



# **Bern Statistik 2**

(Gatto)

## **Aufgaben**

Vertrauensintervalle

- Aufgabe 1: (Approx.) Konfidenzbereiche für einen Binomialparameter  $p$  und deren Aussage
- Aufgabe 2: Approximative Konfidenzintervalle für  $p$  (nach Wald und Wilson)
- Aufgabe 3: Konfidenzintervalle für einen Mittelwert  $\mu$  berechnen (einfachster Fall)
- Aufgabe 4: Empirische Standardabweichung  $s$  mittels Formeln berechnen können
- Aufgabe 5: Konfidenzintervalle für die Differenz zweier Mittelwerte berechnen
- Aufgabe 6: Exakte Konfidenzbereiche für  $p$  (graphisch bestimmen)
- Aufgabe 7: Exakte Konfidenzbereiche für  $p$  (mit R)
- Aufgabe 8: Vertrauensintervall für den Median berechnen
- Aufgabe 9: Wahrscheinlichkeit für Median berechnen
- Aufgabe 10: Ungleichungen für den Median berechnen

<b>III</b>	<b>Schließende Statistik</b>	<b>187</b>
<b>5</b>	<b>Konfidenzbereiche</b>	<b>189</b>
5.1	Allgemeine Beschreibung von Konfidenzbereichen . . . . .	189
5.2	Approximative Konfidenzintervalle nach Wald . . . . .	190
5.2.1	Das allgemeine Prinzip . . . . .	190
5.2.2	Konfidenzschranken für einen Binomialparameter . . . . .	192
5.2.3	Konfidenzschranken für einen Mittelwert . . . . .	195
5.2.4	Vergleich zweier Mittelwerte . . . . .	196
5.3	Exakte Konfidenzbereiche . . . . .	197
5.3.1	Konfidenzbereiche für einen Binomialparameter . . . . .	197
5.3.2	Student- und Chiquadrat-Verteilungen . . . . .	203
5.3.3	Konfidenzbereiche für Quantile . . . . .	208

**Aufgabe 1: (Konfidenzbereiche für einen Binomialparameter  $p$  und deren Aussage)**

Der Gemeinderat von Entenhausen diskutiert, ob der grosse Tresor des kürzlich verstorbenen Dagobert Duck abgerissen oder unter Denkmalschutz gestellt werden soll. Die Bürgerinitiative "Schluss mit Protzen" möchte den Gemeinderat davon überzeugen, dass ein grosser Teil aller Stimmberechtigten für einen Abriss ist. Hierzu plant sie eine Befragung von  $n = 500$  Stimmberechtigten.

a) Hier geht es um den unbekanntem relativen Anteil  $p$  aller Stimmberechtigten, die für einen Abriss sind. Wenn man in der Umfrage  $X$  Befürworter des Abrisses findet, könnte man ein Vertrauensintervall für diesen Parameter  $p$  bestimmen. Welche Art von Konfidenzbereich wäre aus Sicht der Bürgerinitiative angebracht, ein Intervall, eine untere Schranke oder eine obere Schranke für  $p$ ?

b) Angekommen, in der Umfrage fanden sich  $X = 231$  Befürworter des Abrisses.

Berechnen Sie nun den in Teil a) empfohlenen Konfidenzbereich mit Konfidenzniveau  $1 - \alpha = 95\%$ .

Das passende Quantil  $\Phi^{-1}(\gamma)$  finden Sie in der folgenden Tabelle:

$\gamma$	0.75	0.9	0.95	0.975	0.99	0.995
$\Phi^{-1}(\gamma)$	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

c) Formulieren Sie eine Schlussfolgerung in einem Satz.

## Aufgabe 2: (Approximative Konfidenzintervalle für $p$ nach Wald und Wilson)

Um zu klären, ob bei Neugeborenen die relativen Anteile von Mädchen und Knaben unterschiedlich sind, wurden die Daten von  $n = 429'440$  Neugeborenen ausgewertet. Darunter waren  $H = 221'023$  Knaben.

Berechnen Sie mit Wilsons und mit Walds Methode ein 99%-Vertrauensintervall für die Wahrscheinlichkeit  $p$ , dass ein Neugeborenes ein Knabe ist.

Wie beantworten Sie die Ausgangsfrage?

Sie können zur Berechnung der Vertrauensintervalle die folgende Tabelle zu Quantilen der Standardnormalverteilung verwenden.

