- **Luennoija**: Pentti Saikkonen
- **Tyyppi ja laajuus**: Aineopinnot, 5 op
- **Luentoajat**: IV periodi, to 14-16, D123 ja pe 12-14, C124
- **Suoritustapa:** Kurssikoe viikolla 9.5.-13.5. tai yleistentti 24.5.
- **Laskuharjoitukset:** Petri Aaltonen
  - To 12-14, B322 (28.4. C129)
  - Pe 10-12, B322
  - Lisäpisteitä harjoitustehtävien ratkaisemisesta saa 1, 2, 3 tai 4, jos ratkaistujen tehtävien määrä on vastaavasti 20, 40, 60 tai 80 prosenttia.

- **Esitietovaatimukset**
  - Tilastollisen päättelyn kurssi sekä sen vaatimat esitiedot (Todennäköisyyslaskennan kurssi ja perusvalmiudet yhden ja useamman muuttujan differentiaali- ja integraalilaskennassa).
  - Lineaarialgebran ja matriisilaskennan tuntemus on lisäksi **välttämätöntä** (esim. kurssit Lineaarialgebra ja matriisilaskenta I ja II, ja mielellään myös Matriisilaskentaa tilastotieteilijöille)
- **Sisältö:** Kurssilla käydään läpi lineaaristen mallien keskeinen teoria eli parametrien estimointi ja testaus sekä luottamusvälien muodostaminen ja niihin liittyvä jakaumateoria. Lisäksi tarkastellaan varianssianalyysia.

- **Kurssimateriaali**: Luentomoniste (liitteiden materiaali oletetaan pääosin tunnetuksi). Oheislukemistona voi käyttää seuraavia teoksia.
- Matriisialgebra:
  - Abadir, K. ja J. Magnus (2005): Matrix Algebra.
  - Searle, S.R. (1982): Matrix Algebra Useful for Statistics.
  - Schott, J. R. (2005): Matrix Analysis for Statistics.
- Lineaarisen mallin teoria:
  - Seber, G.A.F. ja A.J. Lee (2003): Linear Regression Analysis, 2.laitos.

Lineaarisen mallin määrittely

Yksinkertainen esimerkki

- Peltotilkut $i = 1, ..., n$
- Lannoitemäärät $x_1, ..., x_n$
- Satomäärät $y_1, ..., y_n$
- Mikä on lannoituksen vaikutus odotettavissa olevaan satomäärään?
- Oletetaan satomääriä vastaavista sm:sta
  - $Y_1, ..., Y_n \perp\!\!\!\perp$ eli riippumattomuus
  - $\mu_i := \mathrm{E}(Y_i) = \beta_1 + \beta_2 x_i$ odotusarvon lineaarisuus
  - $\mathrm{Var}(Y_i) = \sigma^2 \ \forall i$ varianssin koko ei riipu lannoitemäärästä
- Tilanne voidaan kuvata yhtälöllä

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \ \ i = 1, ..., n,$$

jossa sm:t $\varepsilon_1, ..., \varepsilon_n$ ovat ei-havaittavia, $\perp\!\!\!\perp$ ja toteuttavat

$$\mathrm{E}(\varepsilon_i) = 0, \ \mathrm{Var}(\varepsilon_i) = \sigma^2 \ \forall i$$

Lineaarisen mallin määrittely

Yksinkertainen esimerkki

- Kuvattaessa lannoituksen vaikutusta odotettavissa olevaan satomäärään päädyttiin siis yhtälöön

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \;\; i = 1, \ldots, n,$$

jossa sm:t $\varepsilon_1, \ldots, \varepsilon_n$ ovat ei-havaittavia, $\perp\!\!\!\perp$ ja toteuttavat

$$\mathrm{E}\left(\varepsilon_i\right) = 0, \;\; \mathrm{Var}\left(\varepsilon_i\right) = \sigma^2 \;\; \forall i$$

- Sm:t $\varepsilon_1, \ldots, \varepsilon_n$ voidaan tulkita havaituissa satomäärissä ilmeneväksi 'puhtaaksi satunnaisvaihteluksi' tai selitysvirheeksi, joka ei selity lannoitemäärällä.
- Edellä esitettyä yhtälöä kutsutaan *yhden selittävän muuttujan lineaariseksi regressiomalliksi*.
- Tiukasti tilastollisen päättelyn näkökulmasta, ei kysymyksessä ole vielä kuitenkaan tilastollinen malli, jollainen vaatii havaintojen yhteistodennäköisyysjakauman ja parametriavaruuden spesifioinnin.

