,\text{Min.}$$

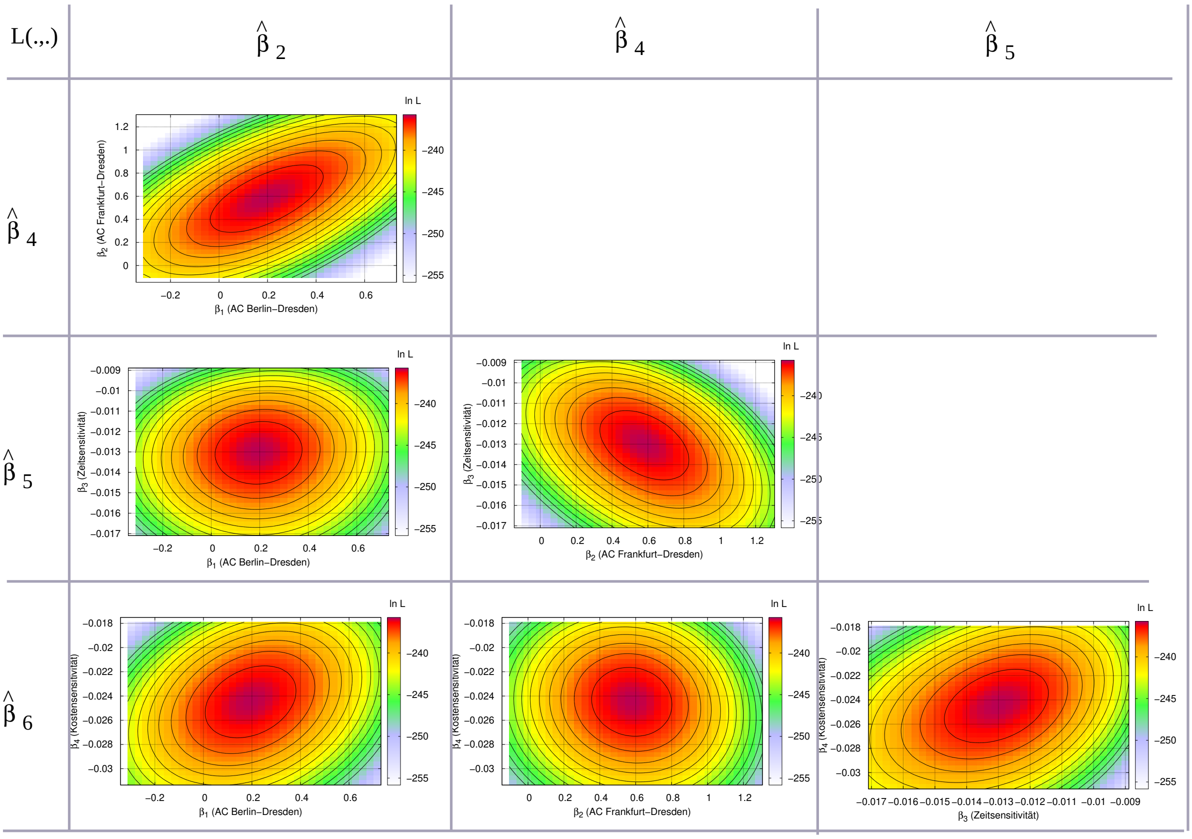
Diese Differenz ist genau der "globale Malus" in Minuten, den Dresden gegenüber Berlin hat, vgl. Teil (c). Die Steigung ist gleich des impliziten Geldwertes $-\beta_4/\beta_3 = -1.88$ Minuten/Euro. Konkret: Falls $T_2 = 120$ Min und $K_2 = 200$ Euro, so ist die Indifferenzgerade $P_1 = P_2$ gegeben durch

$$T_1(K_1) = \frac{\beta_1}{\beta_3} + 120\,\text{Min} - \frac{\beta_4}{\beta_1}(K_1 - 200\,\text{Euro}).$$



Weitere Ergebnisse der Parameterschätzung





Lösungsvorschlag zu Aufgabe 8.2: Maximum-Likelihood-Methode: Grundlagen

(a) (Binomialverteilung)

Geben Sie die analytische Likelihood-Funktion für die beobachteten relativen Häufigkeiten $y = h = 2$ und $y = h = 6$ an.

Einfach die angegebene Binomialverteilung für die Werte $y = 2$ bzw. $y = 6$ sowie den Stichprobenumfang $n = 10$ angeben:

$$h = 2: \quad L(\theta) = \binom{10}{2} \theta^2 (1-\theta)^8 = \underline{\underline{45\ \theta^2 (1-\theta)^8}}.$$

$$h = 6: \quad L(\theta) = \binom{10}{6} \theta^6 (1-\theta)^4 = \underline{\underline{210\ \theta^6 (1-\theta)^4}}.$$

Welcher Wert von θ (auf 0.1 gerundet) ergibt jeweils die Maximum - Likelihood - Parameterschätzung für die beiden Beobachtungswerte gemäß Tabelle?

