**Institut für Angewandte Mathematik**
**Sommersemester 2024**
Andreas Eberle, Stefan Oberdörster

# 2. Übungsblatt „Einführung in die Statistik"

**Abgabe bis Dienstag 23.4, 8.15 Uhr bei Ihrem Tutor.**

---

**1. (Ein sozialwissenschaftliches Experiment)** Im Rahmen einer Fortbildungsveranstaltung nahmen 50 angehende Manager an einem Experiment teil, ohne dies zu wissen. Jeder von ihnen erhielt eine (fiktive) Personalakte und sollte entscheiden, ob die betreffende Person befördert wird oder nicht. Die 50 Personalakten waren identisch bis auf den Namen der Person und wurden rein zufällig verteilt. In 25 Fällen handelte es sich um die Akte von Herrn Müller, und in 25 Fällen ging es um Frau Müller. Das Ergebnis des Experiments fassen wir in der folgenden Vierfeldertafel zusammen:

| | Beförderung | keine Bef. | |
|---|---|---|---|
| Herr Müller | 21 | 4 | 25 |
| Frau Müller | 13 | 12 | 25 |
| | 34 | 16 | 50 |

Bestätigen diese Daten das Vorurteil, dass Männer im Berufsleben gegenüber Frauen bevorzugt werden? Formulieren Sie eine passende Nullhypothese. Werten Sie die Daten dann zum Signifikanzniveau $\alpha = 5\%$ aus. Dabei können Sie folgende Tabelle der hypergeometrischen Verteilung $\mathrm{Hyp}(50, 34, 25)$ verwenden. Sie enthält deren Gewichte $f_{50,34,25}(x)$ auf vier Nachkommastellen gerundet.

| $x$ | $\leq 10$ | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|
| $f_{50,34,25}(x)$ | 0.0000 | 0.0003 | 0.0024 | 0.0134 | 0.0481 | 0.1176 | 0.1995 | 0.2376 |

| $x$ | 18 | 19 | 20 | 21 | 22 | 23 | $\geq 24$ |
|---|---|---|---|---|---|---|---|
| $f_{50,34,25}(x)$ | 0.1995 | 0.1176 | 0.0481 | 0.0134 | 0.0024 | 0.0003 | 0.0000 |

**2. (Eine Monotonieeigenschaft der hypergeometrischen Verteilung)** Bei der Capture-Recapture-Methode verwenden wir die Tatsache, dass die Verteilungsfunktion $F_{N,\ell,n}(x)$ von $\mathrm{Hyp}(N, \ell, n)$ an einer festen Stelle $x \in \mathbb{N}_0$ monoton wachsend ist in $N$. Beweisen Sie diese Tatsache.

*Hinweis: Man kann die Aussage durch wilde Rechnungen beweisen. Eleganter ist aber ein Kopplungsargument: Beschreiben Sie ein Zufallsexperiment mit zwei Zufallsvariablen $X$ und $\tilde{X}$ derart, dass $X \sim \mathrm{Hyp}(N, \ell, n)$, $\tilde{X} \sim \mathrm{Hyp}(N+1, \ell, n)$ und stets $\tilde{X} \leq X$. Denken Sie beispielsweise an eine Urne mit $\ell$ blauen, $N - \ell$ weißen und einer schwarzen Kugel, aus der Sie nacheinander und ohne Zurücklegen $n+1$ Kugeln ziehen ...*

**3. (Individuen mit Kennziffern)** In der Vorlesung betrachteten wir Stichproben $\omega$ vom Umfang $n$ aus $\{1, 2, \ldots, N\}$ und die Kenngrösse $M(\omega) = \max\{\omega_1, \ldots, \omega_n\}$. Es wurde gezeigt, dass für beliebige Zahlen $x \in \mathbb{N}_0$ gilt:

$$\mathbb{P}_N[M \le x] \;=\; F_N(x) := \min\left(\frac{[x]_n}{[N]_n}, 1\right).$$