Klassinen lineaarinen malli perustuu olettaa normaalijakaumaan ja tilastollinen malli saadaan olettamalla

$$Y_1, \ldots, Y_n \perp\!\!\!\perp \quad Y_i \sim \mathrm{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right), \quad \beta_1, \beta_2 \in \mathbb{R}, \quad \sigma^2 > 0$$

tai yhtäpitävästi

$$Y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \; i = 1, \ldots, n, \quad \beta_1, \beta_2 \in \mathbb{R}, \quad \sigma^2 > 0$$

$$\varepsilon_1, \ldots, \varepsilon_n \perp\!\!\!\perp, \quad \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right)$$

Tarkasteltavassa esimerkkitapauksessa oletus $\beta_1 > 0$ tuntuisi järkevältä, mutta johtaa mutkikkaampaa malliin, joka ei sisälly klassiseen lineaariseen malliin.

- Yhden selittävän muuttujan lineaarinen regressiomallissa

$$Y_1, \dots, Y_n \perp\!\!\!\perp \quad Y_i \sim \mathrm{N}\left(\beta_1 + \beta_2 x_i, \sigma^2\right), \quad \beta_1, \beta_2 \in \mathbb{R}, \quad \sigma^2 > 0.$$

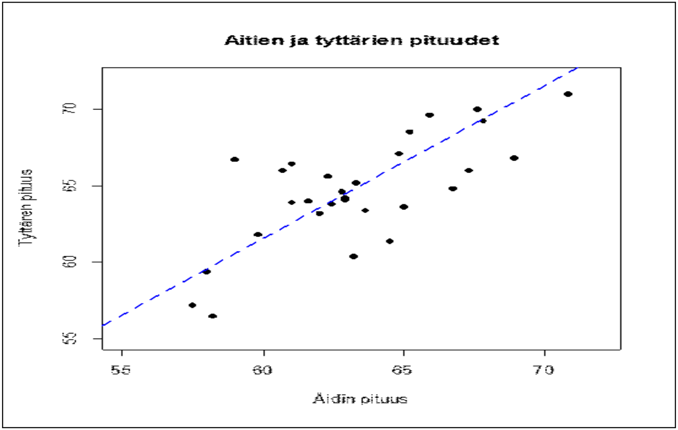
selittävien muuttujien havaintoja $x_1, \dots, x_n$ ei tulkita satunnaisiksi (loogista tarkasteltavassa tapauksessa).

- Jos selittävät muuttujat olisivat satunnaisia, täytyisi tilastollinen malli periaatteessa laajentaa koskemaan sv:ta $\left[Y_i \quad X_i\right]'$, $i = 1, \dots, n$.
- Usein selitettäviä muuttujia $Y_1, \dots, Y_n$ voidaan tarkastella ehdollisesti ehdolla $(X_1, \dots, X_n) = (x_1, \dots, x_n)$, jolloin edellä esitetty malli soveltuu.
- Usein tarvitaan useita selittäviä muuttujia

Lineaarisen mallin määrittely

Yksinkertainen esimerkki

Äitien ja tyttärien pituudet (tuumissa): $n = 30$, katkoviiva 45 asteen suora



Yleinen lineaarinen malli

Tausta: Aineiston muuttujista yksi on luonteeltaan *selitettävä* ja loput sen vaihtelua *selittäviä* muuttujia eli kaaviona

Havainto yksikkö	Selitettävä m:t; y	Selittävät m:t; $x_1, \ldots, x_p$
1	y_1	$x_{11}, \ldots, x_{1p}$
$\vdots$	$\vdots$	$\vdots$
n	y_n	$x_{n1}, \ldots, x_{np}$

Lineaarisessa mallissa selittävän muuttujan vaikutus selitettävään muuttujaan oletetaan (tietyssä mielessä) *lineaariseksi*.

Yleinen lineaarinen malli

- Lineaarisen mallin määritelmä voidaan perustaa yhtälöön

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

  jossa Y_i on havaittava sm, $x_{i1}, ..., x_{ip}$ ovat havaittavia ja ei-satunnaisia (eli kiinteitä), ε_i on ei-havaittava sm ja $\beta_1, ..., \beta_p$ ovat tuntemattomia parametreja.
- ε_i on ns. *virhe(termi)* eli se osa Y_i:stä, jota mallin systemaattinen osa tai rakenne $\beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ ei kykyne selittämään
- Lineaarisuus oletetaan parametrien $\beta_1, ..., \beta_p$ suhteen ja virhetermi lisätään systemaattiseen osaan additiivisesti.
  - Esimerkiksi

$$Y_i = \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i, \quad i = 1, ..., n,$$