$\beta$	0.9	0.95	0.99	0.995	0.999
$\Phi^{-1}(\beta)$	1.2816	1.6449	2.3264	2.5759	3.0903

**Aufgabe 3: (Konfidenzintervalle für einen Mittelwert  $\mu$  berechnen (einfachster Fall))**

Von einer Stichprobe seien die folgenden Grössen bekannt:

$$n = 200, \quad \hat{\mu} = 8, \quad \hat{\sigma} = 20.$$

a) Geben Sie ein 90%-Konfidenzintervall für  $\mu$  an. Verwenden Sie dazu folgende Tabelle von Student-Quantilen:

$t_{k;\gamma}$	$\gamma = 0.85$	0.90	0.95	0.99	0.995
$k = 195$	1.0392	1.2860	1.6528	2.3457	2.6013
$k = 199$	1.0392	1.2859	1.6526	2.3453	2.6008
$k = 200$	1.0392	1.2858	1.6526	2.3452	2.6007
$k = 205$	1.0391	1.2857	1.6524	2.3447	2.6001



**MathCourses**

[www.mathcourses.ch/bern\\_statistik.html](http://www.mathcourses.ch/bern_statistik.html)

**Aufgabe 4: (Empirische Standardabweichung  $s$  mittels Formeln berechnen)**

Angenommen, Sie wissen von einer Stichprobe  $X_1, \dots, X_n$  nur, dass  $n = 10$ ,  $\sum_{i=1}^n X_i = 50$  und  $\sum_{i=1}^n X_i^2 = 300$ . Bestimmen Sie damit  $\bar{X} = \hat{\mu}$ ,  $\hat{\sigma} = S_X^*$  und den Standardfehler  $\hat{\sigma}/\sqrt{n}$  von  $\hat{\mu}$ .

**Aufgabe 5: (Konfidenzintervalle für die Differenz zweier Mittelwerte berechnen)**

Aus zwei Populationen wurden jeweils Stichproben gezogen. Für ein numerisches Merkmal ergaben sich dann die Werte  $X_1, X_2, \dots, X_{n_X}$  beziehungsweise  $Y_1, Y_2, \dots, Y_{n_Y}$ .

a) Angenommen

$$n_X = 90, \quad \sum_{i=1}^{n_X} X_i = 9131.087, \quad \sum_{i=1}^{n_X} X_i^2 = 1013005,$$
$$n_Y = 50, \quad \sum_{i=1}^{n_Y} Y_i = 4580.218, \quad \sum_{i=1}^{n_Y} Y_i^2 = 461619.6.$$

Berechnen Sie hiermit den Punktschätzer  $\hat{\theta} = \hat{\mu}_X - \hat{\mu}_Y$  für den unbekannt Parameter  $\theta = \mu_X - \mu_Y$  sowie seinen Standardfehler  $\hat{\tau}$ .

b) Berechnen Sie ein approximatives 99%-Vertrauensintervall für  $\theta$ .

c) Können Sie aus den Daten ableiten, dass definitiv  $\theta \neq 0$ ? Wenn nein, können Sie dies mit einer Sicherheit von 95%?



