Yksisuuntainen varianssianalyysi

- Jos nollahypoteesi hylätään, on kiinnostavaa tutkia hypoteeseja $\mu_{j_1} - \mu_{j_2} = 0$ tai yleisemmin ns. *kontrasteja* $\mathbf{a}'\boldsymbol{\mu} = 0$, jossa $\mathbf{a}' = [a_1 \ \cdots \ a_p]$ toteuttaa $a_1 + \cdots + a_p = 0$.
- Koska $\mathbf{X}'\mathbf{X} = \text{diag}\,[n_1 \ \cdots \ n_p]$, pätee (HT 3.2, Lause 2.1(i))

$$\hat{\boldsymbol{\mu}} = [\bar{Y}_1 \ \cdots \ \bar{Y}_p]' \sim \mathrm{N}\left(\boldsymbol{\mu}, \sigma^2 \text{diag}\left[\frac{1}{n_1} \ \cdots \ \frac{1}{n_p}\right]\right).$$

- Nollahypoteesin $\mathbf{a}'\boldsymbol{\mu} = 0$ voimassa ollessa pätee siten (Liite A.2.4)

$$\mathbf{a}'\hat{\boldsymbol{\mu}} \sim \mathrm{N}\left(0, \sigma^2 \sum_{j=1}^{p} a_j^2 / n_j\right).$$

- Tämä johtaa F-testisuureeseen

$$F = \frac{\left(\sum_{j=1}^{p} a_j \bar{Y}_j\right)^2}{S^2 \sum_{j=1}^{p} a_j^2 / n_j} \overset{H}{\sim} F_{1,n-p}, \quad S^2 = \frac{1}{n-p} \sum_{j=1}^{p} \sum_{i=1}^{n_j} (Y_{ji} - \bar{Y}_j)^2 .$$

Yksisuuntainen varianssianalyysi

- F-testisuureen

$$F = \frac{\left(\sum_{j=1}^{p} a_j \bar{Y}_j\right)^2}{S^2 \sum_{j=1}^{p} a_j^2 / n_j} \overset{H}{\sim} F_{1,n-p}, \quad S^2 = \frac{1}{n-p} \sum_{j=1}^{p} \sum_{i=1}^{n_j} (Y_{ji} - \bar{Y}_j)^2$$

vaihtoehtona voidaan käyttää testisuuretta (ks. jakso 4.1)

$$T = \frac{\sum_{j=1}^{p} a_j \bar{Y}_j}{S\sqrt{\sum_{j=1}^{p} a_j^2 / n_j}} \overset{H}{\sim} t_{n-p}$$

- Kontrastin $\mathbf{a}'\boldsymbol{\mu}$ luottamusväliksi luottamustasolla $1-\alpha$ saadaan

$$\sum_{j=1}^{p} a_j \bar{y}_j \pm t_{n-p}(\alpha/2)\, s \sqrt{\sum_{j=1}^{p} a_j^2 / n_j}.$$

- Useita luottamusvälejä muodostettaessa on kuitenkin syytä muistaa jaksossa 4.1 mainitut hankaluudet.

Yksisuuntainen varianssianalyysi

Empiirinen esimerkki

- Tutkitaan miten erilaisilla juomilla voidaan vaikuttaa maitohapon kerääntymiseen lihaksiin pitkänmatkan juoksijoilla.
- Kokeessa 50 satunnaisesti valittua juoksijaa jaettiin satunnaisesti viiteen 10 juoksijan ryhmään, joille annettiin eri juomia seuraavasti.

1	Vesi	
2	Urheilujuoma A1	Sama vaikuttava aine eri pitoisuuksina
3	Urheilujuoma A2	
4	Urheilujuoma B1	Sama vaikuttava aine eri pitoisuuksina
5	Urheilujuoma B2	

Juoksumatkan pituus oli 10 mailia ja juomia nautittiin kiinteät määrät ennen juoksua, juoksun puolivälissä ja juoksun päätyttyä. Tämän jälkeen mitattiin maitohappopitoisuudet, joista saatiin tiivistettynä seuraava aineisto (9 juoksijaa keskeytti).

