MATHEMATISCH CENTRUM,
2de Boerhaavestr 49,
Amsterdam-O.
1953-4(2)

Een intervalschatting voor het gemiddelde van een steekproef uit een normale verdeling als een deel-steekproef gegeven is.

door R.Doornbos en Ph.van Elteren.

Vraagstuk

Uit een steekproef van N onderling onafhankelijke waarnemingen uit een normale verdeling met onbekende verwachting μ en onbekende spreiding σ worden aselect n waarnemingen $x_1, \ldots, x_n$ gekozen. Van deze waarnemingen worden het gemiddelde m en de spreiding s $\left(= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$ bepaald. Gevraagd wordt een interval-schatting met gegeven onbetrouwbaarheid α voor het gemiddelde van de steekproef.

Oplossing

Zij m^* het gemiddelde van de N-n niet tot de deel-steekproef behorende waarnemingen, dan is $\underline{m}-\underline{m}^*$ 1) normaal verdeeld met gemiddelde 0 en spreiding

$$\sigma\sqrt{\frac{1}{n}+\frac{1}{N-n}} = \sigma\sqrt{\frac{N}{n(N-n)}}$$

en dan heeft

$$\underline{t} = \frac{\underline{m}-\underline{m}^*}{\underline{s}\sqrt{\frac{N}{n(N-n)}}}$$

een Studentverdeling met n-1 vrijheidsgraden. Als nu $t(\alpha;n-1)$ de positieve waarde van $\underline{t}$ is, die bij een Studentverdeling met n-1 vrijheidsgraden een tweezijdige overschrijdingskans α heeft, dan geldt:

$$\frac{|\underline{m}-\underline{m}^*|}{\underline{s}\sqrt{\frac{N}{n(N-n)}}} \leqq t(\alpha;n-1) \quad \text{spr}\ \alpha. \qquad 2)$$

Hieruit volgt:

$$-t(\alpha;n-1)\,\underline{s}\sqrt{\frac{N}{n(N-n)}} + \underline{m} \leqq \underline{m}^*$$
$$\leqq t(\alpha;n-1)\,\underline{s}\sqrt{\frac{N}{n(N-n)}} + \underline{m} \quad (\text{spr}\ \alpha)$$

Indien wij het gemiddelde van alle N waarnemingen voorstellen door M geldt:

$$M = \frac{1}{N}\left(nm + (N-n)m^*\right)$$

---

1) Stochastische grootheden zullen hier onderstreept worden.
2) spr α betekent: behoudens een waarschijnlijkheid α

dus vinden we als grenzen voor de interval-schatting van M:

$$\underline{m} \pm t(\alpha; n-1)\, \underline{s} \sqrt{\frac{N-n}{nN}}$$

De interpretatie van deze oplossing is analoog aan de interpretatie van betrouwbaarheidsintervallen voor een parameter. Bij herhaalde uitvoering van het gehele experiment — het nemen van de steekproef èn van één deel-steekproef daaruit — zal het aantal foute conclusies op de lange duur ongeveer een fractie α van het totale aantal conclusies zijn.

Het is niet mogelijk iets te zeggen over de kans, dat $\underline{M}$ in genoemd interval zal vallen, als het eenmaal bepaald is