Das Prinzip ist klar:

- Liest man die Tabelle von oben nach unten, ergibt sich die Wahrscheinlichkitsfunktion.

- Liest man die Tabelle von linkds nach rechts, erhält man die Liklihoodfunktion.

Zum Schätzen des Maximum-Likelihood-Wertes

$$\theta^{\text{est}} = \arg\max_{\theta} \left(P(h;\theta)\right)$$

anhand der Tabelle jeweils eine feste Zeile (gemäß der festen Häufigkeit $y = h$) nehmen und bezüglich der Spalten, d.h. des Parameters θ der Binomialverteilung maximieren. Direktes Ablesen ergibt

$$h = 2 \Rightarrow \theta^{\text{est}} = \underline{\underline{0.2}}, \quad h = 6 \Rightarrow \theta^{\text{est}} = \underline{\underline{0.6}},$$

also “modellierte Auswahlwahrscheinlichkeit=beobachtete relative Häufigkeit”, wobei die Auswahlwahrscheinlichkeit bei einer Entscheidung (Bernoulli-Experiment) hier direkt durch den Parameter θ der Binomialverteilung gegeben ist.

Führen Sie die Maximum-Likelihood-Parameterschätzung auch mit der analytischen Likelihood-Funktion durch!

Zur Veranschaulichung der Äquivalenz des Maximierens der Likelihood und der Log-Likelihood-Funktionen wird hier beides vorgeführt:

Direktes Maximieren der Likelihood-Funktion bei beobachteter Häufigkeit h und festem n:

$$L(\theta) = \binom{n}{h} \theta^h (1-\theta)^{n-h}$$

Damit

$$\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}\theta} &= \binom{n}{h} \left[h\theta^{h-1}(1-\theta)^{n-h} + \theta^h (n-h)(1-\theta)^{n-h-1}(-1)\right] \\
&= \binom{n}{h} \theta^{h-1}(1-\theta)^{n-h-1} \left[h(1-\theta) - \theta(n-h)\right] \\
&\overset{!}{=} 0.
\end{aligned}$$

Da der Vorfaktor vor der eckigen Klammer ungleich null ist, muss der Inhalt der eckigen Klammer verschwinden:

$$[h(1-\theta) - \theta(n-h)] = 0 \ \Rightarrow \underline{\underline{\theta = \frac{h}{n}}}.$$

Und damit für $h = 2$ oder $h = 6$ dasselbe Ergebnis wie oben. Die Maximum-Likelihood-Parameterschätzung ergibt also die relative Häufigkeit. Da θ die Wahrscheinlichkeit darstellt, im Bernoulli-Experiment (Einzelentscheidung) den Wert 1 (und nicht 0) zu bekommen, ist dies plausibel: Nach dem Gesetz der großen Zahlen gilt nämlich $f \to \theta$ für $n \to \infty$.

Direktes Maximieren der Log-Likelihood-Funktion bei beobachteter Häufigkeit h und festem n:

Die Log-Likelihood zu obiger Likelihood-Funktion ist gegeben durch

$$\tilde{L}(\theta) = \ln L(\theta) = \ln \begin{pmatrix} n \\ h \end{pmatrix} + h \ln \theta + (n-h)\ln(1-\theta).$$

Ableiten und Nullsetzen ergibt

$$\frac{\mathrm{d}\tilde{L}(\theta)}{\mathrm{d}\theta} = \frac{h}{\theta} + \frac{n-h}{1-\theta}(-1) \stackrel{!}{=} 0,$$

also

$$h(1-\theta) = \theta(n-h) \Rightarrow \theta = \frac{h}{n}.$$

Also – wie erwartet – dasselbe Ergebnis wie oben.

(b) (Poissonverteilung)

Mitarbeiter A tätigte 2 Anrufe, Mitarbeiter B 4 Anrufe. Geben Sie jeweils die analytische Likelihood-Funktion an und führen Sie die Maximum-Likelihood-Parameterschätzung des Parameters μ der Poissonverteilung (i) mit Hilfe der Tabelle, (ii) mit der analytischen Likelihood-Funktion, (iii) mit der Log-Likelihood-Funktion durch!

(i) $y = h = 2$ Anrufe: Maximieren bezüglich der Spalten bei fester Zeile $\Rightarrow \mu = 2$.

$y = h = 4$ Anrufe: Maximieren bezüglich der Spalten $\Rightarrow \mu = 4$.

(ii) Maximierung der Likelihood-Funktion

$$L(\mu) = \frac{\mu^h e^{-\mu}}{h!}$$

bezüglich des Parameters μ:

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}\mu} &= \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{e^{h\ln\mu}e^{-\mu}}{h!}\right) \\ &= \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\frac{e^{h\ln\mu-\mu}}{h!}\right) \\ &= \frac{1}{h!}e^{h\ln\mu-\mu}\left(\frac{h}{\mu}-1\right) \end{aligned}$$

Da der Vorfaktor der runden Klammer immer $\neq 0$ ist, muss die runde Klammer verschwinden $\Rightarrow \underline{\underline{\mu = h}}$.

