Klausur zur Vorlesung Statistik II,
WS 2009/10
für Bachelor-Studenten
Lösungsvorschlag

Aufgabe 1 **(50 Punkte)**

(a) **Glatte Komponente:** Periodendauer 1 Jahr bei Angabe in Quartalen ⇒ Periodenlänge $m = 4$. Damit gilt für die glatte Komponente:

$$\begin{aligned}
G_3 = G_{13} &= \frac{1}{4}\left(\frac{x_1}{2} + x_2 + x_3 + x_4 + \frac{x_5}{2}\right) = \underline{\underline{160}}, \\
G_4 = G_{14} &= \frac{1}{4}\left(\frac{x_2}{2} + x_3 + x_4 + x_5 + \frac{x_6}{2}\right) = \underline{\underline{167.5}}, \\
\cdots & \quad \cdots \\
G_{10} = G_{32} &= \frac{1}{4}\left(\frac{x_8}{2} + x_9 + x_{10} + x_{11} + \frac{x_{12}}{2}\right) = \underline{\underline{210.625}}.
\end{aligned}$$

also insgesamt

	Q1	Q2	Q3	Q4
2007	-	-	160.0	167.5
2008	168.8	172.5	180.6	188.8
2009	201.2	210.6	-	-

(bis auf den ersten Wert ergeben sich durch Rundung die Werte des Aufgabenblattes) Die ersten und letzten beiden Werte der glatten Komponente lassen sich mit symmetrischen gleitenden Mitteln nicht berechnen!

(b) **Saisonkomponente:** Zunächst werden die vorläufigen Saisonanteile durch die Differenz der arithmetischen Mittel der Spalten und der Spaltenmittel der glatten Komponenten berechnet:

$$\begin{aligned}
\tilde{S}_1 &= \frac{x_{21} + x_{31} - G_{21} - G_{31}}{2} = 122.5, \\
\tilde{S}_2 &= \frac{x_{22} + x_{32} - G_{22} - G_{32}}{2} = -57, \\
\tilde{S}_3 &= \frac{x_{13} + x_{23} - G_{13} - G_{23}}{2} = -5, \\
\tilde{S}_4 &= \frac{x_{14} + x_{24} - G_{14} - G_{24}}{2} = -68.5.5.
\end{aligned}$$

Es gilt $s_j = \tilde{S}_j - c$. Der Korrekturterm beträgt

$$c = \frac{1}{4}\sum_{j=1}^{4}\tilde{S}_j = -2,$$

und damit

$$S_1 = \underline{\underline{124.5}}, \quad S_2 = \underline{\underline{-55.0}}, \quad S_3 = \underline{\underline{-3.0}}, \quad S_4 = \underline{\underline{-66.5}}.$$

(c) **Zufallskomponmente** für das Jahr 2008: Allgemein gilt (in der Nomenklatur mit doppeltem Index) $U_{ij} = x_{ij} - G_{ij} - S_j$ also

$$U_{21} = \underline{\underline{-3.5}}, \quad U_{22} = \underline{\underline{2.0}}, \quad U_{23} = \underline{\underline{-8.0}}, \quad U_{24} = \underline{\underline{2.5}}.$$

(d) **Prognose der glatten Komponente für 2010:** Die letzte verfügbare glatte Komponente $G_{10} = G_{32}$ ist für das 2. Quartal 2009. Für 2010 muss man also maximal sechs Quartale prognostizieren:

$$\begin{aligned}
\text{1. Quartal 2010:} \quad G_{41} &= G_{13} = G_{10} + 3(G_{10} - G_9) = \underline{\underline{241}}, \\
\text{2. Quartal 2010:} \quad G_{42} &= G_{14} = G_{10} + 4(G_{10} - G_9) = \underline{\underline{251}}, \\
\text{3. Quartal 2010:} \quad G_{43} &= G_{15} = G_{10} + 5(G_{10} - G_9) = \underline{\underline{261}}, \\
\text{4. Quartal 2010:} \quad G_{44} &= G_{16} = G_{10} + 6(G_{10} - G_9) = \underline{\underline{271}}.
\end{aligned}$$

