1. Модели дисперсионного анализа

1.1. Двуфакторная модель с фиксированными эффектами

Модель с фиксированными эффектами имеет вид:

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \tag{1}$$

$i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K,\ \mu$ - генеральное среднее, α_i - дифференциальный эффект фактора A, β_j - дифференциальный эффект фактора B. Величина $(\alpha\beta)_{ij}$ называется взаимодействием факторов. Эта величина учитывает дифференциальный эффект комбинаций i-го уровня фактора A и j-го уровня фактора B, если он не выражается суммой $\alpha_i+\beta_j+\mu$. Ошибки e_{ijk} предполагаются независимыми и нормально распределенными $\mathcal{N}(0, \sigma)$. При ограничениях на параметры

$$\sum_{i=1}^{I} \alpha_i = 0, \quad \sum_{j=1}^{J} \beta_j = 0,$$

$$\sum_{i=1}^{I} (\alpha\beta)_{ij} = 0, \quad j = 1, 2, \ldots, J; \quad \sum_{j=1}^{J} (\alpha\beta)_{ij} = 0, \quad i = 1, 2, \ldots, I,$$

оценками параметров модели (1) являются:

$\hat{\mu} = \bar{x}$ – общее среднее,

$\hat{\alpha}_i = \bar{x}_{i\cdot} - \bar{x}$, где $\bar{x}_{i\cdot}$ – среднее по i-му значению фактора A;

$\hat{\beta}_j = \bar{x}_{\cdot j} - \bar{x}$, где $\bar{x}_{\cdot j}$ – среднее по j-му значению фактора B;

$\widehat{(\alpha\beta)}_{ij} = \bar{x}_{ij\cdot} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x}$, где $\bar{x}_{ij\cdot}$ – среднее в ячейке ij.

Уровень адреналина при разном вскармливании и содержании

вскармл. с матерью 3.35 ± 0.15			вскармл. без матери 4.48 ± 0.15		
отд.клет.ξ_1	сообщ.клет.ξ_2	общ.клет.ξ_3	отд.клет.ξ_1	сообщ.клет.ξ_2	общ.клет.ξ_3
1.9	4	3.2	3.3	6.3	4.6
2.3	4.6	2.6	4	7.2	4.8
2.2	5.7	2.2	5	4.6	4.6
2	5.7	2.6	3.2	7.2	4.4
2.7	4.8	3.2	2.4	3.8	4.5
2.8	4.8	2.5	3.6	4.4	4.2
2.4	5.4	3	3	4.8	4.4
2.7	3.8	3.3	3	5.8	4.3
$\bar{x}_{11} = 2.38$	$\bar{x}_{12} = 4.85$	$\bar{x}_{13} = 2.83$	$\bar{x}_{21} = 3.44$	$\bar{x}_{22} = 5.51$	$\bar{x}_{23} = 4.48$
$\bar{x}_{\cdot 1} = 2.91$	$\bar{x}_{\cdot 2} = 5.18$	$\bar{x}_{\cdot 3} = 3.65$			

Имеем вектор Y размерности IJK, матрицу плана размерности IJK на $r = 1 + (I-1) + (J-1) + (I-1)(J-1) = IJ$. Например, при $I = 3$ и $J = 4$ она имеет вид:

Y	μ	α_1	α_2	β_1	β_2	β_3	$(\alpha\beta)_{11}$	$(\alpha\beta)_{12}$	$(\alpha\beta)_{13}$	$(\alpha\beta)_{21}$	$(\alpha\beta)_{22}$	$(\alpha\beta)_{23}$
y_{11*}	1	1	0	1	0	0	1	0	0	0	0	0
y_{12*}	1	1	0	0	1	0	0	1	0	0	0	0
y_{13*}	1	1	0	0	0	1	0	0	1	0	0	0
y_{14*}	1	1	0	-1	-1	-1	-1	-1	-1	0	0	0
y_{21*}	1	0	1	1	0	0	0	0	0	1	0	0
y_{22*}	1	0	1	0	1	0	0	0	0	0	1	0
y_{23*}	1	0	1	0	0	1	0	0	0	0	0	1
y_{24*}	1	0	1	-1	-1	-1	0	0	0	-1	-1	-1
y_{31*}	1	-1	-1	1	0	0	-1	0	0	-1	0	0
y_{32*}	1	-1	-1	0	1	0	0	-1	0	0	-1	0
y_{33*}	1	-1	-1	0	0	1	0	0	-1	0	0	-1
y_{34*}	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Вектор параметров: $\beta = (\mu, \alpha_1, \alpha_{I-1}, \beta_1, \ldots, \beta_{J-1}, (\alpha\beta)_{11}, \ldots, (\alpha\beta)_{I-1,J-1})'$. Общее число наблюдений равно IJK. Остаточная сумма квадратов основной модели равна

$$Q_R = R_0^2 = \sum_{i,j,k} (x_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j - (\alpha\hat{\beta})_{ij})^2$$

