Statistik IV für Nebenfachstudierende

2.1 Schätzen von $\underline{\mu}$ und $\underline{\underline{\Sigma}}$

Prof. Dr. Andreas Mayr

Institut für Statistik, LMU München

Sommersemester 2017

Schätzen von $\underline{\mu}$

- Seien $X_1, ..., X_n$ unabhängige Wiederholungen einer $p-$dimensionalen Zufalsvariablen mit
  - $\mathbb{E}(\underline{X}) = \underline{\mu}$
  - Cov$(\underline{X}) = \underline{\underline{\Sigma}}$
- Datenmatrix

$$\underline{\underline{X}} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

- Schätzung von $\underline{\mu}$:

$\underline{\bar{x}} =$

Eigenschaften I

- Schwerpunkt der Daten
- Erwartungstreue
- Genauigkeit des Schätzers

Eigenschaften II

- Asymptotisch normalverteilt
- Bester linearer unverzerrter Schätzer (BLUE)

Schätzen von $\underline{\underline{\Sigma}}$

- Datenmatrix

$$\underline{\underline{X}} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

- Empirische Kovarianzmatrix $\underline{\underline{S}}$:

$$\begin{aligned} \underline{\underline{S}} &= \frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})(\underline{x}_i - \underline{\bar{x}})^\top \\ &= \frac{1}{n-1} \left(\sum_{i=1}^{n} \underline{x}_i \underline{x}_i^\top - n\, \underline{\bar{x}}\, \underline{\bar{x}}^\top \right) \end{aligned}$$

Eigenschaften

- Erwartungstreue
- Unabhängigkeit (wenn $X \sim N(\underline{\mu}, \underline{\underline{\Sigma}})$
- Verteilung (wenn $X \sim N(\underline{\mu}, \underline{\underline{\Sigma}})$

Mehr-Stichprobenfall

- Population zerfällt in mehrere Gruppen, $k = 1, ..., g$
- Einfachster Fall: $g = 2$ (z.B. männlich und weiblich)

► separate Schätzungen von $\underline{\mu}^{(k)}$ und $\underline{\underline{\Sigma}}^{(k)}$ für jede einzelne Gruppe

$$
\begin{aligned}
\underline{\bar{x}}^{(k)} &= \frac{1}{n_k}\sum_{i=1}^{n_k} \underline{x}_i^{(k)} \longrightarrow \underline{\mu}_k \\
\underline{\underline{S}}^{(k)} &= \frac{1}{n_k - 1}\sum_{i=1}^{n_k} (\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)})(\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)})^\top
\end{aligned}
$$

Graphisch (für $p = 2, g = 2$)

Globaler Mittelwert über alle Klassen:

$$\underline{\bar{x}} = \frac{1}{n} \sum_{k=1}^{g} n_k \, \underline{\bar{x}}^{(k)}$$

Streuungszerlegung

$$
\begin{aligned}
\underline{\underline{T}} &= \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\underline{x}_i^{(k)} - \underline{\bar{x}})(\underline{x}_i^{(k)} - \underline{\bar{x}})^\top \\
&= \sum_{k=1}^{g}\sum_{i=1}^{n_k}(\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)})(\underline{x}_i^{(k)} - \underline{\bar{x}}^{(k)})^\top + \sum_{k=1}^{g}(\underline{\bar{x}}^{(k)} - \underline{\bar{x}}^{(k)})(\underline{\bar{x}}^{(k)} - \underline{\bar{x}}^{(k)})^\top
\end{aligned}
$$

Graphisch (für $p = 1, g = 2$)

Graphisch (für $p = 2, g = 2$)