Formelsammlung zur Stichprobentheorie (Jens Mehrhoff)

1. uneingeschränkte Zufallsauswahl

- ***Umfang***
  Grundgesamtheit: N
  Stichprobe: n
- ***Mittelwert***
  Grundgesamtheit: $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$
  Stichprobe: $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$
  Erwartungswert: $E\left(\overline{X}\right) = \mu$
- ***Varianz***
  Grundgesamtheit: $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu\right)^2$
  Stichprobe: $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2$
  Erwartungswert: $E\left(S^2\right) = \sigma^2$
- ***Stichprobenverteilung***
  Auswahlwahrscheinlichkeit (ZmZ): $\frac{n}{N}$
  Anzahl möglicher Stichproben (ZoZ): $\binom{N}{n} = \frac{N!}{\left(N-n\right)!n!}$
  Varianz von $\overline{X}$ (ZoZ): $\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}\frac{N-n}{N-1}$
  approximative Stichprobenverteilung: $\overline{X} \sim N\left(\mu, \sigma_{\overline{X}}^2\right)$
  notwendiger Stichprobenumfang (ZmZ): $n \geq \frac{z^2\sigma^2}{e^2}$
  Stichprobenfehler: $e = z\sigma_{\overline{X}}$
  Standardnormalverteilung: $F_N\left(z\right) = 1 - \frac{\alpha}{2}$
  Hochrechnung: $\sum_{i=1}^{N} x_i = \frac{N}{n}\sum_{j=1}^{n} x_j = N\overline{x}$

2. geschichtete Stichprobe

- ***Stichprobenplan***
  Totalerhebung auf der 1. Stufe (Zerlegung der Grundgesamtheit in K Schichten), Teilerhebung auf der 2. Stufe (Zufallsauswahl aus den Schichten)
- ***Effizienzbedingung***
  homogen innerhalb der Schichten (geringe interne Varianz), heterogen zwischen den Schichten (große externe Varianz)
- ***Umfang***
  Grundgesamtheit: N_k, $\sum_{k=1}^{K} N_k = N$

  Stichprobe: n_k, $\sum_{k=1}^{K} n_k = n$

  Auswahlsatz: $\frac{n_k}{N_K}$
- ***Mittelwert***
  Grundgesamtheit: $\mu = \sum_{k=1}^{K} \frac{N_k}{N} \mu_k$

  Stichprobe: $\overline{X} = \sum_{k=1}^{K} \frac{N_k}{N} \overline{X_k}$
- ***Varianz***
  externe Varianz: $\sigma_{ext}^2 = \sum_{k=1}^{K} \frac{N_k}{N} \left(\mu_k - \mu\right)^2$

  interne Varianz: $\sigma_{int}^2 = \sum_{k=1}^{K} \frac{N_k}{N} \sigma_k^2$

  Gesamtvarianz: $\sigma^2 = \sigma_{ext}^2 + \sigma_{int}^2$
- ***Aufteilung***
  proportional: $\frac{n_k}{n} = \frac{N_k}{N} = \frac{N_k}{\sum_{k=1}^{K} N_k}$

  optimal (ZmZ): $\frac{n_k}{n} = \frac{N_k \sigma_k}{\sum_{k=1}^{K} N_k \sigma_k}$
- ***Stichprobenverteilung***
  Anzahl möglicher Stichproben (ZoZ): $\prod_{k=1}^{K} \binom{N_k}{n_k}$

  Varianz von $\overline{X}$: $\sigma_{\overline{X}}^2 = \sum_{k=1}^{K} \left(\frac{N_k}{N}\right)^2 \sigma_{\overline{X_k}}^2$

  Schichtungsgewinn bei proportionaler Aufteilung (ZmZ): $\frac{1}{n} \sigma_{ext}^2$

  Schichtungsgewinn bei optimaler Aufteilung (ZmZ): $\frac{1}{n}\left(\sigma^2 - \overline{\sigma}^2\right)$

  notwendiger Stichprobenumfang bei proportionaler Aufteilung (ZmZ): $n \geq \frac{z^2 \sigma_{int}^2}{e^2}$

3. Klumpenstichprobe

- ***Stichprobenplan***
  Teilerhebung auf der 1. Stufe (Auswahl von m aus M Klumpen), Totalerhebung auf der 2. Stufe (innerhalb eines Klumpens alle Merkmalträger)
- ***Effizienzbedingung***
  heterogen innerhalb der Klumpen (große Varianz „within"), homogen zwischen den Klumpen (geringe Varianz „between")
- ***Umfang***

  Grundgesamtheit: M Klumpen, N_i Elemente, $\sum_{i=1}^{M} N_i = N$

  Spezialfall: $N_i = \overline{N} \; \forall \, i$, $M\overline{N} = N$

  Stichprobe: $m \le M$ Klumpen, $N_j = N_i$ Elemente, $\sum_{j=1}^{m} N_j = n$

  Spezialfall: $N_j = \overline{N} \; \forall \, j$, $m\overline{N} = n$
- ***Mittelwert***