имеет число степеней свободы $IJK - IJ$. Для проверки гипотезы $H_0 : \alpha_1 = \alpha_1 = \ldots = \alpha_I = 0$ используем $H'\beta = 0$, матрицу H' размерности $I - 1$ на

$r = IJ$ ранга $I-1$, вида $H' = \left[\mathbf{0}_{I-1,1} | \mathbf{I}_{I-1,I-1} | \mathbf{0}_{I-1,IJ-I)}\right]$. Остаточная сумма квадратов усеченной модели равна

$$R_1^2 = \sum_{i,j,k} (x_{ijk} - \hat{\mu} - \hat{\beta}_j - (\alpha\hat{\beta})_{ij})^2 = R_0^2 + Q_A,$$

$$\text{где } Q_A = \sum_{i,j,k} (\hat{\alpha}_i)^2 = JK \sum_{i=1}^{I} (\bar{x}_{i\cdot} - \bar{x})^2$$

Отсюда получаем статистику критерия

$$F = \frac{IJK - IJ}{I-1} \cdot \frac{R_1^2 - R_0^2}{R_0^2} = \frac{IJK - IJ}{I-1} \cdot \frac{Q_A}{Q_R} \;\sim\; F(I-1, IJK - IJ)\,.$$

Аналогично строятся остальные статистики.

Таблица двуфакторного дисперсионного анализа

источник дисперсии	сумма квадратов	степени свободы	средний квадрат
фактор A	$Q_A = JK \sum_{i=1}^{I} (\bar{x}_{i\cdot} - \bar{x})^2$	$\nu_A = I - 1$	$MQ_A = \frac{Q_A}{\nu_A}$
фактор B	$Q_B = IK \sum_{j=1}^{J} (\bar{x}_{\cdot j} - \bar{x})^2$	$\nu_B = J - 1$	$MQ_B = \frac{Q_B}{\nu_B}$
взаимо- действие AB	$Q_{AB} = K \sum_{i=1}^{I} \sum_{j=1}^{J} (\bar{x}_{ij\cdot} -$ $- \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$	$\nu_{AB} =$ $= (I-1)(J-1)$	$MQ_{AB} = \frac{Q_{AB}}{\nu_{AB}}$
остаток (ошибка)	$Q_R =$ $= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (x_{ijk} - \bar{x}_{ij\cdot})^2$	$\nu_R = IJ(K-1)$	$MQ_R = \frac{Q_R}{\nu_R}$
полная	$Q_T = \sum_{i=1}^{I} \sum_{j=1}^{J} s \sum_{k=1}^{K} (x_{ijk} - \bar{x})^2$	$\nu_T = IJK - 1$	

Для проверки гипотезы об отсутствии эффекта взаимодействия

$$H_0 : \text{все } \; (\alpha\beta)_{ij} = 0 \text{ используется статистика } F = \frac{MQ_{AB}}{MQ_R} \;\sim\; F(\nu_{AB}, \nu_R)\,.$$

Для проверки гипотезы об отсутствии эффекта A-фактора

$$H_0 : \text{все } \; \alpha_i = 0 \; \text{ используется отношение } F = \frac{MQ_A}{MQ_R} \;\sim\; F(\nu_A, \nu_R)\,.$$

Для проверки гипотезы об отсутствии эффекта B-фактора

$$H_0 : \text{все} \;\; \beta_j = 0 \;\; \text{берется} \;\; \text{отношение} \; F = \frac{MQ_B}{MQ_R} \; \sim \; F(\nu_B, \nu_R)\,.$$

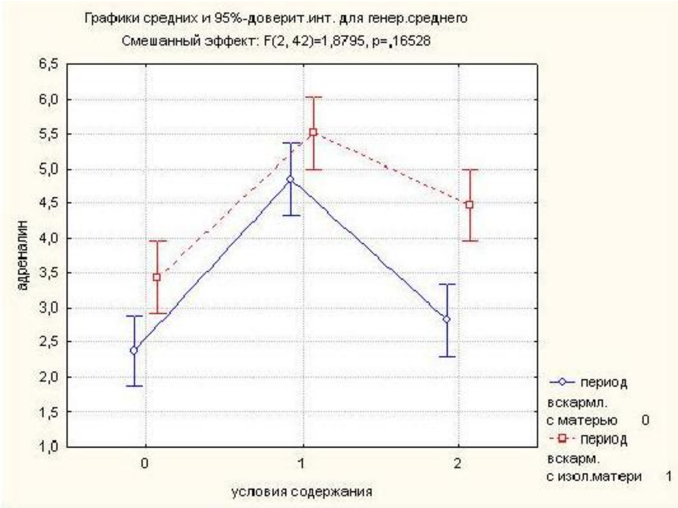


Рис. 1. Иллюстрация отсутствия эффекта взаимодействия факторов вскармливания и содержания.

