Beispielsammlung zur Vorlesung Statistik
142.090 VO

R. Frühwirth

Sommersemester 2012

Teil 1

Deskriptive Statistik

Beispiel 1.1*

Berechnen Sie für die Datenreihe

1.23	0.59	12.38	0.54	2.75	13.96	7.67	3.62
0.26	0.21	11.09	0.18	0.26	4.34	1.34	11.72

a) das Stichprobenmittel;

b) den Stichprobenmedian;

c) die Quartile;

d) die Interquartilsdistanz;

e) den Schiefekoeffizienten;

f) den Wert der empirischen Verteilungsfunktion an der Stelle $x = 1$.

g) Zeichnen sie einen Boxplot der Datenreihe.

Beispiel 1.2*

In einer Ortschaft haben 55% der Haushalte einen PKW, 34% ein Moped/Motorrad, 13% haben beides. Welcher Anteil hat weder PKW noch Moped/Motorrad?

* Die durch * gekennzeichneten Beispiele sind besonders relevant für die Prüfung.

Beispiel 1.3*

a) Ergänzen Sie die folgende Vierfeldertafel und rechnen Sie auf relative Häufigkeiten um:

	B	B'	
A	116		305
A'			
		425	1000

b) Berechnen Sie die bedingte Häufigkeit $f(A|B')$.

c) Berechnen Sie die empirische Vierfelderkorrelation r.

Teil 2

Wahrscheinlichkeitsrechnung

Beispiel 2.1

Eine Münze wird dreimal geworfen. Was ist die Ergebnismenge des Experiments? Welches Ereignis beschreibt die Aussage „Es wird öfter Kopf als Zahl geworfen“?

Beispiel 2.2

Es seien A, B, C drei beliebige Ereignisse. Die folgenden Ereignisse sind durch logische Verknüpfung von A, B und C darzustellen:

a) A tritt nicht ein;

b) Nur A tritt ein;

c) Weder B noch C treten ein;

d) Alle drei Ereignisse treten ein;

e) Genau eines der drei Ereignisse tritt ein;

f) Keines der drei Ereignisse tritt ein;

g) Wenigstens zwei der drei Ereignisse treten ein;

h) Genau zwei der drei Ereignisse treten ein;

i) Höchstens zwei der drei Ereignisse treten ein.

Beispiel 2.3

Es seien A und B zwei Ereignisse und $W(\cdot)$ ein Wahrscheinlichkeitsmaß. Beweisen Sie:

a) $W(A \cup B) \leq W(A) + W(B)$

b) $W(A \cap B) \geq W(A) + W(B) - 1$

c) $W(A \cap B') = W(A) - W(A \cap B)$

d) $W(A' \cap B') = 1 - W(A) - W(B) + W(A \cap B)$

Beispiel 2.4

Beim Bridgespiel wird ein Paket von 52 Karten auf 4 Spieler verteilt, sodass jeder Spieler 13 Karten erhält. Wie groß ist die Wahrscheinlichkeit, dass

a) Sie kein As haben?

b) Sie mindestens ein As haben?

c) Sie alle 4 Asse haben?

d) irgendein Spieler alle 4 Asse hat?

e) Ihr Partner genau 2 Asse hat?

f) Sie gemeinsam mit Ihrem Partner alle 4 Asse haben?

Beispiel 2.5

A und B seien zwei Ereignisse. Zeigen Sie:

a) Wenn A und B unabhängig sind, dann sind auch A und B', A' und B, sowie auch A' und B' unabhängig.

b) Ist $W(A \mid B) = W(A \mid B')$, so sind A und B unabhängig.

Beispiel 2.6*

A und B seien zwei Ereignisse mit $W(A) = 3/4, W(B) = 2/3$. Wie groß muß $W(A \cap B)$ mindestens sein? Wie groß ist $W(A \cap B)$ unter der Annahme der Unabhängigkeit von A und B? Berechnen Sie für diesen Fall die Wahrscheinlichkeiten der folgenden Ereignisse:

a) Keines der beiden Ereignisse tritt ein

b) Genau eines der beiden Ereignisse trit ein

c) Beide Ereignisse treten ein

d) Mindestens eines der beiden Ereignisse tritt ein

e) Höchstens eines der beiden Ereignisse tritt ein

Beispiel 2.7*

In einem Experiment wurde der Trigger so aufgesetzt, dass nur 1% der gewünschten Ereignisse verworfen wird. Die Analyse der aufgezeichneten Ereignisse ergab, dass davon 96% vom erwünschten Typ sind. Insgesamt wurden 92% aller Ereignisse aufgezeichnet.

a) Wie groß ist der ursprüngliche Anteil an erwünschten Ereignissen?

b) Wieviel Prozent der Untergrundereignisse wurden vom Trigger verworfen?

c) Wie groß ist der Anteil an erwünschten Ereignissen an den nicht aufgezeichneten?