    on vaaditulla tavalla lineaarinen

Yleinen lineaarinen malli

- Lineaariseen malliin voidaan päätyä lähtemällä oletuksesta

$$Y_1, \ldots, Y_n \perp\!\!\!\perp, \quad Y_i \sim \mathrm{N}\left(\mu_i, \sigma^2\right)$$

  ja olettamalla odotusarvoille $\boldsymbol{\mu}_i$ (esim. taustatiedon perusteella) rakenne $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$, jossa $\mathbf{x}_i' = \begin{bmatrix} x_{i1} & \cdots & x_{ip} \end{bmatrix}$ ja $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 & \cdots & \beta_p \end{bmatrix}'$.
- Miksi vaihtoehtoinen muotoilu $(i = 1, \ldots, n)$

$$Y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_1, \ldots, \varepsilon_n \perp\!\!\!\perp, \quad \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right)$$

  ja virhetermi ε_i?
- Yksi syy on, että tämä on toisinaan luonteva muotoilu.
- Toinen, että käsitteet *residuaali* $\hat{\boldsymbol{\varepsilon}}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ ja *sovite* $\hat{\mu}_i = \mathbf{x}_i'\hat{\boldsymbol{\beta}} = \hat{y}_i$ ovat hyödyllisiä ($\hat{\boldsymbol{\beta}}$ on parametrin $\boldsymbol{\beta}$ estimaatti).
- Residuaali on teoreettisen virheen $\varepsilon_i = Y_i - \mathbf{x}_i'\boldsymbol{\beta}$ empiirinen vastine ja sovite on mallin systemaattisen osan $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$ empiirinen vastine.

Yleinen lineaarinen malli

- Residuaalit $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}$ sisältävät ilmeisestikin informaatiota virheiden $\varepsilon_i = Y_i - \mathbf{x}_i'\boldsymbol{\beta}$ varianssista eli parametrista $\sigma^2 = \mathrm{Var}\left(\varepsilon_i\right) = \mathrm{Var}\left(Y_i\right)$.
- Residuaalien avulla voidaan myös tutkia mallin oletusten paikkansapitävyyttä yleensä helpommin kuin alkuperäisteen havaintojen $y_1, ..., y_n$ avulla.
- Tutkittavia oletuksia:
  - Onko $\mathrm{Var}\left(Y_i\right) = \mathrm{Var}\left(\varepsilon_i\right)$ vakio kuten oletetaan?
  - Päteekö oletettu riippumattomuus $Y_1, ..., Y_n \perp\!\!\!\perp \Leftrightarrow \varepsilon_1, ..., \varepsilon_n \perp\!\!\!\perp$?
  - Päteekö oletettu normaalisuus $Y_i \sim \mathrm{N}\left(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2\right) \Leftrightarrow \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right)$?

Yleinen lineaarinen malli

- Mallin matriisiesitys

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ & \vdots & \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

eli

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0,$$

jossa $\varepsilon_1, ..., \varepsilon_n \perp\!\!\!\perp, \ \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right) \ \Leftrightarrow \ \boldsymbol{\varepsilon} \sim \mathrm{N}\left(0, \sigma^2 \mathbf{I}_n\right)$.

- Yhtäpitävä muotoilu

$$\mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0.$$

- Lisäoletus: Matriisi $\mathbf{X}$ $(n \times p)$ täyttä sarakeastetta eli $\mathrm{r}\left(\mathbf{X}\right) = p$.

- Riippumaton otos normaalijakaumasta (vrt. tilastollinen päättely).

$$Y_1, \ldots, Y_n \;\perp\!\!\!\perp\;, \quad Y_i \sim \mathrm{N}\left(\mu, \sigma^2\right)$$

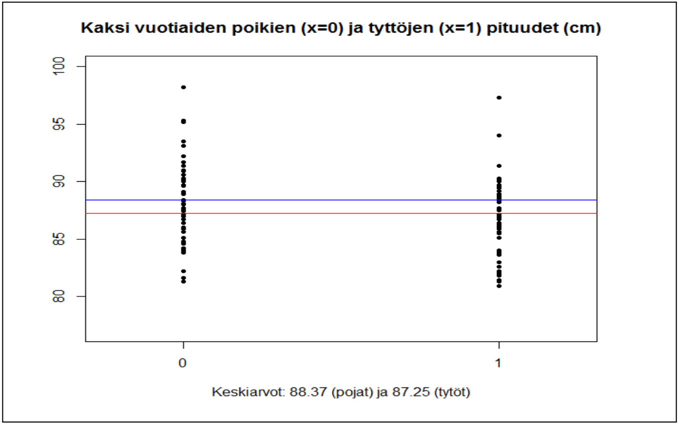
- $$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \mu + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

  eli

$$\mathbf{Y} = \mathbf{1}_n \mu + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right) \quad \Leftrightarrow \quad \mathbf{Y} \sim \mathrm{N}\left(\mathbf{1}_n \mu, \sigma^2 \mathbf{I}_n\right)$$