ПРИМЕР (продолжение) Проверяем гипотезы.

H_0 : все $\alpha_i = 0$. $p = 0.000003$ – следовательно, влияние фактора вскармливания значимо для уровня адреналина, средний уровень адреналина 3.35 ± 0.15 в группе, до конца вскармливания находящейся с матерью, значимо ниже среднего уровня адреналина 4.48 ± 0.15 в группе, изолированной от матери.

H_0 : все $\beta_j = 0$. $p < 0.000001$ – следовательно, влияние фактора условия содержания значимо для уровня адреналина: средние 2.91 ± 0.18, 5.18 ± 0.18 и 3.65 ± 0.18 в группах, отличающихся условиями содержания, значимо различаются.

H_0 : все $(\alpha\beta)_{ij} = 0$. $p = 0.16$ – следовательно, фактор взаимодействия не значим.

1.2. Модель со случайными эффектами

Если подпопуляции выбираются случайно из большого (бесконечного) числа подпопуляций, то фактор относится ко второй модели, которая имеет вид:

$$x_{ij} = \mu + a_i + e_{ij}, \quad i = 1, \ldots, r, \;\; j = 1, \ldots, n_i, \tag{2}$$

где a_i распределены по $\mathcal{N}(0,\sigma_a)$, e_{ij} распределены по $\mathcal{N}(0,\sigma)$, все a_i и e_{ij} в совокупности независимы.

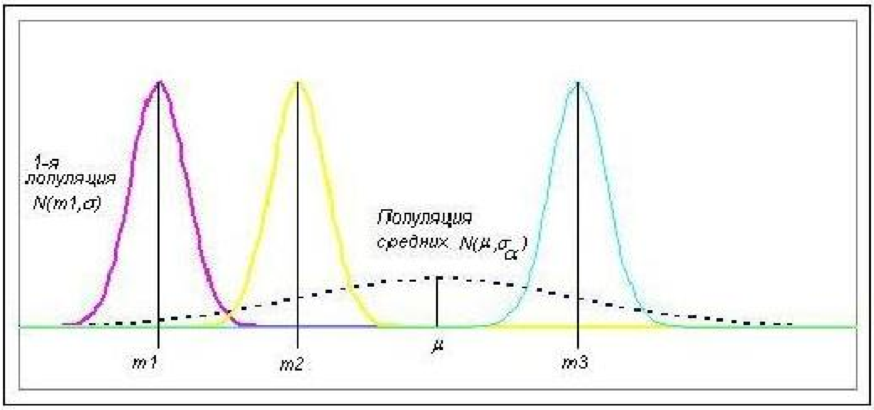


Рис. 2. Модель со случайными эффектами. Проверяется гипотеза $H_0 : \sigma_a^2 = 0, m_1, \ldots, m_r$ выбираются случайно.

При повторении эксперимента мы скорее всего будем иметь дело со случайными выборками из других популяций. В этой модели нас интересует оценка дисперсии σ_a^2 распределения дифференциальных эффектов. Гипотеза $H_0 : \sigma_a^2 = 0$ означает, что фактор не вносит значимого вклада в дисперсию.

Для проверки гипотезы $H_0 : \sigma_a^2 = 0$ используются выражения для математических ожиданий средних квадратов $\frac{Q_1}{r-1}$ и $\frac{Q_2}{n-r}$. Непосредственными вычислениями можно убедиться, что

$$\mathbf{E}\left(\frac{Q_2}{n-r}\right) = \sigma^2, \quad \mathbf{E}\left(\frac{Q_1}{r-1}\right) = \sigma^2 + k\sigma_a^2, \quad \text{где}$$

$$k = \frac{1}{r-1}\left(n - \frac{1}{n}\sum_{i=1}^{r} n_i^2\right). \tag{3}$$

Используем выражение $\bar{x}_i = \mu + a_i + \bar{\epsilon}_i$ для вычисления EQ_2:

$$Q_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\mu + a_i + \epsilon_{ij} - \mu - a_i - \bar{\epsilon}_i)^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\epsilon_{ij}^2 - \sum_{i=1}^{r} n_i\bar{\epsilon}_i^2, \quad \sum_{i=1}^{r} n_i = n,$$

$$EQ_2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i} E\epsilon_{ij}^2 - \sum_{i=1}^{r} n_i E\bar{\epsilon}_i^2 = n\sigma^2 - \sum_{i=1}^{r} n_i \frac{\sigma^2}{n_i} = (n-r)\sigma^2\,.$$

