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-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 residuaali-neliösummamuotoa $(q = p - 1)$

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

- Testisuureeksi saadaan (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, mutta mahdollista)

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

johtamalla *yksittäistä selittäjää koskeva testi, jossa* nollahypoteesina on

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

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)

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

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)

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

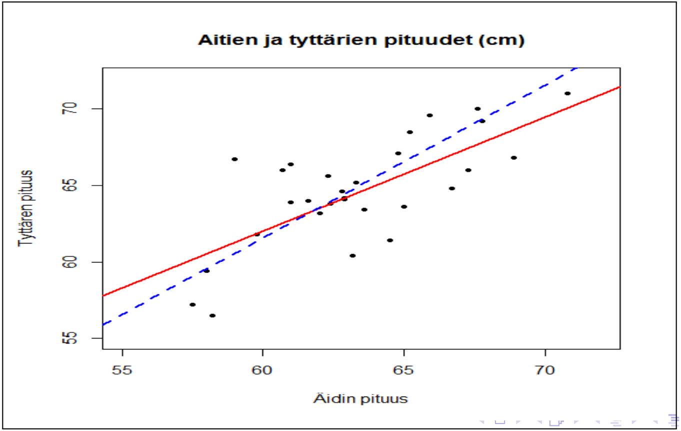
- Eri hypoteeseihin $H_j : \beta_j = 0$ liittyvät t-testit eivät ole riippumattomia, joten "yhdistetyn" testin P-arvon laskeminen on hankalaa.
- Yksittäisiä hypoteeseja H_j ei ilmeisestikään kannata testata, ellei yleishypoteesia $H : \beta_2 = \cdots = \beta_p = 0$ (jossa ei yo hankaluutta) ole ole 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 $y = 1.53 + x$. $H_0 : \beta_1 = 1$, $H_1 : \beta_1 \neq 0$.

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



Luottamusvälit

- Oletetaan malli $\mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\right)$ $\left(\mathrm{r}\left(\mathbf{X}\right) = p\right)$.
- Tyypillisin esimerkki luottamusväleistä liittyy parametrivektorin $\boldsymbol{\beta}$ yksittäisiin komponenteihin β_j $\left(1 \leq j \leq p\right)$.
- Tämä on erikoistapaus parametrivektorin $\boldsymbol{\beta}$ lineaarikombinaatiosta $\mathbf{a}'\boldsymbol{\beta} = a_1\beta_1 + \cdots + a_p\beta_p$, jossa $\mathbf{a} \neq \mathbf{0}$ $\left(p \times 1\right)$ on tunnettu
  - $\mathbf{a}' = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \Rightarrow \mathbf{a}'\boldsymbol{\beta} = \beta_j$
- Muita tyypillisiä erikoistapauksia lineaarikombinaatiosta $\mathbf{a}'\boldsymbol{\beta}$:
  - $\mathbf{a}' = \begin{bmatrix} x_1^* & \cdots & x_p^* \end{bmatrix} \Rightarrow \mathbf{a}'\boldsymbol{\beta} = \beta_1 x_1^* + \cdots + \beta_p x_p^* = Y$:n odotusarvo, kun selittäville muuttujille annetaan arvot $x_1^*, \ldots, x_p^*$
  - odotusarvojen erotus $\mu_1 - \mu_2$ kahden riippumattoman normaalisen otoksen mallissa tai
  - vastaavat erotukset $\mu_j - \mu_k$ $\left(j \neq k\right)$ eli ns. *kontrastit* yleisemmässä yksisuuntaisessa varianssianalyysimallissa
- Kuten tilastollisen päättelyn kurssilla todetaan, voidaan luottamusvälejä muodostaa testien avulla.

Luottamusväli lineaarikombinaatiolle

- T-testi nollahypoteesille

$$H : \mathbf{a}'\boldsymbol{\beta} = \mathbf{a}'\boldsymbol{\beta}_0, \quad \boldsymbol{\beta}_0 \, (p \times 1) \text{ tunnettu}$$

saadaan kuten jaksossa 3.2 tapauksessa $\mathbf{a}' = [0 \cdots 0 \; 1 \; 0 \cdots 0]$:

$$T(\mathbf{Y}) = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}_0}{S\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}} \overset{H}{\sim} t_{n-p}.$$

- Vaihtoehtoa $\mathbf{a}'\boldsymbol{\beta} \neq \mathbf{a}'\boldsymbol{\beta}_0$ vastaava hylkäysalue merk. tasolla α:

$$C_\alpha\left(\mathbf{a}'\boldsymbol{\beta}_0\right) = \left\{\mathbf{y} : |T(\mathbf{y})| \geq t_{n-p}(\alpha/2)\right\}, \quad \mathrm{P}\left(|T_{n-p}| \geq t_{n-p}(\tfrac{\alpha}{2})\right) = \alpha$$

