

Медицинская статистика - 3

Докладчик:

д.ф.ф.-м.н. Исмаил Гафаров

Корреляционный анализ

Корреляция

(от лат. *correlatio* «взаимосвязь»)
или корреляционная зависимость —
статистическая взаимосвязь двух или
более случайных величин (либо
величин, которые можно с некоторой
допустимой степенью точности считать
таковыми)

Корреляционный анализ

Линейная или нелинейная корреляция

А	Б
2	7
4	10
5	9
6	12
10	16
14	20
20	25
22	27
29	35
32	36
40	45

А	Б
2	5
4	15
5	25
6	35
10	100
14	200
20	402
22	450
29	950
32	1100
40	1550

Корреляционный анализ

Прямая (+) или обратная (-) корреляция

Уровень гемоглобина	Содержание железа
10,2	4
10,3	4,1
10,3	4,2
10,5	4,3
10,5	4,5
10,6	4,6
10,6	4,6
10,7	4,6
10,9	4,8
11	4,9
11,2	5
11,3	5,1
11,5	5,2

$$r = 0,094$$

Возраст новорожденного, дни	ЧСС
1	175
2	170
3	165
5	160
8	160
10	158
12	155
15	149
29	135
32	132
40	135

$$r = - 0,096$$

Корреляционный анализ

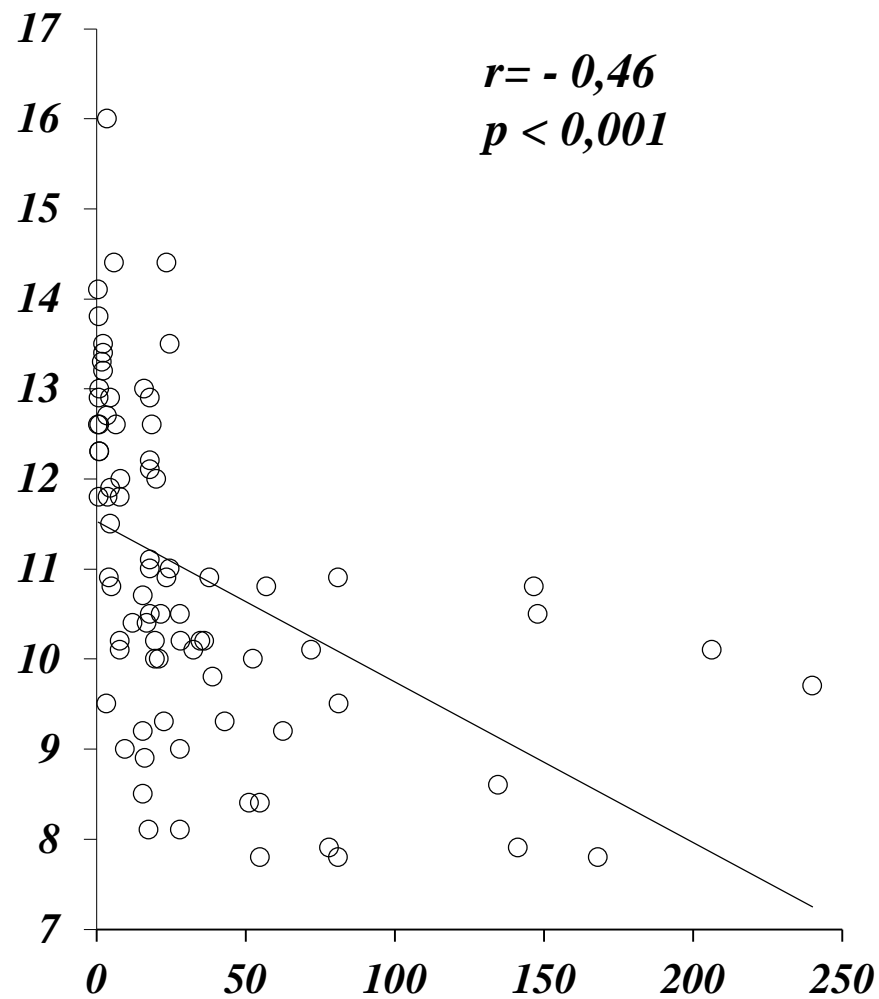
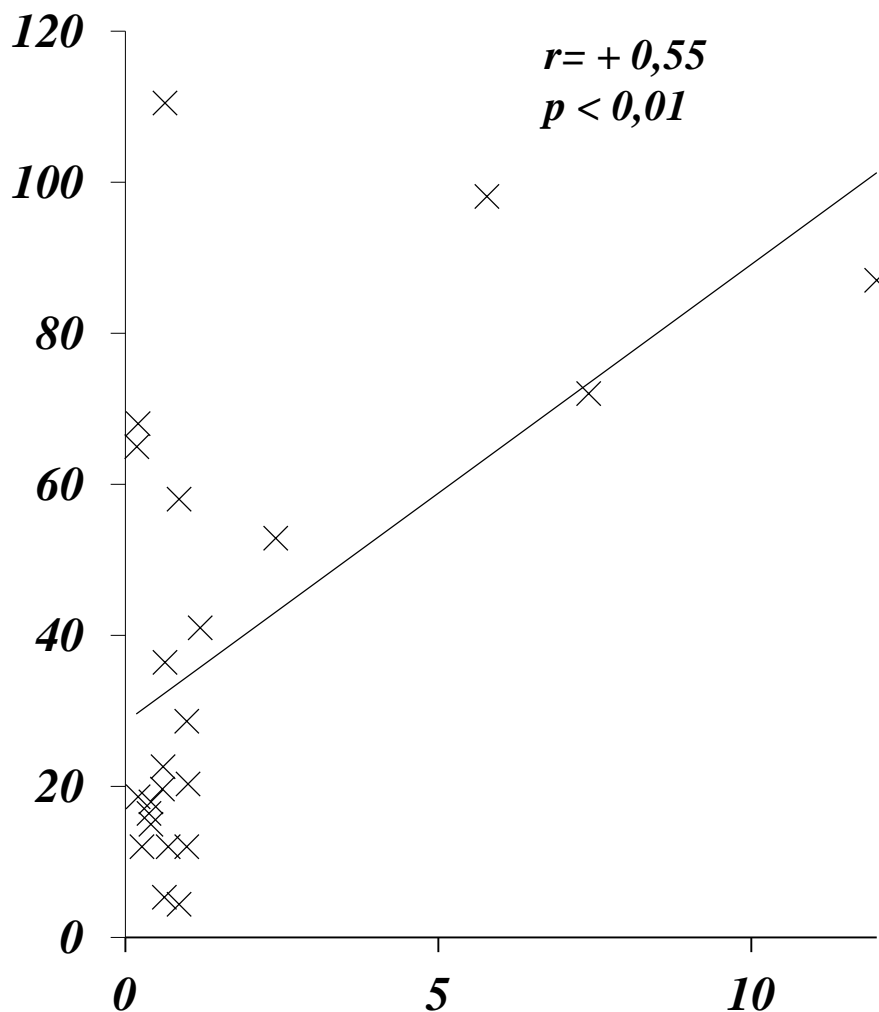
Количественная или качественная корреляция