Для вычисления EQ_1 используем выражение

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n}\sum_{i=1}^{r} n_i\bar{x}_i = \mu + \frac{1}{n}\sum_{i=1}^{r} n_i a_i + \bar{\epsilon},$$

$$Q_1 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left((a_i + \bar{\epsilon}_i) - \left(\frac{1}{n}\sum_{i=1}^{r} n_i a_i + \bar{\epsilon}\right)\right)^2 =$$

$$= \sum_{i=1}^{r} n_i(a_i + \bar{\epsilon}_i)^2 - n\left(\frac{1}{n}\sum_{i=1}^{r} n_i a_i + \bar{\epsilon}\right)^2;$$

Считаем математическое ожидание:

$$EQ_1 = \sum_{i=1}^{r} n_i E(a_i + \bar{\epsilon}_i)^2 - nE\left(\frac{1}{n}\sum_{i=1}^{r} n_i a_i + \bar{\epsilon}\right)^2 =$$

$$= \sum_{i=1}^{r} n_i D(a_i + \bar{\epsilon}_i) - nD\left(\frac{1}{n}\sum_{i=1}^{r} n_i a_i + \bar{\epsilon}_i\right) =$$

$$= \sum_{i=1}^{r} n_i\left(\sigma_a^2 + \frac{\sigma^2}{n_i}\right) - n\left(\sigma_a^2\frac{1}{n^2}\sum_{i=1}^{r} n_i^2 + \frac{\sigma^2}{n}\right) =$$

$$= (r-1)\sigma^2 + \sigma_a^2\left(n - \frac{1}{n}\sum_{i=1}^{r} n_i^2\right) = (r-1)\left(\sigma^2 + k\sigma_a^2\right),$$

$$\text{где } k = \frac{1}{r-1}\left(n - \frac{1}{n}\sum_{i=1}^{r} n_i^2\right).$$

Если используется одинаковое число повторений в каждой подпопуляции, то есть $n_1 = \ldots = n_r = m$, то $k = m$. При справедливости гипотезы

$H_0 : \sigma_a^2 = 0$ статистики $\frac{Q_2}{n-r}$ и $\frac{Q_1}{r-1}$ имеют одинаковые средние σ^2, поэтому можно построить отношение Фишера

$$F = \frac{n-r}{r-1} \cdot \frac{Q_1}{Q_2} \ \sim \ F(r-1, n-r)\,.$$

Значимость критерия равна $\alpha_* = P\{F > F_*\}$, где F_* наблюдаемое значение статистики. Гипотеза об отсутствии влияния фактора отвергается при $\alpha_* < \alpha = 0.05$. что в случае однофакторного дисперсионного анализа значимости моделей со случайными и фиксированными эффектами совпадают. Рассматривая в качестве k выражение (3), получаем оценку дисперсии σ_a^2:

$$\widehat{\sigma}_a^2 = \frac{1}{k}\left(\frac{Q_1}{r-1} - \frac{Q_2}{n-r}\right).$$

1.3. Двуфакторная модель со случайными эффектами

Модель со случайными эффектами имеет вид:

$$x_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}, \tag{4}$$

где $i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K,\ \mu$ - генеральное среднее, случайные дифференциальные эффекты A-фактора a_i независимы и распределены по $\mathcal{N}(0; \sigma_a)$, случайные дифференциальные эффекты B-фактора b_j независимы и распределены по $\mathcal{N}(0; \sigma_b)$, эффекты взаимодействия $(ab)_{ij}$ независимы и распределены по $\mathcal{N}(0; \sigma_{ab})$. Величины e_{ijk} независимы и распределены по $\mathcal{N}(0; \sigma)$. Все величины $a_i, b_j,\ (ab)_{ij},\ e_{ijk}$ независимы в совокупности.

