Gewichtungseffizienz und statistische Power

Beeinträchtigung der statistischen Test-Power durch Gewichtung

Martin Förster

Abteilung Zentrale Methodenlehre
Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften
Europa-Universität Flensburg

6. Januar 2020

Gewichtungseffizienz

Seien w_i normalisierte Gewichte, sodass $\sum w_i = n$, dann ist die Gewichtungseffizienz $WE = \frac{n}{\sum w_i^2}$.

Mithin: $0 < WE \leq 1$, wobei $WE = 1$ die maximale Gewichtungseffizienz kennzeichnet. Je kleiner die WE, desto größer die Varianz und die Standardfehler der Schätzungen in der Stichprobe.

Effektive Stichprobengröße (effective sample size)

Die effektive Stichprobengröße $\Xi = n \times WE$ beschreibt die Größe einer ungewichteten Stichprobe, welche die gleiche Varianz aufweist wie die gewichtete Stichprobe.

Also: je kleiner die WE, desto kleiner Ξ, und desto größer die Varianz (Standardabweichung s und Standardfehler SE).

Statistische Hypothesen und Signifikanztests

Sei γ ein (unbekannter) Parameter in der Population, $\hat{\gamma}$ die konkrete Punktschätzung für γ aus einer Stichprobe, und ζ ein Vergleichswert für γ, dann überprüft ein Signifikanztest die Nullhypothese $H_0 : \gamma = \zeta$. Weiterhin wird eine Alternativhypothese formuliert, welche quasi das Gegenteil der Nullhypothese behauptet; z.B. $H_A : \gamma \neq \zeta$.

Da γ unbekannt ist, muss H_0 mit $\hat{\gamma}$ überprüft werden. Allerdings ist $\hat{\gamma}$ die Realisierung einer Zufallsvariable, mithin besitzt der Schätzer eine Varianz (d.h. $P(\hat{\gamma} = \gamma) < 1$). Somit $\hat{\gamma} \neq \zeta \nLeftrightarrow \gamma \neq \zeta$.

Für konsistente Schätzer ist ihre Varianz abhängig von der Stichprobengröße: $P(\hat{\gamma} = \gamma) = \lim_{n \to \infty} 1$.

α-Fehler

Im Kern berechnet ein statistischer Test eine Prüfgröße $\hat{\theta}$, die quasi ein Maß dafür ist, wie stark sich $\hat{\gamma}$ und ζ unterscheiden. Der Erwartungswert $E\left(\hat{\theta}\right)$ der Prüfgröße ist der Wert, welcher für $\hat{\theta}$ erwartet wird, wenn H_0 zutrifft.

Da aber $\hat{\gamma}$ eine Zufallsvariable ist, gilt dies auch für die Prüfgröße: $\hat{\theta}$ ist eine Zufallsvariable. Daher $P\left(\hat{\theta} \neq E\left(\hat{\theta}\right) | \gamma = \zeta\right) > 0$ und $P\left(\hat{\theta} \neq E\left(\hat{\theta}\right) | \gamma = \zeta\right) = \lim_{n \to \infty} 0$. Weil und insofern die Verteilung von $\hat{\theta}$ bekannt ist, lässt sich diese Wahrscheinlichkeit (p) für konkrete Schätzungen von θ ermitteln.

Wird H_0 abgelehnt, gibt es also immer die Wahrscheinlichkeit p, einen Fehler zu begehen (da H_0 eigentlich zutrifft). Dieser Fehler wird als α-Fehler bezeichnet.

β-Fehler

Darüber hinaus ist noch ein zweiter Fehler möglich: H_0 anzunehmen, obwohl H_0 in Wirklichkeit nicht zutrifft (also H_A abzulehnen, obwohl H_A eigentlich zutrifft). Dieser Fehler wird als β-Fehler bezeichnet.

