

Formeloversigt til stikprøvetæori

	Alternativ variation	Kontinuert variation
Simpel tilfældig udvælgelse ^{note 1}	Punkt 1	Punkt 2
Generel stratificeret udvælgelse	Punkt 3	Punkt 4
Proportional stratificeret udvælgelse	Punkt 5	Punkt 6
Optimal stratificeret udvælgelse	Punkt 7	Punkt 8

1. Simpel tilfældig udvælgelse med alternativ variation

- Estimat af *populationsandelen* ^{note 2} $\bar{\mu} = \theta : \hat{\theta} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \frac{\text{antal "mærkede"}}{\text{stikprøvestørrelse}}$

- Estimat af *populationstotalen* $N \cdot \theta : N \cdot \hat{\theta}$

- Estimat af *populationsvariansen* $\tau^2 : s^2 = \frac{n}{n-1} \cdot \hat{\theta} \cdot (1 - \hat{\theta})$

- Estimat af *variansen af estimatet på populationsandelen*

$$\text{vâr}(\hat{\theta}) = \frac{s^2}{n} \cdot (1 - f) = \frac{\hat{\theta} \cdot (1 - \hat{\theta})}{n - 1} \cdot (1 - f) \quad , f \equiv \frac{n}{N} \text{ note 3}$$

- (Approksimativt) *konfidensinterval*: $\hat{\theta} \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\hat{\theta})}$

- Hvis der ønskes et *konfidensinterval*, hvor $\text{var}(\hat{\theta}) \leq \sigma_0^2$:

$$n \geq \frac{\theta(1-\theta)}{\sigma_0^2 + \frac{\theta(1-\theta)}{N}} \quad (\text{Worst case } \theta = 1/2: \quad n \geq \frac{\frac{1}{4}}{\sigma_0^2 + \frac{1}{4N}})$$

- Hvis der ønskes et *konfidensinterval*, hvor bredden af dette $\leq L_0$:

$$n \geq \frac{\theta(1-\theta)}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{\theta(1-\theta)}{N}} \quad (\text{Worst case } \theta = 1/2: \quad n \geq \frac{\frac{1}{4}}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{1}{4N}})$$

2. Simpel tilfældig udvælgelse med kontinuert variation

- Estimat af *populationsgennemsnittet* $\bar{\mu} : \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$

¹ Der antages uden tilbagelægning, ergo befinder vi os i en hypergeometrisk fordeling.

² Andelen af enheder i populationen, der besidder en given egenskab.

³ *f* kaldes udvalgsbrøken, da den repræsenterer den brøkdæl af den samlede population, der er udvalgt til stikprøven.

- Estimat af *populationstotalen* $N \cdot \bar{\mu} : N \cdot \bar{x}$
- Estimat af populationsvariansen $\tau^2 : s^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2$
- Estimat af variansen af estimatet på populationsgennemsnittet
 $\text{vâr}(\bar{x}) = \frac{s^2}{n} \cdot (1-f) \quad , f \equiv \frac{n}{N}$
- (Approksimativt) konfidensinterval: $\bar{x} \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\bar{x})}$

- Hvis der ønskes et konfidensinterval, hvor $\text{var}(\bar{x}) \leq \sigma_0^2 :$

$$n \geq \frac{\tau^2}{\sigma_0^2 + \frac{\tau^2}{N}} \quad (\text{Worst case: Sæt } \tau^2 \text{ højt})$$

- Hvis der ønskes et konfidensinterval, hvor bredden af dette $\leq L_0 :$

$$n \geq \frac{\tau^2}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{\tau^2}{N}} \quad (\text{Worst case: Sæt } \tau^2 \text{ højt})$$

3. Generel stratificeret udvælgelse med alternativ variation

Vi ser på m strata hver indeholdende en population på N_j , der summer til N . Fra stratum j udtages en stikprøve på n_j enheder. For at anvende stratificeret udvælgelse skal samtlige N_j og N kendes.