(e) **Prognose der Übernachtungszahlen für 2010:** Hier addiert man alle regelmäßigen Komponenten, also G und S. Damit

$$\begin{aligned}
\text{1. Quartal 2010:} \quad x_{41} &= G_{41} + S_1 = \underline{\underline{365.5}}, \\
\text{2. Quartal 2010:} \quad x_{42} &= G_{42} + S_2 = \underline{\underline{196.0}}, \\
\text{3. Quartal 2010:} \quad x_{43} &= G_{43} + S_3 = \underline{\underline{258.0}}, \\
\text{4. Quartal 2010:} \quad x_{44} &= G_{44} + S_4 = \underline{\underline{204.5}}.
\end{aligned}$$

(f) **Fehler bei direkter unbedingter Prognose:** Prognostiziert man die Übernachtungszahlen direkt durch die Fortschreibung der Zahlen selbst, erhält man starke Fehler dadurch, das der meiste Anteil des "Trends" in Wirklichkeit vom Auf- oder Abstieg des Saisomanteils kommt! Hier gilt[1]

$$\begin{aligned}
\text{1. Quartal 2010:} \quad h_{13} &= \underline{\underline{20}}, \\
\text{2. Quartal 2010:} \quad h_{14} &= \underline{\underline{-90}}, \\
\text{3. Quartal 2010:} \quad h_{15} &= \underline{\underline{-200}}, \\
\text{4. Quartal 2010:} \quad h_{16} &= \underline{\underline{-310}}.
\end{aligned}$$

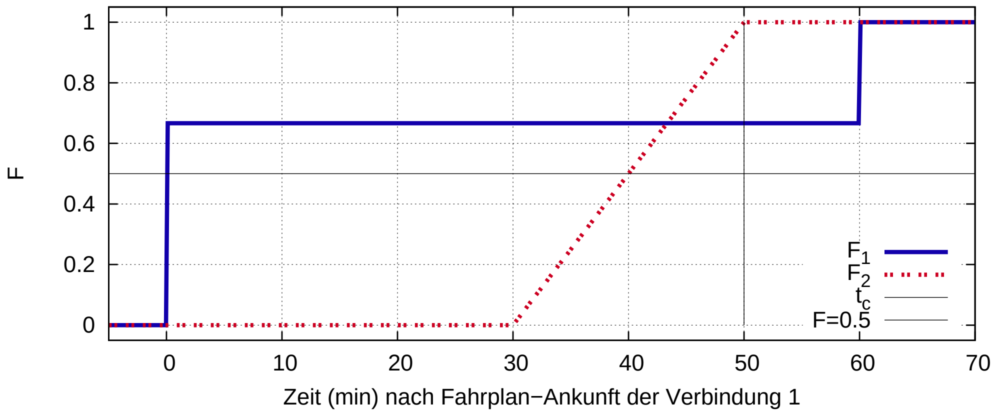
[1] Es gilt folgende Regel bei der Bedienung von Statistikprogrammen: "Junk in - Junk out".

Aufgabe 2 (30 Punkte)

(a) **Verteilungsfunktionen der Ankunftszeiten:** Die Verteilungsfunktionen der Verbindung 1 (mit Umsteigen) und Verbindung 2 (Direktverbindung) lauten [nicht verlangt, nur zeichnen]

$$F_1(T) = \begin{cases} 0 & T < 0 \\ 2/3 & 0 \le T \le 60 \\ 1 & \text{sonst.} \end{cases}, \quad F_2(T) = \begin{cases} 0 & T < 30 \\ \frac{T-30}{20} & 30 \le T \le 50 \\ 1 & \text{sonst.} \end{cases}.$$

Der Wert 2/3 in der Verteilungsfunktion ist gleich der Wahrscheinlichkeit p_{happy}, den Anschlusszug *nicht* zu verpassen. Mit 20 min fahrplanmäßige Umsteigezeit, einer $G(0, 30)$-gleichverteilten Verspätung der Ankunft des ersten Zuges und pünktlichen Anschlusszug wird der Anschluss verpasst, wenn die Verspätung größer als 20 min ist, was mit der Wahrsch. $1 - p_{\text{happy}} = 1/3$ geschieht.



(b) **Median** (vgl. die obige Abbildung):

$$\text{Verbindung 1: } T_{0.5} = \underline{\underline{0}}, \qquad \text{Verbindung 2: } T_{0.5} = 30 + 0.5 * 20 = \underline{\underline{40}}$$

Erwartungswert:

$$\begin{aligned} &\text{Verbindung 1:} \quad && E(T) = 2/3 * 0 + 1/3 * 60 = \underline{\underline{20}}, \\ &\text{Verbindung 2:} \quad && E(T) = 30 + 0.5 * 20 = \underline{\underline{40}}. \end{aligned}$$

(c) **Wahl der Verbindung**

(i) Falls das Ziel eine maximale Zeitersparnis im Mittel ist, sollte die Verbindung mit den kleineren Erwartungswert gewählt werden, also Verbindung 1.