  Matriiseiksi $\mathbf{X}$ tulee siten vektori $\mathbf{1}_n = [1 \ \cdots \ 1]' \ \ (n \times 1)$.
- Kiinnostava hypoteesi $H_0 : \mu = \mu_0$

Kaksivuotiaiden poikien ($x = 0$) ja tyttöjen ($x = 1$) pituudet (cm)
$n_1 = 66$, $n_2 = 70$. Kysymys: Eroavatko tyttöjen ja poikien keskipituudet?



- Kahden odotusarvoltaan (mahdollisesti) poikkeavan riippumattoman normaalisen otoksen malli eli

$$Y_1, \ldots, Y_n \perp\!\!\!\perp, \quad Y_i \sim \begin{cases} \mathrm{N}\left(\mu_1, \sigma^2\right), & \text{kun } i = 1, \ldots, n_1 \\ \mathrm{N}\left(\mu_2, \sigma^2\right), & \text{kun } i = n_1 + 1, \ldots, n_1 + n_2 = n. \end{cases}$$

- $$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_{1+1}} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_1} \\ \varepsilon_{n_{1+1}} \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

  eli

$$\mathbf{Y} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{n_2} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}_n\right); \quad H_0 : \mu_1 = \mu_2$$

Yleinen lineaarinen malli

Lineaarisen mallin erikoistapauksia

- *Yksisuuntainen varianssianalyysimalli*. Havaintoina on p riippumatonta otosta jakaumista $\mathrm{N}(\mu_j, \sigma^2)$ $(j = 1, ..., p)$.
- Esimerkiksi vehnälajikkeiden $A_1, ..., A_p$ satoisuuden tutkiminen, kun niitä viljellään samoissa olosuhteissa.
- Kiinnostuksen kohteena on odotusarvoissa $\mu_1, .., \mu_p$ mahdollisesti ilmenevät erot; esim. $H_0 : \mu_1 = \cdots = \mu_p$
- *Kaksisuuntainen varianssianalyysimalli*. Esimerkiksi vehnälajikkeita A_1 ja A_2 lannoitetaan kahta eri lannoitetta B_1 ja B_2 käyttäen, jolloin havainnot peräisin neljästä ryhmästä.
- Mallin avulla voidaan tutkia onko satomäärissä eroja eri ryhmien välillä ja johtuvatko mahdolliset erot vehnälajikkeesta, lannoitteesta vai niiden yhteisvaikutuksesta.
- Näissä malleissa selittävät muuttujat ovat ryhmää osoittavia indikaattoreita.

- *Usean selittäjän lineaarinen regressiomalli*

$$Y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, ..., n, \quad \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0.$$

  β_1 ns. *(regressio)vakio* ja β_j $(j = 2, .., p)$ ns. *regressiokerroin*, joka kuvaa paljonko selitettävä muuttuja (tai sen odotusarvo) muuttuu, kun j:nnen selittäjän arvo muuttuu yhden yksikön ja muiden selittäjien arvot pysyvät ennallaan.

- Saadaan yleisestä mallista valitsemalla $x_{i1} = 1$, joten

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & \cdots & x_{1p} \\ & & \vdots & \\ 1 & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

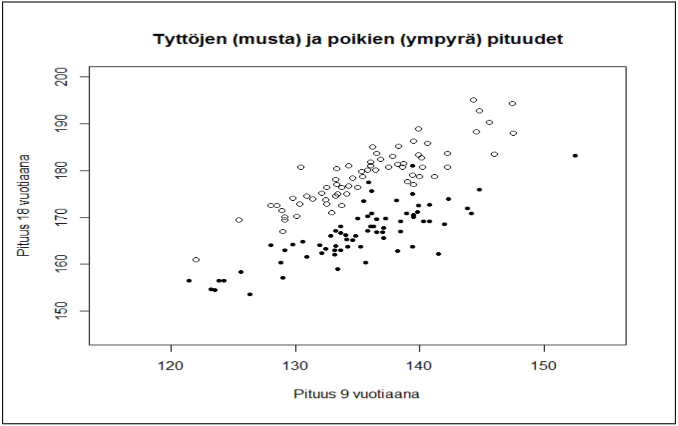
- Samassa mallissa voi olla sekä kvantitatiivisia selittäjiä että kvalitatiivisia ryhmää osoittavia indikaattoreita.
- Tutkitaan esimerkiksi vehnälajikkeiden A_1 ja A_2 satoisuutta, kun molempia lannoitetaan samaa lannoitetta käyttäen:

