Problem

V populaciji ljudi Ω (trenutnih in bodočih) starih nad 18 let, ki imajo previsok nivo holesterola LDL v krvi, želimo primerjati dve terapiji oziroma zdravili, recimo jima A in B, zaradi kontrole pa želimo hkrati primerjati še psihosomatske učinke terapije s placebom, recimo ji C. Tedaj bi nas zanimale razlike nivojev holesterola pred in po terapijah A, B, C. Te razlike bi bile slučajne spremenljivke X_1, X_2, X_3.

Osnovni privzetek („model“)

Verjamemo, da gre za normalno porazdeljene slučajne spremenljivke, kjer je odklon σ neodvisen od zdravljenja in enak za vse tri slučajne spremenljivke, sicer pa neznan. Torej $X_1 \sim N(\mu_1, \sigma)$, $X_2 \sim N(\mu_2, \sigma)$, $X_3 \sim N(\mu_3, \sigma)$.

Temeljno vprašanje

Temeljno vprašanje je, ali lahko statistično ugotovimo, da je med vsem tremi terapijami (oziroma med obema terapijama ter placebom) sploh kakšna statistično zaznavna razlika.

Če želimo dokazati, da gre za razliko,

želimo zavrniti ničelno hipotezo $\mu_1 = \mu_2 = \mu_3$.

Privzemimo, da imamo m terapij.

- Izberemo naključnega bolnika ω.
- Izberemo terapijo $j \in \{1, 2, 3, \ldots, m\}$ (naključno).
- Beležimo rezultat eksperimenta $X_j(\omega)$.

Ureditev in predstavitev podatkov

- Naj bo n_1 število tistih bolnikov, ki so prejeli terapijo 1.
- Naj bo n_2 število tistih bolnikov, ki so prejeli terapijo 2.
- $\ldots$
- Naj bo n_m število tistih bolnikov, ki so prejeli terapijo m.

Velja

$$n_1 + n_2 + \cdots + n_m = n.$$

Popolnoma randomizirani model
Ureditev

Rezultate predstavimo v tabeli:

$i \backslash j$	1	2	3	$\ldots$	m
1	X_{11}	X_{12}	X_{13}	$\ldots$	X_{1m}
2	X_{21}	X_{22}	X_{23}	$\ldots$	X_{2m}
3	X_{31}	X_{32}	X_{33}	$\ldots$	X_{3m}
$\vdots$					
n_2	$\ldots$	$X_{n_2,2}$	$\ldots$	$\ldots$	$\ldots$
n_1	$X_{n_1,1}$		$\ldots$	$\ldots$	$\ldots$
n_m			$\ldots$	$\ldots$	$X_{n_m,m}$
n_3			$X_{n_3,3}$		

V j-tem stolpcu navedemo rezultate terapije št. j.

Oznaki

- $\overline{X}_{\bullet j}$ = povprečje j-tega stolpca = $\frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$,
- $\overline{X}_{\bullet\bullet}$ = povprečje celotnega vzorca
  $= \frac{1}{n}\sum_{j=1}^{m}\sum_{i=1}^{n_j} X_{ij} = \frac{1}{n}\sum_{j=1}^{m} n_j \overline{X}_{\bullet j}$

Naivna testna statistika

$$\frac{\sum_{j=1}^{m} n_j \overbrace{(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2}^{\text{variacija skupine } j \text{ glede na celotno povprečje}}}{\sum_{j=1}^{m} \underbrace{\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_{\bullet j})^2}_{(n_j-1)\cdot S_j^2:\ \text{variacija znotraj skupine } j}} = \frac{\mathrm{VMS}}{\mathrm{VZS}}.$$

Dejanska testna statistika

$$F = \frac{\mathrm{VMS}/(m-1)}{\mathrm{VZS}/(n-m)} = \frac{\sum_{j=1}^{m} n_j(\overline{X}_{\bullet j} - \overline{X}_{\bullet\bullet})^2/(m-1)}{\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_{\bullet j})^2/(n-m)}.$$

Porazdelitev testne statistike

Če velja ničelna hipoteza $H_0\colon \mu_1 = \mu_2 = \cdots = \mu_m$ (pravimo ji tudi hipoteza homogenosti), ima F *Fisherjevo* porazdelitev z $m-1$ ter $n-m$ prostostnimi stopnjami.

Test

$$\begin{cases} \mathrm{H}_0 \text{ zavrnemo}, & \text{če } F > C \\ \mathrm{H}_0 \text{ ne zavrnemo}, & \text{če } F \leqslant C, \end{cases}$$

kjer $C = F_{m-1,n-m;\alpha}$, zgornji α-percentil F-porazdelitve z $m-1$ in n prostostnimi stopnjami. Test ima eksaktno značilnost α.

Oglejmo si $\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_{\bullet j})^2$

$i \backslash j$	1	2	3	$\dots$	j	$\dots$	m
1	X_{11}	X_{12}	X_{13}	$\dots$	X_{1j}	$\dots$	X_{1m}
2	X_{21}	X_{22}	X_{23}	$\dots$	X_{2j}	$\dots$	X_{2m}
3	X_{31}	X_{32}	X_{33}	$\dots$	X_{3j}	$\dots$	X_{3m}
$\vdots$							
n_2	$\dots$	$X_{n_2,2}$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
n_1	$X_{n_1,1}$		$\dots$	$\dots$	$X_{n_j j}$	$\dots$	$\dots$
n_m			$\dots$	$\dots$		$\dots$	$X_{n_m,m}$
n_3			$X_{n_3,3}$				

- j-ti stolpec lahko gledamo kot n_j neodvisnih replikacij eksperimenta $N(\mu_j, \sigma)$.
- Torej je $\frac{1}{\sigma^2}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_{\bullet j})^2$ porazdeljena kot $\chi^2_{n_j-1}$.

Oglejmo si $\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2$

- Od prej: $\frac{1}{\sigma^2}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2$ je porazdeljena kot $\chi^2_{n_j-1}$.
- Pripomnimo, da se $C_j=\frac{1}{\sigma^2}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2$ tiče samo j-tega stolpca. Torej so $C_1,\ldots,C_m$ paroma neodvisne.
- Torej je $\frac{1}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2$ porazdeljena po zakonu χ^2 z $\sum_{j=1}^{m}(n_j-1)=n-m$ prostostnimi stopnjami.

Naj velja H_0, t.j. $\mu_1 = \mu_2 = \cdots = \mu_m =: \mu$

Velja dekompozicija

$$\frac{1}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet\bullet})^2 = \frac{1}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2+\frac{1}{\sigma^2}\sum_{j=1}^{m} n_j(\overline{X}_{\bullet j}-X_{\bullet\bullet})^2.$$

- Leva stran je porazdeljena po zakonu χ^2_{n-1} (ker velja H_0).
- Po zgornjem neodvisno od H_0 vedno velja $\frac{1}{\sigma^2}\sum_{j=1}^{m}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2 \sim \chi^2_{n-m}$.
- Izkaže se, da sta sumanda na desni neodvisna.
- Sledi $\frac{1}{\sigma^2}\sum_{j=1}^{m} n_j(\overline{X}_{\bullet j}-X_{\bullet\bullet})^2 \sim \chi^2_{m-1}$.