источник дисперсии	средний квадрат	математическое ожидание
A	MQ_A	$\sigma^2 + K\sigma_{ab}^2 + JK\sigma_a^2$
B	MQ_B	$\sigma^2 + K\sigma_{ab}^2 + IK\sigma_b^2$
AB	MQ_{AB}	$\sigma^2 + K\sigma_{ab}^2$
R	MQ_R	σ^2

Отсюда получаем выражения для оценок дисперсий:

$$\sigma_a^2 = \frac{MQ_A - MQ_{AB}}{JK}, \quad \sigma_b^2 = \frac{MQ_B - MQ_{AB}}{IK}, \quad \sigma_{ab}^2 = \frac{MQ_{AB} - MQ_R}{K}.$$

Для проверки гипотезы $H_0 : \sigma_{ab}^2 = 0$ используется отношение $F = \frac{MQ_{AB}}{MQ_R}$, для $H_0 : \sigma_a^2 = 0$ используется отношение $F = \frac{MQ_A}{MQ_{AB}}$, для $H_0 : \sigma_b^2 = 0$ отношение $F = \frac{MQ_B}{MQ_{AB}}$ с соответствующими степенями свободы.

ПРИМЕР (продолжение) При проверке гипотезы $H_0 : \sigma_{ab}^2 = 0$ об отсутствии взаимодействия получаем тот же критерий, что и в случае модели с фиксированными эффектами. $p = 0.16$, взаимодействие незначимо.

При проверке гипотезы $H_0 : \sigma_a^2 = 0$ об отсутствии эффекта условий вскармливания получаем значимость $p = 0.059$, при проверке гипотезы $H_0 : \sigma_b^2 = 0$ получаем значимость $p = 0.044$. Значимости отличаются от значимостей в модели с фиксированными эффектами, сохраняется только соотношение между ними: фактор условия содержания более значим, чем фактор вскармливания.

В *StatSoft* наиболее удобной для практического использования дисперсионного анализа с фиксированными и случайными эффектами является программа: *Statistics* - *Advanced Linear/ Nonlinear Models* - *Variance Components*. В *Variables* указываем нужные переменные. Результаты счета находятся в таблице, спрятанной под кнопкой *Summary* : *Components of Variance*.

Средние, соответствующие различным уровням факторов и их пересечениям, можно посмотреть во вкладке *Estimation* за кнопкой *Margimal means*.

1.4. Двуфакторная модель со смешанными эффектами

Модель со смешанными эффектами имеет вид:

$$x_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + e_{ijk}, \tag{5}$$

где $i = 1, \ldots, I,\ j = 1, \ldots, J,\ k = 1, \ldots, K,$

μ - генеральное среднее,

α_i – i-й дифференциальный эффект A-фактора,

b_j – случайные эффекты B-фактора независимы и нормально распределены $\mathcal{N}(0; \sigma_b)$, эффекты взаимодействия

$(\alpha b)_{ij}$ независимы и нормально распределены $\mathcal{N}(0;\sigma_{ab})$.

Величины e_{ijk} независимы распределены по $\mathcal{N}(0;\sigma)$. Все величины b_j, $(\alpha b)_{ij}$, e_{ijk} независимы в совокупности.

источник дисперсии	средний квадрат	математическое ожидание
A	MQ_A	$\sigma^2 + K\sigma_{ab}^2 + \frac{JK\sum\limits_{i=1}^{I}\alpha_i^2}{I-1}$
B	MQ_B	$\sigma^2 + K\sigma_{ab}^2 + IK\sigma_b^2$
AB	MQ_{AB}	$\sigma^2 + K\sigma_{ab}^2$
R	MQ_R	σ^2

Для проверки гипотезы

$H_0 : \sigma_{ab}^2 = 0$ используется отношение $F = \frac{MQ_{AB}}{MQ_R}$, для

$H_0 : \sigma_b^2 = 0$ используется отношение $F = \frac{MQ_B}{MQ_{AB}}$, для

$H_0 :$ все $\alpha_i = 0$ отношение $F = \frac{MQ_A}{MQ_{AB}}$ с соответствующими степенями свободы. Заметим, что в случае двуфакторного плана результаты дисперсионного анализа в случае хотя бы одного случайного фактора совпадают.

1.5. Двуфакторная модель с группировкой

ПРИМЕР. Пусть имеется фактор диеты A с $r = 4$ уровнями. Зависимой переменной является дважды ($K = 2$) измеряемое количество выдыхаемого азота x_{ijk}, $i = 1,\ldots,I$; $j = 1,\ldots,J$, $k = 1,\ldots,K$. Фактор B является фактором испытуемого, $J = 4$.