Hieraus ergaben sich $(1-\alpha)$-Konfidenzschranken

$$\begin{aligned} a_\alpha(M) &:= \min\{N \ge x : F_N(M-1) < 1-\alpha\}, \\ b_\alpha(M) &:= \max\{N \ge x : F_N(M) > \alpha\} \end{aligned}$$

für die unbekannte tatsächliche Populationsgrösse $N$.

a) Wie verändern sich der Wertebereich von $M$ sowie $F_N(x)$, wenn man die Stichprobe $\omega$ mit Zurücklegen aus $\{1, 2, \ldots, N\}$ zieht? Welche Parameter $N$ sind nun möglich?

b) Geben Sie in diesem neuen Szenario explizite Formeln für die analogen Konfidenzschranken $a_\alpha^{\mathrm{neu}}(M)$ und $b_\alpha^{\mathrm{neu}}(M)$ an.

c) Wie verhält sich $F_N^{\mathrm{neu}}(x)$ zu $F_N(x)$? Welche Ungleichungen ergeben sich daraus für $a_\alpha(M)$ und $b_\alpha(M)$ in der herkömmlichen Situation?

**4. (Populationsgröße und erwartungstreue Schätzung)** Ein Capture-Recapture-Experiment liefere $H \sim \mathrm{Hyp}(N, \ell, n)$ mit bekannten Parametern $\ell, n \ge 1$ und unbekanntem Parameter $N \ge \max(\ell, n)$.

a) Begründen Sie, dass kein Schätzer der Form $\hat{N} = g(H)$ mit einer reellwertigen Funktion $g$ auf $\{0, 1, \ldots, \min(\ell, n)\}$ den Parameter $N$ erwartungstreu schätzen kann.

b) Bestimmen Sie den systematischen Fehler des Schätzers $\hat{N} := (\ell+1)(n+1)/(H+1)$, und zeigen Sie, dass stets $\mathrm{Bias}_N(\hat{N}) \le 1$.

**5. (Erwartungswert des Maximums)** Wie in der Vorlesung sei $\mathbb{P}_N$ die Gleichverteilung auf der Menge $\Omega_N$ aller Tupel $\omega = (\omega_1, \ldots, \omega_n)$ von $n$ verschiedenen Zahlen aus $\{1, 2, \ldots, N\}$, und $M(\omega) := \max(\omega_1, \ldots, \omega_n)$. Beweisen Sie die Formel

$$\mathbb{E}_N[M] = n(N+1)/(n+1)$$

noch einmal mithilfe der folgenden Symmetrieüberlegung:
Für $\omega \in \Omega_N$ seien $1 \le \omega_{(1)} < \omega_{(2)} < \cdots < \omega_{(n)} \le N$ die der Größe nach sortierten Einträge; insbesondere ist $M(\omega) = \omega_{(n)}$. Mit $\omega_{(0)} := 0$ und $\omega_{(n+1)} := N+1$ definieren wir den Zufallsvektor $\mathbf{Z} = (Z_i)_{i=1}^{n+1}$ mit Einträgen $Z_i(\omega) := \omega_{(i)} - \omega_{(i-1)}$:

$$0,\ \underbrace{1, \ldots, \omega_{(1)}}_{Z_1(\omega)},\ \underbrace{\omega_{(1)}+1, \ldots, \omega_{(2)}}_{Z_2(\omega)},\ \ldots,\ \underbrace{\omega_{(n-1)}+1, \ldots, \omega_{(n)}}_{Z_n(\omega)},\ \underbrace{\omega_{(n)}+1, \ldots, N+1}_{Z_{n+1}(\omega)}.$$

Zeigen Sie, dass $\mathbf{Z}$ gleichverteilt ist auf der Menge $\mathcal{Z}_N := \left\{\mathbf{z} \in \mathbb{N}^{n+1} : \sum_{i=1}^{n+1} z_i = N+1\right\}$. Leiten Sie aus dieser Tatsache und Symmetrieeigenschaften von $\mathcal{Z}_N$ ab, dass die Zufallsvariablen $Z_1, Z_2, \ldots, Z_{n+1}$ identisch verteilt sind, und bestimmen Sie $\mathbb{E}_N[Z_1]$ ...