Уровень гемоглобина	Содержание железа
10,2	4
10,3	4,1
10,3	4,2
10,5	4,3
10,5	4,5
10,6	4,6
10,6	4,6
10,7	4,6
10,9	4,8
11	4,9
11,2	5
11,3	5,1
11,5	5,2

Возрастная группа	Средний САД
1 (до 20 лет)	115,1
2 (21-30 лет)	120,2
3 (31-40 лет)	122,6
4 (41-50 лет)	128,3
5 (51-60 лет)	130,1
6 (61-70 лет)	132,3
7 (71-80 лет)	135,4
8 (> 80 лет)	139,9

Корреляционный анализ

Прямая (+) или обратная (-) корреляция



Корреляционный анализ

«Миф» о коэффициенте корреляции

Если $0,30 < r < 0,50$ – то слабая корреляция

Если $0,50 < r < 0,70$ – то умеренная корреляция

Если $0,70 < r$ – то сильная корреляция

Абсолютная ложь

Коэффициент корреляции должен сопровождаться статистической значимостью, которая оценивается односторонним или двусторонним критерием значимости:

Например:

при $n = 1000$; $r = 0,07$; $p = 0,027$ – стат. достоверная

при $n = 10$; $r = 0,70$; $p = 0,051$ – стат. недостоверная

Корреляционный анализ

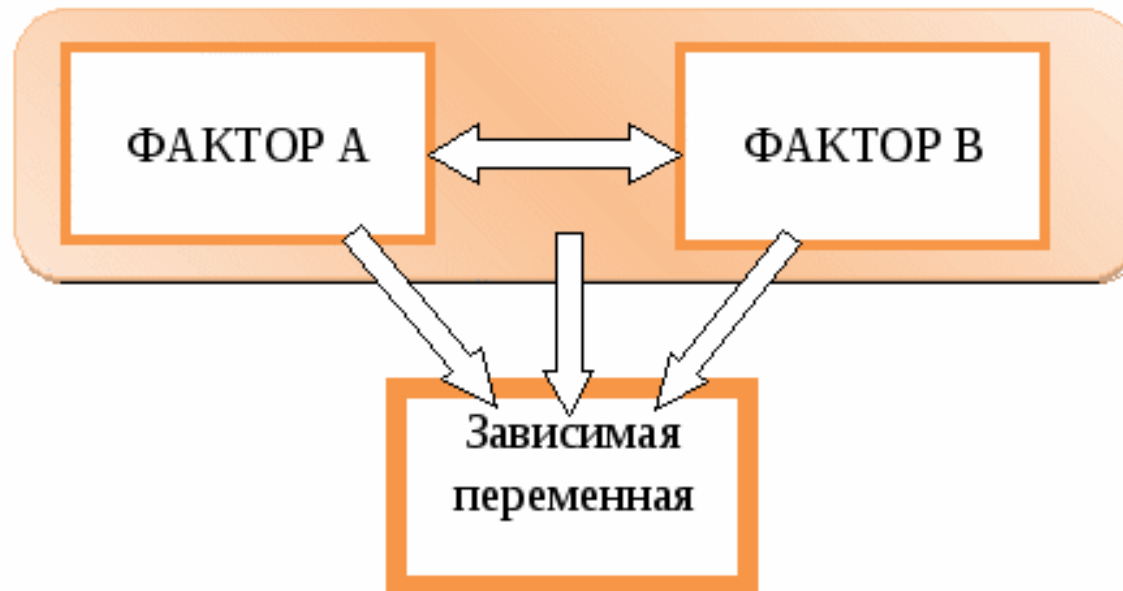
- **Линейная корреляция (с Z-преобразованием Фишера)**
- **Нелинейная корреляция**
- **Корреляция Пирсона (для качественных данных)**
- **Корреляция Тай-б-Кэндала**
- **Корреляция Спирмена**
- **и т.д.**



Дисперсионный анализ

Задачей дисперсионного анализа является изучение влияния одного или нескольких факторов на рассматриваемый признак.

