

# Versuchsplanung für Nebenfachstudierende

Erwin Grüner

FB Psychologie Uni Marburg

23.11.2005

Schichtenstichprobe

Quotenstichprobe

Mehrstufige Auswahlverfahren

Klumpenstichprobe

Schichtenstichprobe

Quotenstichprobe

Mehrstufige

Auswahlverfahren

Klumpenstichprobe

## Stichprobenverfahren II

- Die Population wird in Teilpopulationen eingeteilt.
- Die Einteilung muss disjunkt und vollständig sein.
- Aus jeder Teilpopulation wird getrennt eine Stichprobe gezogen (uneingeschränkte Zufallsauswahl!).

- Bei gleichem Stichprobenumfang ist eine höhere Genauigkeit möglich als bei der uneingeschränkten Zufallsauswahl.
- Oder: für die gleiche Genauigkeit ist ein kleinerer Stichprobenumfang nötig.

- Das Schichtmerkmal muss mit erhoben werden.

- Die Zusammensetzung der Schichten in der Population muss bekannt sein.
- Das Schichtmerkmal sollte mit dem zu untersuchenden Merkmal zusammenhängen.
- Die Schichtenstichprobe ist von Vorteil bei möglichst großer Heterogenität zwischen den Schichten und möglichst großer Homogenität innerhalb der Schichten.

Mit  $k$  bezeichnen wir die Anzahl der Schichten.

Wir betrachten erst den Fall **endlicher Populationen**:

- $N$ : Gesamtumfang der Population
- $N_j$ : Umfang der  $j$ -ten Schicht ( $\sum N_j = N$ )
- $g_j$ : Anteil der  $j$ -ten Schicht ( $g_j = \frac{N_j}{N}$ ) an der Gesamtpopulation
- $\mu_j$ : Mittelwert des zu untersuchenden Merkmals in der  $j$ -ten Schicht
- $\sigma_j^2$ : Varianz des zu untersuchenden Merkmals in der  $j$ -ten Schicht

Für den Fall **unendlicher Populationen** haben wir:

- $g_j$ : Anteil der  $j$ -ten Schicht (es gilt:  $\sum g_j = 1$ ) an der Gesamtpopulation
- $\mu_j$ : Mittelwert des zu untersuchenden Merkmals in der  $j$ -ten Schicht
- $\sigma_j^2$ : Varianz des zu untersuchenden Merkmals in der  $j$ -ten Schicht

Für die Stichprobe verwenden wir die folgende Notation:

- $n$ : Gesamtstichprobenumfang
- $n_j$ : Stichprobenumfang der  $j$ -ten Schicht
- $\bar{X}_j$ : Mittelwert des zu untersuchenden Merkmals in der  $j$ -ten Schicht
- $s_j^2$ : Varianz des zu untersuchenden Merkmals in der  $j$ -ten Schicht

$$\mu = \sum g_j \mu_j$$

$\mu$  ist also ein “gewichteter Mittelwert”.

# Schätzung des Mittelwerts $\mu_j$ der $j$ -ten Schicht

Versuchs-  
planung

Erwin Grüner

Schichtenstichprobe

Quotenstichprobe

Mehrstufige

Auswahlverfahren

Klumpenstichprobe

$$\hat{\mu}_j = \bar{X}_j = \frac{\sum_{i=1}^{n_j} X_i^{(j)}}{n_j}$$

# Schätzung des Varianz $\sigma_j^2$ der $j$ -ten Schicht

Versuchs-  
planung

Erwin Grüner

Schichtenstichprobe

Quotenstichprobe

Mehrstufige

Auswahlverfahren

Klumpenstichprobe

$$\hat{\sigma}_j^2 = s_j^2 = \frac{\sum_{i=1}^{n_j} (X_i^{(j)} - \bar{X}_j)^2}{n_j - 1}$$

$$\hat{\mu} = \bar{X} = \sum g_j \bar{X}_j$$

# Standardfehler von $\bar{X}$

$$\sigma_{\bar{X}}^2 = \sum g_j^2 \frac{\sigma_j^2}{n_j}$$

Also:

$$\sigma_{\bar{X}} = \sqrt{\sum g_j^2 \frac{\sigma_j^2}{n_j}}$$

# Schätzung des Standardfehlers von $\bar{X}$

Versuchs-  
planung

Erwin Grüner

Schichtenstichprobe

Quotenstichprobe

Mehrstufige

Auswahlverfahren

Klumpenstichprobe

$$\hat{\sigma}_{\bar{X}} = \sqrt{\sum g_j^2 \frac{s_j^2}{n_j}}$$

$$Pr\{\bar{X} - z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}} \leq \mu \leq \bar{X} + z_{1-\frac{\alpha}{2}}\sigma_{\bar{X}}\} = 1 - \alpha$$

bzw.

$$Pr\{\bar{X} - t_{(1-\frac{\alpha}{2}), (n-k)}\hat{\sigma}_{\bar{X}} \leq \mu \leq \bar{X} + t_{(1-\frac{\alpha}{2}), (n-k)}\hat{\sigma}_{\bar{X}}\} = 1 - \alpha$$

Das Vorgehen ist völlig analog zum Vorgehen bei der uneingeschränkten Zufallsauswahl, d.h. wir substituieren

- $\mu$  durch  $\pi$
- $\bar{X}$  durch  $p$

und verwenden die obigen Formeln für geschichtete Stichproben.

