Peatükk 5

Friedmani test

Vahel on tarvis võrrelda rohkem kui kahte töötlust/populatsiooni, ning vahel tuleb sellist võrdlust teha olukorras, kus mõõtmised on tehtud samadel objektidel. Näiteks katsetati k erinevat viljasorti n põllul, lasti k erinevat toodet hinnata n inimesel vms. Kontrollimaks, kas erinevad töötlused annavad erinevaid tulemusi või mitte, saab kasutada Friedmani testi. Vahel on Friedmani testi kutsutud ka 2-faktorilise dispersioonanalüüsi mitteparameetriliseks alternatiiviks.

Kasutame alljärgnevalt ka näidet — kolme sorti vilja kasvatati 4-l põllul. Saadud saagid on toodud alljärgnevas tabelis:

põld	sort 1	sort 2	sort 3
1	989	670	780
2	860	123	870
3	600	345	450
4	520	480	490

Kuidas konstrueerida sobivat mitteparameetrilist testi? Kõigepealt võime igal põllul saadud saagid asendada astakutega — mitmes paremuselt antud sort antud põllul oli:

põld	sort 1	sort 2	sort 3
A	1	3	2
B	2	3	1
C	1	3	2
D	1	3	2

Edaspidistes aruteludes kasutame i. sordi astakut j. põllul tähega R_{ij}. Leiame iga sordi jaoks keskmise astaku — i. sordi keskmise astaku tähistame sümboliga $\overline{R_i}$:

$$\overline{R_i} := \frac{1}{n}\sum_{j=1}^{n} R_{ij}$$

Juhul, kui kõik sordid oleksid samahead, peaks astakute jaotuseks olema ühtlane jaotus ning $E\overline{R_i} = \frac{k+1}{2}$. Sestap peaksid vahed $\overline{R_i} - \frac{k+1}{2}$ olema nullhüpoteesi kehtides nullilähedased. Siis aga peaks nullhüpoteesi kehtides olema suhteliselt väike ka suurus

$$\sum_{i=1}^{k}\left(\overline{R_i} - \frac{k+1}{2}\right)^2.$$

Friedmani teststatistik, Q, saadakse nimetatud avaldise läbikorrutamisel sobivaid omadusi tagava konstandiga (miks selline konstant on valitud, uurime veidi hiljem):

$$Q = \frac{12n}{k(k+1)}\sum_{i=1}^{k}\left(\overline{R_i} - \frac{k+1}{2}\right)^2.$$

Toodud näite puhul oleksid $\overline{R_1} = 5/4 = 1,25$, $\overline{R_2} = 3$ ja $\overline{R_3} = 7/4 = 1,75$ ning

$$\begin{aligned} Q &= \frac{12\cdot 4}{3\cdot 4}\left((1,25-2)^2 + (3-2)^2 + (1,75-2)^2\right) \\ &= 4\cdot 1.625\ (=6,5). \end{aligned}$$

5.1 Friedmani teststatistiku jaotus nullhüpoteesi kehtides

Palju on erinevaid võimalusi astakuid ümber paigutada? Igas grupis (põllul) on $k!$ võimalikku astakute järjestust, kusjuures nullhüpoteesi kehtides (sortidel pole mingit vahet) on kõik nimetatud järjestused võrdtõenäolised. Kuna gruppe (põlde) on n tükki, siis on kokku $k!^n$ erinevat võimalust paigutada astakuid ringi. Kõik need võimalused on nullhüpoteesi kehtides võrdtõenäolised, st. esinevad tõenäosusega $1/(k!)^n$. Meie näites eksisteerib seega $6^4 = 1296$ erinevat astakute järjestust. Arvuti abil peaks olema suhteliselt lihtne kontrollida, et neist selliseid, mille puhul Q väärtus tuleb samasuur kui meil või veel suurem, on 54 (6-juhtu, kus kõik sordid saavad kõigil põldudel sama astaku; siis olukord, kus muidu kõik sordid saavad sama astaku, aga ühel põllul on astakud 1 ja 2 vahetusse läinud — 4 põldu * 6 võimalust järjestada astakuid ühel põllul; siis juhtum kus ühel põllul on vahetusse läinud astakud 3 ja 2 — 4*6=24 võimalust; kokku 54 astakute järjestust). Seega toodud näite puhul on Friedmani teststatistiku olulisus $54/1296 = 0,042$ ning meil on võimalik (olulisuse nivool 0,05) kummutada nullhüpotees — kõik sordid pole samahead.