Alternativ mit der Produktregel:

$$\begin{aligned} \frac{\mathrm{d}L}{\mathrm{d}\mu} &= \frac{1}{h!}\left[h\mu^{h-1}e^{-\mu} + \mu^h(-1)e^{-\mu}\right] \\ &= \frac{\mu^{h-1}e^{-\mu}}{h!}[h-\mu] \end{aligned}$$

(iii)

$$\tilde{L}(\mu) = -\ln(h!) + h \ln \mu - \mu$$

also

$$\frac{\mathrm{d}\tilde{L}(\mu)}{\mathrm{d}\mu} = \frac{h}{\mu} - 1 \Rightarrow \underline{\underline{\mu = h}}.$$

(c) **Gaußverteilung**:

Gegeben:

- Daten: wöchentliche Umsatz y_i (in Woche i)
- Modell: Gaußverteilung der Umsätze (Zentraler Grenzwertsatz): $Y_i \sim N(\mu, \sigma^2)$.

Likelihood-Funktiion:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}.$$

Log-Likelihood-Funktion:

$$\begin{aligned} \tilde{L}(\mu, \sigma^2) = \ln(L(\mu, \sigma^2)) &= \sum_{i=1}^{n} \left(\frac{-\ln 2\pi - \ln \sigma^2}{2} - \frac{(y_i - \mu)^2}{2\sigma^2} \right) \\ &= \text{const.} - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \end{aligned}$$

Ableiten und nullsetzen:[1]

$$\frac{\partial l}{\partial \mu} = -\frac{1}{2} \sum_{i=1}^{n} \left(-2 \frac{y_i - \mu}{\sigma^2} \right) \stackrel{!}{=} 0, \tag{6}$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2} \sum_{i=1}^{n} \left(\frac{(y_i - \mu)^2}{\sigma^4} \right) \stackrel{!}{=} 0. \tag{7}$$

Aus (6) erhält man

$$\sum_{i=1}^{n} y_i = n\mu \;\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} y_i = \underline{\underline{\bar{y}}}$$

und aus (7)

$$n = \sum_{i=1}^{n} \frac{(y_i - \mu)^2}{\sigma^2} \Rightarrow \hat{\sigma}^2 = \underline{\underline{\frac{1}{n} \sum_{i=1}^{n} (y_i - \mu)^2)}}.$$

Vergleichen Sie das Ergebnis mit der linearen Regression des "Trivialmodells" $\hat{y} = \mu =$ const.

[1] Man beachte, dass nach $\sigma^2 := u$ abgeleitet wird und nicht nach σ. Deshalb ist $\frac{\mathrm{d}(1/\sigma^2)}{\mathrm{d}\sigma^2} = \frac{\mathrm{d}(1/u)}{\mathrm{d}u} = -1/u^2 = -1/\sigma^4$.

Schätzt man den Parameter a der trivialen Regressionsfunktion $\hat{y}(x) = a$ anhand der Daten (x_i, y_i) ($x_i = i =$Woche i und dazugehöriger Umsatz y_i) mit der Regressionsrechnung ab, erhält man das Ergebnis

$$\hat{a} = \bar{y},$$

also stimmt die Schätzung der Regressionsrechnung mit der der Maximum-Likelihood-Methode überein. Bestimmt man nun mit den Mitteln der induktiven Statistik die Residualvarianz (nichterklärte Varianz) von $\hat{y}(x)$, so ist diese hier gleich des Schätzers für die Gesamtvarianz (da das Trivialmodell ja keinerlei Änderungen "erklärt"), also

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2.$$

Die Maximum-Likelihood-Varianzschätzer und Regressions-Varianzschätzer stimmen also formal *nicht* genau überein, wenn die Varianz nicht bekannt ist, also im Normalfall: Einerseits hängt der Maximum-Likelihood-Schätzer von μ und der Regressionsschätzer von $\hat{\mu} = \bar{y}$ ab, andererseits steht n statt $n-1$ im Nenner. Setzt man insbesondere im Maximum-Likelihood-Schätzer $\hat{\mu}$ für den wahren Wert μ ein, erhält man einen nicht-erwartungstreuen Schätzer![2] Der Regressionsschätzer ist also im Falle von Gaußverteilungen tendenziell vorzuziehen, dafür ist die Maximum-Lieklihood-Schätzung in vielen Situationen möglich, wo die Regression versagt, z.B. im Falle bekannter Korrelationen oder auch bei diskreten Wahlmodellen.

[2] Wäre allerdings μ in der Tat exakt bekannt, dann wäre auch $\hat{\sigma}^2$ erwartungstreu.