СТАТИСТИКА БОЛЬШИХ ДАННЫХ – ИДЗ 2

Студент: Эспинола Ривера, Хольгер Элиас

Тема: Статистический анализ динамики изменений наблюдаемой характеристики с течением времени

Для когорты индивидов проведены измеения исследуемой характеристики Y в моменты времени t1, …, td (Visit), расстояние между которыми одинакого. Помимо исследуемой характеристики измерялись сопутствующие факторы A (не меняющаяся со времени характеристика, измеряемая в момент начала исследования) и B (меняющаяся со временем характеристика, измерение которой проводилось в каждой временой точке). Из таблицы данных следует выбрать значения описанных характеристики, с персональным значением переменной “Variant”.

Вопросы

1. Загрузить данные и отфильтрировать строки с персональным значением переменной “Variant”.

Применение филтра для выбора данных для варианта 9:

```
> # collect the data which corresponds to variant n° 9
> datavar9 <- dataset[dataset$Variant == 9, -1]
> print(head(datavar9))
       ID  Y         A  B  Visit
19579  1   54.16743  1  1  0
19580  1   33.62468  1  1  1
19581  1   16.39175  1  2  2
19582  1   40.94266  1  2  3
19583  1   31.66774  1  2  4
19584  2   41.43027  0  1  0
> print("Dimensions dataset - Variant 9")
```

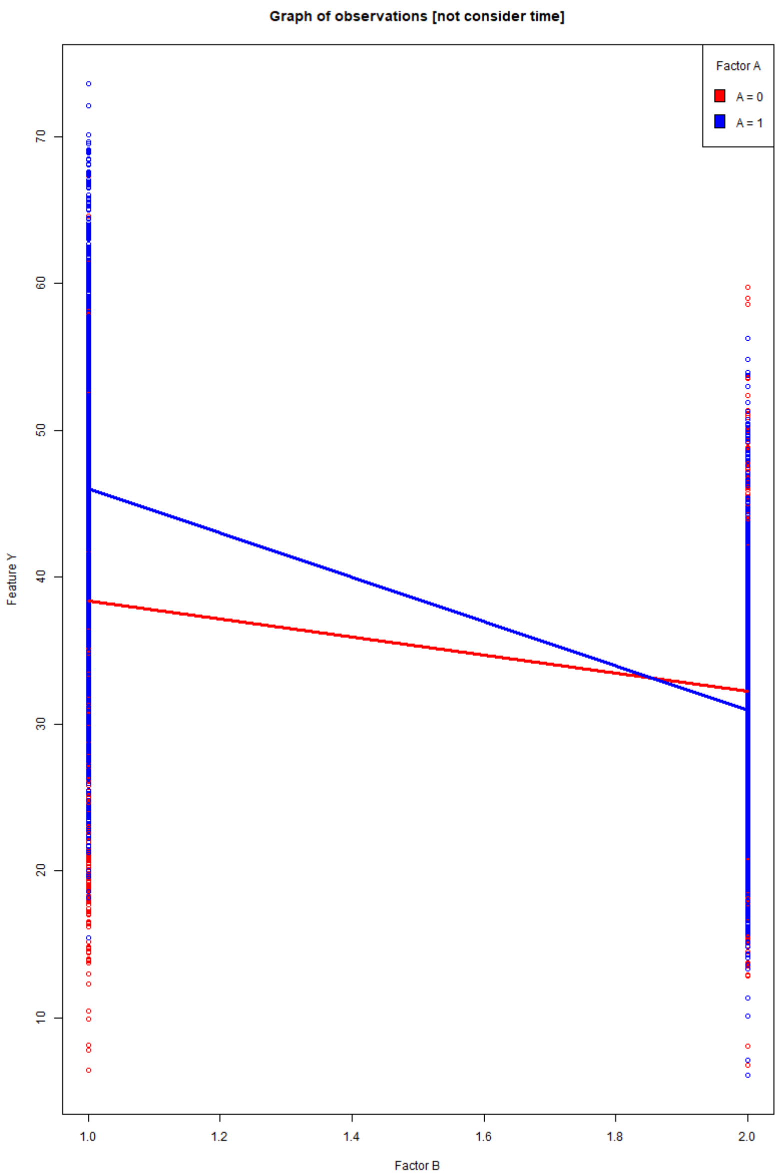
[1] "Dimensions dataset - Variant 9"

> print(dim(datavar9))

[1] 4405   5

2. Изобразить графически результаты наблюдений без учета времени измерений исследуемой характеристики. В предположении нормальности значений наблюдаемого признака провести двухфакторный дисперсионный анализ зависимости наблюдаемого признака от значений факторов А и В.

А) График построена:



B) Двухфакторный дисперсионный анализ

Шаг 01: Модель взаймодействия между факторов A и B

$$\eta_{ij} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB}$$

```
> qab <- lm(Y ~ as.factor(A) * as.factor(B), data = datavar9)
> qab

Call:
lm(formula = Y ~ as.factor(A) * as.factor(B), data = datavar9)
Coefficients:
        (Intercept)          as.factor(A)1
            38.354                 7.679
    as.factor(B)2  as.factor(A)1:as.factor(B)2
            -6.149                -8.971
```

Шаг 02: Аддитивные модели

- Фактроры A и B:

$$\eta_{ij} = \mu + \alpha_i^A + \alpha_j^B$$

```
> qadd <- lm(Y ~ as.factor(A) + as.factor(B), data = datavar9)
> qadd

Call:
lm(formula = Y ~ as.factor(A) + as.factor(B), data = datavar9)
Coefficients:
 (Intercept)       as.factor(A)1       as.factor(B)2
     40.49              3.11              -10.33
```

- Только Фактор А:

$$\eta_{ij} = \mu + \alpha_i^A$$

```
> qa <- lm(Y ~ as.factor(A), data = datavar9)
> qa
Call:
lm(formula = Y ~ as.factor(A), data = datavar9)
Coefficients:
 (Intercept)   as.factor(A)1
      35.217           3.128
```

- Только фактор В:

$$\eta_{ij} = \mu + \alpha_j^B$$

```
> qb <- lm(Y ~ as.factor(B), data = datavar9)
> qb
Call:
lm(formula = Y ~ as.factor(B), data = datavar9)
Coefficients:
 (Intercept)   as.factor(B)2
       41.94          -10.34
```

- Независимые от факторов А и В:

$$\eta_{ij} = \mu$$

```
> q0 <- lm(Y ~ 1, data = datavar9)
> q0
Call:
lm(formula = Y ~ 1, data = datavar9)
Coefficients:
(Intercept)
      36.68
```

Шаг 03: Проверка гипотез для двухфакторный дисперсионный анализ

Пусть модель: $\eta_{ij} = \mu + \alpha_i^A + \alpha_j^B + \alpha_{ij}^{AB}$

А) Гипотеза отсутствия взаймодействий

$H_{AB} : \alpha_{ij}^{AB} = 0$

> anova(qadd, qab)

Analysis of Variance Table

Model 1: Y ~ as.factor(A) + as.factor(B)

Model 2: Y ~ as.factor(A) * as.factor(B)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4402	401345				
2	4401	379295	1	22050	255.85	< 2.2e-16 ***

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

Гипотеза $\mathrm{H}_{\boldsymbol{AB}}$ **отвергает**, следовательно, есть влияние взаимодействия между A и B.