Beispiel 2.8*

Ein Experiment verwendet ein große Zahl von ICs. Es bezieht diese von drei verschiedenen Herstellern A, B und C, und zwar von A und B je 25%, und von C 50%. Die Wahrscheinlichkeit, dass ein IC mindestens 40000 Stunden fehlerfrei arbeitet, beträgt für die drei Hersteller 0.92, 0.95 und 0.97.

a) Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählter IC mindestens 40000 Stunden arbeitet?

b) Eine IC fällt vor Ablauf der 40000 Stunden aus. Mit welcher Wahrscheinlichkeit stammt er von A, B oder C?

Beispiel 2.9 (M)

Für zwei Merkmale A und B gilt: $W(A) = 0.65$, $W(A|B) = 0.53$, $W(B|A) = 0.34$. Erzeugen Sie 1000 Vierfeldertafeln mit jeweils 250 Einträgen gemäß diesen Wahrscheinlichkeiten und stellen Sie die Verteilung der Vierfelderkorrelation in einem Histogramm dar. Wie ändert sich die Verteilung, wenn die Anzahl der Einträge verdoppelt wird?

Teil 3

Zufallsvariable

Beispiel 3.1*

Lieferungen eines Produktionsbetriebs, bestehend aus Serien zu je 100 Stück, werden vom Empfänger kontrolliert. Es werden Stichproben vom Umfang 5 gezogen, und die Serie wird zurückgewiesen, wenn mindestens ein Stück der Stichprobe mangelhaft ist. Wie groß ist die Wahrscheinlichkeit, dass eine Serie mit 5% Ausschuß zurückgewiesen wird? Rechnen Sie mit und ohne Zurücklegen.

Beispiel 3.2*

Sie entnehmen einer Lieferung von Widerständen zufällig n Stück, wobei Sie jedes Stück wieder zurücklegen. Sie wissen, dass die Fehlerquote bei der Produktion durch die Zahl p gegeben ist.

a) Wie groß ist die Wahrscheinlichkeit, dass Sie k fehlerhafte Widerstände ziehen?

b) Wie groß ist die zu erwartende Anzahl von fehlerhaften Stücken, und wie groß ist ihre Varianz?

c) Wie klein muß p sein, damit die Wahrscheinlichkeit einer fehlerlosen Stichprobe vom Umfang 10 über 99% steigt?

Beispiel 3.3*

Eine Blasenkammer expandiert während einer Sekunde, in der sie den einfallenden Strahl hat, 30 mal. Die Wahrscheinlichkeit, dass während einer Expansion eine Wechselwirkung stattfindet, ist $p = 0.2$. Jede Wechselwirkung wird photographiert.

a) Wie ist die Anzahl der Bilder pro Sekunde verteilt?

b) Wie groß ist die mittlere Anzahl und die Standardabweichung?

c) Wie groß ist die Wahrscheinlichkeit, pro Sekunde mindestens ein Bild zu nehmen?

d) Wie groß ist die Wahrscheinlichkeit, in einer Sekunde alle 30 Bilder zu nehmen?

Beispiel 3.4

Die Lebensdauer eines elektrischen Bauteils ist exponentialverteilt mit dem Mittel τ. Sie schalten N gleichartige Bauteile gleichzeitig ein. Wie ist die Wartezeit bis zum ersten Ausfall verteilt? Wie ist die Wartezeit bis zum letzten Ausfall verteilt?

Beispiel 3.5

Sie messen die Aktivität einer Quelle mit einer mittleren Zerfallsrate μ. Die Ansprechwahrscheinlichkeit Ihres Detektors beträgt jedoch nur 80%. Wie ist die Zahl der beobachteten Zerfälle verteilt? Wie groß ist ihr Mittelwert? Wie ist die Wartezeit zwischen zwei beobachteten Zerfällen verteilt?

Beispiel 3.6

Sie führen einen Alternativversuch mit Erfolgswahrscheinlichkeit p solange durch, bis sie einen Erfolg erzielt haben. Wie ist ist die Anzahl n der dazu notwendigen Versuche verteilt? Was ist der Erwartungswert von n?

Beispiel 3.7*

Wie ist die Summe von zwei unabhängigen Poisson-verteilten Zufallsvariablen verteilt? Wie hängt die mittlere Wartezeit des Summenprozesses von den mittleren Wartezeiten der Einzelprozess ab?

