

4. Übungsblatt „Einführung in die Statistik“

Abgabe bis Montag 2.5., 16 Uhr.

Hinweis zu R: Mit den Befehlen

```
binom.test(x = H, n = n),  
binom.test(x = H, n = n, conf.level = 1 -  $\alpha$ )
```

erhalten Sie das in der Vorlesung beschriebene 95%- bzw. $(1 - \alpha)$ -Konfidenzintervall $[a_{\alpha/2}(H), b_{\alpha/2}(H)]$ für den Parameter p im Modell $H \sim \text{Bin}(n, p)$. Mit

```
binom.test(..., alternative = 'less'),  
binom.test(..., alternative = 'greater')
```

erhalten Sie das Intervall $[0, b_{\alpha}(H)]$ bzw. $[a_{\alpha}(H), 1]$.

1. (Beispiele zu Konfidenzbereichen für einen Binomialparameter) Definieren Sie für die folgenden Anwendungssituationen jeweils einen geeigneten Wahrscheinlichkeitsparameter p und überlegen Sie, ob hierfür eine untere Konfidenzschranke, eine obere Konfidenzschranke oder ein Konfidenzintervall besonders geeignet wäre. Berechnen Sie dann diesen Konfidenzbereich mit $\alpha = 5\%$. Dabei können Sie entweder (i) exakte Schranken oder (ii) Wilson's Methode verwenden.

a) Wie verbreitet ist Flugangst? Anlässlich eines spektakulären “Fluchtversuches” eines Flugpassagiers kurz vor dem Start äußerten sich 240 Schweizerinnen und Schweizer zu der Frage, ob sie unter Flugangst leiden. Ergebnis: 41 Personen antworteten mit ‘ja’.

b) Ein Anbieter eines WLAN-Routers möchte untermauern, dass die meisten Kunden mit der neuen Installationssoftware gut zurechtkommen. Zu diesem Zweck recherchiert er über sein Callcenter, wie viele von 1400 Neukunden die Service-Hotline in Anspruch nahmen. Ergebnis: 11 Kunden ließen sich wegen Problemen bei der Installation beraten.

c) Möchte die Mehrheit der Wahlberechtigten gerne per Internet abstimmen? Man fragte 88 Personen, ob sie den Gang zur Urne, eine Briefwahl oder eine Online-Wahl bevorzugen würden. Ergebnis: 60 Personen bevorzugten die Online-Wahl.

2. (McNemar's Test) Sei $\mathbf{H} \sim \text{Mult}(n, \mathbf{p})$ mit unbekanntem Wahrscheinlichkeitsvektor $\mathbf{p} = (p_k)_{k=1}^K$. Die Frage ist nun, ob $p_1 \leq p_2$ (Nullhypothese) oder $p_1 > p_2$ (Alternativhypothese). Anstelle eines statistischen Tests konstruieren wir nun eine geeignete Konfidenzschranke für p_1/p_2 :

a) Zeigen Sie, dass H_1 bei gegebener Summe $H_1 + H_2$ binomialverteilt ist mit Parametern $H_1 + H_2$ und $\rho := p_1/(p_1 + p_2)$. Das heißt, für beliebige Zahlen $m \in \{0, 1, \dots, n\}$ und $x \in \{0, 1, \dots, m\}$ ist

$$\mathbb{P}(H_1 = x \mid H_1 + H_2 = m) = \binom{m}{x} \rho^x (1 - \rho)^{m-x}.$$

b) Beschreiben Sie nun, wie man mit Hilfe von Konfidenzschranken für einen Binomialparameter Konfidenzschranken für den Quotienten p_1/p_2 angeben kann.

c) Werten Sie das folgende Datenbeispiel aus: Für den Nachweis einer bestimmten Krankheit gibt es zwei konkurrierende medizinische Tests A und B . Die Arbeitshypothese lautet, dass Test A sensiver ist als Test B . Das heißt, bei einer erkrankten Person ist $\mathbb{P}(\text{Test } A \text{ positiv})$ größer als $\mathbb{P}(\text{Test } B \text{ positiv})$. Nun werden bei insgesamt $n = 80$ erkrankten Personen beide Tests angewendet. Bei 77 Personen war Test A positiv, bei 70 Personen war Test B positiv, bei 69 Personen waren beide Tests positiv. Belegen diese Daten, dass Test A sensiver ist als Test B ?

Hinweis: Bei jeder Person sind vier verschiedene Ausgänge denkbar. Benennen Sie diese vier Ausgänge und formulieren Sie die Arbeitshypothese mit Hilfe der entsprechenden Wahrscheinlichkeiten. Wenden Sie dann eine der einseitigen Konfidenzschranken aus Teil b) oder einen geeigneten Binomialtest an.

3. (Individuen mit Kennziffern) Wir ziehen rein zufällig und ohne Zurücklegen eine Stichprobe ω vom Umfang $n \geq 2$ aus der "Population" $\{a + 1, a + 2, \dots, b\}$. Also folgt ω der Gleichverteilung $\mathbb{P}_{a,b}$ auf der Menge

$$\Omega_{a,b} := \{\omega \in \{a + 1, a + 2, \dots, b\}^n : \omega_i \neq \omega_j \text{ falls } i \neq j\}.$$

Dabei sind a, b uns unbekannte ganze Zahlen mit $b - a \geq n$. Nun interessiert uns die Populationsgröße $N = b - a$. Hierfür betrachten wir die Kenngröße

$$X(\omega) := \max(\omega) - \min(\omega).$$

a) Zeigen Sie, dass X genauso verteilt ist wie $\sum_{i=1}^{n-1} Z_i$, wobei $\mathbf{Z} = (Z_i)_{i=1}^{n+1}$ ein Zufallsvektor ist, der gleichverteilt ist auf $\mathcal{Z}_N := \{\mathbf{z} \in \mathbb{N}^{n+1} : \sum_{i=1}^{n+1} z_i = N + 1\}$.

b) Bestimmen Sie mit Hilfe von Teil a) einen erwartungstreuen Schätzer $\hat{N}$ für N .

c) Aus a) ergibt sich, dass die Verteilungsfunktion von X nur von $N = b - a$ abhängt, nämlich $\mathbb{P}_{a,b}(X \leq x) = F_N(x)$ mit der Verteilungsfunktion F_N von $\sum_{i=1}^{n-1} Z_i$. Zeigen Sie

$$F_N(x) = 1 - F_{N,x,n}(n - 2),$$

wobei $F_{N,\ell,n}$ die Verteilungsfunktion von $\text{Hyp}(N, \ell, n)$ ist.

Vorschlag: Überzeugen Sie sich davon, dass mit den sortierten Elementen $\omega_{(1)} < \omega_{(2)} < \dots < \omega_{(n)}$ von ω gilt: $F_N(x) = \mathbb{P}_{0,N}(\omega_{(n-1)} \leq x)$. Betrachten Sie nun die Zufallsvariable $H(\omega) := \#\{i \leq n : \omega_i \leq x\}$ im Falle von $a = 0, b = N, \dots$

d) Begründen Sie, dass $F_N(x)$ für festes x monoton fallend in $N \geq n$ ist. Zeigen Sie auch, dass $\lim_{N \rightarrow \infty} F_N(x) = 0$. Beschreiben Sie dann, wie man eine untere bzw. obere $(1 - \alpha)$ -Konfidenzschranke für N definieren kann.