Б) Гипотеза отсутствие влияния фактора А

$H_{AB} : \alpha_{ij}^{AB} = 0$; $H_A : \alpha_i^A = 0$

> anova(qa, qab)

Analysis of Variance Table

Model 1: Y ~ as.factor(A)

Model 2: Y ~ as.factor(A) * as.factor(B)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4403	518919				
2	4401	379295	2	139624	810.03	< 2.2e-16 ***

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

Гипотезы $\mathrm{H}_{\boldsymbol{AB}}$ **и** $\mathrm{H}_{\boldsymbol{A}}$ **отвергают**, следовательно, есть аддитивный влияние фактора А.

В) Гипотеза отсутствие влияния фактора B

$H_{AB}:\alpha_{ij}^{AB}=0$; $H_B:\alpha_j^B=0$

> anova(qb, qab)

Analysis of Variance Table

Model 1: Y ~ as.factor(B)

Model 2: Y ~ as.factor(A) * as.factor(B)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4403	411951				
2	4401	379295	2	32656	189.45	< 2.2e-16 ***

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

Гипотезы $\mathrm{H}_{\boldsymbol{AB}}$ **и** $\mathrm{H}_{\boldsymbol{B}}$ **отвергают**, следовательно, есть аддитивный влияние фактора B.

Г) Гипотеза независимости Y от факторов A, B.

$H_{AB}:\alpha_{ij}^{AB}=0$; $H_A:\alpha_i^A=0$; $H_B:\alpha_j^B=0$

> anova(q0, qab)

Analysis of Variance Table

Model 1: Y ~ 1

Model 2: Y ~ as.factor(A) * as.factor(B)

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	4404	529645				
2	4401	379295	3	150350	581.51	< 2.2e-16 ***

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

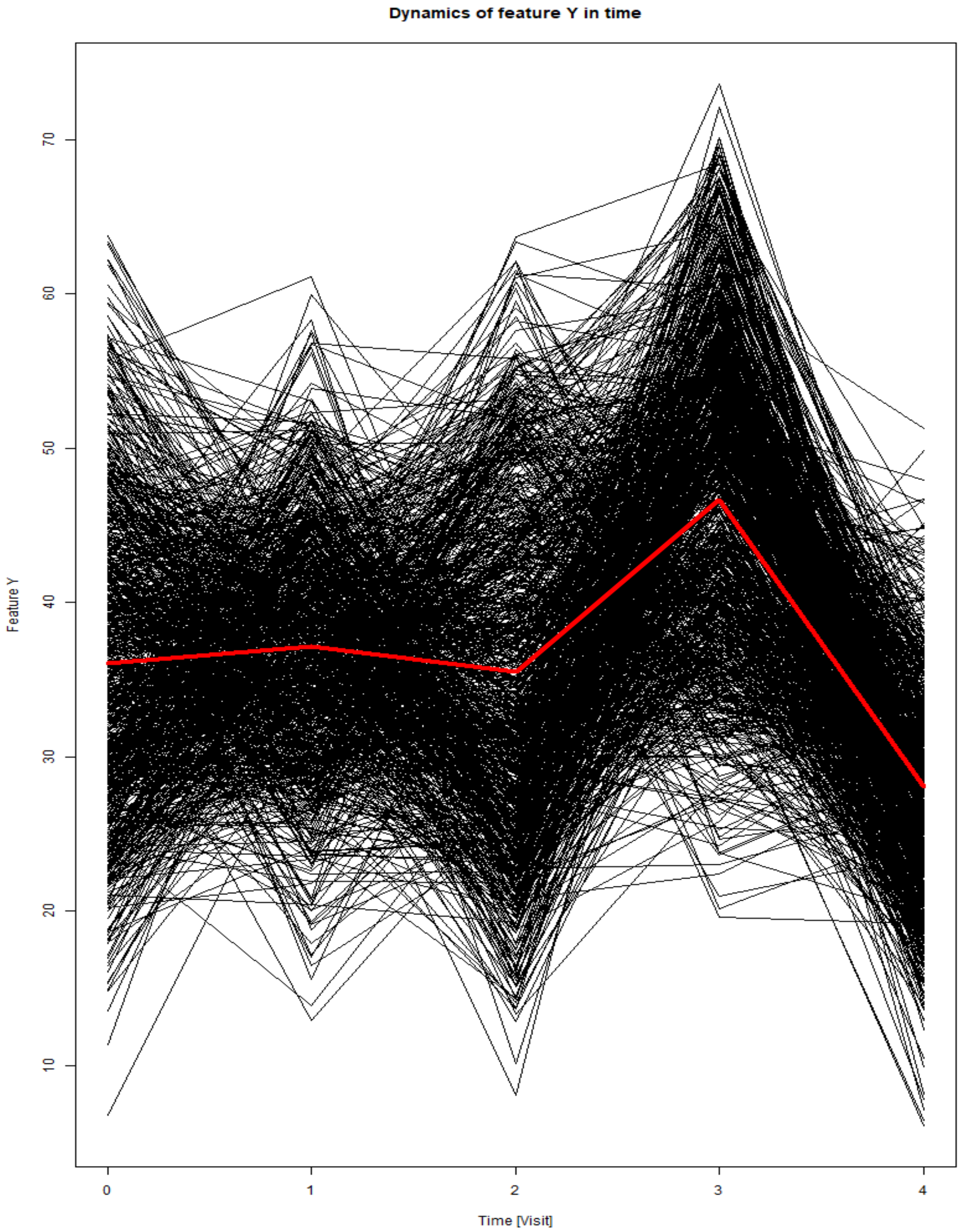
Гипотезы H_{AB} , H_A и H_B отвергают

Общий вывод:

Есть влиянгия взаймодейство и аддитивнное для факторов A и B с отношениями наблюдаемого признака.

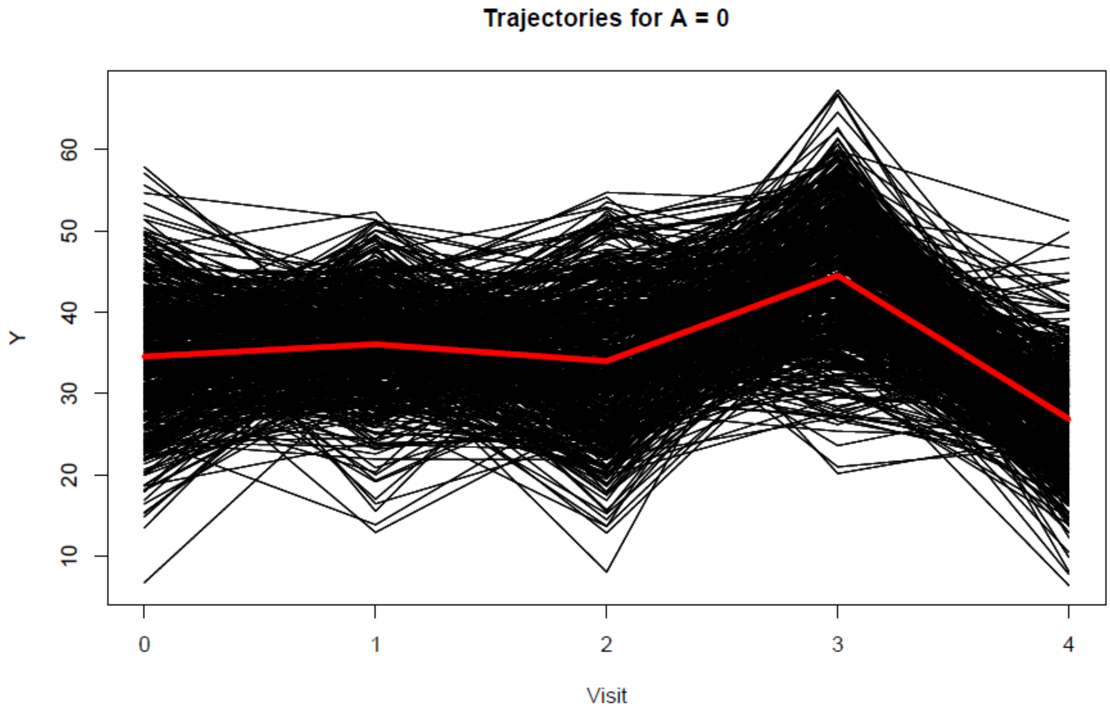
3. Преставить визуально динамику изменений наблюдаемого признака в виде траекторий и оценить наличие влияния постоянного фактора А на значения измеряемого признака. Оценить корреляции значений наблюдаемого признака в различные моменты времени при каждом значений признака А без учета влияния признака В.