Beispiel 3.8*

Sie messen den inversen Impuls $q = 1/p$ eines Teilchens mit einem relativen Fehler von 10%. Berechnen Sie mit Fehlerfortpflanzung den relativen Fehler des Impulses p.

Beispiel 3.9*

Sie messen an einem Widerstand eine Spannung von $U = 12\,\mathrm{V}$ und eine Stromstärke von $I = 0.2\,\mathrm{A}$, mit einem Fehler von $0.1\,\mathrm{V}$ bzw. $1\,\mathrm{mA}$. Berechnen Sie mit Fehlerfortpflanzung den relativen Fehler des Widerstands $R = U/I$.

Teil 4

Schätzung

Beispiel 4.1*

Eine Stichprobe vom Umfang $n = 500$ stammt aus einer Gammaverteilung mit $a = 3, b = 2$. Bestimmen Sie die Standardabweichung

a) des Stichprobenmittels

b) des Stichprobenmedians (asymptotischer Wert)

Beispiel 4.2*

Eine bestimmte Operation wurde N mal ausgeführt und ist dabei n mal gelungen. Wie würden Sie die Erfolgswahrscheinlichkeit schätzen? Welche Eigenschaften hat die Schätzung?

Beispiel 4.3*

Um die Wahrscheinlichkeit p eines seltenen Ereignisses zu schätzen, wiederholen Sie einen Versuch so lange, bis das Ereignis eintritt. Schätzen Sie p aus der Anzahl n der benötigten Versuche. Welche Eigenschaften hat die Schätzung? Welche Funktion von p kann erwartungstreu geschätzt werden?

Beispiel 4.4*

Sie wollen die Aktivität (=mittlere Zerfallsrate) einer Quelle bestimmen und messen dazu N mal die Anzahl der Zerfälle pro Sekunde. Ihr Detektor hat eine Ansprechwahrscheinlichkeit von 75%. Bestimmen Sie die Maximum-Likelihood-Schätzung der Aktivität.

Beispiel 4.5*

Die Stichprobe $x_1, \dots, x_n$ entstammt einer Gammaverteilung mit bekanntem Formparameter a und unbekanntem Skalenparameter b. Berechnen Sie den ML-Schätzer von b, bestimmen Sie seine Verteilung und berechnen Sie die Fisher-Information der Stichprobe. Welche Eigenschaften hat der Schätzer?

Beispiel 4.6*

Sie messen die Masse m einer Probe n mal und erhalten die Messreihe $x_1, \dots, x_n$. Schätzen Sie m unter der Annahme, dass die Messfehler unabhängig und normalverteilt sind, und zwar

a) mit der bekannten Varianz σ^2,

b) mit unbekannter Varianz,

c) mit den bekannten Varianzen $\sigma_1^2, \dots, \sigma_n^2$.

Geben Sie im Fall b) auch eine Schätzung der unbekannten Varianz an.

Beispiel 4.7*

Eine Messreihe der Länge $n = 50$ stammt aus einer Exponentialverteilung mit unbekanntem Mittelwert τ. Die Summe aller Beobachtungen ist gleich $T = 102.37$.

a) Bestimmen Sie den Maximum-Likelihood-Schätzwert $\hat{\tau}$ von τ.

b) Bestimmen Sie das symmetrische 95%-Konfidenzintervall $[T_1, T_2]$ für den unbekannten Mittelwert τ.

Beispiel 4.8 (M)

Erzeugen Sie 1000 normalverteilte Stichproben der Größe $n = 50$ mit Mittel $\mu = 5$ und Varianz $\sigma^2 = 2$.

a) Bestimmen Sie für jede Stichprobe das symmetrische 95%-Konfidenzintervall $[M_1, M_2]$ für den unbekannten Mittelwert μ, unter der Annahme unbekannter Varianz. Zählen Sie, wie oft der wahre Wert vom Konfidenzintervall überdeckt wird.

b) Bestimmen Sie für jede Stichprobe das 95%-Konfidenzintervall $[V_1, V_2]$ für die unbekannte Varianz σ^2. Zählen Sie, wie oft der wahre Wert vom Konfidenzintervall überdeckt wird.

Beispiel 4.9

Die Stichprobe $x_1, \ldots, x_n$ entstammt einer Gleichverteilung im Intervall $[0,a]$ mit unbekannter oberer Grenze a.

a) Berechnen Sie den ML-Schätzer von a und untersuchen Sie seine Eigenschaften.

b) Korrigieren Sie den ML-Schätzer so, dass er unverzerrt ist.

c) Suchen Sie einen unverzerrten Schätzer, der auf dem Stichprobenmittel basiert.

d) Vergleichen Sie die Varianzen der beiden Schätzer.