- Vastaava hyväksymisalue muodostuu aineistoista, joilla

$$-t_{n-p}(\alpha/2) < T(\mathbf{y}) < t_{n-p}(\alpha/2)$$

tai yhtäpitävästi (käyttäen $T(\mathbf{y})$:n lauseketta)

$$\mathbf{a}'\hat{\boldsymbol{\beta}} - t_{n-p}(\tfrac{\alpha}{2}) s\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}} < \mathbf{a}'\boldsymbol{\beta}_0 < \mathbf{a}'\hat{\boldsymbol{\beta}} + t_{n-p}(\tfrac{\alpha}{2}) s\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}$$

Luottamusväli lineaarikombinaatiolle

- Todennäköisyys, että (nyt satunnaisiksi tulkitut) epäyhtälöt

  $$\mathbf{a}'\hat{\boldsymbol{\beta}} - t_{n-p}(\tfrac{\alpha}{2})S\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}} < \mathbf{a}'\boldsymbol{\beta}_0 < \mathbf{a}'\hat{\boldsymbol{\beta}} + t_{n-p}(\tfrac{\alpha}{2})S\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}$$

  pätevät kaikilla $\boldsymbol{\beta}_0$ ja σ^2 on $1-\alpha$.
- Siis, lineaarikombinaation $\mathbf{a}'\boldsymbol{\beta}$ luottamusväli luottamustasolla $1-\alpha$ on

  $$\mathbf{a}'\hat{\boldsymbol{\beta}} \pm t_{n-p}\left(\alpha/2\right)s\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}.$$

  - Koska $\mathrm{Var}(\mathbf{a}'\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}$ (Lause 2.1(i)), on $s\sqrt{\mathbf{a}'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{a}}$ yllä estimaattorin $\mathbf{a}'\hat{\boldsymbol{\beta}}$ keskivirhe.
- Jos $\mathbf{a}'\boldsymbol{\beta} = \beta_j$, saadaan luottamusväli

  $$\hat{\beta}_j \pm t_{n-p}\left(\alpha/2\right)s\sqrt{m^{jj}},$$

  jossa $m^{jj} = \left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right]_{jj}$ on matriisin $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ j. diagonaalialkio.

Luottamusjoukot

- Johdetaan luottamusjoukko parametrivektorille $\boldsymbol{\beta}$ kokonaisuudessaan.
- Tarkastellaan nollahypoteesia $H : \boldsymbol{\beta} = \boldsymbol{\beta}_0$, jossa $\boldsymbol{\beta}_0$ $(p \times 1)$ annettu.
- F-testisuure ($\mathbf{A} = \mathbf{I}_p$ ja $\mathbf{c} = \boldsymbol{\beta}_0$)

$$F = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) / pS^2 \overset{H}{\sim} F_{p,n-p}.$$

- Jos $F_{p,n-p}(\alpha)$ toteuttaa $\mathrm{P}\left(F_{p,n-p} \geq F_{p,n-p}(\alpha)\right) = \alpha$, kaikilla $\boldsymbol{\beta}_0$ ja σ^2 pätee

$$\mathrm{P}_{\boldsymbol{\beta}_0,\sigma^2}\left((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) / pS^2 < F_{p,n-p}(\alpha)\right) = 1 - \alpha.$$

- $\boldsymbol{\beta}$:n luottamusjoukko luottamustasolla $1 - \alpha$ on siten

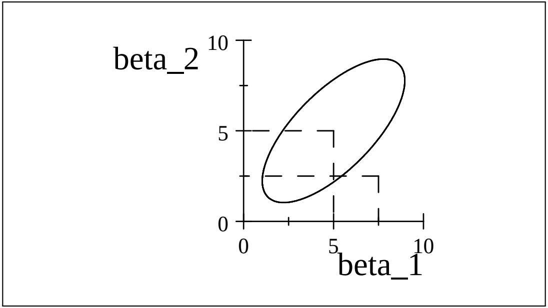
$$\left\{\boldsymbol{\beta} \in \mathbb{R}^p : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) / ps^2 < F_{p,n-p}(\alpha)\right\}.$$

Luottamusjoukot

- $\boldsymbol{\beta}$:n luottamusjoukko luottamustasolla $1 - \alpha$ on

$$\left\{ \boldsymbol{\beta} \in \mathbb{R}^p : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \mathbf{X}' \mathbf{X} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) / ps^2 < F_{p,n-p}(\alpha) \right\}$$

- Sen rajoittama pinta on $\mathbb{R}^p$:n ellipsoidi, jonka keskipiste on $\hat{\boldsymbol{\beta}}$ ja muodon määrää matriisi $\mathbf{X}'\mathbf{X}$.



Kuva 4.1. $\boldsymbol{\beta} = (\beta_1, \beta_2)$:n luottamusellipsi. PNS-estimaatti $\hat{\boldsymbol{\beta}} = (5, 5)$.