(ii) Falls maximal 50 Minuten Verspätung erlaubt sind: Es sollte die Verbindung gewählt werden, für die dies mit höherer Wahrscheinlichkeit

$$P(T \le 50) = F(50)$$

erfüllt ist:

$$\text{Verbindung 1: } F(50) = \underline{\underline{2/3}}, \qquad \text{Verbindung 2: } F(50) = \underline{\underline{1}},$$

Also Verbindung 2.

(d) **Wahrscheinlichkeit und Streuparameter:**

– Standardabweichung:

$$\sigma_1^2 = 60^2 \left[\frac{2}{3} \left(\frac{1}{3}\right)^2 + \frac{1}{3} \left(\frac{2}{3}\right)^2 \right] = 60^2 \frac{6}{27}, \qquad \sigma = 20\sqrt{2} = \underline{\underline{28.2}},$$

$$\sigma_2^2 = \frac{20^2}{12}, \qquad \sigma_2 = \frac{20}{\sqrt{12}} = \underline{\underline{5.8}}.$$

– Wahrscheinlichkeit $P(T < 45) = F(45)$:

Verbindung 1: $F(45) = \underline{\underline{2/3}}$, Verbindung 2: $F(45) = \underline{\underline{3/4}}$.

– Spannweite:

Verbindung 1: $R = \underline{\underline{60}}$, Verbindung 2: $R = \underline{\underline{20}}$.

Aufgabe 3 **(40 Punkte)**

(a) **Schätzer:** Es sei X die Geschwindigkeit in Metern pro Sekunde:

$$\hat{\mu} = \bar{x} = \frac{1}{16}\sum_{i=1}^{16} x_i = \underline{\underline{1.526}},$$

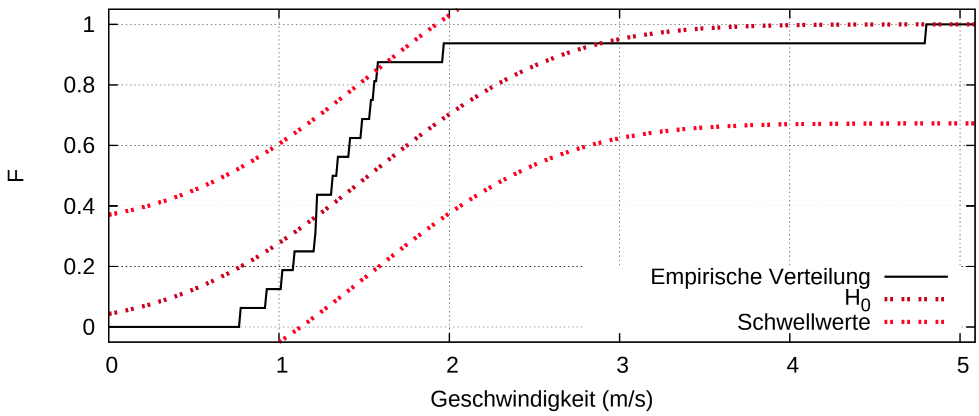
$$\hat{\sigma}^2 = \frac{1}{15}\sum_{i=1}^{16}(x_i - \bar{x})^2 = 0.793, \quad \hat{\sigma} = \underline{\underline{0.890}}.$$

(b) **Kolmo-Test:** Die Realisierung d der Test-Statistik D des Kolmogorow-Smirnow-Tests ist durch den maximalen vertikalen Abstand $\max_x(|F(x) - F_0(x)|)$ der empirischen Verteilung bezüglich der H_0-Verteilung gegeben. Aus der Abbildung der Aufgabenstellung liest man ab: $d = 0.35$ ($d = 0.4$ würde auch gehen, ebenfalls $d = 0.3$ mit falscher Schlussfolgerung).

Das $1-\alpha$ Quantil $d_{n,1-\alpha}$ der D-Statistik ist für $\alpha = 5\%$ und $n = 16$ nach der Vorlesung angenähert gegeben durch

$$d_{0.95,16} = \frac{1.358}{\sqrt{n} + 0.12 + 0.11/\sqrt{n}} = \underline{\underline{0.33}}.$$

Also kann H_0 abgelehnt werden, wenn auch knapp,[2] siehe Abbildung (die nicht verlangt war).

