Проверка гипотез в дисперсионном анализе повторных наблюдений

Н.П. Алексеева

СПбГУ, математико-механический факультет

2020 г.

Структура данных с повторностями

Пусть в i-й группе, $i = 1, 2, \ldots, I$ имеется ν_i индивидов, $\nu_1 + \ldots + \nu_I = n$, которые набдюдаются в моменты времени $t = 1, 2, \ldots, T$. Выборку наблюдений обозначим через x_{ijt} (ТШХ - тест шестиминутной ходьбы).

Модель

$$x_{ijt} = \mu + \alpha_i + e^1_{ij} + \beta_t + \gamma_{it} + e_{ijt}$$

- μ параметр генерального среднего,
- α_i дифф.эффекты групп (пол),
- β_t дифф. эффекты фактора времени (при поступлении в стационар, при выписке и через 3 месяца),
- γ_{it} дифф. эффекты взаимодействия факторов времени и разделения на группы,
- $e^1_{ij} \sim \mathcal{N}(0, \sigma_1)$, $e_{ijt} \sim \mathcal{N}(0, \sigma)$ независимые случайные ошибки.

Оценка параметров модели

В случае полных данных устраняется индивидуальное влияние, измеряемое индивидуальным средним $x_{ij\cdot}$ и модель разделяется на две модели:

$$x_{(i)j\cdot} = \mu + \alpha_i + e^1_{ij}, \quad x_{ijt} - x_{(i)j\cdot} = \beta_t + \gamma_{it} + e_{ijt}$$

Общий источник вариации Q с числом степеней свободы, равным $\nu = nt - 1$, имеет вид:

$$Q = \sum_{j=1}^{n} \sum_{k=1}^{t} (x_{(i)jk} - \bar{x}_{\dots})^2 .$$

Главные эффекты

Оценкой влияния фактора A является разность

$$\widehat{\alpha} = \bar{x}_{i..} - \bar{x}_{...},$$

где $x_{i..}$ есть среднее по каждому полу, $x_{...}$ – общее среднее. Источник вариации, обусловленный влиянием фактора A, с числом степеней свободы, равным $\nu_A = r - 1$, имеет вид:

$$Q_A = t\sum_{i=1}^{r} n_i(\bar{x}_{i..} - \bar{x}_{...})^2.$$

Источник вариации, обусловленный различием индивидов

Вариация Q_A является частью вариации Q_1, обуславливающей различие индивидов, с числом степеней свободы, равным $\nu_1 = n - 1$:

$$Q_1 = t\sum_{j=1}^{n}(\bar{x}_{(i)j\cdot} - \bar{x}_{\cdots})^2 =$$

$$= t\sum_{i=1}^{r} n_i(\bar{x}_{i\cdot\cdot} - \bar{x}_{\cdots})^2 + t\sum_{j=1}^{n}(\bar{x}_{(i)j\cdot} - \bar{x}_{i\cdot\cdot})^2 = Q_A + Q_{1e}\,,$$

где усредненный показатель по каждому индивиду вычисляется как

$$\bar{x}_{(i)j\cdot} = \frac{1}{t}\sum_{k=1}^{t} x_{(i)jk}\,.$$

Ошибка Q_{1e} имеет число степеней свободы, равное $\nu_{1e} = n - r$.

Проверка значимости дифференциального эффекта A

Для проверки гипотезы о том, что все дифференциальные эффекты фактора A равны нулю, вычисляется статистика

$$\mathrm{F} = \frac{\mathrm{MQ_A}}{\mathrm{MQ_{1e}}} = \frac{\mathrm{Q_A}/\nu_\mathrm{A}}{\mathrm{Q_{1e}}/\nu_\mathrm{1e}},$$

которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_\mathrm{A} = \mathrm{r} - 1$ и $\nu_\mathrm{1e} = \mathrm{n} - \mathrm{r}$.

Источник вариации после устранения индивидуальных различий

$$Q_2 = Q - Q_1 = \sum_{j=1}^{n}\sum_{k=1}^{t}(x_{(i)jk} - \bar{x}_{(i)j\cdot})^2\,,$$

$$\nu_2 = (nt-1) - (n-1) - nt - n = n(t-1).$$

Источник вариации по фактору B

$$Q_B = n\sum_{k=1}^{t}(x_{\cdot\cdot k} - \bar{x}_{\cdot\cdot\cdot})^2\,, \quad \nu_B = t-1\,.$$

Источник вариации по взаимодействию AB

$$Q_{AB} = \sum_{i=1}^{r} n_i \sum_{k=1}^{t}(\bar{x}_{i\cdot k} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot k} + \bar{x}_{\cdot\cdot\cdot})^2\,,$$

$$\nu_{AB} = (tr-1) - (t-1) - (r-1) = (r-1)(t-1)\,.$$

Ошибка и ее степени свободы

$$Q_{err} = Q_2 - Q_B - Q_{AB} =$$
$$= \sum_{j=1}^{n} \sum_{k=1}^{t} (x_{(i)jk} - \bar{x}_{(i)j\cdot} - \bar{x}_{i\cdot k} + \bar{x}_{i\cdot\cdot})^2 ,$$
$$\nu_{err} = (nt - n) - (t - 1) - (t - 1)(r - 1) = (t - 1)(n - r).$$