Иногда ее называют «тест ANOVA»



Дисперсионный анализ

- Однофакторный дисперсионный анализ
- Двухфакторный дисперсионный анализ
- Многофакторный дисперсионный анализ

$$F_{расч} = \frac{d_{факт}}{d_{ост}}$$

Дисперсионный анализ (тест ANOVA)

<i>ANOVA</i>						
		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
<i>Area</i>	<i>Between Groups</i>	349417,653	1	349417,653	1,842	,122
	<i>Within Groups</i>	150804,896	214	704,696		
	<i>Total</i>	500222,548	215			
<i>Feret Mean</i>	<i>Between Groups</i>	285,495	1	285,495	6,291	,023
	<i>Within Groups</i>	615,324	214	2,875		
	<i>Total</i>	900,819	215			
<i>Shape Factor</i>	<i>Between Groups</i>	5,309	1	5,309	12,688	,000
	<i>Within Groups</i>	,879	214	,004		
	<i>Total</i>	6,188	215			



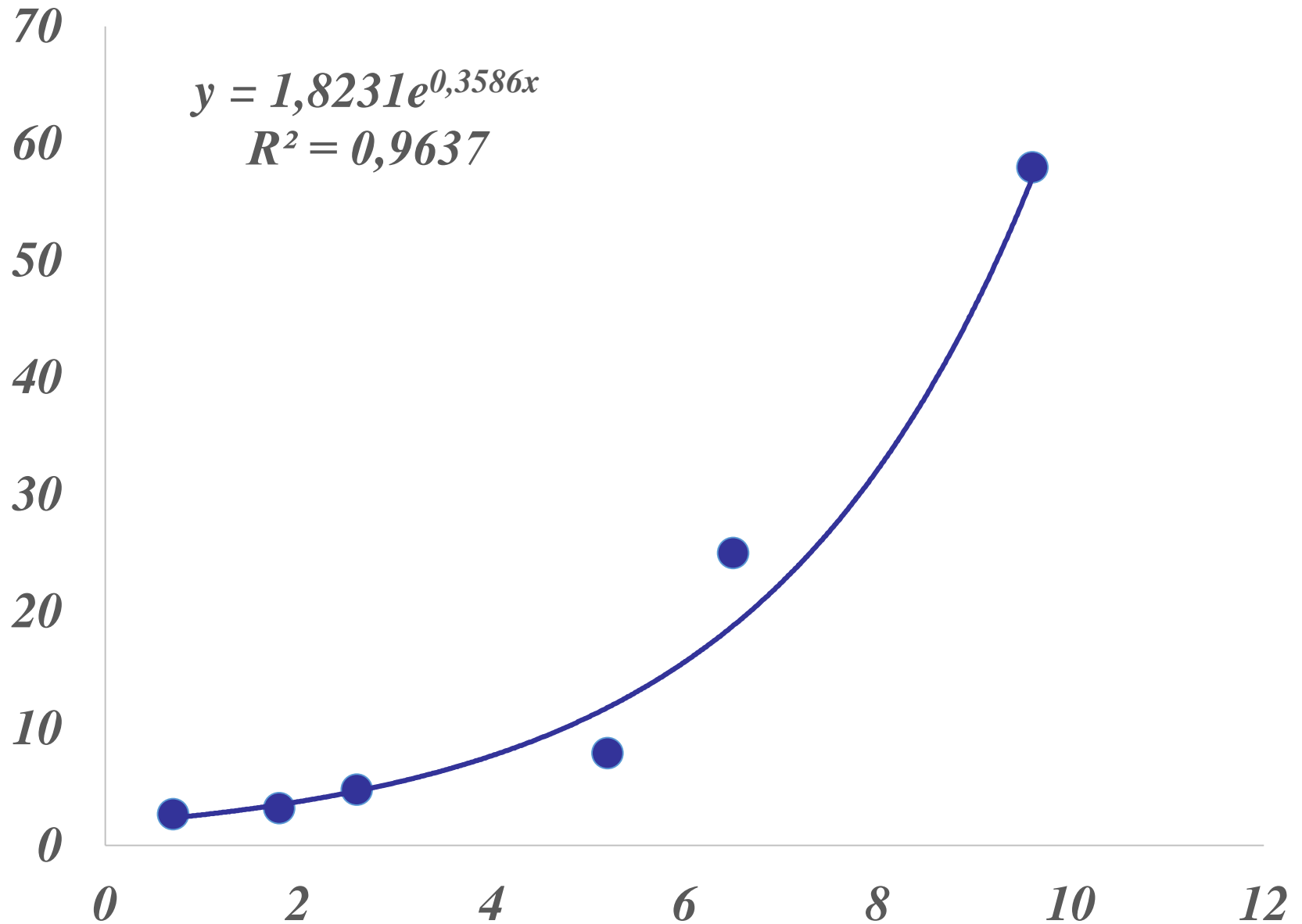
Регрессионный анализ

Регрессионный анализ статистический метод исследования влияния одной или нескольких независимых переменных на зависимую переменную

Виды регрессионных анализов

- Линейная
- Нелинейная
- Порядковая
- Категориальная
- Логистическая
- Мультиномиальная
- Метод наименьших квадратов
- Регрессия Кокса и т.д.

Регрессионный анализ



Регрессионный анализ

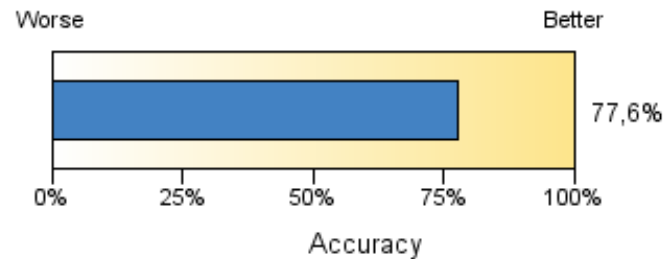
Automatic Linear Modeling

Case Processing Summary		
	N	Percent
Included	238	100,0%
Excluded	0	0,0%
Total	238	100,0%

Model Summary

Target	Neonolum
Automatic Data Preparation	On
Model Selection Method	Forward Stepwise
Information Criterion	1 411,999

The information criterion is used to compare to models. Models with smaller information criterion values fit better.