Beispiel 4.10 (M)

Schätzen Sie den Parameter a einer Gammaverteilung

$$f(x) = \frac{x^{a-1}\, e^{-x}}{\Gamma(a)}, \quad x \geq 0$$

mit Hilfe der Maximum-Likelihood-Methode aus einer simulierten Stichprobe vom Umfang 250 (a=2). Bestimmen Sie den Fehler der Schätzung durch 2000-malige Wiederholung der Stichprobe. Vergleichen Sie die resultierende Verteilung des Schätzwertes mit einer individuellen Likelihoodfunktion.

Teil 5

Testen von Hypothesen

Beispiel 5.1*

Von 600 befragten Personen geben 123 eine Präferenz für Partei X an.

a) Schätzen Sie das Wählerpotential p für diese Partei.

b) Geben Sie ein Konfidenzintervall für das Wählerpotential an.

c) Testen Sie, ob sich das Wählerpotential im Vergleich zur letzten Wahl ($p_0 = 0.18$) signifikant vergrößert hat ($\alpha = 0.05$).

Beispiel 5.2*

Bei einer Meinungsumfrage mit 800 Befragten geben 291 Personen an, Produkt X zu kennen.

a) Geben sie ein 95%-iges symmetrisches Konfidenzintervall für den Bekanntheitsgrad des Produkts an (Bootstrapmethode).

b) Testen Sie die Hypothese, dass der Bekanntheitsgrad mindestens 40% beträgt ($\alpha = 0.05$). Welchen Wert hat die Testgröße T? Muss die Hypothese verworfen werden? (Benützen Sie die Näherung durch Normalverteilung!)

Beispiel 5.3*

Eine Stichprobe vom Umfang $n = 50$ aus einer Normalverteilung hat das Stichprobenmittel $\bar{x} = 150.45$ und die Stichprobenvarianz $S^2 = 4.12$.

a) Berechnen Sie ein 95%-iges symmetrisches Konfidenzintervall für den unbekannten Mittelwert der Verteilung.

b) Berechnen Sie ein 95%-iges Konfidenzintervall für die unbekannte Varianz der Verteilung.

c) Testen Sie die Hypothese, dass der unbekannte Mittelwert höchstens 150 ist ($\alpha = 0.01$). Welchen Wert hat die Testgröße? Muss die Hypothese verworfen werden?

Beispiel 5.4*

Die folgende Datenreihe stammt aus einer Poissonverteilung mit unbekanntem Mittel λ:

24	20	27	28	15	29	17	19	27	26
30	36	29	32	18	16	34	27	20	20

a) Berechnen Sie den Maximum-Likelihood-Schätzer von λ.

b) Geben Sie ein 95%-iges Konfidenzintervall für λ an. (Benützen Sie die Näherung durch Normalverteilung!)

c) Testen Sie die Hypothese $\lambda \leq \lambda_0 = 24$ ($\alpha = 0.01$). Welchen Wert hat die Testgröße T? Muss die Hypothese verworfen werden? (Benützen Sie die Näherung durch Normalverteilung!)

Beispiel 5.5 (M)

Erzeugen Sie 1000 Poisson-verteilte Stichproben der Größe $n = 20$ mit Mittel $\lambda = 8$.

a) Testen Sie die Hypothese $H_0 : \mu \leq 8$ mit einem Signifikanzniveau von $\alpha = 0.05$. Zählen Sie, wie oft H_0 abgelehnt wird.

b) Testen Sie die Hypothese $H_1 : \mu \geq 10$ mit einem Signifikanzniveau von $\alpha = 0.05$. Zählen Sie, wie oft H_1 abgelehnt wird.

Beispiel 5.6*

Das Ergebnis des 1000-maligen Würfelns eines Würfels sind folgende Häufigkeiten:

1: 170, 2: 152, 3: 164, 4: 179, 5: 158, 6: 177.

Testen Sie mit dem χ^2-Test, ob der Würfel symmetrisch ist ($\alpha = 0.01$).

Beispiel 5.7 (M)

Stellen Sie die Dichte- bzw. Verteilungsfunktionen der folgenden Verteilungen graphisch dar:

a) Exponentialverteilung mit $\tau = 2$

b) Normalverteilung mit $\mu = 0$ und $\sigma = 1, 2, 3$

c) Cauchyverteilung

Untersuchen Sie anhand von Stichproben der Größe $n = 100, 400, 1600$ die Konvergenz der empirischen Verteilungsfunktion $F_n(x)$ gegen die wahre Verteilungsfunktion $F(x)$. Überprüfen Sie die Übereinstimmung mit dem Kolmogorov-Smirnov-Test.

