

Kansrekening & statistiek

Examen Theorie augustus 23-24

VRAAG 1

Indien je over een steekproef beschikt, wat is dan de relatieve en de cumulatieve frequentie ?
Hoe kun je ze grafisch voorstellen (eventueel op meerdere manieren) ? Wat is een empirische verdelingsfunctie ?

Relatieve frequentie: f_i / n , de dichtheidsschaal, som van hoogtes = 1.

Cumulatieve frequentie: som van de tot-nu-toe behandelde waarden.

Met behulp van een **histogram** voorstellen. Ook met **polygoon** en **ogieven**.

De **empirische verdelingsfunctie** stelt het aantal metingen voor met een waarde kleiner dan of gelijk aan een bepaalde waarde (x) gedeeld door get totale aantal metingen:

$$F_n(x) = \frac{\text{aantal metingen} \leq x}{n}$$

VRAAG 2

Geef de definitie van een statistiek, van een schatter en van een zuivere schatter. Geef tevens een zuivere schatter voor de populatievariantie en bewijs zijn zuiverheid.

Een **statistiek** is een stochastische variabele die alleen een functie is van de steekproef en niet van onbekende parameters (zoals μ en σ).

Een **schatter** is een statistiek die gebruikt wordt om een onbekende parameter te benaderen. Een schatting is de getalwaarde van de schatter in een concreet experiment.

Een schatter T van een parameter λ heet **zuiver** (Eng. unbiased), als $E[T] = \lambda$, dus als de verwachtingswaarde van de schatter gelijk is aan de (gezochte) parameter.

De steekproefvariantie is een zuivere schatter van de populatievariantie.

Bewijs:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1} \sum_{i=1}^n ((X_i - \mu) - (\bar{X}_n - \mu))^2 \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n (X_i - \mu)^2 - \sum_{i=1}^n (\bar{X}_n - \mu)^2 \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n (X_i - E[X_i])^2 - n(\bar{X}_n - E[\bar{X}_n])^2 \right] \end{aligned}$$

Dus:

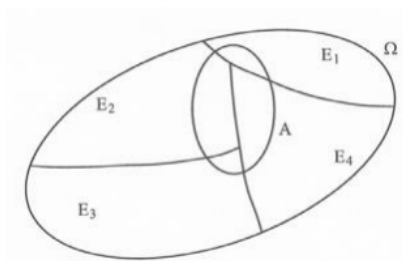
$$\begin{aligned} E[S_n^2] &= \frac{1}{n-1} \left[\sum_{i=1}^n E[(X_i - E[X_i])^2] - nE[(\bar{X}_n - E[\bar{X}_n])^2] \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n \text{Var}(X_i) - n\text{Var}(\bar{X}_n) \right] \\ &= \frac{1}{n-1} \left[n\sigma^2 - n\frac{\sigma^2}{n} \right] = \sigma^2 \end{aligned}$$

VRAAG 3

Wet der totale kans

Onderstel dat $E_1, E_2, \dots, E_n$ een partitie is van het universum Ω (wat wil zeggen dat $E_i \cap E_j = \emptyset$ voor $i \neq j$ en dat $\bigcup_{i=1}^n E_i = \Omega$). Neem nu een willekeurige gebeurtenis A , dan geldt:

$$P(A) = \sum_{i=1}^n P(A|E_i)P(E_i)$$



Bewijs:

$$\begin{aligned}
 A &= A \cap [E_1 \cup E_2 \cup \dots \cup E_n] = [A \cap E_1] \cup \dots \cup [A \cap E_n] \\
 \uparrow & \quad \quad \quad \text{— } E_i \text{'s zijn disjunct} \\
 A &= A \cap \Omega = \quad \quad \quad (A \cap E_i) \text{ 2 aan 2 disjunct} \\
 P(A) &= P(A \cap E_1) + P(A \cap E_2) + \dots + P(A \cap E_n) \\
 &= P(A|E_1)P(E_1) + \dots + P(A|E_n)P(E_n) \\
 &= \sum_{i=1}^n P(A|E_i)P(E_i)
 \end{aligned}$$

VRAAG 4

Hoe kun je het lineaire kleinste kwadraten model verifiëren? Leg uit.

- residuële plots: scatter plot van e_i in functie van y_i of x_i
- de variantieanalyse (F-toets): analysis of variance (ANOVA) en de determinatiecoëfficiënt R^2
- t-toets op $\text{rico} = 0$ (betrouwbaarheidsintervallen)

Wat is de determinatiecoëfficiënt bij lineaire regressie? Hoe kun je de significantie van je regressie toetsen?

De determinatiecoëfficiënt R^2 wordt nu gedefinieerd als de proportie van de totale variatie in de afhankelijke variabele y die door de regressie verklaard wordt:

$$R^2 = \frac{SS_{\text{Reg}}}{SS_T} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$