А) График построена динамику изменений наблюдаемого признака

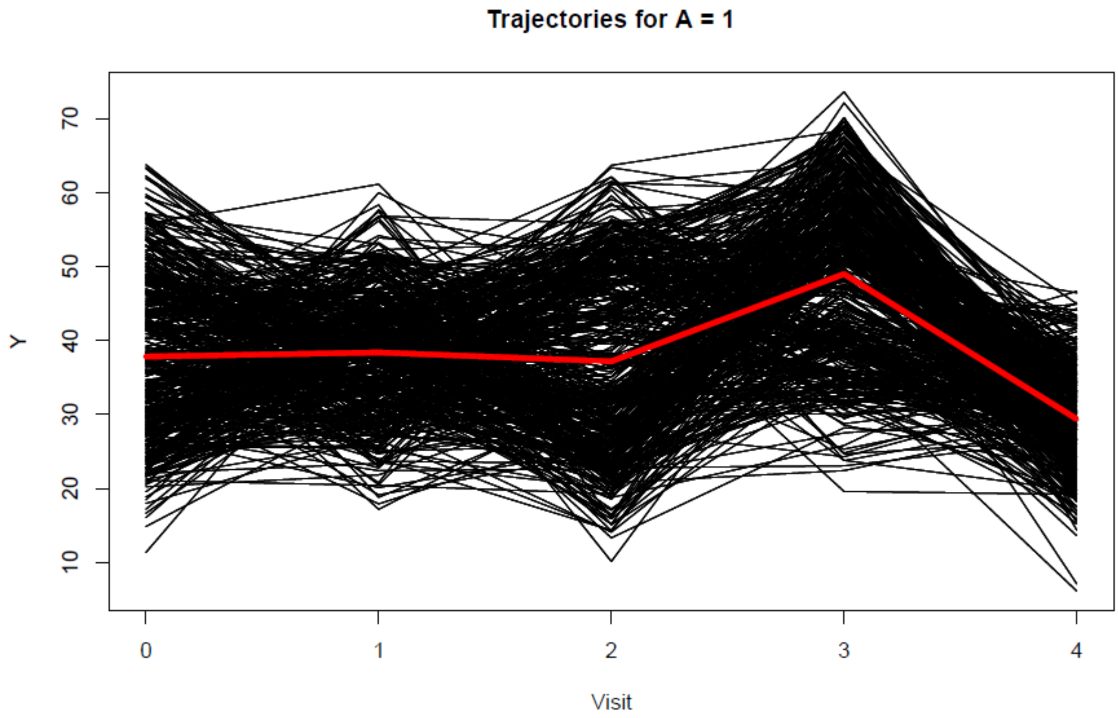


Б) Оценить наличие влияния постоянного фактора А на значения Y.

- Траектория динамику Y (во времени) для A = 0



- Траектория динамику Y (во времени) для A = 1



B) Оценить корреляции значения Y (во времени) при каждом значения A.

Выявляем возможные значения переменной A

> lvA = levels(as.factor(datavar9$A))

[1] "0" "1"

Корреляционная матрица для Y (во времени) при A = 0

> corr_A0

$`A=0`$corr

	Y_V0	Y_V1	Y_V2	Y_V3	Y_V4
Y_V0	1.00000000	-0.026984902	-0.036062662	-0.04480147	0.05599858
Y_V1	-0.02698490	1.000000000	0.008961128	0.09223862	0.02129035
Y_V2	-0.03606266	0.008961128	1.000000000	-0.07581508	0.05553220
Y_V3	-0.04480147	0.092238617	-0.075815077	1.00000000	-0.05049627
Y_V4	0.05599858	0.021290351	0.055532204	-0.05049627	1.00000000

Корреляционная матрица для Y (во времени) при A = 1

> corr_A1

$`A=1`$corr

	Y_V0	Y_V1	Y_V2	Y_V3	Y_V4
Y_V0	1.00000000	0.017432803	0.01894922	0.034281278	-0.02109358
Y_V1	0.01743280	1.000000000	-0.04929156	-0.001735304	0.06703278
Y_V2	0.01894922	-0.049291562	1.00000000	-0.039914042	0.02080742
Y_V3	0.03428128	-0.001735304	-0.03991404	1.000000000	-0.06104107
Y_V4	-0.02109358	0.067032779	0.02080742	-0.061041066	1.00000000

4. Построит оценки средних значений наблюдаемой характеристики при различных значенях факторов A и B и центрировать исходные наблюдения. С использованием семивариограммы, оценить зависимость корреляции значений центрированного процесса от времени в предположении его стационарности при каждом значении пары факторов A и B.

Шаг 01: Для оценить средних значений наблюдаемой характеристики при различных значениях факторов A и B, получиться смещанную линейную модель по каждому пары значений факторов A и B.

```
> # generate the linear mixed effect model for each pair values A, B
> m0 <- NULL
> for(i in names(d9)){
  m0[[i]] <- lme(Y ~ as.factor(Visit), random = ~1|ID, data = d9[[i]])
 }
```

Шаг 02: Построить семивариограммы по каждому пары значений факторов A и B.

Таблички:

```
> # check the table of semi-variogram distances and variogram values
```

$`a=0_b=1` $`a=0_b=2`

	variog	dist	n.pairs
1	1.0975740	1	445
2	0.9219419	2	330
3	0.9705741	3	232
4	1.0422249	4	121

$`a=1_b=1`

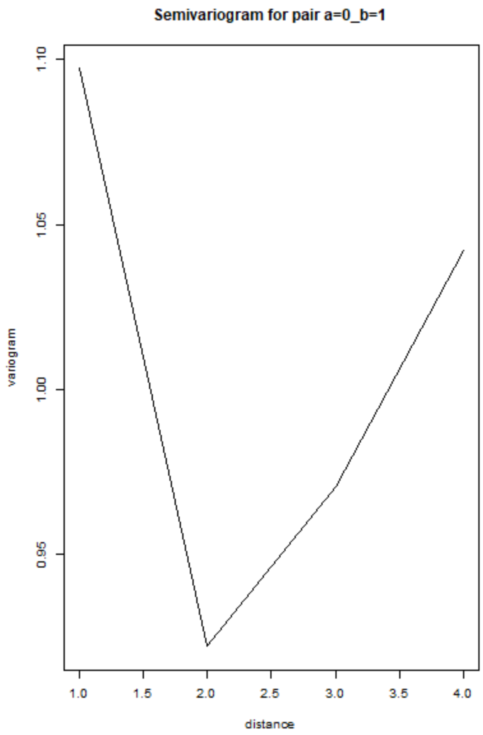
	variog	dist	n.pairs
1	1.0479601	1	398
2	0.9884016	2	304
3	0.9778377	3	210
4	0.8359854	4	97