Регрессионный анализ

Automatic Data Preparation

Target: Neonolum

Field	Role	Actions Taken
(Agleke_transformed)	Predictor	Trim outliers Replace missing values
(Akrosian_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Albumin_transformed)	Predictor	Trim outliers Replace missing values
(ALT_transformed)	Predictor	Trim outliers Replace missing values
(Anem_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(AOD_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Apqar1_transformed)	Predictor	Trim outliers Replace missing values
(Apqar2_transformed)	Predictor	Trim outliers Replace missing values
(asim_transformed)	Predictor	Change measurement level from continuous to ordinal
(AST_transformed)	Predictor	Trim outliers Replace missing values
(ASTALT_transformed)	Predictor	Trim outliers Replace missing values

Регрессионный анализ

Automatic Data Preparation

Target: Neonolum

Field	Role	Actions Taken
(Ca_transformed)	Predictor	Trim outliers Replace missing values
(Cift_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(dinam_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Dsay_transformed)	Predictor	Trim outliers Replace missing values
(DT_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Duzbil_transformed)	Predictor	Trim outliers Replace missing values
(FMA_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Gx_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Hb_transformed)	Predictor	Trim outliers Replace missing values
(Hepatom_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Hiptut_transformed)	Predictor	Change measurement level from continuous to ordinal
(Hofn_transformed)	Predictor	Trim outliers

If the original field name is X, then the transformed field is displayed (X_transformed). The original field is excluded from the analysis and the transformed field is included instead.

Регрессионный анализ

Automatic Data Preparation

Target: Neonolum

Field	Role	Actions Taken
(Agleke_transformed)	Predictor	Trim outliers Replace missing values
(Akrosian_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Albumin_transformed)	Predictor	Trim outliers Replace missing values
(ALT_transformed)	Predictor	Trim outliers Replace missing values
(Anem_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(AOD_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Apqar1_transformed)	Predictor	Trim outliers Replace missing values
(Apqar2_transformed)	Predictor	Trim outliers Replace missing values
(asim_transformed)	Predictor	Change measurement level from continuous to ordinal
(AST_transformed)	Predictor	Trim outliers Replace missing values
(ASTALT_transformed)	Predictor	Trim outliers Replace missing values
(BAIR_transformed)	Predictor	Trim outliers Replace missing values
(BDI_transformed)	Predictor	Trim outliers Replace missing values
(BDIL_transformed)	Predictor	Change measurement level from continuous to ordinal
(Beyodem_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values

If the original field name is X, then the transformed field is displayed as (X_transformed). The original field is excluded from the analysis and the transformed field is included instead.

Automatic Data Preparation

Target: Neonolum

Field	Role	Actions Taken
(Ca_transformed)	Predictor	Trim outliers Replace missing values
(Cift_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(dinam_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Dsay_transformed)	Predictor	Trim outliers Replace missing values
(DT_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Duzbil_transformed)	Predictor	Trim outliers Replace missing values
(FMA_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values Merge categories to maximize association with target
(Gx_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Hb_transformed)	Predictor	Trim outliers Replace missing values
(Hepatom_transformed)	Predictor	Change measurement level from continuous to ordinal Replace missing values
(Hiptut_transformed)	Predictor	Change measurement level from continuous to ordinal
(Hofn_transformed)	Predictor	Trim outliers

If the original field name is X, then the transformed field is displayed as (X_transformed). The original field is excluded from the analysis and the transformed field is included instead.