Данные для двуфакторного плана с группировкой

	диета 1		диета 2		диета 3		диета 4	
B	1 набл.	2 набл.	1 набл.	2 набл.	1 набл.	2 набл.	1 набл.	2 набл.
1	4.079	4.859	4.368	5.668	4.169	5.709	4.928	5.608
2	3.541	5.047	3.752	5.848	4.416	5.666	4.941	5.291
3	3.298	4.679	3.802	4.844	4.123	5.059	4.674	5.038
4	2.871	4.648	3.578	5.393	4.403	4.496	4.905	5.208

Пусть фактор A является случайным фактором. Тогда модель имеет вид:

$$x_{ijk} = \mu + a_i + b_{j(i)} + e_{ijk},$$

где μ - генеральное среднее, величины a_i независимы и распределены по $\mathcal{N}(0;\sigma_a)$, величины $b_{j(i)}$ - независимы и распределены по $\mathcal{N}(0;\sigma_{b(a)})$. Величины e_{ijk} независимы распределены по $\mathcal{N}(0;\sigma)$. Все величины $a_i, b_{j(i)}, e_{ijk}$ независимы в совокупности. $\hat{b}_{j(i)} = \bar{x}_{ij\cdot} - \bar{x}_{i\cdot}$.

Таблица двуфакторного плана с группировкой

источник дисперсии	сумма квадратов	степени свободы	средний квадрат
фактор A	$Q_A = JK\sum\limits_{i=1}^{I}(\bar{x}_{i\cdot} - \bar{x})^2$	$\nu_A = I - 1$	$MQ_A = \frac{Q_A}{\nu_A}$
фактор B (внутриА)	$Q_{B(A)} = K\sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}(\bar{x}_{ij\cdot} - \bar{x}_{i\cdot})^2$	$\nu_{B(A)} = I(J-1)$	$MQ_{B(A)} = \frac{Q_{B(A)}}{\nu_{B(A)}}$
остаток (ошибка)	$Q_R = \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}\sum\limits_{k=1}^{K}(x_{ijk} - \bar{x}_{ij\cdot})^2$	$\nu_R = IJ(K-1)$	$MQ_R = \frac{Q_R}{\nu_R}$
полная	$Q_T = \sum\limits_{i=1}^{I}\sum\limits_{j=1}^{J}\sum\limits_{k=1}^{K}(x_{ijk} - \bar{x})^2$	$\nu_T = IJK - 1$	

Смешанный двуфакторный план с группировкой можно записать в виде:

$$x_{ijk} = \mu + \alpha_i + b_{j(i)} + e_{ijk},$$

α_i – дифференциальный эффект, определяемый i-м уровнем фактора A.

Приведем ожидания средних квадратов для двуфакторного плана с группировкой в случае модели со случайными эффектами:

$$\mathbf{E}\left(MQ_A\right) = \sigma^2 + K\sigma^2_{b(a)} + KJ\sigma^2_a,$$

$$\mathbf{E}\left(MQ_{B(A)}\right) = \sigma^2 + K\sigma^2_{b(a)}, \quad \mathbf{E}\left(MQ_R\right) = \sigma^2,$$

и в случае модели со смешанными эффектами:

$$\mathbf{E}\left(MQ_A\right) = \sigma^2 + K\sigma^2_{b(a)} + \frac{JK\sum\limits_{i=1}^{I}\alpha_i^2}{I-1},$$
$$\mathbf{E}\left(MQ_{B(A)}\right) = \sigma^2 + K\sigma^2_{b(a)}, \quad \mathbf{E}\left(MQ_R\right) = \sigma^2.$$

Для проверки гипотез используем отношения

$$H_0 : \sigma^2_{b(a)} = 0 \qquad F = \frac{MQ_{B(A)}}{MQ_R},$$
$$H_0 : \sigma^2_a = 0 \text{ или} H_0 : \text{все } \ \alpha_i = 0 \ \ F = \frac{MQ_A}{MQ_{B(A)}}.$$

В *Statistics - Advanced Linear/ Nonlinear Models - Variance Components*. Фактор A можно считать фиксированным или случайным эффектом (в зависимости от модели). Фактор B считать случайным эффектом. В *Model* выбрать *Hierarchically nested design* (и указать вложенность фактора B кнопкой *codes identify levels within other factors*.