  Grundgesamtheit: $\mu = \sum_{i=1}^{M} \frac{N_i}{N} \mu_i$

  Spezialfall: $\mu = \frac{\sum_{i=1}^{M} \mu_i}{M}$

  Stichprobe: $\overline{x} = \sum_{j=1}^{m} \frac{N_j}{n} \overline{x_j} = \sum_{j=1}^{m} \frac{N_j}{n} \mu_j$, $\overline{x_j} = \mu_j$

  Spezialfall: $\overline{x} = \frac{\sum_{j=1}^{m} \mu_j}{m}$
- ***Varianz***

  Varianz between: $\sigma_b^2 = \frac{1}{M} \sum_{i=1}^{M} (\mu_i - \mu)^2 = \frac{1}{M\overline{N}^2} \sum_{i=1}^{M} \left(\sum_{k=1}^{\overline{N}} x_{ik} - \mu \right)^2$

  Intraklass-Kovarianz: $\sigma_{kl} = \frac{1}{M\overline{N}\left(\overline{N}-1\right)} \sum_{i=1}^{M} \sum_{k} \sum_{l} (x_{ik} - \mu)(x_{il} - \mu)$, $k \ne l$

  Intraklass-Korrelation: $\rho = \frac{\sigma_{kl}}{\sigma^2}$
- ***Totalwert***

  Grundgesamtheit: $X_{(M)} = \sum_{i=1}^{M} \sum_{k=1}^{N_i} x_{ik} = N\mu$

  Stichprobe: $X_{(m)} = \sum_{j=1}^{m} \sum_{l=1}^{N_j} x_{jl} = n\overline{x}$

- ***Stichprobenverteilung***

  Varianz von $\overline{X}$ (ZoZ): $\sigma_{\overline{X}}^2 = \frac{\sigma_b^2}{m}\frac{M-m}{M-1}$

  approximativer Varianzaufblähungsfaktor: $\sigma_{BL}^{2\,*} = 1 + \left(\overline{N} - 1\right)\rho$

  Varianzaufblähungsfaktor (ZoZ): $\sigma_{BL}^2 = \frac{\sigma_{\overline{X}}^{2\,(K)}}{\sigma_{\overline{X}}^{2\,(Z)}} = \frac{M - \frac{1}{\overline{N}}}{M-1}\sigma_{BL}^{2\,*}$

  Totalwert: $\hat{X}_{(M)} = \frac{N}{n} X_{(m)} = N\overline{x}$

  Spezialfall: $\hat{X}_{(M)} = \frac{M}{m} X_{(m)}$

Formelsammlung zur Wahrscheinlichkeitstheorie

- ***momenterzeugende Funktion***

  Definition: $M_X(t) = E\left(e^{tX}\right) = \begin{cases} \sum_X e^{tX} f(X) & X \ diskret \\ \int e^{tX} f(X)\,dX & X \ stetig \end{cases}$

  e-Funktion: $e^X = \sum_{i=0}^{\infty} \frac{X^i}{i!}$

  k-tes Anfangsmoment: $E\left(X^k\right) = M_X^{(k)}(0)$

  k-tes Zentralmoment: $E(X-\mu)^k = \sum_{i=0}^{k} (-1)^i \binom{k}{k-i} E\left(X^{k-i}\right) E(X)^i$

- ***charakteristische Funktion***

  Definition: $\Psi_X(t) = E\left(e^{itX}\right) = \begin{cases} \sum_X e^{itX} f(X) & X \ diskret \\ \int e^{itX} f(X)\,dX & X \ stetig \end{cases}$

  Zusammenhang: $\Psi_X^{(k)}(0) = i^k M_X^{(k)}(0) = i^k E\left(X^k\right)$

- ***wahrscheinlichkeitserzeugende Funktion***

  Definition: $W_X(t) = E\left(t^X\right) = \sum_X t^X f(X)$ (X *diskret*)

  k-tes faktorielles Moment: $E\left\{\prod_{i=1}^{k}\left[X - (i-1)\right]\right\} = W_X^{(k)}(1)$

- ***Reproduktivität von Verteilungen und Faltung von Zufallsvariablen***

  Reproduktivität: $f(X) = \prod_{i=1}^{n} f_i(X_i)$, $X = \sum_{i=1}^{n} X_i$

  Faltung: $M_X(t) = \prod_{i=1}^{n} M_{X_i}(t)$, $\Psi_X(t) = \prod_{i=1}^{n} \Psi_{X_i}(t)$, $W_X(t) = \prod_{i=1}^{n} W_{X_i}(t)$