Kapitel 8

Punktskattningar och Konfidensintervall

UPPSALA UNIVERSITET

Vad innebär statistisk inferens?

Ett av huvudmålen i statistiska sammanhang är att utifrån informationen i ett stickprov (i något avseende) dra slutsatser angående den bakomliggande populationen.

En population karakteriseras ofta av en eller flera *parametrar* vilket innebär att det blir naturligt att *skatta* dessa.

Dessa skattningar baseras vanligtvis på den information som finns i *statistikor*.

Eftersom en statistika är en slumpvariabel får dess *samplingfördelning* ligga till grund för den kvalitetsbedömning som görs för en statistika beträffande skattning av en populationsparameter.

Grundläggande begrepp

Definition. En *punktskattare* (estimator) är en regel som anger hur man utifrån stickprovsinformationen bestämmer (parameter)skattningens värde.

Vad vill vi då att en bra punktskattare ska ha för egenskaper?

Definition. Låt $\hat{\theta}$ vara en punktskattare av parametern θ.

$$E\left(\widehat{\theta}\right) = \theta$$

sägs $\hat{\theta}$ vara en *väntevärdesriktig*, eller *förväntningsriktig* (eng. *unbiased*) punktskattare av θ.

UPPSALA UNIVERSITET

Grundläggande begrepp

Definition. En punktskattare som inte är väntevärdesriktig sägs vara *biased*. En punktskattares *bias* ges av

$$B\left(\widehat{\theta}\right) = E\left(\widehat{\theta}\right) - \theta$$

Definition. Medelkvadratfelet för $\hat{\theta}$ ges av

$$MSE\left(\widehat{\theta}\right) = E\left[\left(\widehat{\theta} - \theta\right)^2\right]$$

Det går att visa att (se uppgift 8.1)

$$MSE\left(\widehat{\theta}\right) = V\left(\widehat{\theta}\right) + \left[B\left(\widehat{\theta}\right)\right]^2$$

Konfidensintervall

Definition. En *intervallskattare* är en regel som anger hur man utifrån stickprovsinformationen bestämmer intervallgränserna $\hat{\theta}_L$ och $\hat{\theta}_U$.

Definition. Intervallskattaren $(\hat{\theta}_L,\hat{\theta}_U)$ sägs vara ett *konfidensintervall* för θ med *konfidensgraden* 1-α om

$$\Pr\left(\widehat{\theta}_L \leq \theta \leq \widehat{\theta}_U\right) = 1 - \alpha$$

Definition. Intervallskattarna $(\hat{\theta}_L,\infty)$ och $(-\infty,\hat{\theta}_U)$ vara *ensidiga konfidensintervall* för θ med konfidensgraden 1-α om

$$\Pr\left(\widehat{\theta}_L \leq \theta\right) = \Pr\left(\theta \leq \widehat{\theta}_U\right) = 1 - \alpha$$

Konfidensintervall

För att finna ett konfidensintervall för θ används med fördel den sk *pivotmetoden*. Metoden går ut på att finna en sk *pivotkvantitet* vilken ska ha två egenskaper.

1. Den ska vara en funktion av <u>både</u> stickprovet och θ.

2. Dess sannolikhetsfördelning ska <u>inte</u> bero på θ.

Förutom dessa båda egenskaper är det även önskvärt om pivotkvantiteten har följande egenskap.

3. Pivotkvantiteten ska helst ha en sannolikhetsfördelning som är någorlunda enkel att hantera.

Konfidensintervall i samband med normalfördelningen

Låt $X_1, X_2, \ldots, X_n$ vara olfsv $N(\mu, \sigma)$ med σ känd.

$$\overline{X} \text{ är } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$ följer att

vilket innebär att Z är en pivotkvantitet.

$$\Pr\left(\widehat{\mu}_L \leq \mu \leq \widehat{\mu}_U\right) = 0.95$$

och eftersom Z är en pivotkvantitet gör vi detta relativt enkelt via

$$\Pr\left(-1.96 \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$$

som leder till att

$$\Pr\left(\underbrace{\overline{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_L} \leq \mu \leq \underbrace{\overline{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}}_{\widehat{\mu}_U}\right) = 0.95$$

Uppgift 8.43 a

Låt $Y_1, Y_2, \ldots, Y_n$ vara oberoende och likafördelade $U(0,\theta)$. Låt vidare $Y_{(n)}=\max(Y_1, Y_2, \ldots, Y_n)$ och betrakta $U=(1/\theta)Y_{(n)}$.

Eftersom $U=(1/\theta)Y_{(n)}$ är en funktion både av stickprovet och θ är punkt 1 uppfylld.

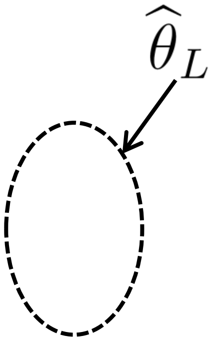
$$F_U(u) = \Pr(U \leq u)$$

Eftersom fördelningen för *U* inte beror på θ är även punkt 2 uppfylld. U är således en pivotkvantitet.

Uppgift 8.43 b

Vi söker $\hat{\theta}_L$ i uttrycket

$$0.95 = \Pr\left(\widehat{\theta}_L \leq \theta\right)$$



Detta följer utifrån sambandet

$$F_U(u) = \Pr(U \leq u)$$

Så vad ska *u* vara? Detta finner vi via

$$F_U(u) = u^n = 0.95 \qquad \text{dvs}$$

och således gäller att

$$\widehat{\theta}_L = \frac{Y_{(n)}}{0.95^{1/n}}$$

vilket betyder att intervallet ges av