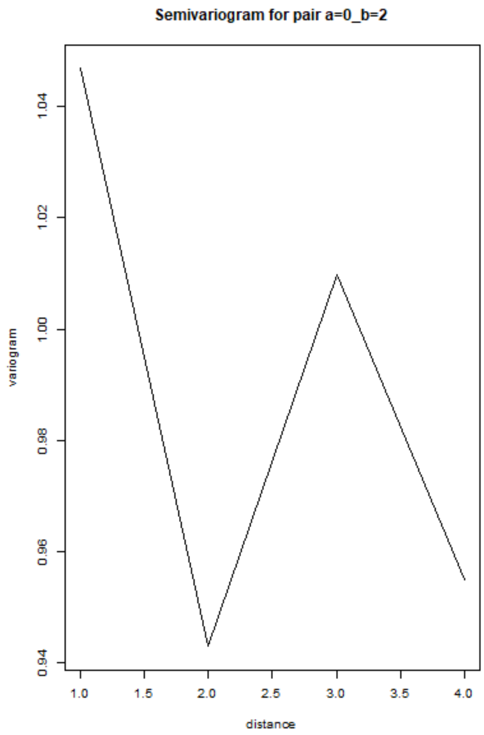
	variog	dist	n.pairs
1	1.0469750	1	504
2	0.9429442	2	367
3	1.0097347	3	243
4	0.9548429	4	110

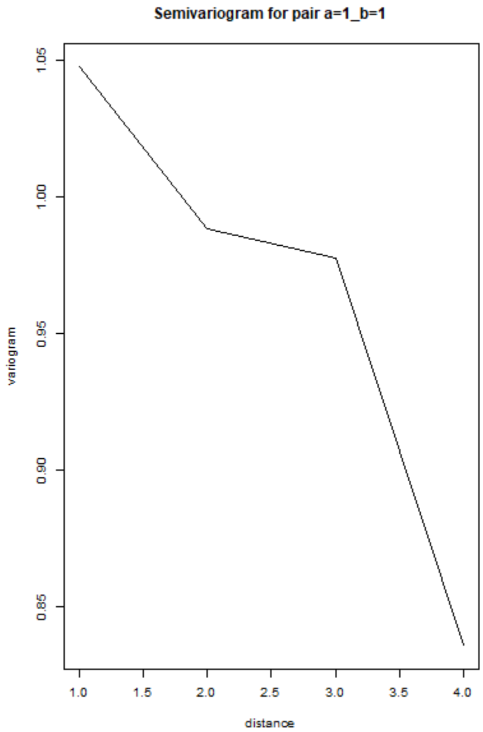
$`a=1_b=2`

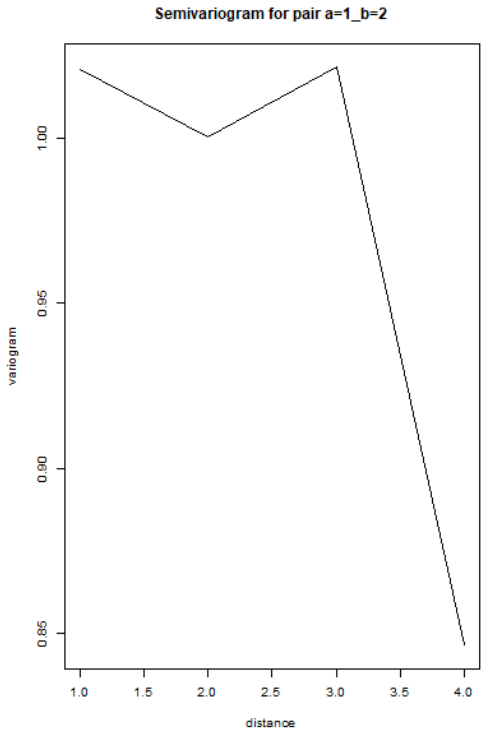
	variog	dist	n.pairs
1	1.0206838	1	419
2	1.0002901	2	336
3	1.0216352	3	213
4	0.8464707	4	111

Графики семивариограммы:









Шаг 03: Оценить зависимость корреляции значений центрированного процесса от времени в предположении его стационарности при каждом значении пары факторов A и B

```
> print(corr)
```

```
$`a=0_b=1`$corr
        [Visit = 0]  [Visit = 1]  [Visit = 2]  [Visit = 3]  [Visit = 4]
[1,]  1.00000000  -0.09757398   0.07805808   0.02942590  -0.04222494
[2,] -0.09757398   1.00000000  -0.09757398   0.07805808   0.02942590
[3,]  0.07805808  -0.09757398   1.00000000  -0.09757398   0.07805808
[4,]  0.02942590   0.07805808  -0.09757398   1.00000000  -0.09757398
[5,] -0.04222494   0.02942590   0.07805808  -0.09757398   1.00000000
$`a=0_b=2`$corr
        [Visit = 0]   [Visit = 1]   [Visit = 2]   [Visit = 3]   [Visit = 4]
[1,]  1.000000000  -0.046974985   0.05705581  -0.009734725   0.045157074
[2,] -0.046974985   1.000000000  -0.04697499   0.057055806  -0.009734725
[3,]  0.057055806  -0.046974985   1.00000000  -0.046974985   0.057055806
[4,] -0.009734725   0.057055806  -0.04697499   1.000000000  -0.046974985
[5,]  0.045157074  -0.009734725   0.05705581  -0.046974985   1.000000000
$`a=1_b=1`$corr
        [Visit = 0]  [Visit = 1]  [Visit = 2]  [Visit = 3]  [Visit = 4]
[1,]  1.00000000  -0.04796008   0.01159842   0.02216233   0.16401459
[2,] -0.04796008   1.00000000  -0.04796008   0.01159842   0.02216233
[3,]  0.01159842  -0.04796008   1.00000000  -0.04796008   0.01159842
[4,]  0.02216233   0.01159842  -0.04796008   1.00000000  -0.04796008
[5,]  0.16401459   0.02216233   0.01159842  -0.04796008   1.00000000
$`a=1_b=2`$corr
        [Visit = 0]   [Visit = 1]   [Visit = 2]   [Visit = 3]   [Visit = 4]
[1,]  1.00000000   -0.02068379  -0.00029010  -0.02163522   0.15352928
[2,] -0.02068379    1.00000000  -0.02068379  -0.00029010  -0.02163522
[3,] -0.00029010   -0.02068379   1.00000000  -0.02068379  -0.00029010
[4,] -0.02163522   -0.00029010  -0.02068379   1.00000000  -0.02068379
[5,]  0.15352928   -0.02163522  -0.00029010  -0.02068379   1.00000000
```

5. Пострить смещанную линейную с простым эффектом индивида, а также смещанную линейную модель, допускающую линейную зависимость эффекта индивида от времени.