- Estimat af *stratumandelen* $\bar{\mu}_j = \theta_j : \hat{\theta}_j = \frac{1}{n_j} \cdot \sum_{i=1}^{n_j} x_{ij} = \frac{\text{antal "mærkede" i stikprøve } j}{\text{stratumstikprøvestørrelse}}$
- Estimat af *stratumtotalen* $N_j \cdot \theta_j : N_j \cdot \hat{\theta}_j$
- Estimat af *populationsandelen* $\bar{\mu} = \theta_s : \hat{\theta}_s = \sum_{j=1}^m W_j \cdot \hat{\theta}_j \quad , W_j \equiv \frac{N_j}{N}$ note 4
- Estimat af *stratumvariansen* $\tau_j^2 : s_j^2 = \frac{n_j}{n_j - 1} \cdot \hat{\theta}_j \cdot (1 - \hat{\theta}_j)$

⁴ W_j kaldes for stratumvægtene, altså den vægt hvormed det enkelte stratum indgår i populationen.

- Estimat af variansen af estimatet på stratumandelen

$$\text{vâr}(\hat{\theta}_j) = \frac{s_j^2}{n_j} \cdot (1 - f_j) = \frac{\hat{\theta}_j \cdot (1 - \hat{\theta}_j)}{n_j - 1} \cdot (1 - f_j) \quad , f_j \equiv \frac{n_j}{N_j}$$

- (Approksimativt) konfidensinterval for stratumandelen: $\hat{\theta}_j \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\hat{\theta}_j)}$

- Estimat af variansen af estimatet på populationsandelen

$$\text{vâr}(\hat{\theta}_s) = \sum_{j=1}^m W_j^2 \cdot \text{vâr}(\hat{\theta}_j) = \sum_{j=1}^m W_j^2 \cdot \frac{s_j^2}{n_j} \cdot (1 - f_j) = \sum_{j=1}^m W_j^2 \cdot \frac{\hat{\theta}_j \cdot (1 - \hat{\theta}_j)}{n_j - 1} \cdot (1 - f_j)$$

- (Approksimativt) konfidensinterval for populationsandelen: $\hat{\theta}_s \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\hat{\theta}_s)}$

4. Generel stratificeret udvælgelse med kontinuert variation

Vi ser på m strata hver indeholdende en population på N_j , der summer til N . Fra stratum j udtages en stikprøve på n_j enheder. For at anvende stratificeret udvælgelse skal samtlige N_j og N kendes.

- Estimat af *stratumgennemsnittet* $\bar{\mu}_j$: $\bar{x}_j = \frac{1}{n_j} \cdot \sum_{i=1}^{n_j} x_{ij}$

- Estimat af *stratumtotalen* $N_j \cdot \bar{\mu}_j$: $N_j \cdot \bar{x}_j$

- Estimat af *populationsgennemsnittet* $\bar{\mu}$: $\bar{x}_s = \sum_{j=1}^m W_j \cdot \bar{x}_j$, $W_j \equiv \frac{N_j}{N}$

- Estimat af *stratumvariansen* τ_j^2 : $s_j^2 = \frac{1}{n_j - 1} \cdot \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$

- Estimat af variansen af estimatet på stratumgennemsnittet

$$\text{vâr}(\bar{x}_j) = \frac{s_j^2}{n_j} \cdot (1 - f_j) \quad , f_j \equiv \frac{n_j}{N_j}$$

- (Approksimativt) konfidensinterval for stratumgennemsnittet: $\bar{x}_j \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\bar{x}_j)}$:

- Estimat af variansen af estimatet på populationsandelen

$$\text{vâr}(\bar{x}_s) = \sum_{j=1}^m W_j^2 \cdot \text{vâr}(\bar{x}_j) = \sum_{j=1}^m W_j^2 \cdot \frac{s_j^2}{n_j} \cdot (1 - f_j)$$

- (Approksimativt) konfidensinterval for populationsgennemsnittet:

$$\bar{x}_s \pm u_{1-\alpha/2} \cdot \sqrt{\hat{\text{vâr}}(\bar{x}_s)}$$

5. Proportional stratificeret udvælgelse med alternativ variation

- For at finde estimater for stratumandel og populationsandel følges blot formlerne fra punkt 3.
- Der hvor den proportionale allokering skal i brug, det er ved bestemmelse af stikprøvestørrelse. Den proportionale allokering udnytter variationen mellem strata til at mindske variansen på estimatoren.
- Under proportional allokering er udvalgsbrøkerne f_j ens for alle strata:

$f_j = \frac{n_j}{N_j} = \frac{n}{N}$, og dermed kan fodtegn på n og f i formlen for variansen på estimatoren fjernes.