Результаты дисперсионного анализа

источник дисперсии	сумма квадратов	число ст.св.	средний квадрат	F	p
Диета	3.711	3	1.237	8.14	0.0032
Объект(диета)	1.828	12	0.152	0.20	0.9960
Остаток	12.020	16	0.751		
Полная	17.559	31			

1.6. Дисперсионный анализ для зависимых выборок

Рассмотрим данные вида $x_{(i)jk}$, где $i = 1, \ldots, r$ уровни фактора A (пол) с фиксированными эффектами, $j = 1, \ldots, n$ – индивиды (фактор B), $k = 1, \ldots, t$ уровни временного фактора C со случайными эффектами. Поскольку фактор A является следствием фактора индивида, то его индекс при наличии индекса j заключаем в скобки, $n_1 + n_2 + \ldots + n_r = n$. Требуется выяснить влияние факторов A и C на переменную X (количество пройденных метров по ТШХ в три момента времени: при поступлении в стационар, при выписке и через три месяца). Построим модель:

$$x_{(i)jk} = \mu + \alpha_i + c_k + e^1_{(i)j} + (\alpha c)_{ik} + e_{(i)jk},$$

где $e^1_{(i)j}$ – ошибка, обусловленная влиянием индивида, $e_{(i)jk}$ – ошибка наблюдения. Оценкой влияния фактора A является разность

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

где $x_{i..}$ есть среднее по каждому полу, $x_{...}$ – общее среднее. Общий источник вариации Q с числом степеней свободы, равным $\nu = nt - 1$, имеет вид:

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

Источник вариации, обусловленный влиянием фактора A, с числом степеней свободы, равным $\nu_A = r - 1$, имеет вид:

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

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

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

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

где усредненный показатель по каждому индивиду вычисляется как

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

Ошибка Q_{1e} имеет число степеней свободы, равное $\nu_{1e} = n - r$. Из оставшегося после исключения влияния индивида источник вариации

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

выделим вариацию, обуславливающую различие по времени. Источник вариации, обусловленный влиянием временного фактора C, с числом степеней свободы, равным $\nu_C = t - 1$, имеет вид:

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

а источник вариации, обусловленный влиянием взаимодействия фактора A и фактора времени C, с числом степеней свободы, равным $\nu_{AC} = (t - 1)(r - 1)$, имеет вид:

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

Ошибка имеет вид:

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

Ее число степеней свободы равно $\nu_{err} = (t - 1)(n - r)$.

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

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

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

Для проверки гипотезы о том, что случайные эффекты временного фактора имеют нулевую дисперсию, вычисляется статистика

$$F = \frac{MQ_C}{MQ_{AC}} = \frac{Q_C/\nu_C}{Q_{AC}/\nu_{AC}},$$

которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_C = t-1$ и $\nu_{AC} = (t-1)(r-1)$.

В случае предположения фиксированных эффектов временного фактора C вычисляется статистика

$$F = \frac{MQ_C}{MQ_{err}} = \frac{Q_C/\nu_C}{Q_{err}/\nu_{err}},$$

которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_C = t-1$ и $\nu_{err} = (t-1)(n-r)$.

Для проверки эффекта взаимодействия вычисляется статистика

$$F = \frac{MQ_{AC}}{MQ_{err}} = \frac{Q_{AC}/\nu_{AC}}{Q_{err}/\nu_{err}},$$

которая в случае справедливости нулевой гипотезы имеет распределение Фишера с числом степеней свободы $\nu_{AC} = (t-1)(r-1)$ и $\nu_{err} = (t-1)(n-r)$.

Для обработки в $StatSoft$ данных о ТХШ, дифференцируемых по признаку пол, в трех временных точках нужно представить исходные данные в виде четырех столбцов: sex, $m1$, $m2$ и $m3$. В *Statistics* выбираем $ANOVA$ - *Repeated measures* $ANOVA$. В $Variables$ выбираем переменные: в *Dependent variable list* отмечаем $m1, m2, m3$, в *Categorial predictors* ($factors$) отмечаем sex. Обязательно нажать кнопку $Within\ effects$, включением которой данные о ТХШ в трех временных точках собираются в одну переменную, обычно ее называют $R1$. В противном случае, вместо нужной процедуры пойдет многомерный дисперсионный анализ. Кроме того, для получения указанных ранее сумм квадратов, в опциях $Sum\ of\ squares$ нужно указать последовательный тип *Type* 1 *sequental*.

Исходные данные: $n = 37$, $r = 2$, $t = 3$. В результате счета получаем, что $Q_A = 173604$, $\nu_A = 1$, $Q_{1e} = 721179$, $\nu_{1e} = 35$, для проверки значимости фактора пола используем статистику $F = \frac{173604/1}{721179/35} = 8.42$, $p = 0.006364$, по которой видно, что различие в средних 310 метров у женщин и 406 метров у мужчин является значимым.