Die Wahrscheinlichkeit, den β-Fehler zu begehen, lässt sich jedoch nicht unmittelbar aus den Daten bestimmen, da H_A nicht hinreichend spezifisch dafür ist, d.h. es gibt keinen konkreten Erwartungswert für $\hat{\theta}$ unter der Bedingung, dass H_A zutrifft.

Für spezifizierte Szenarien kann die Wahrscheinlichkeit β allerdings ermittelt werden. Die Power eines statistischen Tests ist definiert als $1 - \beta$.

Europa-Universität Flensburg
Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften
Abteilung Zentrale Methodenlehre

Daten und Test

generierte Daten:

- $n = 100$
- Variable X, annähernd normalverteilt
- Gewicht $W1$, normalisiert auf $\bar{x}_{W1} = 1$ und $\sum w1_i = n$
- Gewicht $W2$, verstärkte $W1$-Gewichte, $w2_i = \left((w1_i - 1)^2 \frac{w1_i - 1}{|w1_i - 1|}\right) + 1$, somit ebenfalls $\bar{x}_{W2} = 1$ und $\sum w2_i = n$, aber $min(W2) < min(W1)$ und $max(W2) > max(W1)$

Test:

- Parameter ist Erwartungswert, Schätzer ist arithmetisches Mittel $\hat{\mu}$
- Es wird getestet auf $H_0 : \mu = \zeta$, wobei $\zeta = -0.5$

Europa-Universität Flensburg
Interdisziplinäres Institut für Umwelt-, Sozial- und Humanwissenschaften
Abteilung Zentrale Methodenlehre

Szenario 0: Test mit ungewichteten Daten

$\hat{\mu}_0 = \frac{1}{n}\sum x_i = -0.1569$
Varianz X: $\hat{\sigma}_0^2 = \frac{1}{n-1}\sum(x_i - \hat{\mu}_0) = 1.0359$
Standardfehler $\hat{\mu}$: $SE_0 = \sqrt{\frac{\hat{\sigma}_0^2}{n}} = 0.1018$
$\hat{\theta}_0 = \frac{\hat{\mu}_0 - \zeta}{SE_0} = 3.371 \qquad E\left(\hat{\theta}\right) = 0 \qquad \hat{\theta} \sim N\left(E\left(\hat{\theta}\right), 1\right)$
$p_0 = 0.0004$
Für die Bestimmung der statistischen Power muss eine maximal hinnehmbare Wahrscheinlichkeit α des α-Fehlers gesetzt werden; d.h. wenn $p \leq \alpha$, dann soll H_0 abgelehnt werden. Es wird hier festgelegt $\alpha = 0.05$. Weiterhin wird spezifiziert $H_A : \mu \neq \zeta$.
Für das so spezifizierte Szenario ergibt sich eine statistische Power $1 - \beta_0 = \Phi\left(\hat{\theta} - \theta_{1-\alpha/2}\right) = 0.92$.
Für einen akzeptablen Test wird $1 - \beta \geq 0.8$ erwartet.

Szenario 1: Test mit $W1$-gewichteten Daten

$$WE_1 = 0.9251 \qquad \Xi_1 = 92.51$$

$$\hat{\mu}_1 = -0.1959$$

$$SE_1 = 0.1026$$

$$\hat{\theta}_1 = 2.9654$$

$$p_1 = 0.0015 < \alpha$$

$$1 - \beta_1 = 0.84$$

Szenario 2: Test mit $W2$-gewichteten Daten

$$WE_2 = 0.7553 \qquad \Xi_2 = 75.53$$

Eine Gewichtungseffizienz von unter 0.9 gilt als problematisch!

$$\hat{\mu}_2 = -0.2348$$

$$SE_2 = 0.1032$$

$$\hat{\theta}_2 = 2.5698$$

$$p_2 = 0.0051 < \alpha$$

$$1 - \beta_2 = 0.73 < 0.8$$

Fazit

Je gringer die Gewichtungseffizienz (ceteris paribus), desto geringer die Power statistischer Tests.