Likelihood ratio test
Antag model $Y \sim \mathcal{N}_n(\mu, \sigma^2 I)$, hvor $\mu \in \Omega_1 \subset \mathbb{R}^n$, $\dim(\Omega_1) = m_1$. Lad $\Omega_0 \subset \Omega_1$ være et lineært underrum af Ω_1 med $\dim(\Omega_0) = m_0$.
Vi tester hypotesen

$$\begin{aligned} \mathcal{H}_0 : && \mu \in \Omega_0 \\ \mathcal{H}_1 : && \mu \in \Omega_1 n \Omega_0 \end{aligned}$$

med en likelihood ratio test. $\mathcal{H}_0$ betyder at én eller flere parametre kan fjernes.

Sætning 6.3
Likelihood ratio

$$\lambda(y) = \frac{\sup_{\mu \in \Omega_0} L(\mu, \sigma^2;)}{\sup_{\mu \in \Omega_1} L(\mu, \sigma^2; y)}$$

er en strengt aftagende funktion af

$$F(y) = \frac{D(p_1(y); p_0(y))/(m_1 - m_0)}{D(y; p_1(y))/(n - m_1)}$$

Under $\mathcal{H}_0$, så

$$F(Y) \sim F(m_1 - m_0, n - m_1)$$

Bevis
Under

$$\mathcal{H}_1 : \qquad \hat{\mu}_1 = p_1(y), \qquad \hat{\sigma^2}_1 = \frac{D(y; p_1(y))}{n}$$

$$\begin{aligned} l((\hat{\mu}_1, \hat{\sigma}^2); y) &= -\frac{n}{2} \log(\hat{\sigma}_1^2) - \frac{1}{2\hat{\sigma}_1^2} F(y; p_1(y)) \\ &= -\frac{n}{2} \log(\hat{\sigma}_1^2 - \frac{n}{2} \end{aligned}$$

Der fås tilsvarende udtryk under $\mathcal{H}_0$. Derfor fås

$$\begin{aligned}
\lambda(y) &= \exp\left(l(\hat{\mu}_0, \hat{\sigma}_0), y)l((\hat{\mu}_1, \hat{\sigma}_1); y)\right) \\
&= \exp\left(-\frac{n}{2}(\log(\hat{\sigma}_0^2) - \log(\hat{\sigma}_1^2))\right) \\
&= \left(\frac{D(y; p_1(y)}{D(y; p_0(y)}\right)^{\frac{n}{2}} \\
&\text{Pythagoras: } y - p_1(y) \perp \Omega_1 \quad \text{og} \quad p_1(y) - p_0(y) \in \Omega_1 \\
&= \left(\frac{D(y; p_1(y))}{D(y; p_1(y)) + D(p_1(y); p_0(y))}\right) P\frac{n}{2} \\
&= \left(\frac{1}{1 + \frac{D(p_1(y); p_0(y))}{D(y; p_1(y))}}\right)^{\frac{n}{2}} \\
&= \left(\frac{1}{1 + \frac{m_1 - m_0}{n - m_1} F(y)}\right)^{\frac{n}{2}}
\end{aligned}$$

Dvs. $\lambda(y)$ er en strengt aftagende funktion af $F(y)$. De er tilsvarende.

Desuden

$$F(y) = \frac{D(p_1(y); p_0(y))/(m_1 - m_0)}{D(y; p_1(y))/(n - m_1)} \sim \underbrace{\frac{\chi^2(m_1 - m_0)/(m_1 - m_0)}{\chi^2(n - m_1)/(n - m_1)}}_{\text{uafhængige } \chi^2\text{-fordelninger}} = F(m_1 - m_0, n - m_1)$$

Da små værdier af $\lambda(y)$ er kritiske for $\mathcal{H}_0$, så er store værdier af $F(y)$ kritiske for $\mathcal{H}_0$. Dvs. F-testen er højresidet – hypotesen forkastes, hvis der bevæges for langt ud mod højre.

Determinationskoefficienter

Lad

$$\Omega_{null} = \text{span}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

svarende til $y = \beta_0 + \varepsilon_i$ og $p_{null}(y)$ er den tilsvarende projektion. Så haves determinationskoefficient

$$R^2 = \frac{D(p_1(y); p_0(y))}{D(y; p_{null}(y))} = 1 - \frac{D(y; p_1(y))}{D(y; p_{null}(y))}$$

Bemærk:

- $0 \leq R^2 \leq 1$
- Ω_{null} svarer til at alle x'er er ukendte, dvs $D(y; p_{null}(y))$ udtrykker den totale variation i data.
- $D(p_1(y); p_{null}(y))$ udtrykker forbedring i variation ved brug af Ω_1 i stedet for Ω_0.
- Dvs. R^2 er andelen af variation i data som er forklaret af modellen med Ω_1.
- R^2 vokser altid for mere komplicerede modeller.

Justeret determinationskoefficient

$$R^2_{adj} = 1 - \frac{D(y; p_1(y))/(n-k)}{D(y; p_{null}(y))/(n-1)}$$

Bemærk:

- R^2_{adj} er justeret for modelkompleksitet.
- Kan bruges til at sammenligne modeller.
- Højere R^2_{adj} betyder bedre model.

Multiple modelreduktioner
Betragt en kæde af underrum af $\mathbb{R}^n$.

$$\Omega_m \subset \Omega_{m-1} \subset \ldots \subset \Omega_2 \subset \Omega_1 \subset \mathbb{R}^n$$

Der kan opstilles hypoteser for hvert underrum:

$$\begin{aligned} \mathcal{H}_0^i : &\quad \mu \in \Omega_i \\ \mathcal{H}_1^i : &\mu \in \Omega_{i-1} n \Omega_i \end{aligned}$$

Sætning 3.7

$$D(p_1(y); p_M(y)) = \sum_{i=1}^{M-1} D(p_{i+1}(y); p_1(y))$$

hvor

$$D(p_{i+1}(y); p_i(y)) = D(y; p_{i+1}(y)) - D(y; p_i(y))$$

svarende til stigning i devians fra model i til $i+1$, hvor model $i+1$ er en model på et mindre rum end model i.

Bevis
Af Pythagoras fås

$$D(y; p_M(y)) = D(y; p_1(y)) + D(P_1(y); p_M(y))$$

Derfor:

$$\begin{aligned} D(p_1(y); p_M(y)) &= D(y; p_M(y)) - D(y; p_1(y)) \\ &= \sum_{i=1}^{M-1} (D(y; P_{i+1}(y)) - D(y; p_i(y))) \end{aligned}$$

Der kan altså laves en kæde af F-tests startende med den mest komplicerede model, og så fjernes et eller flere led af gangen.

Bemærk:

- Testene afhænger af rækkefølge af led.
- I backward elimination vælges rækkefølge efter størrelse af p-værdier.