F-testi yleiselle lineaariselle hypoteesille

- Oletetaan lin. malli $\mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\right)$ $(\mathrm{r}(\mathbf{X}) = p)$, ja tarkastellaan nollahypoteesia

$$H : \mathbf{A}\boldsymbol{\beta} = \mathbf{c}, \quad \mathbf{A} \ (q \times p) \ \text{ ja } \ \mathbf{c} \ (q \times 1) \ \text{ tunnettuja}, \ \mathrm{r}(\mathbf{A}) = q.$$

- Testi on luontevaa perustaa erotukseen $\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}$, jossa $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ on $\boldsymbol{\beta}$:n (vapaa) PNS-estimaattori tai yhtäpitävästi (vapaa) SU-estimaattori.

- Koska $\hat{\boldsymbol{\beta}}$ estimoi $\boldsymbol{\beta}$:aa tehokkaasti olipa H tosi vai ei, pätee aina $\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c} \approx \mathbf{A}\boldsymbol{\beta} - \mathbf{c}$. Erotus $\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}$ saa siten tyypillisesti "pieniä" arvoja, kun H on tosi ja "suuria" arvoja, kun H ei ole tosi.

- Lausetta 2.1 käyttäen voidaan johtaa testisuure, jonka avulla erotuksen $\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c}$ suuruutta voidaan arvioida.

- Saatava testi perustuu tilastollisen päättelyn kurssilla esitetyn Waldin testin periaatteeseen.

F-testi yleiselle lineaariselle hypoteesille

- Oletetaan lin. malli $\mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\beta, \sigma^2\mathbf{I}_n\right)$ $(\mathrm{r}\left(\mathbf{X}\right) = p)$, ja tarkastellaan nollahypoteesia

$$H : \mathbf{A}\beta = \mathbf{c}, \quad \mathbf{A} \;\; (q \times p) \;\; \text{ja} \;\; \mathbf{c} \;\; (q \times 1) \;\; \text{tunnettuja}, \;\; \mathrm{r}\left(\mathbf{A}\right) = q.$$

- Ns. F-testi hypoteesille H perustuu β:n (vapaaseen) PNS-estimaattoriin $\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}$ ja σ^2:n harhattomaan estimaattoriin $S^2 = \left(n-p\right)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$.
- Testisuureena on ns. F-testisuure

$$F = (\mathbf{A}\hat{\beta} - \mathbf{c})'(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{c})/qS^2 \overset{H}{\sim} F_{q,n-p}$$

- Suuret F:n arvot ovat kriittisiä nollahypoteesin kannalta. P-arvot:

$$\mathrm{P} = \mathrm{P}_H\left(F\left(\mathbf{Y}\right) \geq F\left(\mathbf{y}\right)\right) = \mathrm{P}\left(F_{q,n-p} \geq F\left(\mathbf{y}\right)\right),$$

jossa $F_{q,n-p}$ on $F_{q,n-p}$-jakaumaa noudattava satunnaismuuttuja.

F-testi yleiselle lineaariselle hypoteesille

- F-testisuure

$$F = (\mathbf{A}\hat{\beta} - \mathbf{c})'(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\beta} - \mathbf{c})/qS^2 \overset{H}{\sim} F_{q,n-p},$$

jossa $S^2 = (n-p)^{-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) = \mathrm{S}(\hat{\beta})/(n-p)$ ja $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ voidaan kirjoittaa myös

$$F = \frac{(\mathrm{S}(\hat{\beta}_H) - \mathrm{S}(\hat{\beta}))/q}{\mathrm{S}(\hat{\beta})/(n-p)} \overset{H}{\sim} F_{q,n-p}.$$

- Siis, F-testi asettaa nollahypoteesin epäilyksen alaiseksi, jos rajoitettu residuaalineliösumma $\mathrm{S}(\hat{\beta}_H) = (\mathbf{Y} - \mathbf{X}\hat{\beta}_H)'(\mathbf{Y} - \mathbf{X}\hat{\beta}_H)$ on "kohtuuttoman paljon" suurempi kuin vapaa residuaalineliösumma $\mathrm{S}(\hat{\beta}) = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$.