Beispiel 5.8 (M)

Vergleichen Sie die Gütefunktionen des Vorzeichentests und des Vorzeichenrangtests für Cauchy-verteilte Stichproben der Größe $n = 100$.

Teil 6

Regressionsanalyse

Beispiel 6.1*

Eine Versicherung interessiert sich für die Abhängigkeit der Schadenswahrscheinlichkeit von der Fahrleistung. Folgende Tabelle zeigt Daten einer Erhebung der jährlichen Fahrleistung x (in 1000 km) und der Schadenshäufigkeit y (in %):

x	5	10	15	20	25	30	40	50
y	9	10	14	18	22	24	29	29

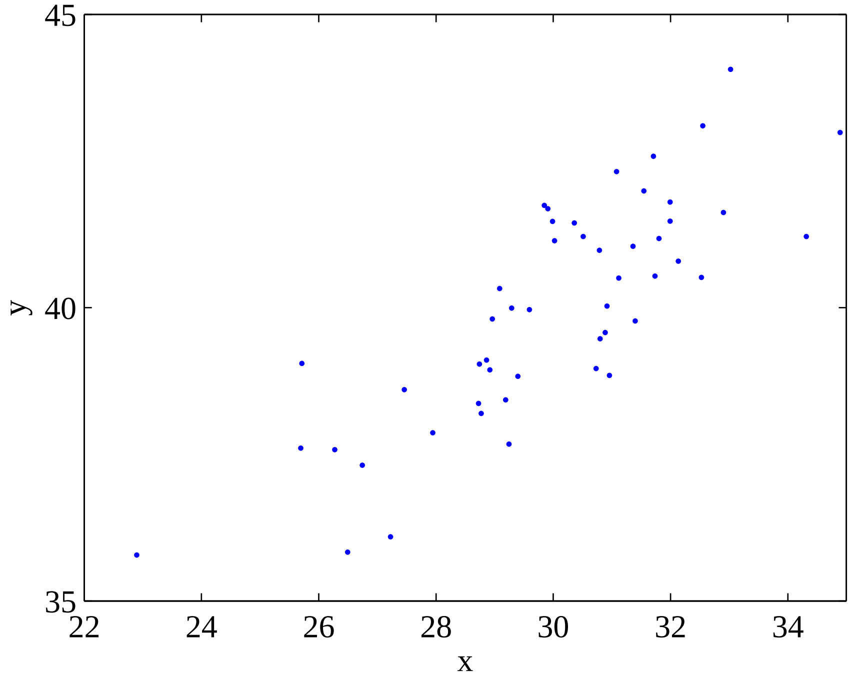
a) Stellen Sie die Punkte in einem Streudiagramm dar

b) Ermitteln Sie die Regressionsgerade

c) Welche Schadenshäufigkeit ist für Fahrer mit 35000 km jährlicher Fahrleistung zu erwarten?

Beispiel 6.2*

Die Abbildung zeigt einen Datensatz mit 50 Paaren (x, y).



Folgende Größen sind gegeben:

$$\sum x_i = 1499.02, \sum y_i = 1998.53,$$
$$\sum x_i^2 = 45203.15, \sum y_i^2 = 80058.24,$$
$$\sum x_i y_i = 60091.29$$

a) Schätzen Sie die Koeffizienten der Regressionsgeraden $y = \alpha + \beta x$.

b) Berechnen Sie das Bestimmtheitsmaß der Regression.

Beispiel 6.3 (M)

Die Lebensdauer einer Maschinenkomponente hängt von der Betriebsspannung x_1, der Motordrehzahl x_2 und der Betriebstemperatur x_3 ab. Ein Experiment ergibt die folgende Lebensdauer in Stunden:

y	x_1	x_2	x_3
2145	110	750	140
2155	110	850	180
2220	110	1000	140
2225	110	1100	180
2260	120	750	140
2266	120	850	180
2334	120	1000	140
2340	130	1000	180
2212	115	840	150
2180	115	880	150

a) Passen Sie ein lineares Modell an die Daten an.

b) Bestimmen Sie die Kovarianzmatrix der Regressionsparameter.

c) Welche Regressionsparameter sind signifikant von Null verschieden?

d) Bestimmen Sie ein 95%-iges Konfidenzintervall für die erwartete Lebensdauer, wenn $x_1 = 125$, $x_2 = 900$ und $x_3 = 160$.

(Aus: Sheldon M. Ross, Statistik für Ingenieure und Naturwissenschafter)