Automatic Data Preparation

Target: Neonolum

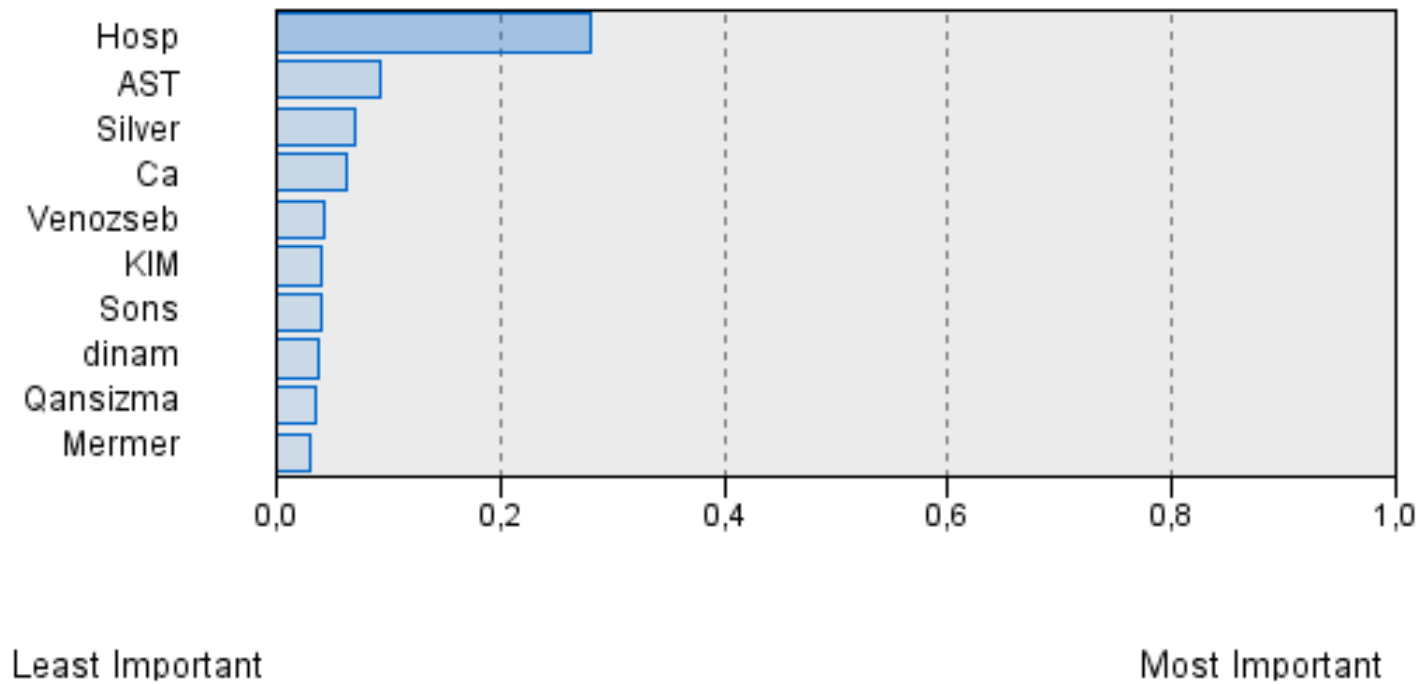
Field	Role	Actions Taken
(Hosp_transformed)	Predictor	Trim outliers Replace missing values
(Hsay_transformed)	Predictor	Trim outliers Replace missing values
(HY_transformed)	Predictor	Trim outliers
(IFABPser_transformed)	Predictor	Trim outliers Replace missing values
(IFABPur_transformed)	Predictor	Trim outliers Replace missing values
(K_transformed)	Predictor	Trim outliers Replace missing values
(KIM_transformed)	Predictor	Trim outliers Replace missing values
(Kreatinin_transformed)	Predictor	Trim outliers Replace missing values
(Kutle_transformed)	Predictor	Trim outliers
(LBP_transformed)	Predictor	Trim outliers Replace missing values
(Leyk_transformed)	Predictor	Trim outliers Replace missing values
(LFABPser_transformed)	Predictor	Trim outliers Replace missing values
(LFABPur_transformed)	Predictor	Trim outliers Replace missing values
(LII_transformed)	Predictor	Trim outliers Replace missing values
(Lkaiser_transformed)	Predictor	Trim outliers Replace missing values

If the original field name is X, then the transformed field is displayed as (X_transformed). The original field is excluded from the analysis and the transformed field is included instead.