- Estimat af variansen på populationsandelen bliver ved proportional allokering givet ved:

$$\hat{\text{vâr}}(\hat{\theta}_p) = \frac{1-f}{n} \cdot \sum_{j=1}^m W_j \cdot s_j^2$$

- (Approksimativt) konfidensinterval for populationsandelen under proportional allokering: $\hat{\theta}_p \pm u_{1-\alpha/2} \cdot \sqrt{\hat{\text{vâr}}(\hat{\theta}_p)}$
- Den mindst mulige stikprøvestørrelse kan, hvis kriteriet er, at $\text{var}(\hat{\theta}) \leq \sigma_0^2$ findes ved

$$n \geq \frac{\sum_{j=1}^m W_j \cdot \theta_j (1-\theta_j)}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \theta_j (1-\theta_j)} \quad (\text{Worst case } \theta_j = 1/2 \forall j \text{ giver } n \text{ som i punkt 1})^{\text{note 5}}$$

⁵ I formlen er der anvendt følgende approksimation: $\tau_j^2 = \frac{N_j}{N_j - 1} \cdot \theta \cdot (1-\theta) \approx \theta \cdot (1-\theta)$

- Hvis der ønskes et konfidensinterval, hvor bredden af dette $\leq L_0$ ^{note 5}:

$$n \geq \frac{\sum_{j=1}^m W_j \cdot \theta_j (1 - \theta_j)}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}} \right)^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \theta_j (1 - \theta_j)} \quad (\text{Worst case } \theta_j = 1/2 \forall j \text{ giver } n \text{ som i punkt 1})$$

- Når n er fundet kan stikprøvestørrelserne i den enkelte strata nemt findes ved:

$$n_j = n \cdot \frac{N_j}{N} = n \cdot W_j$$

6. Proportional stratificeret udvælgelse med kontinuert variation

- For at finde estimater for stratumgennemsnit og populationsgennemsnit følges blot formlerne fra punkt 4.
- Der hvor den proportionale allokering skal i brug, det er ved bestemmelse af stikprøvestørrelse. Den proportionale allokering udnytter variationen mellem strata til at mindske variansen på estimatoren.
- Under proportional allokering er udvalgsbrøkerne f_j ens for alle strata:

$f_j = \frac{n_j}{N_j} = \frac{n}{N}$, og dermed kan fodtegn på n og f i formlen for variansen på estimatoren fjernes.

- Estimat af variansen på populationsgennemsnittet bliver ved proportional allokering givet ved:

$$\hat{\text{var}}(\bar{x}_p) = \frac{1-f}{n} \cdot \sum_{j=1}^m W_j \cdot s_j^2$$

- (Approksimativt) konfidensinterval for populationsgennemsnittet under proportional allokering:

$$\bar{x}_p \pm u_{1-\alpha/2} \cdot \sqrt{\hat{\text{var}}(\bar{x}_p)}$$

- Den mindst mulige stikprøvestørrelse kan, hvis kriteriet er, at $\text{var}(\bar{x}) \leq \sigma_0^2$ findes ved

$$n \geq \frac{\sum_{j=1}^m W_j \cdot \tau_j^2}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \tau_j^2} \quad (\text{Worst case: Høj } \tau_j^2)$$

- Hvis der ønskes et konfidensinterval, hvor bredden af dette $\leq L_0$:

$$n \geq \frac{\sum_{j=1}^m W_j \cdot \tau_j^2}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \tau_j^2} \quad (\text{Worst case: Høj } \tau_j^2)$$

- Når n er fundet kan stikprøvestørrelserne i den enkelte strata nemt findes ved:

$$n_j = n \cdot \frac{N_j}{N} = n \cdot W_j$$

7. Optimal stratificeret udvælgelse med alternativ variation

- For at finde estimater for stratumandel, populationsandel og varianser på disse følges blot formlerne fra punkt 3.
- Der hvor den optimale allokering skal i brug, det er ved bestemmelse af stikprøvestørrelse. Den optimale allokering udnytter for det første ligesom den proportionale allokering variationen mellem strata, men udnytter også forskelle i varianser indefor de enkelte strata til at mindske variansen på estimatoren.
- Estimat af variansen på populationsandelen bliver ved optimal allokering givet ved:

$$\text{vâr}(\hat{\theta}_o) = \frac{1}{n} \cdot \left(\sum_{j=1}^m W_j \cdot s_j \right)^2 - \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot s_j^2$$