$$Q_{err} + Q_B + Q_{AB} = \sum_{kj} (x_{(i)jk} - \bar{x}_{(i)j\cdot} - \bar{x}_{i\cdot k} + \bar{x}_{i\cdot\cdot})^2 +$$
$$+ \sum_{kj} (x_{\cdot\cdot k} - \bar{x}_{\cdot\cdot\cdot})^2 + \sum_{kj} (\bar{x}_{i\cdot k} - \bar{x}_{i\cdot\cdot} - \bar{x}_{\cdot\cdot k} + \bar{x}_{\cdot\cdot\cdot})^2 =$$
$$= \sum_{kj} (x_{(i)jk} - \bar{x}_{(i)j\cdot})^2 = Q_2$$

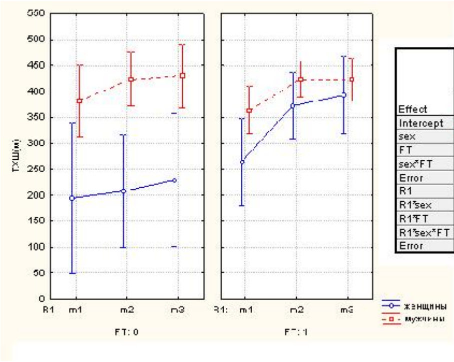
Проверка значимости эффекта B

- Для проверки гипотезы о том, что случайные эффекты временного фактора имеют нулевую дисперсию $H_0 : \sigma_b^2 = 0$, вычисляется статистика
$$F = \frac{MQ_B}{MQ_{AB}} = \frac{Q_B/\nu_B}{Q_{AB}/\nu_{AB}},$$
которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_B = t - 1$ и $\nu_{AB} = (t-1)(r-1)$.
- В случае предположения фиксированных эффектов временного фактора B вычисляется статистика
$$F = \frac{MQ_B}{MQ_{err}} = \frac{Q_B/\nu_B}{Q_{err}/\nu_{err}},$$
которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_B = t - 1$ и $\nu_{err} = (t-1)(n-r)$.

Проверка значимости эффекта взаимодействия

$$F = \frac{MQ_{AB}}{MQ_{err}} = \frac{Q_{AB}/\nu_{AB}}{Q_{err}/\nu_{err}} \sim \mathbb{F}(\nu_{AB}, \nu_{err}),$$

$\nu_{AB} = (t-1)(r-1)$ и $\nu_{err} = (t-1)(n-r)$.



Effect	Repeated Measures Analysis of Variance (kutuso Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	678713	1	678713	349,646	0,00000
sex	24821	1	24821	12,787	0,00110
FT	5596	1	5596	2,882	0,09894
sex*FT	7223	1	7223	3,721	0,06235
Error	64057	33	1941		
R1	5050	2	2525	11,061	0,00007
R1*sex	191	2	95	0,419	0,65900
R1*FT	950	2	475	2,081	0,13286
R1*sex*FT	494	2	247	1,083	0,34432
Error	15065	66	228		

Figure: Динамика ТХШ в зависимости от пола и FT лечебной физкультуры.

Пример кода в R и результата вычислений

```
dat.AR.T <- data.frame( stack(dat.AR[,-1]),
                  sub=as.factor(rep(seq(nrow(dat.AR)),m)),
                  gr=as.factor(rep(dat.AR[,1],m)))
dat.AR.T $values <- as.numeric(dat.AR.T$values)
formula<-values gr*ind+Error(sub/ind)
aov.out <- aov(formula, data=dat.AR.T)
summary(aov.out)
```

——————————————-

Error: sub

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gr	2	3377	1688.6	3.503	0.0507

Residuals 19 9159 482.1

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Error: sub:ind

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	2	1041	520.7	4.139	0.0229 *

Residuals 42 5283 125.8

—

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Построение графика взаимодействий

```
Names<-names(table(dat.AR.T$gr)); Names; K<-length(Names)
interaction.plot(x.factor=dat.AR.T$ind,
                 trace.factor=dat.AR.T$gr,
                 response=dat.AR.T$values,
                 fun = mean,
                 type = "b", legend = FALSE,
                 trace.label ="group",
                 xlab = "",ylab = "",
                 lty = seq(K), col = seq(K), pch = 20, lwd=2 )
legend(top,Names,lty = seq(K), col =seq(K), cex=0.7,pch=20)
```

График эффектов взаимодействия факторов группы и времени

Считаем, что данные о принадлежности к группе находятся в первом столбце файла dat.AR

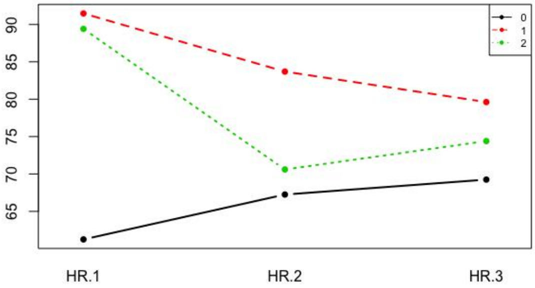


Figure: Динамика ЧСС у больных ААС.