Регрессионный анализ

Predictor Importance

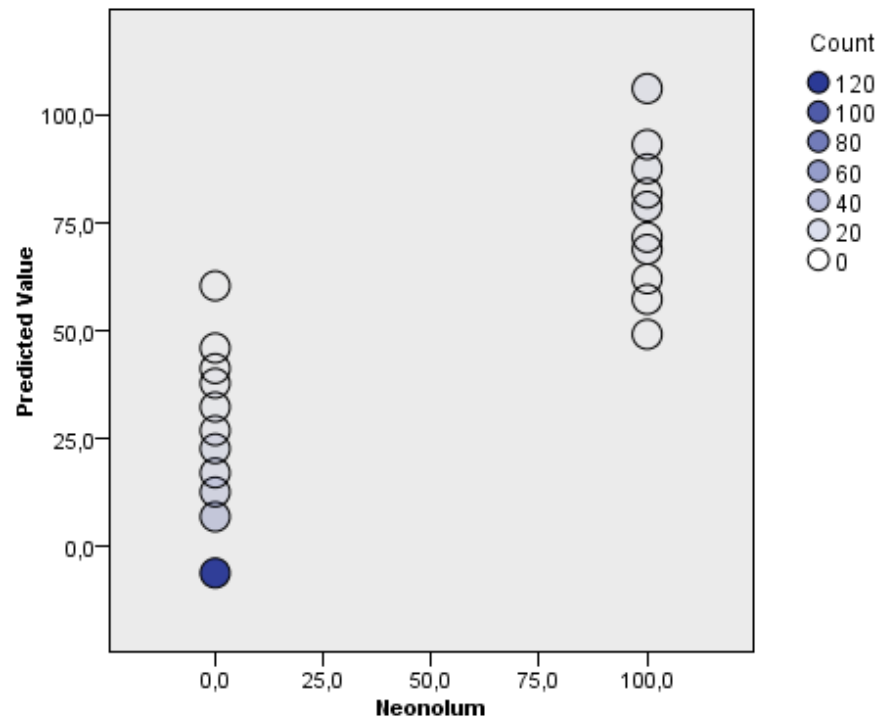
Target: Neonolum



Регрессионный анализ

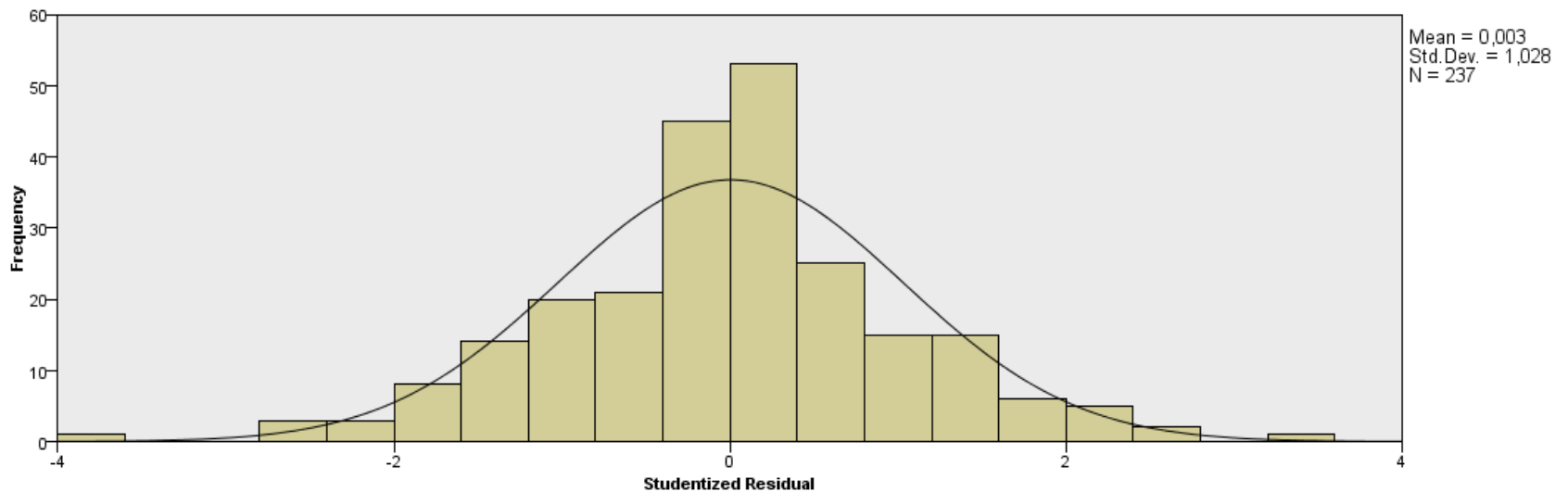
Predicted by Observed

Target: Neonolum



Регрессионный анализ

Residuals
Target: Neonolum

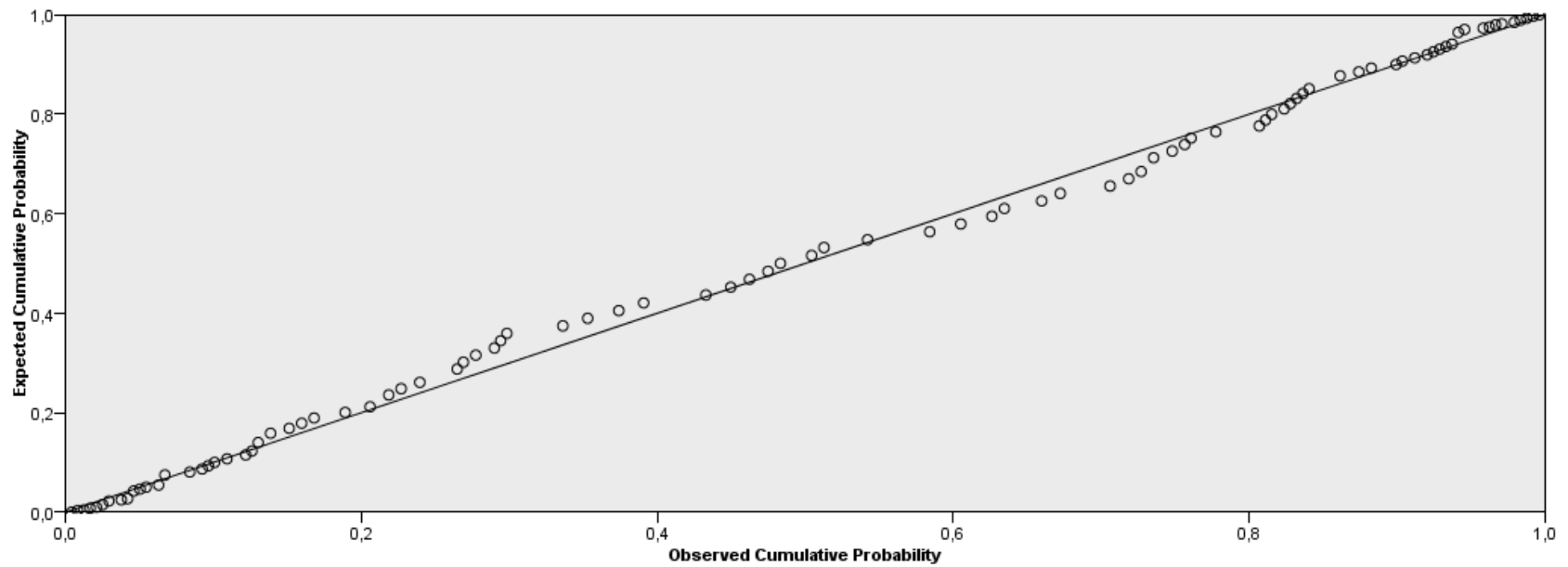


The histogram of Studentized residuals compares the distribution of the residuals to a normal distribution. The smooth line represents the normal distribution. The closer the frequencies of the residuals are to this line, the closer the distribution of the residuals is to the normal distribution.

Регрессионный анализ

Residuals

Target: Neonolum



The P-P plot of Studentized residuals compares the distribution of the residuals to a normal distribution. The diagonal line represents the normal distribution. The closer the observed cumulative probabilities of the residuals are to this line, the closer the distribution of the residuals is to the normal distribution.

Регрессионный анализ

Outliers

Target: Neonolum

Record ID	Neonolum	Cook's Distance
211	100,0	0,135
198	100,0	0,094
72	0,0	0,092
53	0,0	0,070
238	100,0	0,051
224	100,0	0,051
221	100,0	0,051
101	0,0	0,048
201	100,0	0,040
45	0,0	0,039
46	0,0	0,037
22	0,0	0,035
118	0,0	0,035
235	100,0	0,033
111	0,0	0,032
19	0,0	0,031
121	0,0	0,030
94	0,0	0,029

Records with large Cook's distance values are highly influential in the mod computations. Such records may distort the model accuracy.