Если эффекты времени фиксированы, то влияние времени значимо:

$$Q_C = 100805, \nu_C = (r-1)(t-1) = 2,$$

$$Q_{err} = 160441,\ \nu_{err} = (t-1)(n-r) = 2 \cdot (37-2) = 70.$$

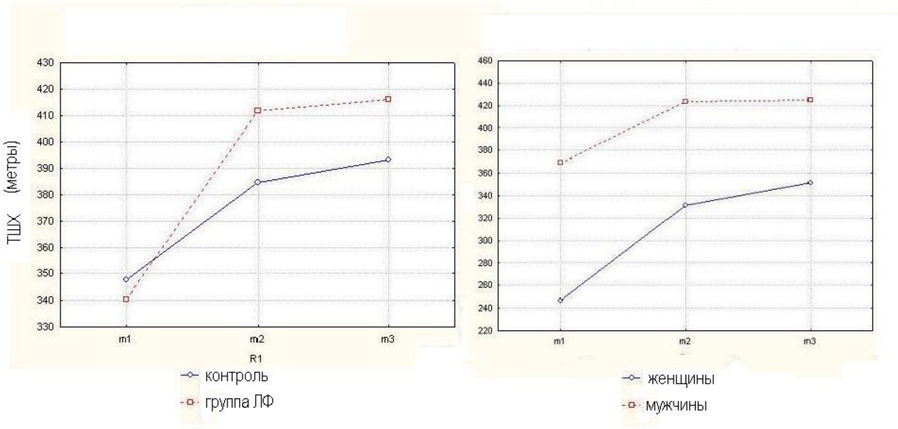
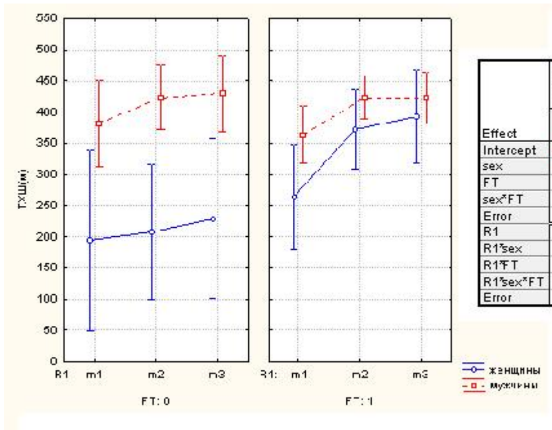


Рис. 3. Эффект взаимодействия (по полу $p = 0.199$ и по ЛФ $p = 0.32$).



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	248218	1	248218	12,7872	0,00110
FT	55960	1	55960	2,8828	0,09894
sex*FT	72239	1	72239	3,7215	0,06235
Error	640577	33	19411		
R1	50500	2	25250	11,0615	0,00007
R1*sex	1916	2	958	0,4197	0,65900
R1*FT	9502	2	4751	2,0814	0,13286
R1*sex*FT	4947	2	2473	1,0836	0,34432
Error	150657	66	2283		

Рис. 4. Динамика ТХШ в зависимости от пола и лечебной физкультуры.

$F = 21.99$, $p < 0.0001$. Если эффекты считать случайными, то $p = 0.06$. Графики средних указывают на значимое отличие пройденных метров по ТХШ при поступлении в стационар и при выписке. Эффект взаимодействия не значим, $p = 0.199$.

Аналогичный счет при исследовании влияния фактора лечебной физкультуры на динамику ТХШ показал, что влияние ЛФК не значимо ($p = 0.67$), влияние взаимодействия также не значимо ($p = 0.32$). Значимо влияние фактора времени ($p < 0.001$).

ANOVA Repeated Measures в R

Предположим, что имеются данные $dat.AR$ в виде $k+1$ столбца, в первом фактор или группирующая переменная, в остальных данные наблюдений у одного и того же индивида в k моментах времени. Сначала нужно преобразовать данные, например, при помощи следующего кода.

```
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)
```

Затем применяем встроенную функцию.

```
aov.out <- aov(values gr*ind+Error(sub/ind), data=dat.AR.T)

summary(aov.out)
```

Результат должен состоять из двух таблиц дисперсионного анализа.

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

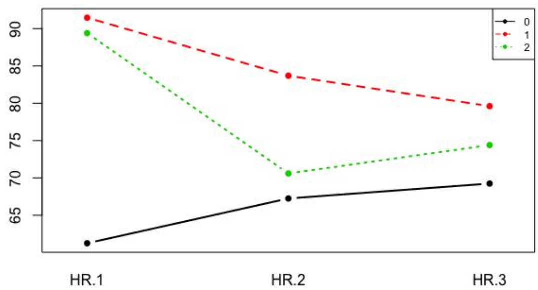


Рис. 5. Динамика ЧСС у больных ААС.

Для построение графика эффектов взаимодействия можно поступить следующим образом.

```
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)
```