Auch hier gilt wiederum:

$$\sigma_j^2 = \pi_j(1 - \pi_j)$$

und

$$\hat{\sigma}_j^2 = p_j(1 - p_j)$$

Wir haben also:

$$\hat{\sigma}_p^2 = \sum g_j^2 \frac{p_j(1 - p_j)}{n_j}$$

und

$$\hat{\sigma}_p = \sqrt{\sum g_j^2 \frac{p_j(1 - p_j)}{n_j}}$$

Wir erhalten also als Näherungsformel:

$$Pr\{p - z_{1-\frac{\alpha}{2}}\hat{\sigma}_p \leq \pi \leq p + z_{1-\frac{\alpha}{2}}\hat{\sigma}_p\} = 1 - \alpha$$

# Aufteilung des Stichprobenumfangs auf die Schichten

Im Prinzip gibt es viele mögliche Aufteilungen des Gesamtstichprobenumfangs.

Zwei Spezialfälle sind hervorzuheben:

- Die proportionale Aufteilung
- Die bestmögliche Aufteilung

Unabhängig von der Art der Aufteilung bleiben die Formeln für die Schätzung des Mittelwerts und das Konfidenzintervall natürlich allesamt gültig<sup>1</sup>.

---

<sup>1</sup>Durch Einsetzen und Umformen vereinfachen sie sich allenfalls.

$$n_j = ng_j$$

Die Zahlenverhältnisse bei der Stichprobe entsprechen jenen bei der Population.

$$n_j = n \frac{g_j \sigma_j}{\sum g_i \sigma_i}$$

- Hier wird neben den Proportionen auch noch die Varianz des Merkmals innerhalb der Schichten berücksichtigt, d.h. für Schichten mit größerer Heterogenität ist die Wahrscheinlichkeit, gezogen zu werden, höher.
- Die bestmögliche Aufteilung bringt die höchste Schätzgenauigkeit.
- Sind die Varianzen in den Schichten alle gleich, so ergibt sich die proportionale Aufteilung.

- In der Markt- und Meinungsforschung ist dieses Stichprobenverfahren stark verbreitet.
- Wie bei der Schichtenstichprobe wird die Population nach verschiedenen Merkmalen<sup>2</sup> aufgeteilt.

---

<sup>2</sup>Z.B. Geschlecht, Alter, Bildung

- Die Auswahl ist nicht zufällig, sondern bleibt dem Interviewer überlassen.
- Bei mehr als einem Auswahlmerkmal werden nur die zu erreichenden Randhäufigkeiten und nicht die verbundenen Häufigkeiten vorgegeben <sup>3</sup>.

---

<sup>3</sup>Dadurch kann es theoretisch vorkommen, dass bestimmte Kombinationen der Auswahlmerkmale in der Stichprobe nicht vertreten oder unterrepräsentiert sind.

- Kostenvorteile
- Organisatorische Vorteile (Verfügbarkeit, Freiwilligkeit)

- Statistische Aussagen (Schätzwerte und deren Genauigkeit) fraglich
- Mangelnde Repräsentativität

Bei einem mehrstufigen Auswahlverfahren wird folgendermaßen vorgegangen:

- Die Population wird in “Primäreinheiten” eingeteilt (z.B. Länder, Betriebe)
- Einige Primäreinheiten werden nach dem Zufall gezogen.
- Die Primäreinheiten können weiter aufgeteilt werden in “Sekundäreinheiten” (z.B. Städte, Abteilungen).
- Von diesen werden wiederum einige nach dem Zufall gezogen.
- usw.

- Bei einem mehrstufigen Auswahlverfahren können auf der untersten Stufe alle Mitglieder untersucht werden oder nur eine Zufallsstichprobe.
- Werden auf der untersten Stufe alle Mitglieder untersucht (Vollerhebung), so spricht man von einer **Klumpenstichprobe**.

# Voraussetzungen für die Anwendung des Verfahrens

- Alle möglichen Einheiten bzw. Klumpen müssen zugänglich sein und für die Untersuchung prinzipiell zur Verfügung stehen.
- Das Merkmal, nach dem die Aufteilung der Populationen erfolgt, sollte mit dem zu untersuchenden Merkmal **nicht** zusammenhängen.
- Die Einheiten bzw. Klumpen sollten in sich möglich heterogen sein.
- Die Klumpenstichprobe ist von Vorteil bei möglichst großer Homogenität zwischen den Schichten und möglichst großer Heterogenität innerhalb der Teilpopulationen.

- Organisatorische Vorteile
- Kostenvorteil
- Zuweilen die einzige Möglichkeit <sup>4</sup>

---

<sup>4</sup>Eine uneingeschränkte Zufallsauswahl bzw. Schichtenstichprobe ist oft nicht einsetzbar bzw. auch nicht sinnvoll, beispielsweise bei Schulklassen oder Betrieben.

- Mangelnde Repräsentativität?
- Gesamtstichprobenumfang zufallsabhängig

# Versuchs- planung

Erwin Grüner

Schichtenstichprobe

Quotenstichprobe

Mehrstufige

Auswahlverfahren

Klumpenstichprobe