	Vesi	A1	A2	B1	B2	
$\bar{y}_j$	33.3	32.6	30.9	29.0	26.1	$\bar{y} = 30.7$
s_j^2	13.1	14.2	12.2	13.9	14.2	$s^2 = 13.37$
n_j	10	7	10	8	6	$n = 41$

- Olettaen, että havainnot voidaan tulkita riippumattomiksi ja että ryhmästä j saadut havainnot noudattavat $\mathrm{N}(\mu_j, \sigma^2)$-jakaumaa $(j = 1, ..., 5)$, voidaan soveltaa varianssianalyysia ja testata nollahypoteesia $H : \mu_1 = \cdots = \mu_5$.
- Tutkitaan ensin vakiovarianssioletusta laskemalla

$$\max_{1 \leq j \leq 5} s_j^2 / \min_{1 \leq j \leq 5} s_j^2 = 14.2/12.2 = 1.16.$$

- Koska $\mathrm{P}\left(F_{6,9} > 1.16\right) \approx 0.6$, voidaan vakiovarianssioletusta pitää kohtuullisena (vrt. Lause 2.1, mutta huomaa, ettei F-jakauman käyttö ole tässä tarkkaan ottaen oikein).
- $H : \mu_1 = \cdots = \mu_5 \Rightarrow F = 4.55$ ja $\mathrm{P} = \mathrm{P}\left(F_{4,36} \geq 4.55\right) = 0.004$, joten nollahypoteesi hylätään.

Seuraavaksi on kiinnostavaa hakea vastauksia (ainakin) seuraaviin vertailuihin.

1. Vesi vastaan urheilujuomat
2. Urheilujuoma A vastaan urheilujuoma B
3. Urheilujuoma A1 vastaan urheilujuoma A2
4. Urheilujuoma B1 vastaan urheilujuoma B2

Näihin liittyvät hypoteesit ovat kontrastien avulla esitettyinä

$$H_1 : \mu_1 - \tfrac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5) = 0$$
$$H_2 : \tfrac{1}{2}(\mu_2 + \mu_3) - \tfrac{1}{2}(\mu_4 + \mu_5) = 0$$
$$H_3 : \mu_2 - \mu_3 = 0$$
$$H_4 : \mu_4 - \mu_5 = 0.$$

Valitsemalla vektorin $\mathbf{a} = \begin{bmatrix} a_1 & \cdots & a_5 \end{bmatrix}'$ komponentit sopivasti voidaan näitä hypoteeseja testata testisuureella

$$F = \frac{\left(\sum_{j=1}^{5} a_j \bar{y}_j\right)^2}{s^2 \sum_{j=1}^{5} a_j^2 \,/\, n_j},$$

jonka saamia arvoja verrataan $F_{1,36}$-jakauman prosenttipisteisiin. Tämä johtaa seuraaviin tuloksiin.

$H_1 : F = 7.46, \quad \mathrm{P} = \mathbf{0.01}$ (vesi vastaan urheilujuomat)
$H_2 : F = 9.87, \quad \mathrm{P} = \mathbf{0.003}$ (urheilujuoma A vastaan urheilujuoma B)
$H_3 : F = 0.89, \quad \mathrm{P} = 0.35$ (urheilujuoma A1 vastaan urheilujuoma A2)
$H_4 : F = 2.15, \quad \mathrm{P} = 0.15$ (urheilujuoma B1 vastaan urheilujuoma B2)

- Tuloksista voidaan tehdä seuraavat johtopäätökset.
- Hypoteeseihin H_1 ja H_2 liittyvät P-arvot ovat niin pieniä, että nämä hypoteesit voidaan hylätä. Sen sijaan hypoteeseja H_3 ja H_4 vastaan ei ole näyttöä.
- Tulosten mukaan veden ja urheilujuomien välillä on siis eroa ja samoin urheilujuomien A ja B välillä on eroa. Sen sijaan sillä ei ole merkitystä kumpaa kahdesta tutkitusta pitoisuudesta urheilujuomasta A tai B käytetään.