Шаг 01: Построение смешанную линейную модель с простым эффектом индивида.

```
> # Build linear mixed effect model (lme) with simple individual effect
> q51 <- lme(fixed = Y ~ as.factor(Visit), random =~ 1|ID,
+              data = datavar9, method = "ML")
> print(q51)
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -16039.02
  Fixed: Y ~ as.factor(Visit)
     (Intercept)      as.factor(Visit)1     as.factor(Visit)2     as.factor(Visit)3
      36.0884731        1.0633315          -0.5996778            10.5155339
as.factor(Visit)4
      -8.0404377

Random effects:
 Formula: ~1 | ID
                (Intercept)   Residual
StdDev:         1.447286      9.118381

Number of Observations: 4405
Number of Groups: 881
```

Шаг 02: Построение смешанную линейную модель с линейную зависимость эффекта индивида от времени

```
> # Linear mixed model with linear dependency between individual effect and time
> q52 <- lme(fixed = Y ~ as.factor(Visit), random =~ (1 + Visit)|ID,
+            data = datavar9, method = "ML")
> print(q52)
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -16039.02
  Fixed: Y ~ as.factor(Visit)
     (Intercept)   as.factor(Visit)1   as.factor(Visit)2   as.factor(Visit)3
      36.0884731        1.0633315          -0.5996778          10.5155339
as.factor(Visit)4
      -8.0404377

Random effects:
 Formula: ~(1 + Visit) | ID
 Structure: General positive-definite, Log-Cholesky parametrization
               StdDev          Corr
(Intercept)    1.4472896795    (Intr)
Visit          0.0004759267    -0.001
Residual       9.1183808651

Number of Observations: 4405
Number of Groups: 881
```

Шаг 03: Чтобы определить если целесообразность введения коэффициента наклона случайного эффекта, мы пользуем критерия cAIC.

```
> # apply cAIC criteria to establish the best model
> tblcAIC <- data.frame(models = c(model1, model2),
+                       cAICvalues = c(q51_cAIC$caic, q52_cAIC$caic))
>
> # table of cAIC values
> print(tblcAIC)
          models              cAICvalues
1     y ~ Visit + (1|ID)       32087.86
2     y ~ Visit + (1+Visit|ID) 32099.03
```

Модель 1 является лучшей моделью, так как она имеет более низкое значение индекса cAIC.

6. С учетом результатов п.5, пострить смешанную модель зависимости наблюдаемого признака от значений признаков A и B с учетом времени наблюдения. При каждом значений фактора A проверить гипотезы аддитивности влияния фактора B и времени наблюдения а также гипотезы отсутствия влияния каждого из факторов B и времени наблюдения. Включить в модель фактор A и оценить вляние фактора A на значение наблюдаемой характеристики.

Шаг 01: Смешанная модель зависимости A, B и времени наблюдения с простым эффектом индивида: $\eta_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^T + \alpha_{ij}^{AB} + \alpha_{ik}^{AT} + \alpha_{jk}^{BT} + \alpha_{ijk}^{ABT}$

```
> # mixed model: (Y|A, B, Visit)
> # model: nijk = u + aA_i + aB_j + aT_k +
> #             aA_i * aB_j + aA_i * aT_k + aB_j * aT_k + aABT_ijk
> q61 <- lme(fixed = Y ~ as.factor(A) * as.factor(B) * as.factor(Visit),
+            random =~ 1|ID, data = datavar9, method = "ML")
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Fixed: Y ~ as.factor(A) * as.factor(B) * as.factor(Visit)
```

```
       (Intercept)
        39.6987827
      as.factor(A)1
         7.6186228
      as.factor(B)2
       -10.2483269
  as.factor(Visit)1
        -3.1754012
  as.factor(Visit)2
         0.9867510
  as.factor(Visit)3
        10.9074422
  as.factor(Visit)4
       -14.1641351
```

```
     as.factor(A)1:as.factor(B)2
                      -7.5762389
 as.factor(A)1:as.factor(Visit)1
                      -0.6507950
 as.factor(A)1:as.factor(Visit)2
                       0.3873670
 as.factor(A)1:as.factor(Visit)3
                      -0.1745915
 as.factor(A)1:as.factor(Visit)4
                      -1.1514323
 as.factor(B)2:as.factor(Visit)1
                       9.4057480
 as.factor(B)2:as.factor(Visit)2
                      -2.3935396
 as.factor(B)2:as.factor(Visit)3
                      -1.3219062
 as.factor(B)2:as.factor(Visit)4
                      13.0831640
```

```
as.factor(A)1:as.factor(B)2:as.factor(Visit)1
                                   -1.6467547
as.factor(A)1:as.factor(B)2:as.factor(Visit)2
                                   -1.2982832
as.factor(A)1:as.factor(B)2:as.factor(Visit)3
                                   -0.3697754
as.factor(A)1:as.factor(B)2:as.factor(Visit)4
                                   -0.4951229
```

```
Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:   0.7247436 6.431324

Number of Observations: 4405
Number of Groups: 881
```

Шаг 02: Модель для взаймодействия между фактора B и времени наблюдения по каждому значения фактора A: $\eta_{jk} = \mu + \alpha_j^B + \alpha_k^T + \alpha_{jk}^{BT}$

> # interaction model for factor B and time (for each value of A)

> # model: nij = u + aB_j + aT_k + aBT_jk

> q6_intBT

```
$`A=0`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -7719.06
  Fixed: Y ~ as.factor(B) * as.factor(Visit)
                      (Intercept)                    as.factor(B)2
                       39.7013396                      -10.2534407
                 as.factor(Visit)1                as.factor(Visit)2
                       -3.1777745                        0.9825991
                 as.factor(Visit)3                as.factor(Visit)4
                       10.9028203                      -14.1659785
as.factor(B)2:as.factor(Visit)1 as.factor(B)2:as.factor(Visit)2
                        9.4105095                       -2.3854028
as.factor(B)2:as.factor(Visit)3 as.factor(B)2:as.factor(Visit)4
                       -1.3128631                       13.0867806

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:   0.5251229 6.439798

Number of Observations: 2350
Number of Groups: 470
```

```
$`A=1`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -6756.805
  Fixed: Y ~ as.factor(B) * as.factor(Visit)
                      (Intercept)                    as.factor(B)2
                        47.316026                       -17.821977
                 as.factor(Visit)1                as.factor(Visit)2
                        -3.823295                         1.376311
                 as.factor(Visit)3                as.factor(Visit)4
                        10.739890                       -15.302130
as.factor(B)2:as.factor(Visit)1 as.factor(B)2:as.factor(Visit)2
                         7.753412                        -3.695932
as.factor(B)2:as.factor(Visit)3 as.factor(B)2:as.factor(Visit)4
                        -1.706449                        12.561395

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:   0.9018753 6.421413

Number of Observations: 2055
Number of Groups: 411
```

Шаг 03: Аддитивный модель для фактора B и времени наблюдения по каждому значения фактора A: $\eta_{jk} = \mu + \alpha_j^B + \alpha_k^T$

> # additive model for factor B and time (for each value of A)

> # model: njk = u + aB_j + aT_k

> q6_addBT

```
$`A=0`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -7966.97
  Fixed: Y ~ as.factor(B) + as.factor(Visit)
      (Intercept)     as.factor(B)2 as.factor(Visit)1 as.factor(Visit)2
       37.8217600        -6.4942815         1.6477217        -0.3800583
as.factor(Visit)3 as.factor(Visit)4
       10.1168904        -7.8408942

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:   0.3187778 7.172675

Number of Observations: 2350
Number of Groups: 470
```

```
$`A=1`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -6965.95
  Fixed: Y ~ as.factor(B) + as.factor(Visit)
      (Intercept)     as.factor(B)2 as.factor(Visit)1 as.factor(Visit)2
       45.7284596       -14.8425713         0.1919292        -0.6092956
as.factor(Visit)3 as.factor(Visit)4
       10.1498450        -8.9119120

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:    1.218979 7.077988

Number of Observations: 2055
Number of Groups: 411
```

Шаг 04: Аддитивный модель только для фактора времени по каждому значения фактора А: $\eta_{jk} = \mu + \alpha_k^T$

> # additive model for time (for each value of A)

