

8327 NL

MATHEMATISCH CENTRUM,
 2e Boerhaavestraat 49,
 AMSTERDAM. - O. STATISTISCHE AFD.
 S 145A(M 53)

Een schatting en betrouwbaarheidsgrenzen voor de waarde van de onafhankelijke variabele bij een gegeven verwachting van de afhankelijke variabele indien tussen beide een lineair verband bestaat. ¹⁾

Onderstellingen. Bij een aantal, k , bekende x - waarden (b.v. verschillende doses van een praeparaat toegediend aan ratten) worden groepjes waarnemingen voor y ²⁾ verricht (b.v. het percentage van het praeparaat, dat na enkele uren in de rat wordt teruggevonden). In een schema:

x- waarden	x_1	x_2		x_k
waarnemingen	y_{11} y_{1n_1}	y_{21} y_{2n_2}		y_{k1} y_{kn_k}

Hierbij wordt de onderstelling gemaakt, dat de waarnemingen y_{ij} alle onderling onafhankelijk normaal verdeeld zijn met dezelfde (onbekende) spreiding σ en met verwachtingen:

$$(1) \quad \mathcal{E} y_{ij} = \alpha + \beta x_i \quad (i = 1, \dots, k ; j = 1, \dots, n_i),$$

waarin α en β onbekende parameters zijn.

Gevraagd wordt nu een punt- en een interval schatting te geven voor de waarde van x , waarbij een gegeven waarde van de verwachting van y behoort. (b.v. in ons voorbeeld

- 1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
- 2) Door onderstreping van de letter wordt aangegeven, dat de grootte een waarschijnlijkheidsverdeling bezit.

- 2 -

de dosis van het praeparaat, waarbij $\sum y = 50\%$).
 Geven we de gegeven verwachting van y aan met η en de
 bijbehorende onbekende x waarde met ξ_η , dan geldt dus:

$$2) \quad \eta = \alpha + \beta \xi_\eta \quad \text{of} \quad \xi_\eta = \frac{\eta - \alpha}{\beta}$$

Een puntschatting voor ξ_η , aangegeven door x_η , verkrijgen we door in (2) de kleinste kwadraten-schattingen voor α en β ($\underline{a}$ resp. $\underline{b}$ te vinden door $\sum_i \sum_j (y_{ij} - \alpha - \beta x_i)^2$ naar α en β te minimaliseren) in te vullen. Deze schattingen zijn:

$$\underline{b} = \frac{\sum n_i \bar{y}_i (x_i - \bar{x})}{\sum n_i (x_i - \bar{x})^2},$$

waarin

$$\bar{y}_i = \frac{1}{n_i} \sum_j y_{ij} \quad \text{en} \quad \bar{x} = \frac{\sum n_i x_i}{\sum n_i},$$

en

$$\underline{a} = \bar{y} + \underline{b} \bar{x},$$

waarin

$$\bar{y} = \frac{\sum n_i \bar{y}_i}{\sum n_i} = \frac{\sum_i \sum_j y_{ij}}{\sum_i n_i}.$$

Uit (2) vinden we dus

$$(3) \quad x_\eta = \bar{x} + \frac{\eta - \bar{y}}{\underline{b}}.$$

Om een interval schatting voor ξ_η te vinden, kunnen we de methode van FIELLER toepassen (zie [1], [2], of [3]), waarmee betrouwbaarheidsgrenzen voor $\xi_\eta - \bar{x}$ berekend kunnen worden. Hierbij zullen we gebruik maken van de volgende schatting voor σ^2 (de variantie van de grootheden y_{ij}):

$$s^2 = \frac{\sum_i (n_i - 1) \underline{s}_i^2}{\sum_i (n_i - 1)} = \frac{\sum_i (n_i - 1) \underline{s}_i^2}{N - k}$$

als we $\sum n_i$ door N voorstellen en $\underline{s}_i^2$ de schatting van σ^2 uit $y_{i1}, \dots, y_{in_i}$ is:

$$\underline{s}_i^2 = \frac{1}{n_i - 1} \sum_j (y_{ij} - \bar{y}_i)^2.$$

Hieruit volgt dan voor de schattingen van de varianties van resp. $\bar{y}$, $\bar{y}_i$ en $\bar{\ell}$:

$$\frac{s^2}{\bar{y}} = \frac{s^2}{\sum n_i} \quad ; \quad \frac{s^2}{\bar{y}_i} = \frac{s^2}{n_i} \quad ; \quad \frac{s^2}{\bar{\ell}} = \frac{s^2}{\sum n_i (x_i - \bar{x})^2} .$$

Verder kan men op grond van de veronderstelde normaliteit van de waarnemingen y_{ij} bewijzen, dat de grootheden $\eta - \bar{y}$ en $\bar{\ell}$ onderling onafhankelijk verdeeld zijn.