**Aufgabe 6: (Exakte Konfidenzbereiche für  $p$  graphisch bestimmen)**

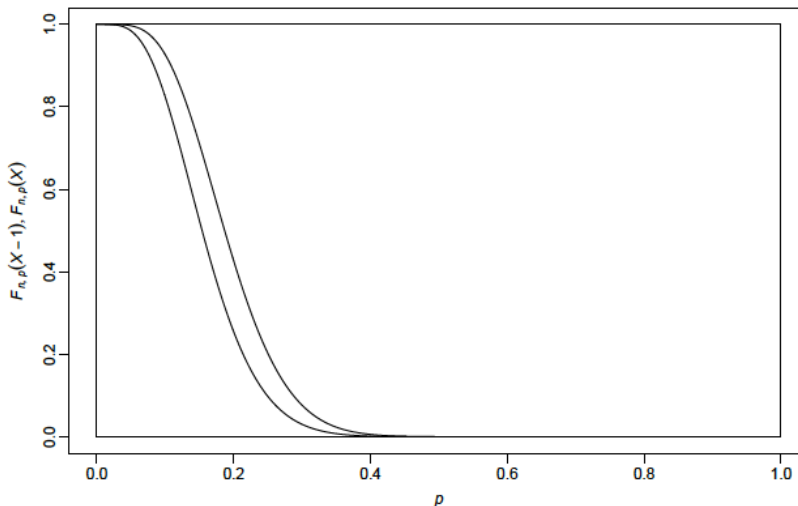
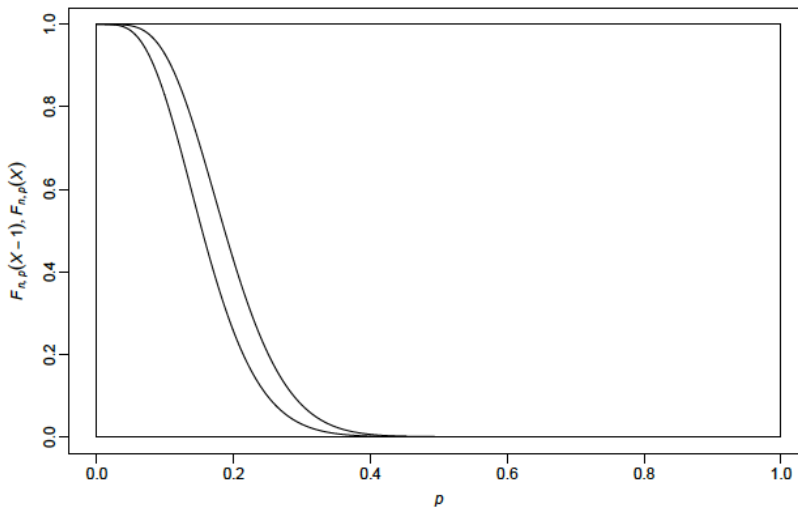
Bei einer kleinen Umfrage zur geplanten Umwandlung einer Ortsstraße in eine Tempo-30-Zone wurden 30 Einwohner des Dorfes gefragt, ob sie gegen diese Umwandlung seien. Fünf Befragte bejahten diese Frage.

- a) Geben Sie den Punktschätzer für den relativen Anteil  $p$  von Gegnern der Umwandlung an.
- b) Könnte man hier mit den approximativen Methoden (Wald oder Wilson) Konfidenzintervalle für  $p$  berechnen?
- c) Nachfolgend sehen Sie zweimal die Graphen der Funktionen

$$p \mapsto F_{30,p}(4), F_{30,p}(5).$$

Zeichnen Sie im ersten Bild die einseitigen 80%-Vertrauensschranken für  $p$  ein. Zeichnen Sie im zweiten Bild die Schranken des 80%-Vertrauensintervalls für  $p$  ein.

- d) Wie würden Sie einem Laien knapp erläutern, was das 80%-Konfidenzintervall zu bedeuten hat?



## Aufgabe 7: (Exakte Konfidenzbereiche für $p$ mit R)

Entscheiden Sie für die folgenden Teilaufgaben, wie Sie jeweils die Parameter  $n \in \mathbb{N}, p \in [0, 1]$  und die Zufallsvariable  $X \sim \text{Bin}(n, p)$  definieren würden, und welche Art von Konfidenzbereich (untere oder obere Schranke oder Intervall) für  $p$  interessant ist. Bestimmen Sie dann jeweils den gewünschten Konfidenzbereich mit Hilfe der drei R-Outputs und formulieren Sie eine Schlussfolgerung.

- a) Eine politische Partei möchte unterstreichen, dass ihre Initiative in der Bevölkerung breite Unterstützung genießt. Dazu hat sie eine Umfrage unter 420 Personen durchgeführt. Davon gaben 290 Personen an, die Initiative zu befürworten.

$$\text{binom.test}(290, 420, \text{alternative} = 'g')\$conf.int \rightsquigarrow [0.6352, 1.0000]$$

$$\text{binom.test}(290, 420, \text{alternative} = 'l')\$conf.int \rightsquigarrow [0.0000, 0.7118]$$

$$\text{binom.test}(290, 420)\$conf.int \rightsquigarrow [0.6278, 0.7186]$$

- b) Eine Zeitung macht eine Reportage zum Thema Flugangst und möchte wissen, wie verbreitet Flugangst in der Bevölkerung ist. Auf die Frage 'Haben Sie Flugangst?' haben 44 von 113 Personen mit 'Ja' geantwortet.

$$\text{binom.test}(44, 113, \text{alternative} = 'g')\$conf.int \rightsquigarrow [0.3124, 1.0000]$$

$$\text{binom.test}(44, 113, \text{alternative} = 'l')\$conf.int \rightsquigarrow [0.0000, 0.4708]$$

$$\text{binom.test}(44, 113)\$conf.int \rightsquigarrow [0.2990, 0.4857]$$

- c) Eine Firma möchte nachweisen, dass der relative Anteil fehlerhafter Produkte unter 1% liegt. In der letzten Qualitätskontrolle hatten 8 von 863 geprüften Produkten Mängel.