> # model: njk = u + aT_k

> q6_addT

```
$`A=0`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -8185.204
  Fixed: Y ~ as.factor(Visit)
      (Intercept) as.factor(Visit)1 as.factor(Visit)2 as.factor(Visit)3
       34.5746193         1.5095455        -0.5596874         9.9510789
as.factor(Visit)4
       -7.6889004

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.002015962 7.878417

Number of Observations: 2350
Number of Groups: 470
```

```
$`A=1`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -7709.621
  Fixed: Y ~ as.factor(Visit)
      (Intercept) as.factor(Visit)1 as.factor(Visit)2 as.factor(Visit)3
       37.8196442         0.5530623        -0.6454089        11.1610178
as.factor(Visit)4
       -8.4424389

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.002904375 10.30575

Number of Observations: 2055
Number of Groups: 411
```

Шаг 05: Аддитивный модель только для фактора B по каждому значения фактора А: $\eta_{jk} = \mu + \alpha_j^B$

> # additive model for factor B (for each value of A)

>   # model: njk = u + aB_j

> q6_addB

```
$`A=0`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -8545.669
  Fixed: Y ~ as.factor(B)
  (Intercept) as.factor(B)2
    38.354372     -6.149092

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.0004598311 9.184493

Number of Observations: 2350
Number of Groups: 470
```

```
$`A=1`
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9[datavar9$A == idx, ]
  Log-likelihood: -7517.637
  Fixed: Y ~ as.factor(B)
  (Intercept) as.factor(B)2
    46.03372     -15.12015

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.0004920496 9.386562

Number of Observations: 2055
Number of Groups: 411
```

Шаг 06: Проверка гипотезы

H_BT: Гипотеза аддитивности влияния фактора B и времени наблюдения

$H_{BT} : \alpha_{jk}^{BT} = 0$

H_B: Гипотеза отсутствие влияния фактора B

$H_{BT} : \alpha_{jk}^{BT} = 0$; $H_B : \alpha_j^B = 0$

H_T: Гипотеза отсутствие влияния фактора времени наблюдения

$H_{BT} : \alpha_{jk}^{BT} = 0$; $H_T : \alpha_k^T = 0$

> # check the p-values to make the hypothesis inference

> print(tbl_hipothesis())

	A = 0	A = 1
H_BT: bBT_ij = 0	5.372022e-106	3.106080e-89
H_B: bBT_ij = 0 and bB_i = 0	2.733439e-199	0.000000e+00
H_T: bBT_ij = 0 and bT_j = 0	0.000000e+00	2.964394e-323

Шаг 07: Заключение

Для A = 0 и A = 1:

Гипотеза H_{BT} **отвергает**, тогда есть влияние взаймодействия между фактора B и времени наблюдения

Гипотезы H_{BT} **и** H_{B} **отвергают**, тогда есть аддитивности влияния фактора B

Гипотезы H_{BT} **и** H_{T} **отвергают**, тогда есть аддитивности влияния фактора времени наблюдения

Шаг 08: Построить модель с отсутсвием влияния фактора А

$$\eta_{ijk} = \mu + \alpha_j^B + \alpha_k^T + \alpha_{jk}^{BT}$$

```
> # model without factor A
> q62 <- lme(fixed = Y ~ as.factor(B) * as.factor(Visit),
+                random =~ 1|ID, data = datavar9, method = "ML")
```

```
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -14793.27
  Fixed: Y ~ as.factor(B) * as.factor(Visit)
                     (Intercept)                    as.factor(B)2
                       43.111498                       -13.628380
               as.factor(Visit)1                as.factor(Visit)2
                       -3.286819                         1.285314
               as.factor(Visit)3                as.factor(Visit)4
                       11.193019                       -14.607156
as.factor(B)2:as.factor(Visit)1 as.factor(B)2:as.factor(Visit)2
                        8.441592                        -3.140770
as.factor(B)2:as.factor(Visit)3 as.factor(B)2:as.factor(Visit)4
                       -1.860544                        12.693484

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev:    1.600061 6.785727

Number of Observations: 4405
Number of Groups: 881
```

Шаг 09: Оценить влияние фактора А на значение наблюдаемой характеристика Y

Пусть модели:

Модель 1: $\eta_{ijk} = \mu + \alpha_j^B + \alpha_k^T + \alpha_{jk}^{BT}$

Модель 2: $\eta_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^T + \alpha_{ij}^{AB} + \alpha_{ik}^{AT} + \alpha_{jk}^{BT} + \alpha_{ijk}^{ABT}$

Гипотеза H_A: $H_A : \alpha_i^A = \alpha_{ij}^{AB} = \alpha_{ik}^{AT} = \alpha_{ijk}^{ABT} = 0$

```
> # study the influence of factor A
> anov_addA <- anova(q62, q61)
> print(anov_addA)
```

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
q62	1	12	29610.53	29687.22	-14793.27			
q61	2	22	28996.08	29136.67	-14476.04	1 vs 2	634.452	<.0001

```
> print(paste("p-value = ", anov_addA$"p-value"[2]))
[1] "p-value =  7.26550582148688e-130"
```

Шаг 10: Заключение

Гипотеза H_A **отвергает**, тогда есть влияние фактора A на значение наблюдаемой характеристики.

```
> # check cAIC values
> q62_aic <- cAIC(q62)$caic
> print(q62_aic)
[1] 29586.8
> q61_aic <- cAIC(q61)$caic
> print(q61_aic)
[1] 28996.23
```

С критерием AIC, тоже видем что модель с влияния фактора A – это лучшее модель чем модель с отсутствием влияния фактора A.

8. Построить смешанную модель ковариационного анализа в предположении полиномиальной зависимости второго порядка наблюодаемого признака от времени. Проверить гипотезу линейности зависимости средного значения наблюдаемого признака от времени в пристствии факторов A и B.

Шаг 01: Пострить смешанную модель для полиномиальной зависимости второго порядка наблюдаемого признака от времени

$$\eta_{ijkl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^T \cdot T + \alpha_l^{T^2} \cdot T^2 + \alpha_{ij}^{AB} + \alpha_{ik}^{AT} \cdot T + \alpha_{jk}^{BT} \cdot T + \alpha_{il}^{AT^2} \cdot T^2 + \alpha_{jl}^{BT^2} \cdot T^2 + \alpha_{ijk}^{ABT} \cdot T + \alpha_{ijl}^{ABT^2} \cdot T^2$$

```
> # build model mixed-analysis for 2nd polynomial order in time
> # model: nijkl = u + aA_i + aB_j + aT_k * T + aT^2_l *T^2 +
> #           aAB_ij + aAT_ik * T + aBT_jk * T + aAT^2_il * T^2 + aBT^2_jl * T^2 +
> #           aABT_ijk * T + aABT^2_ijl * T^2
> q8.abt_square <- lme(fixed = Y ~ as.factor(A) * as.factor(B) * Visit +
+                  as.factor(A) * as.factor(B) * Visit2,
+                  random =~ 1|ID, data = datavar9, method = "ML")
> print(q8.abt_square)
```

```
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -15637.05
  Fixed: Y ~ as.factor(A) * as.factor(B) * Visit + as.factor(A) * as.factor(B) *      Visit2
                       (Intercept)                      as.factor(A)1
                        36.1461399                          6.8813034
                     as.factor(B)2                              Visit
                        -6.4161795                          9.4532502
                            Visit2          as.factor(A)1:as.factor(B)2
                        -2.7506763                         -7.3154050
              as.factor(A)1:Visit                 as.factor(B)2:Visit
                         1.1078696                         -5.0909960
             as.factor(A)1:Visit2                as.factor(B)2:Visit2
                        -0.2627845                          1.6993514
 as.factor(A)1:as.factor(B)2:Visit as.factor(A)1:as.factor(B)2:Visit2
                        -1.9759779                          0.4198586

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.0006608732 8.422465

Number of Observations: 4405
Number of Groups: 881
```

```
> # check the significance for each coefficient
> anova(q8.abt_square)
```

```
> anova(q8.abt_square)
                                    numDF denDF  F-value p-value
(Intercept)                             1  3514 83301.34  <.0001
as.factor(A)                            1   879   150.79  <.0001
as.factor(B)                            1  3514  1652.90  <.0001
Visit                                   1  3514    67.47  <.0001
Visit2                                  1  3514   653.11  <.0001
as.factor(A):as.factor(B)               1  3514   294.38  <.0001
as.factor(A):Visit                      1  3514     0.15  0.7017
as.factor(B):Visit                      1  3514    76.13  <.0001
as.factor(A):Visit2                     1  3514     0.02  0.8914
as.factor(B):Visit2                     1  3514   155.45  <.0001
as.factor(A):as.factor(B):Visit         1  3514     0.68  0.4110
as.factor(A):as.factor(B):Visit2        1  3514     1.90  0.1683
```