F-testi yleiselle lineaariselle hypoteesille

- Suoraviivaisella laskulla nähdään, että uskottavuusosamäärän testisuure on

$$
\begin{aligned}
r(\mathbf{y}) &= 2\left[l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2; \mathbf{y}) - l(\hat{\boldsymbol{\beta}}_H, \hat{\sigma}_H^2; \mathbf{y})\right] \\
&= 2\left[\frac{n}{2}\log \hat{\sigma}_H^2 + \frac{n}{2} - \frac{n}{2}\log \hat{\sigma}^2 - \frac{n}{2}\right] \\
&= n \log\left(\frac{q}{n-p}F + 1\right).
\end{aligned}
$$

- Koska $r(\mathbf{y})$ on monotonisesti kasvava funktio F-testisuureesta, määrittelevät molemmat testisuureet saman testin eli samat P-arvot.

- Voidaan osoittaa, että myös Raon pistemäärätestisuure ja F-testisuure määrittelevät saman testin.

F-testin erikoistapauksia

- Sovelletaan F-testiä usean selittäjän regressiomalliin

$$Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, ..., n,$$

johtamalla ns. *yleistesti*, joka koskee nollahypoteesia

$$H : \beta_2 = \cdots = \beta_p = 0 \quad \text{eli} \quad [\mathbf{0} \;\; \mathbf{I}_{p-1}]\,\boldsymbol{\beta} = \mathbf{0}.$$

Nollahypoteesin voimassa ollessa mallilla ei ole mitään virkaa.

- Tässä tapauksessa kannattaa käyttää testiruureen residuaalineliösummamuotoa $(q = p - 1)$

$$F = \frac{(\mathrm{S}(\hat{\boldsymbol{\beta}}_H) - \mathrm{S}(\hat{\boldsymbol{\beta}}))/(p-1)}{\mathrm{S}(\hat{\boldsymbol{\beta}})/(n-p)} \overset{H}{\sim} F_{p-1,n-p}.$$

- Testisuure voidaan kirjoittaa (välivaiheiden jälkeen)

$$F = \frac{(\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\bar{Y}^2)/(p-1)}{(\mathbf{Y}'\mathbf{Y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y})/(n-p)} \overset{H}{\sim} F_{p-1,n-p}.$$

F-testin erikoistapauksia

- Sovelletaan F-testiä malliin ($x_{i1} = 1 \ \forall \ i$ ei välttämätöntä)

$$Y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, ..., n,$$

johtamalla ns. *yksittäistä selittäjää koskeva testi, jossa* nollahypot. on

$$H_j : \beta_j = 0, \quad 1 \leq j \leq p \quad \text{eli} \quad \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \boldsymbol{\beta} = 0.$$

Nollahypoteesi merkitsee, että muiden selittäjien ollessa mallissa tutkitaan onko selittäjä x_j tarpeen.

- Tässä tapauksessa kannattaa käyttää testiruureen "alkuperäistä" muotoa $(q = 1)$

$$F = (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})'(\mathbf{A}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{A}')^{-1}(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})/S^2 \overset{H}{\sim} F_{1,n-p}.$$

- Testisuureeksi saadaan (välivaiheiden jälkeen)

$$F = \hat{\beta}_j^2 / S^2 m^{jj} \overset{H}{\sim} F_{1,n-p}, \quad \left(\mathbf{X}'\mathbf{X}\right)^{-1} = \mathbf{M}^{-1} = \left[m^{ab}\right], \quad a, b = 1, ..., p.$$

F-testin erikoistapauksia

- Vaihtoehto testisuureelle

$$F = \hat{\beta}_j^2 / S^2 m^{jj} \overset{H}{\sim} F_{1,n-p}, \quad (\mathbf{X}'\mathbf{X})^{-1} = \mathbf{M}^{-1} = \left[m^{ab} \right], \quad a, b = 1, ..., p.$$