$$Y_1, \ldots, Y_n \perp\!\!\!\perp, \quad Y_i \sim \begin{cases} \mathrm{N}\left(\beta_1 + \beta_3 x_i, \sigma^2\right), & \text{kun } i = 1, \ldots, n_1 \\ \mathrm{N}\left(\beta_2 + \beta_3 x_i, \sigma^2\right), & \text{kun } i = n_1 + 1, \ldots, n_1 + n_2 \end{cases}$$

eli

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_{n_1} \\ Y_{n_1+1} \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_{n_1} \\ 0 & 1 & x_{n_1+1} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n_1} \\ \varepsilon_{n_1+1} \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Tyttöjen (musta) ja poikien (ympyrä) pituudet 9 ja 18 v. ikäisinä (cm) $n_1 = 66$, $n_2 = 70$. Kysymys: Onko tyttöjen ja poikien kasvuvauhdissa eroja?



- Joskus epälineaarinen malli voi olla *linearisoituva*:

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

  jossa muuttujat oletetaan positiivisiksi.

- Ottamalla (luonnollinen) logaritmi puolittain päädytään yhtälöön

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

  eli (merkintöjä vaille) usean selittäjän lineaarinen regressiomalli.

- Tällaiset ns. multiplikatiiviset mallit ovat tavallisia taloudellisissa sovelluksissa mallinnettaessa esimerkiksi jonkin tuotteen kysyntää.

Yleinen lineaarinen malli

Yhteenveto lineaarisen mallin määrittelystä

- Lineaarisen mallin määritelmä voidaan perustaa yhtälöön

$$\begin{aligned} Y_i &= \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \\ &= \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \end{aligned}$$

  jossa Y_i on havaittava sm, $\mathbf{x}_i = \begin{bmatrix} x_{i1} & \cdots & x_{ip} \end{bmatrix}'$ on havaittava ja ei-satunnainen (eli kiinteitä), ε_i on ei-havaittava sm ja $\beta_1, \ldots, \beta_p$ ovat tuntemattomia parametreja.

- ε_i on ns. *virhe(termi)* eli se osa Y_i stä, jota mallin systemaattinen osa tai rakenne $\beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ ei kykyne selittämään.

- Tilastollista mallia varten oletetaan

$$\varepsilon_1, \ldots, \varepsilon_n \;\underline{\perp\!\!\!\perp}\;, \quad \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right) \Leftrightarrow Y_1, \ldots, Y_n \;\underline{\perp\!\!\!\perp}\; Y_i \sim \mathrm{N}\left(\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2\right)$$

Yleinen lineaarinen malli

Yhteenveto lineaarisen mallin määrittelystä

- Mallin matriisiesitys

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ & \vdots & \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

  eli

$$\mathbf{Y} = \mathbf{X}\beta + \boldsymbol{\varepsilon},$$

  jossa $\varepsilon_1, ..., \varepsilon_n \perp\!\!\!\perp, \ \varepsilon_i \sim \mathrm{N}\left(0, \sigma^2\right) \ \Leftrightarrow \ \boldsymbol{\varepsilon} \sim \mathrm{N}\left(0, \sigma^2 \mathbf{I}_n\right).$

- Nyt malli voidaan määritellä lyhyesti kirjoittamalla

$$\mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\beta, \sigma^2 \mathbf{I}_n\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0.$$

- Lisäoletus: Matriisi $\mathbf{X}$ $(n \times p)$ täyttä sarakeastetta eli $\mathrm{r}\left(\mathbf{X}\right) = p$.

- Lineaarinen malli voidaan siis määritellä kahdella yhtäpitävällä tavalla:

$$(1) \qquad \mathbf{Y} \sim \mathrm{N}\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0.$$

tai

$$(2) \qquad Y_1, ..., Y_n \perp\!\!\!\perp Y_i \sim \mathrm{N}\left(\mathbf{x}_i'\boldsymbol{\beta}, \sigma^2\right), \quad \boldsymbol{\beta} \in \mathbb{R}^p, \ \sigma^2 > 0$$

- Uskottavuusfunktio voidaan johtaa kumpaa tahansa käyttäen.
- Lisäoletus, että matriisi $\mathbf{X}$ $(n \times p)$ on täyttä sarakeastetta eli $\mathrm{r}\left(\mathbf{X}\right) = p$ ei vaikuta uskottavuusfunktion johtoon.