Шаг 02: Постриить смешанную модель с линейности зависимости средного значения наблдаемого признака от времени в присутствии факторов А и В

$$\eta_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^T \cdot T + \alpha_{ij}^{AB} + \alpha_{ik}^{AT} \cdot T + \alpha_{jk}^{BT} \cdot T + \alpha_{ijk}^{ABT} \cdot T$$

```
> # build model mixed-analysis for linear dependency
> # model: nijk = u + aA_i + aB_j + aT_j * T + aAB_ij + aAT_ik * T + aBT_jk +
                   aABT_ijk * T
> q8.abt_linear <- lme(fixed = Y ~ as.factor(A) * as.factor(B) * Visit,
+                 random =~ 1|ID, data = datavar9, method = "ML")
> print(q8.abt_linear)
```

```
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -15999.27
  Fixed: Y ~ as.factor(A) * as.factor(B) * Visit
                    (Intercept)                     as.factor(A)1
                    41.58498944                        7.60481559
                  as.factor(B)2                             Visit
                    -9.73325341                       -1.60139570
      as.factor(A)1:as.factor(B)2               as.factor(A)1:Visit
                    -8.40876955                        0.05699598
             as.factor(B)2:Visit as.factor(A)1:as.factor(B)2:Visit
                     1.77965432                       -0.30380007

Random effects:
 Formula: ~1 | ID
         (Intercept) Residual
StdDev: 0.0006331532 9.144316

Number of Observations: 4405
Number of Groups: 881
```

Шаг 03: Проверка гипотеза линейности зависимости

$$H_{T^2} : \alpha_l^{T^2} = \alpha_{il}^{AT^2} = \alpha_{jl}^{BT^2} = \alpha_{ijl}^{ABT^2} = 0$$

```
> # proof of hypothesis for linear dependency
> # H_T^2: aABT^2_ijl = 0 and aAT^2_il = 0 and aBT^2_jl = 0 and aT^2_l = 0
```

```
> q8_anova1 <- anova.lme(q8.abt_linear, q8.abt_square)
> print(q8_anova1)
              Model   df    AIC       BIC       logLik    Test   L.Ratio   p-value
q8.abt_linear     1   10  32018.55  32082.45  -15999.27
q8.abt_square     2   14  31302.10  31391.57  -15637.05 1 vs 2 724.4463  <.0001
>
> # check the p-value for H_T^2
> q8_pvalue1 <- q8_anova1$"p-value"[2]
> print(q8_pvalue1)
[1] 1.772828e-155
```

Шаг 04: Заключение:

Гипотеза отвергает, тогда есть влияние по полиномиальной зависимости второго порядка.

```
> # check the AIC values
> q8square_caic <- cAIC(q8.abt_square)$caic
> print(q8square_caic)
[1] 31308.41
> q8linear_caic <- cAIC(q8.abt_linear)$caic
> print(q8linear_caic)
[1] 32029.95
```

С пользованием условного информационного критерия Акайке (cAIC), видем что модель второго порядка для времени наблюдения – это лучшее чем просто линейной зависимости

9. С использованием информационных критериев AIC и BIC, выбрать найлучшую модель для неслучайного эффекта в рамках смешанной модели с простым эффектом индивида. Провести исследование влияния факторов A, B и времени наблюдения на значение исследуемой характеристики в рамках данной модели.

Сравнить результаты, полученные с использованием AIC, и выбор найлучшей модели с учетом случайнного эффекта при помощи cAIC.

Шаг 01: Проверяем значения информационных критериев AIC и BIC для каждой рассматриваемой модели.

> # build the information-criteria table which contains AIC and BIC values

> IC_table <- data.frame(model = func, AIC = aic_values, BIC = bic_values)

> print(IC_table)

```
                                      model      AIC      BIC
1                              Y ~ A*B*Visit 32018.55 32082.45
2                  Y ~ A*B + A*Visit + B*Visit 32017.15 32074.66
3                        Y ~ A*B + A*Visit 32085.20 32136.32
4                        Y ~ A*B + B*Visit 32015.39 32066.52
5                    Y ~ A*Visit + B*Visit 32273.62 32324.74
6                          Y ~ A*Visit + B 32337.57 32382.30
7                          Y ~ A + B*Visit 32271.64 32316.37
8                          Y ~ A*B + Visit 32083.33 32128.06
9                              Y ~ A*Visit 33487.26 33525.60
10                             Y ~ B*Visit 32389.20 32427.54
11                       Y ~ A + B + Visit 32335.57 32373.91
12                           Y ~ A + Visit 33485.34 33517.29
13                           Y ~ B + Visit 32449.83 32481.78
14                               Y ~ Visit 33574.13 33599.69
15             Y ~ A*B*Visit + A*B*Visit2 31302.10 31391.57
16       Y ~ A*B*Visit + A*Visit2 + B*Visit2 31302.00 31385.08
17                 Y ~ A*B*Visit + A*Visit2 32018.55 32082.45
18                 Y ~ A*B*Visit + B*Visit2 31300.10 31376.79
19 Y ~ A*B + A*Visit + A*Visit2 + B*Visit+ B*Visit2 31300.68 31377.37
20               Y ~ A*B + A*Visit + A*Visit2 31522.85 31586.75
21               Y ~ A*B + B*Visit + B*Visit2 31297.06 31360.96
22                   Y ~ A*B + Visit + Visit2 31519.00 31570.13
23     Y ~ A*Visit + A*Visit2 + B*Visit + B*Visit2 31584.30 31654.59
24                   Y ~ A*Visit + A*Visit2 33096.96 33148.08
25                   Y ~ B*Visit + B*Visit2 31712.15 31763.27
26                             Y ~ A*Visit2 33398.01 33436.35
27                             Y ~ B*Visit2 32209.50 32247.84
28               Y ~ A + B + Visit + Visit2 31788.69 31833.42
29                   Y ~ A + Visit + Visit2 33093.42 33131.76
30                   Y ~ B + Visit + Visit2 31918.13 31956.48
31                       Y ~ Visit + Visit2 33190.61 33222.56
```

Шаг 02: Выбираем лучшую модель по критериям AIC и BIC.

> # take the best model for AIC criteria

> best_modelAIC <- IC_table[IC_table$AIC == min(IC_table$AIC),]

> print(best_modelAIC)

id	model	AIC	BIC
21	Y ~ A*B + B*Visit + B*Visit2	31297.06	31360.96

> # take the best model for BIC criteria

> best_modelBIC <- IC_table[IC_table$BIC == min(IC_table$BIC),]

> print(best_modelBIC)

id	model	AIC	BIC
21	Y ~ A*B + B*Visit + B*Visit2	31297.06	31360.96

Шаг 03: Провести исследование влияния факторов A, B и времени наблюдения на значение времени наблюдаемой характеристики в самой лучшей модели.