Outliers

Target: Neonolum

Record ID	Neonolum	Cook's Distance
2	0,0	0,028
43	0,0	0,027
65	0,0	0,025
225	100,0	0,024
236	100,0	0,023
117	0,0	0,022
227	100,0	0,021
237	100,0	0,020
205	100,0	0,020

Records with large Cook's distance values are highly influential in the mod computations. Such records may distort the model accuracy.

Регрессионный анализ

Outliers

Target: Neonolum

Record ID	Neonolum	Cook's Distance
211	100,0	0,135
198	100,0	0,094
72	0,0	0,092
53	0,0	0,070
238	100,0	0,051
224	100,0	0,051
221	100,0	0,051
101	0,0	0,048
201	100,0	0,040
45	0,0	0,039
46	0,0	0,037
22	0,0	0,035
118	0,0	0,035
235	100,0	0,033
111	0,0	0,032
19	0,0	0,031
121	0,0	0,030
94	0,0	0,029

Records with large Cook's distance values are highly influential in the mod computations. Such records may distort the model accuracy.

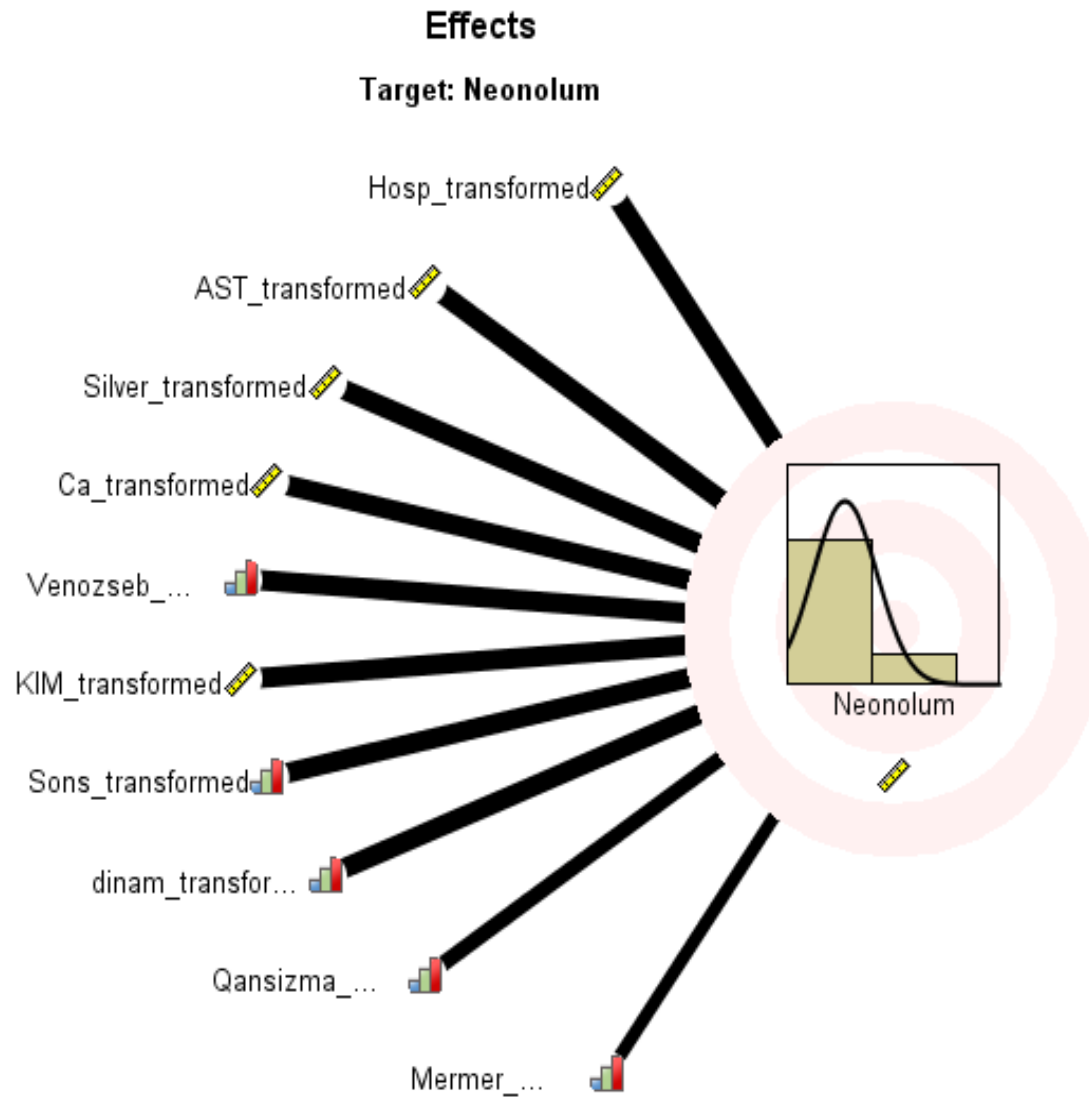
Outliers

Target: Neonolum

Record ID	Neonolum	Cook's Distance
2	0,0	0,028
43	0,0	0,027
65	0,0	0,025
225	100,0	0,024
236	100,0	0,023
117	0,0	0,022
227	100,0	0,021
237	100,0	0,020
205	100,0	0,020

Records with large Cook's distance values are highly influential in the mod computations. Such records may distort the model accuracy.