- Jos $Z \sim \mathrm{N}(0,1)$ ja $Z \perp\!\!\!\perp \chi_k^2$, niin tunnetusti $T_k = Z / \sqrt{\frac{1}{k}\chi_k^2} \sim t_k$ (t-jakauma vapausastein k) ja $T_k^2 \sim F_{1,k}$. Nollahypoteesia voidaan siten testata myös testisuureella

$$T = t(\mathbf{Y}) = \hat{\beta}_j / S\sqrt{m^{jj}} \overset{H}{\sim} t_{n-p}.$$

- Tämän testin P-arvot:

$$\mathrm{P} = \mathrm{P}_{H_j}(|t(\mathbf{Y})| \geq |t(\mathbf{y})|) = \mathrm{P}(|T_{n-p}| \geq |t(\mathbf{y})|), \quad T_{n-p} \sim t_{n-p}$$

- Tässä vaihtoehto on kaksisuuntainen eli $\beta_j \neq 0$. Yksisuuntaisen vaihtoehdon $\beta_j > 0$ $[$tai $\beta_j < 0]$ tapauksessa P-arvot lasketaan kaavalla $\mathrm{P}(T_{n-p} \geq t(\mathbf{y}))$ $[$tai $\mathrm{P}(T_{n-p} \leq t(\mathbf{y}))]$.

F-testin erikoistapauksia

Käytännössä käytetään yleensä t-testisuuretta ja tuloksista voidaan laatia seuraavanlainen taulukko ($\hat{\beta}_j$:n keskivirhe = $\hat{\beta}_j$:n estimoitu hajonta)

Parametri	Estimaatti	Keskivirhe	t-suhde
β_1	$\hat{\beta}_1$	$s\sqrt{m^{11}}$	$\hat{\beta}_1 / s\sqrt{m^{11}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
β_p	$\hat{\beta}_p$	$s\sqrt{m^{pp}}$	$\hat{\beta}_p / s\sqrt{m^{pp}}$

tai yhtälö ($s.e.(\hat{\beta}_j)$ on $\hat{\beta}_j$:n keskivirhe)

$$y_i = \underset{\left(s.e.(\hat{\beta}_1)\right)}{\hat{\beta}_1} x_{i1} + \cdots + \underset{\left(s.e.(\hat{\beta}_p)\right)}{\hat{\beta}_p} x_{ip} + \hat{\varepsilon}_i, \quad s^2 = \ldots,$$

- Eri hypoteeseihin $H_j : \beta_j = 0$ liittyvät t-testit eivät ole riippumattomia, joten "yhdistetyn" testin P-arvon laskeminen on hankalaa.
- Yksittäisiä hypoteeseja H_j ei ilmeisestikään kannata testata, ellei yleishypoteesia $H : \beta_2 = \cdots = \beta_p = 0$ (jossa ei yo hankaluutta) ole ole hylätty.
- Mahdollista, että yleishypoteesi hylätään ja kaikki yksittäiset hypoteesit jäävät voimaan.
- Jos H_j:ta ei hylätä, on j. selittäjä tarpeeton. Selittäjien poistaminen ei ole kuitenkaan yksiselitteistä. Esimerkiksi lopputulos voi riippua yksittäisten hypoteesien testausjärjestyksestä

Esimerkki

Äitien ja tyttärien pituudet (tuumissa): $n = 30$, punainen suora perustuu PNS-estimointiin $y = \underset{(9.07)}{17.29} + \underset{(0.14)}{0.75}\, x + \hat{\varepsilon}_t$, $s^2 = 6.43$, $R^2 = 0.49$.

Sininen katkoviiva $y = 1.53 + x$. $H_0 : \beta_2 = 1$, $H_1 : \beta_2 \neq 0$.

$t(\mathbf{y}) = (0.75 - 1)/0.14 = -1.79$, $\mathrm{P} \approx 0.085$.