$$\text{binom.test}(8, 863, \text{alternative} = 'g')\$conf.int \rightsquigarrow [0.0046, 1.0000]$$

$$\text{binom.test}(8, 863, \text{alternative} = 'l')\$conf.int \rightsquigarrow [0.0000, 0.0167]$$

$$\text{binom.test}(8, 863)\$conf.int \rightsquigarrow [0.00401, 0.01819]$$

- d) Sie stellen fest, dass ein Freund bei einem Brettspiel ungewöhnlich oft eine Sechs würfelt. Daraufhin bitten Sie ihn, den Würfel 20-mal zu werfen. Von diesen 20 Würfeln waren 7 eine Sechs.

$$\text{binom.test}(7, 20, \text{alternative} = 'g')\$conf.int \rightsquigarrow [0.1773, 1.0000]$$

$$\text{binom.test}(7, 20, \text{alternative} = 'l')\$conf.int \rightsquigarrow [0.0000, 0.5581]$$

$$\text{binom.test}(7, 20)\$conf.int \rightsquigarrow [0.1539, 0.5922]$$

**Aufgabe 8: (Vertrauensintervall für den Median berechnen)**

Nachfolgend sehen Sie einen Datensatz zur Lebensdauer von  $n = 30$  Hauskatzen in Monaten:

145.4	71.7	147.9	104.9	118.1	105.6	139.3	119.0	138.9	147.7
145.1	80.2	125.6	188.2	169.6	114.2	198.0	96.4	112.6	94.9
167.2	216.6	162.8	165.9	225.7	162.0	89.7	107.0	172.9	66.6

- a) Geben Sie Punktschätzer für die drei Quartile der Lebensdauer einer Hauskatze an.
- b) Berechnen Sie das 90% Vertrauensintervall für den Median der Lebensdauer einer Hauskatze für die obigen Daten. Verwenden Sie dazu folgende Tabelle der Verteilungsfunktion  $B_{n,0.5}$  von  $\text{Bin}(n, 0.5)$ .

$x$	6	7	8	9	10	11	12	13	14
$B_{30,0.5}(x)$	0.001	0.003	0.008	0.021	0.049	0.100	0.181	0.292	0.428
$x$	15	16	17	18	19	20	21	22	23
$B_{30,0.5}(x)$	0.572	0.708	0.819	0.900	0.951	0.979	0.992	0.997	0.999

Formulieren Sie die Bedeutung dieses Intervalls.

**Aufgabe 9: (Wahrscheinlichkeit für Median  $q_{0.5}$  berechnen)**

Angenommen Sie haben einen Datensatz mit  $n = 15$  Beobachtungen  $X_1, \dots, X_n \in \mathbb{R}$ .

- a) Welche Ordnungsstatistik benötigen Sie, um den Boxplot zu zeichnen?  
b) Mit welcher Wahrscheinlichkeit liegt  $q_{0.5}$  im Bereich der Box?

Verwenden Sie dazu folgende Tabelle:

$x$	0	1	2	3	4	5	6	7
$F_{15,0.5}(x)$	0.000	0.000	0.004	0.018	0.059	0.151	0.304	0.500
$x$	8	9	10	11	12	13	14	15
$F_{15,0.5}(x)$	0.696	0.849	0.941	0.982	0.996	1.000	1.000	1.000

**Aufgabe 10: (Ungleichungen für den Median berechnen)**

Seien  $X_1, X_2, \dots, X_n$  stochastisch unabhängige Zufallsvariablen mit unbekanntem Median  $q_{0.5}$ , das heisst,  $P(X_i < q_{0.5}) \leq 0.5 \leq P(X_i \leq q_{0.5})$ .

a) Zeigen Sie, dass für die Ordnungsstatistiken  $X_{(1)}, X_{(2)}, \dots, X_{(n)}$  gilt:

$$P(X_{(n)} < q_{0.5}) \leq 0.5^n,$$

$$P(X_{(1)} > q_{0.5}) \leq 0.5^n,$$

$$P(X_{(1)} < q_{0.5} X_{(n)}) \leq 1 - 2^{1-n},$$

Hinweis: Aussagen der Form  $X_{(n)} < c$  oder  $X_{(1)} > c$  sind gleichbedeutend damit, dass alle Beobachtungen  $X_i$  kleiner bzw. grösser als  $c$  sind.

b) Interpretieren Sie die letzte Aussage mit dem Box-Plot der Daten  $X_1, X_2, \dots, X_n$ .

c) Wie gross muss  $n$  sein, damit die Schranke für das Intervall aus a) mindestens 0.99 beträgt?