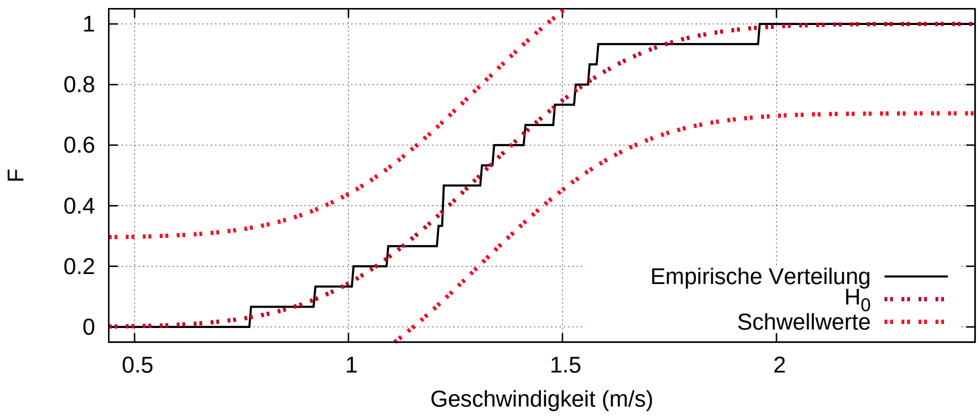


(c) **Ausreißer** Der Fußgänger mit $x = 4.8\,\mathrm{m/s}$ (etwa $17\,\mathrm{km/h}$) war ein Jogger (eher sogar ein veritabler Läufer). Da die Bemessung von Ampel-Freigabezeiten sich nach den langsamsten Fußgängern richten sollte, ist der Läufer offensichtlich irrelevant und sollte, ja muss aus den Daten eliminiert werden.[3]

Nach Eliminierung bekommt man folgenden Kolmo-Test (nicht verlangt):

[2] Falls man bei einer falschen Schätzung von d eine andere Schlussfolgerung bekommt, die Schlussfolgerung aber an sich logisch richtig ist, ergibt das volle Punktzahl.

[3] Ironischerweise würde die Berücksichtigung des Joggers sogar zu einer *längeren* Mindestfreigabezeit führen. Was da dahintersteckt, wird evtl. in einer der nächsten Klausuren gefragt.



(d) **Perzentile der Geschwindigkeiten:** Hierzu benötigt man die Quantilstabelle der Standardnormalverteilung. Hier ist bei Gültigkeit von H_0 $Z = (X - \hat{\mu})/\hat{\sigma} \sim N(0, 1)$ standardnormalverteilt, also

$$\begin{aligned} x_{0.9} &= \hat{\mu} + \hat{\sigma} z_{0.9} \overset{z_{0.9}=1.2815}{=} \underline{\underline{1.678}}, \\ x_{0.99} &= \hat{\mu} + \hat{\sigma} z_{0.99} \overset{z_{0.9}=2.326}{=} \underline{\underline{1.980}}. \end{aligned}$$

(e) **Perzentile der Überquerungszeiten:**2 Die Überquerungszeit $Y = b/X$ mit $b =$ 10 m ist *micht* normalverteilt: Dies gilt zwar für Summen, Differenzen und Linearkombinationen, aber *nicht* für nichtlineare Funktionen (beispielsweise ist $Z^2 \sim \chi^2(1)$ verteilt). Für auf Quantile basierte Untersuchungen ist dies jedoch egal. Da die Funktion $Y(X)$ streng monoton fallend ist, kehren sich die Quantile allerdings um: Die größte Geschwindigkeit führt zur kleinsten Überquerungszeit, also

$$y_q = \frac{b}{x_{1-q}} = \frac{b}{\hat{\mu} + \hat{\sigma} z_{1-q}} = \frac{b}{\hat{\mu} - \hat{\sigma} z_q}$$

Die letzte Umformung ist nötig, da nur Quantile $q \geq 0.5$ in der Quantilstabelle aufgeführt sind. Also hier (Überquerungszeit-Quantile y_q in Sekunden):

$$\begin{aligned} y_{0.9} &= \frac{b}{\hat{\mu} - \hat{\sigma} z_{0.9}} = \underline{\underline{10.7}}, \\ y_{0.99} &= \frac{b}{\hat{\mu} - \hat{\sigma} z_{0.99}} = \underline{\underline{15.8}}. \end{aligned}$$