- (Approximativt) konfidensinterval for populationsandelen under proportional allokering: $\hat{\theta}_o \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\hat{\theta}_o)}$

- Den mindst mulige stikprøvestørrelse kan, hvis kriteriet er, at $\text{var}(\hat{\theta}) \leq \sigma_0^2$ findes ved ^{note 6}:

$$n \geq \frac{\left(\sum_{j=1}^m W_j \cdot \tau_j^2 \right)^2}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \tau_j^2} = \frac{\left(\sum_{j=1}^m W_j \cdot \sqrt{\theta_j(1-\theta_j)} \right)^2}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \theta_j(1-\theta_j)} \quad (\text{Worst case } \theta_j = 1/2 \forall j)$$

- Hvis der ønskes et konfidensinterval, hvor bredden af dette $\leq L_0$ ^{note 6}:

$$n \geq \frac{\sum_{j=1}^m W_j \cdot \theta_j (1 - \theta_j)}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \theta_j (1 - \theta_j)} \quad (\text{Worst case } \theta_j = 1/2 \forall j \text{ giver } n \text{ som i punkt 1})$$

- Den optimale stikprøvestørrelse for stratum j kan findes ved ^{note 6}

$$n_j \geq \frac{W_j \cdot \theta_j \cdot (1 - \theta_j)}{\sum_{k=1}^m W_k \cdot \theta_k (1 - \theta_k)} \cdot n \quad (\text{Worst case } \theta_j = 1/2)$$

8. Optimal stratificeret udvælgelse med kontinuert variation

- For at finde estimater for stratumgennemsnit og populationsgennemsnit følges blot formlerne fra punkt 4.
- Der hvor den optimale allokering skal i brug, det er ved bestemmelse af stikprøvestørrelse. Den optimale allokering udnytter for det første ligesom den proportionale allokering variationen mellem strata, men udnytter også forskelle i varianser indefor de enkelte strata til at mindske variansen på estimatoren.
- Estimat af variansen på populationsgennemsnittet bliver ved optimal allokering givet ved:

$$\text{vâr}(\bar{x}_o) = \frac{1}{n} \cdot \left(\sum_{j=1}^m W_j \cdot s_j \right)^2 - \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot s_j^2 = \sum_{j=1}^m W_j^2 \cdot \frac{s_j^2}{n_j} \cdot (1 - f_j)$$

- (Approximativt) konfidensinterval for populationsandelen under proportional allokering: $\bar{x}_o \pm u_{1-\alpha/2} \cdot \sqrt{\text{vâr}(\bar{x}_o)}$
- Den mindst mulige stikprøvestørrelse kan, hvis kriteriet er, at $\text{var}(\bar{x}) \leq \sigma_0^2$ findes ved:

⁶ I formelen er der anvendt følgende approximation: $\tau_j^2 = \frac{N_j}{N_j - 1} \cdot \theta \cdot (1 - \theta) \approx \theta \cdot (1 - \theta)$

$$n \geq \frac{\left(\sum_{j=1}^m W_j \cdot \tau_j^2\right)^2}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \tau_j^2} \approx \frac{\left(\sum_{j=1}^m W_j \cdot s_j^2\right)^2}{\sigma_0^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot s_j^2} \quad (\text{Worst case: Høj } \tau_j^2)$$

- Hvis der ønskes et konfidensinterval, hvor bredden af dette $\leq L_0$ ^{note 6}:

$$n \geq \frac{\left(\sum_{j=1}^m W_j \cdot \tau_j^2\right)^2}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot \tau_j^2} \approx \frac{\left(\sum_{j=1}^m W_j \cdot s_j^2\right)^2}{\left(\frac{L_0}{2 \cdot u_{1-\alpha/2}}\right)^2 + \frac{1}{N} \cdot \sum_{j=1}^m W_j \cdot s_j^2} \quad (\text{Worst case } \theta_j = 1/2 \forall j)$$

giver n som i punkt 1)

- Den optimale stikprøvestørrelse for stratum j kan findes ved ^{note 7}

$$n_j \geq \frac{W_j \cdot s_j}{\sum_{k=1}^m W_k \cdot s_k} \cdot n$$

⁷ I formelen er der anvendt følgende approksimation: $\tau_j^2 = \frac{N_j}{N_j - 1} \cdot \theta \cdot (1 - \theta) \approx \theta \cdot (1 - \theta)$