Lösung zu Aufgabe 21

Mit der Tabelle aus dem Hinweis erhält man die folgenden beobachteten und erwarteten Sequenzhäufigkeiten (Hochrechnen der günstigen Fälle auf $N = 40003600$)

Sequenzlänge	beobachtet	erwartet (Formel)	erwartet (numerisch)	$\frac{(O_i-E_i)^2}{E_i}$
1	20190890	$N\frac{\binom{44}{6}}{\binom{49}{6}}$	$20193879,31$	$0,442509046$
2	17904796	$N\frac{\binom{44}{5}\binom{5}{1}+\binom{44}{4}\binom{4}{2}+\binom{44}{3}}{\binom{49}{6}}$	$1783404,78$	$0,534219134$
3	1782830	$N\frac{\binom{44}{4}\binom{4}{1}+\binom{44}{3}\binom{3}{2}+\binom{44}{2}}{\binom{49}{6}}$	$1783404,78$	$0,185247627$
4	119558	$N\frac{\binom{44}{3}\binom{3}{1}\binom{2}{1}+\binom{44}{2}\binom{1}{1}}{\binom{49}{6}}$	$119074,0672$	$1,966767076$
5	5384	$N\frac{\binom{44}{2}\binom{2}{1}}{\binom{49}{6}}$	$5412,457601$	$0,149624277$
6	142	$N\frac{\binom{44}{1}}{\binom{49}{6}}$	$125,8711070$	$2,066726794$
$\hat{\chi}^2$				$5,3451$

Das ergibt die Chi-Quadrat-Statistik

$$\hat{\chi}^2 = 5,345093953$$

mit Stichprobenverteilung χ^2_5 und p-value $\int_{\hat{\chi}^2}^{\infty} p(y)dy = 0,375230952$. Der Schwellenwert (bspw.) für $\alpha = 0.05$ ist $\chi_{0.95}(5) \approx 11.07$

Die Hypothese dass der Quicktipp-Generator Sequenzlängen gemäß einer Laplace-Urnenziehung erzeugt, kann zu keinem sinnvollen Signifikanzniveau abgelehnt werden.

Lösung zu Aufgabe 22

	Verletzungsschwere				
Gurt benutzt	keine	leicht	mäßig	schwer	Gesamt
nein	65963 (66191,8)	4000 (3904,7)	2642 (2521,6)	303 (289,9)	72908
ja	12813 (12584,2)	647 (742,3)	359 (479,4)	42 (55,1)	13861
Gesamt	78776	4647	3001	345	86769

(in Klammern: Erwartete Zahlen)

Beispielsweise Eintrag links oben: $\frac{72908\times78776}{86769} \approx 66191,8$

Die Chi-Quadrat-Statistik lautet dann

$\chi^2 = \frac{(65963-66191,8)^2}{66191,8} + \frac{(12813-12584,2)^2}{12584,2} + \frac{(4000-3904,7)^2}{3904,7} + \frac{(647-742,3)^2}{742,3} + \frac{(2642-2521,6)^2}{2521,6} + \frac{(359-479,6)^2}{479,6} + \frac{(303-289,9)^2}{289,9} + \frac{(42-55,1)^2}{55,1} \approx 59,224$

Unter der Nullhypothese „unabhängige Merkmale“besitzt die Prüfgröße eine $\chi^2((2-1)(4-1))$-Verteilung mit Verteilungsfunktion $F_{\chi^2(3)}$

Als Schwellenwert für den Chi-Quadrat-Test zum Niveau α dient $F^{-1}_{\chi^2(3)1-\alpha}(3)$. Die Werte betragen 16.266236 für $\alpha = 0.001$, 11.344867 für $\alpha = 0.01$, 7.814728 für $\alpha = 0.05$ und 6.251389 für $\alpha = 0.1$. χ^2 liegt also für gängige Werte von α oberhalb dieses Schwellenwertes, die Nullhypothese muss also zu jedem gängigen Signifikanzniveau abgelehnt werden.

Dies erkennt man auch an dem p-value $1 - F_{\chi^2(3)}(59, 224) \approx 8,6 \times 10^{13}$, der unterhalb jedes gängigen Signifikanzniveaus liegt.

Lösung zu Aufgabe 23

a)

j	x_j	$z_j = \frac{x_j - \bar{x}}{s}$	$F(x_j) = \Phi(z_j)$	$\frac{j-1}{5}$	$\lvert F(x_j) - \frac{j-1}{5} \rvert$	$\frac{j}{5}$	$\lvert F(x_j) - \frac{j}{5} \rvert$
1	-1.00	-1.3316	0.0915	0.0	0.0915	0.2	0.1085
2	-0.20	-0.5696	0.2845	0.2	0.0845	0.4	0.1155
3	0.45	0.0495	0.5198	0.4	0.1198	0.6	0.0802
4	1.05	0,6210	0.7327	0.6	0.1327	0.8	0.0673
5	1.69	1,2306	0.8907	0.8	0.0907	1.0	0.1093

Spalte 4 ergibt die Werte der theoretischen Verteilungsfunktion. In Spalte 6 und 8 stehen die zu vergleichenden Werte, also ist das Maximum $D_5 = 0.1327$. Die KS-Statistik wird noch mit $\sqrt{5}$ multipliziert, also $\sqrt{5}D_5 = 0.2967$. Der entsprechende Schwellenwert der KS-Quantiltabelle ist $d_{0.95}(5) = 0.76$. Weil der Wert der KS-Statistik unterhalb liegt, wird die Nullhypothese nicht verworfen (der KS-Test entscheidet recht häufig so, er hat einen recht großen Fehler 2. Art. Zum Test z.B. der grundsätzlichen NV-Annahme gibt es bessere Tests, z.B. den Shapiro-Wilk-Test).

b) Da die EVF stückweise konstant ist, setzt sich die Differenz $F(x) - \hat{F}(x)$ aus stückweise, d.h. auf den Intervallen $[x_{(j)}, x_{(j+1)}[$ monoton wachsenden Funktionen zusammen. Die Infima und Suprema der Differenzen werden daher stets an den Grenzen der Intervalle angenommen, je nach Lage der beiden VF zueinander als

- rechtsseitige Limiten $\lim\limits_{x \downarrow x_{(j)}} F(x) - \hat{F}(x) = F(x_{(j)}) - \hat{F}(x_{(j)}) = F(x_{(j)}) - \frac{j}{n}$
- linksseitige Limiten $\lim\limits_{x \uparrow x_{(j)}} F(x) - \hat{F}(x) = F(x_{(j)}) - \frac{j-1}{n}$.

Die KS-Statistik ist daher das Maximum der Absolutbeträge dieser „Rand-Differenzen“. Zu beachten ist, dass bei den Anfangs- und Endintervallen die Werte $0 = \hat{F}(-\infty) = \frac{1-1}{n}$ bzw. $1 = \hat{F}(\infty) = \frac{n}{n}$ berücksichtigt werden.