Bij toepassing van de methode van FIELLER (zie [1] of [2]) worden nu de grenzen voor het betrouwbaarheidsinterval (met betrouwbaarheid $1 - \varepsilon$) gevonden als de twee oplossingen voor α van:

$$(\eta - \bar{y})^2 - t_\varepsilon^2 s_y^2 - 2\alpha(\eta - \bar{y})\bar{\ell} + \alpha^2(\bar{\ell}^2 - t_\varepsilon^2 s_\ell^2) = 0 ,$$

waarin t_ε de kritieke waarde voor de Student-verdeling met N-k vrijheidsgraden is, met onbetrouwbaarheid ε , dus die waarde waarvoor geldt:

$$P[|t| \leq t_\varepsilon] = 1 - \varepsilon .$$

Het betrouwbaarheidsgebied bestaat uit alle waarden tussen deze grenzen, d.w.z. het is een eindig interval, indien

$$\bar{\ell}^2 - t_\varepsilon^2 s_\ell^2 > 0 ,$$

dus indien volgens de toets van Student met onbetrouwbaarheid ε de hypothese $\bar{\ell} = \beta = 0$ verworpen wordt. De grenzen voor het interval kunnen we schrijven als:

$$\frac{(\eta - \bar{y})\bar{\ell} \pm \sqrt{(\eta - \bar{y})^2 \bar{\ell}^2 - (\bar{\ell}^2 - t_\varepsilon^2 s_\ell^2)[(\eta - \bar{y})^2 - t_\varepsilon^2 s_y^2]}}{\bar{\ell}^2 - t_\varepsilon^2 s_\ell^2} ,$$

of wanneer we $\frac{\bar{\ell}^2}{\bar{\ell}^2 - t_\varepsilon^2 s_\ell^2} = C$ stellen en na enige herleiding:

$$C \frac{\eta - \bar{y}}{\bar{\ell}} \pm \frac{t_\varepsilon \sqrt{C}}{\bar{\ell}} \sqrt{s_y^2 + C \left(\frac{\eta - \bar{y}}{\bar{\ell}} \right)^2 s_\ell^2} .$$

Het gevraagde betrouwbaarheidsinterval voor ξ_η heeft dus als grenzen:

$$\bar{x} + C \frac{\eta - \bar{y}}{\bar{\ell}} \pm \frac{t_\varepsilon \sqrt{C}}{\bar{\ell}} \sqrt{s_y^2 + C \left(\frac{\eta - \bar{y}}{\bar{\ell}} \right)^2 s_\ell^2} .$$

Is $\ell^2 \gg t_\varepsilon^2 s_\ell^2$, dus $C \approx 1$, dan kan dit vereenvoudigd worden tot:

$$\underline{x}_\eta \pm \frac{t_\varepsilon}{\ell} \sqrt{s_{\underline{y}}^2 + \left(\frac{\eta - \bar{y}}{\ell}\right)^2 s_\ell^2}$$

Door de groepen even groot te maken, dus $n_1 = n_2 = \dots = n_k$, wordt verder de berekening van de grootheden $\bar{x}$, ℓ en s^2 belangrijk korter, nl:

$$\bar{x} = \frac{1}{k} \sum x_i \quad ; \quad \ell = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2} \quad ; \quad s^2 = \frac{\sum s_i^2}{k}$$

Literatuur:

- [1] E.C.FIELLER, A fundamental formula in the statistics of biological assay and some applications. Quart.J.of Pharm.and Pharmacology, XVII(1944), pp.117-123.
- [2] G.KLERK-GROBBEN, Een betrouwbaarheidsgebied voor het quotiënt van de verwachtingen van twee stochastische grootheden, die een simultane normale verdeling bezitten.
Rapport S 90(M 36) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam 1952.
- [3] G.GROBBEN en Dr J.HEMELRIJK Jr, Onderzoek ener ijkmethode.
Rapport S 70, Statistische Afdeling, Mathematisch Centrum, Amsterdam 1951.