5.2 Asümptootiline jaotus nullhüpoteesi kehtides

Friedmani teststatistiku asümptootiline jaotus nullhüpoteesi kehtides on hii-ruut jaotus,

$$Q \xrightarrow{\mathcal{D}} \chi^2_{df=k-1}.$$

Üritame seda näidata, kasutades järgmiseid tehnilisi tulemusi.

Kõigepealt mainime (ilma tõestuseta) peamisi tulemusi, mida tõesamiseks kasutame.

A. **Tsentraalne piirteoreem** (valimikeskmise jaotus on normaaljaotus). Kui $\mathrm{E}X_i = \mu$ ja $\mathrm{D}X_i = \Sigma$, siis

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{\mathcal{D}} N(0, \Sigma)$$

B. Kui $X_n \xrightarrow{\mathcal{D}} X$ ja $f(X) \sim F$, kus $f(x)$ on pidev funktsioon, siis

$$f(X_n) \xrightarrow{\mathcal{D}} F.$$

C. Lisaks vajame tulemust, mis ütleb: Kui $X \sim N(0, \Sigma)$ ja Σ on idempotentne, st $\Sigma\Sigma = \Sigma$, siis (ja ainult siis) $X^T X \sim \chi^2_{df=\mathrm{rank}(\Sigma)}$.

Tähistame i. katseüksuse (põld/degustaator) j. töötlusele (väetamine/veinisort) antud astaku R_{ij}. Meenutame kõigepealt, milline on astakute keskväärtus, dispersioon ja kahe erineva (samal põllul mõõdetud) astaku omavahelist kovariatsiooni nullhüpoteesi kehtides (vaata tulemusi Wilcoxoni astaksummatesti käsitlevast osast):

$$\begin{aligned} \mathrm{E}R_{ij} &= \frac{k+1}{2} \\ \mathrm{D}R_{ij} &= \frac{k^2-1}{12} \\ \mathrm{Cov}(R_{ij}, R_{ij'}) &= -\frac{k+1}{12} \end{aligned}$$

Tsentraalsest piirteoreemist järeldub, et

$$\sqrt{n}\left(\overline{R_1}-\frac{k+1}{2},\ldots,\overline{R_k}-\frac{k+1}{2}\right)\xrightarrow{\mathcal{D}}N(0,\Sigma),$$

kus

$$\Sigma=\begin{pmatrix}\frac{k^2-1}{12} & \ldots & -\frac{k+1}{12}\\ \vdots & \ddots & \vdots\\ -\frac{k+1}{12} & \ldots & \frac{k^2-1}{12}\end{pmatrix}.$$

Paraku pole saadud kovariatsioonimaatriks idempotentne (mida on lihtne kontrollida). Proovime olukorda parandada — korrutades kõik astakud läbi konstandiga, $R^*_{ij}=\sqrt{\frac{12}{k(k+1)}}R_{ij}$, saame $\Sigma^*:=\mathrm{D}R^*_{ij}$ väärtuseks

$$\Sigma^*=\begin{pmatrix}\frac{k-1}{k} & \ldots & -\frac{1}{k}\\ \vdots & \ddots & \vdots\\ -\frac{1}{k} & \ldots & \frac{k-1}{k}\end{pmatrix}.$$

On lihtne näha, et mainitud maatriks on idempotentne (korruta ta iseendaga läbi!). Järelikult

$$v:=\sqrt{\frac{12n}{k(k+1)}}\left(\overline{R_1}-\frac{k+1}{2},\ldots,\overline{R_k}-\frac{k+1}{2}\right)\xrightarrow{\mathcal{D}}N(0,\Sigma^*).$$

Aga tehnilistest abitulemustest B ja C järeldub seega, et $v^Tv\xrightarrow{\mathcal{D}}\chi^2_{\mathrm{rank}(\Sigma^*)}$. Σ^* astakut on aga lihtne leida — idempotentse maatriksi puhul maariksi astak langeb kokku maatriksi jäljega. Seega $\mathrm{rank}(\Sigma^*)=\sum_{i=1}^k\frac{k-1}{k}=k-1$. Olemegi näidanud, et H_0 kehtides $Q\xrightarrow{\mathcal{D}}\chi^2_{k-1}$.