Самое лучшей модель: Y ~ A*B + B*Visit + B*Visit2

$$\eta_{ijkl} = \mu + \alpha_i^A + \alpha_j^B + \alpha_k^T + \alpha_l^{T^2} + \alpha_{ij}^{AB} + \alpha_{jk}^{BT} + \alpha_{jl}^{BT^2}$$

> # check the parameters inside of the optimal model

> opt_id <- IC_table[IC_table$AIC == min(IC_table$AIC),]$id

> print(paste("opt-id = ", opt_id))

[1] "opt-id = 21"

> model_opt <- lme(fixed = form[[opt_id]],

+   random =~ 1|ID, data = datavar9, method = "ML")

> print(model_opt)

```
Linear mixed-effects model fit by maximum likelihood
  Data: datavar9
  Log-likelihood: -15638.53
  Fixed: form[[opt_id]]
              (Intercept)            as.factor(A)1
                35.858952                 7.516174
            as.factor(B)2                    Visit
                -5.751165                 9.963029
                   Visit2 as.factor(A)1:as.factor(B)2
                -2.871588                -8.740373
      as.factor(B)2:Visit       as.factor(B)2:Visit2
                -6.010975                 1.894432

Random effects:
 Formula: ~1 | ID
        (Intercept) Residual
StdDev: 0.000660717 8.425295

Number of Observations: 4405
Number of Groups: 881
```

Значимость параметров оптимальной модели:

> # check the significance of the parameters in the optimal model

> anova_opt <- anova(model_opt)

> print(anova_opt)

```
                             numDF denDF  F-value p-value
(Intercept)                      1  3518 83321.19  <.0001
as.factor(A)                     1   879   150.82  <.0001
as.factor(B)                     1  3518  1653.30  <.0001
Visit                            1  3518    67.49  <.0001
Visit2                           1  3518   653.27  <.0001
as.factor(A):as.factor(B)        1  3518   294.45  <.0001
as.factor(B):Visit               1  3518    76.02  <.0001
as.factor(B):Visit2              1  3518   155.39  <.0001
```

- отсутствие влияния фактор А: $H_A : \alpha_i^A = \alpha_{ij}^{AB} = 0$

```
> # study of ausence in influence of factor A
> id_m9A <- as.numeric(which(func == "Y ~ B*Visit + B*Visit2"))
> model_m9A <- lme(fixed = form[[id_m9A]], random =~ 1|ID,
+                     data = datavar9, method = "ML")
> anova_m9A <- anova(model_m9A, model_opt)
> print(anova_m9A)
          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model_m9A     1  8 31712.15 31763.27 -15848.07
model_opt     2 10 31297.06 31360.96 -15638.53 1 vs 2 419.0892  <.0001
> print(anova_m9A$"p-value"[2])   # reject HA
[1] 9.906854e-92
```

Отвергает H_A, тогда есть влияния фактора А в найлучшей модели.

- отсутствие влияния фактор В: $H_B : \alpha_j^B = \alpha_{ij}^{AB} = \alpha_{jk}^{BT} = \alpha_{jl}^{BT^2} = 0$

```
> # study of ausence in influence of factor B
> id_m9B <- as.numeric(which(func == "Y ~ A + Visit + Visit2"))
> model_m9B <- lme(fixed = form[[id_m9B]], random =~ 1|ID,
+                     data = datavar9, method = "ML")
> anova_m9B <- anova(model_m9B, model_opt)
> print(anova_m9B)
          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model_m9B     1  6 33093.42 33131.76 -16540.71
model_opt     2 10 31297.06 31360.96 -15638.53 1 vs 2 1804.358  <.0001
> print(anova_m9B$"p-value"[2])   # reject HB
[1] 0
```

Отвергает H_B, тогда есть влияния фактора В в найлучшей модели.

- отсутствие влияния фактор времени наблюдения:

$H_T : \alpha_k^T = \alpha_l^{T^2} = \alpha_{jk}^{BT} = \alpha_{jl}^{BT^2} = 0$

```
> # study of ausence in influence of time
> model_m9T <- lme(fixed = Y ~ as.factor(A) * as.factor(B), random =~ 1|ID,
+                   data = datavar9, method = "ML")
> anova_m9T <- anova(model_m9T, model_opt)
> print(anova_m9T)
          Model df      AIC      BIC    logLik   Test  L.Ratio p-value
model_m9T     1  6 32139.65 32178.00 -16063.83
model_opt     2 10 31297.06 31360.96 -15638.53 1 vs 2 850.5932  <.0001
> print(anova_m9T$"p-value"[2])   # reject HT
[1] 8.42858e-183
```

Отвергает H_T, тогда есть влияния фактора врмени в найлучшей модели.

Шаг 04: Сравнить результаты и выбор найлучшей модели с учетом случайнного эффеута при помощи cAIC.

Модель: Y ~ A*B + B*Visit + B*Visit2

> # compare results of AIC and cAIC for best model

> aic_best <- AIC(model_opt)

> print(aic_best)

[1] 31297.06

> caic_best <- cAIC(model_opt)

> print(caic_best$caic)

[1] 31303.36

10. Посчитать частные и совместные доверительные интервалы (прнимает с простой эффект индивида).

Шаг 01: Частное доверительные доверительные интервалы: $\alpha = 0.05$

> intervals(q10, which = "fixed", level = 1 - alpha)

```
> # take the confidence interval for the set of model parameters
> print(func[opt_id])
[1] "Y ~ A*B + B*Visit + B*Visit2"
> q10 <- lme(fixed = form[[opt_id]], random =~ 1|ID,
+                   data = datavar9, method = "ML")
> intervals(q10, which = "fixed")
Approximate 95% confidence intervals

 Fixed effects:
                               lower       est.      upper
(Intercept)                35.042048 35.858952 36.675855
as.factor(A)1               6.803030  7.516174  8.229318
as.factor(B)2              -6.897780 -5.751165 -4.604551
Visit                       9.072449  9.963029 10.853609
Visit2                     -3.084347 -2.871588 -2.658829
as.factor(A)1:as.factor(B)2 -9.738608 -8.740373 -7.742138
as.factor(B)2:Visit        -7.252795 -6.010975 -4.769156
as.factor(B)2:Visit2        1.596742  1.894432  2.192121
```

Шаг 02: Совместное доверительные доверительные интервалы

- использовать метод Bonferroni: $\alpha_d = \alpha / d$

```
> alpha2 <- alpha / length(q10$coefficients[[1]])
> print(paste("alpha for bonferroni = ", alpha2))
[1] "alpha for bonferonii =  0.00625"
```

```
> intervals(q10, which = "fixed", level = 1 - alpha2)
```

```
 Fixed effects:
                                  lower       est.      upper
(Intercept)                   34.718986 35.858952 36.998918
as.factor(A)1                  6.520228  7.516174  8.512120
as.factor(B)2                 -7.351234 -5.751165 -4.151097
Visit                          8.720249  9.963029 11.205809
Visit2                        -3.168488 -2.871588 -2.574688
as.factor(A)1:as.factor(B)2 -10.133382 -8.740373 -7.347364
as.factor(B)2:Visit           -7.743901 -6.010975 -4.278050
as.factor(B)2:Visit2           1.479014  1.894432  2.309849
```