Регрессионный анализ



Регрессионный анализ

Effects
Target: Neonolum

Source	Sum of Squares	df	Mean Square	F	Sig.
Corrected Model ▶	272 928,846	30	9 097,628	28,344	,000
Residual	66 440,902	207	320,971		
Corrected Total	339 369,748	237			

Регрессионный анализ

Effects

Target: Neonolum

Source	Sum of Squares	df	Mean Square	F
Corrected Model ▼	272 928,846	30	9 097,628	28,344
Hosp_transformed	38 516,921	1	38 516,921	120,001
AST_transformed	12 707,878	1	12 707,878	39,592
Silver_transformed	9 662,931	1	9 662,931	30,105
Ca_transformed	8 491,564	1	8 491,564	26,456
Venozseb_transformed	5 800,880	1	5 800,880	18,073
KIM_transformed	5 572,078	1	5 572,078	17,360
Sons_transformed	5 340,919	1	5 340,919	16,640
dinam_transformed	5 249,908	1	5 249,908	16,356
Qansizma_transformed	4 853,882	2	2 426,941	7,561
Mermer_transformed	4 041,652	2	2 020,826	6,296
Residual	66 440,902	207	320,971	
Corrected Total	339 369,748	237		

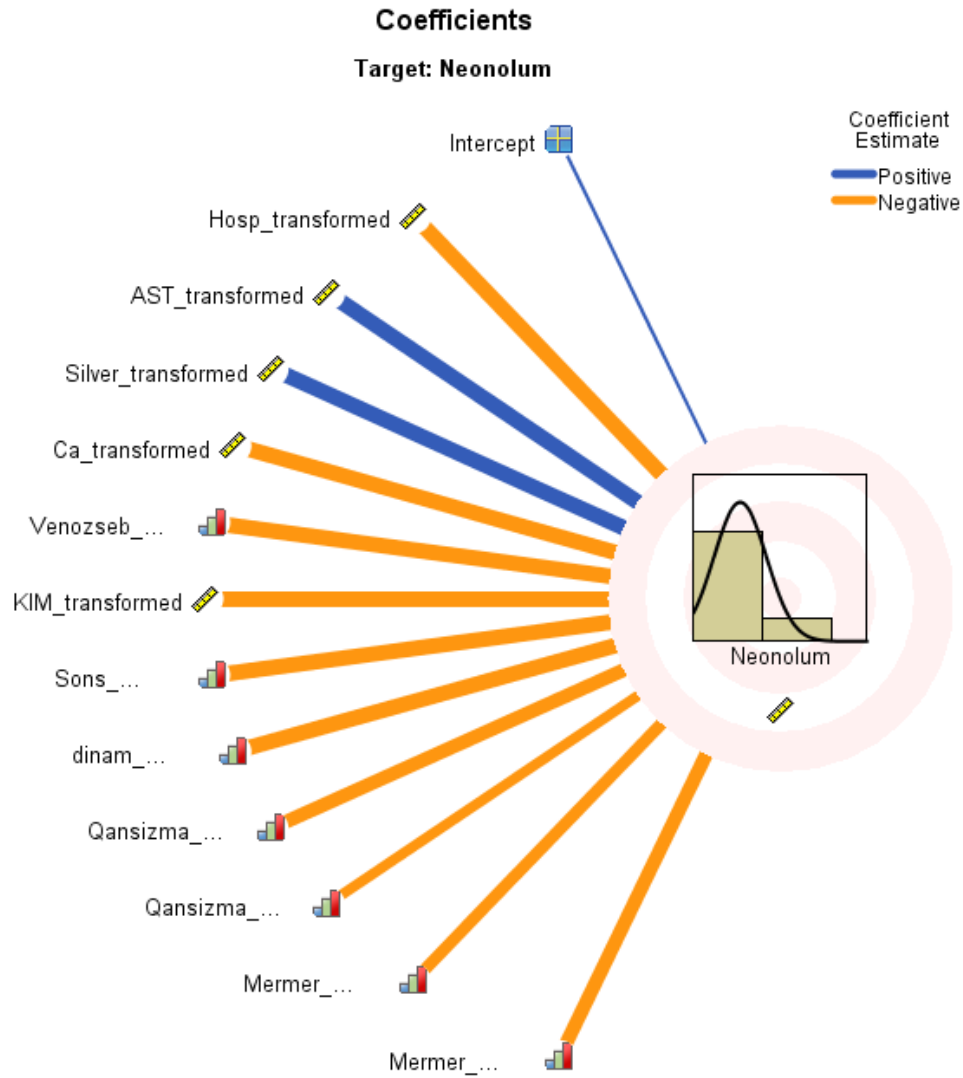
Регрессионный анализ

Effects

Target: Neonolum

Source	Sig.	Importance
Corrected Model ▼	,000	
Hosp_transformed	,000	0,281
AST_transformed	,000	0,093
Silver_transformed	,000	0,071
Ca_transformed	,000	0,062
Venozseb_transformed	,000	0,042
KIM_transformed	,000	0,041
Sons_transformed	,000	0,039
dinam_transformed	,000	0,038
Qansizma_transformed	,001	0,035
Mermer_transformed	,002	0,030
Residual		
Corrected Total		

Регрессионный анализ



Регрессионный анализ

Coefficients
Target: Neonolum

Model Term	Coefficient	Sig.	Importance
Intercept	91,597	,177	
Hosp_transformed	-3,330	,000	0,281
AST_transformed	348,117	,000	0,093
Silver_transformed	3,895	,000	0,071
Ca_transformed	-28,402	,000	0,062
Venozseb_transformed=0	-83,518	,000	0,042
Venozseb_transformed=1	0,000 [¶]		0,042
KIM_transformed	-11,616	,000	0,041
Sons_transformed=0	-16,781	,000	0,039
Sons_transformed=1	0,000 [¶]		0,039
dinam_transformed=0	-15,056	,000	0,038
dinam_transformed=1	0,000 [¶]		0,038
Qansizma_transformed=0	-16,486	,000	0,035
Qansizma_transformed=1	-13,044	,001	0,035
Qansizma_transformed=2	0,000 [¶]		0,035
Mermer_transformed=0	-49,205	,001	0,030
Mermer_transformed=1	-51,942	,000	0,030
Mermer_transformed=2	0,000 [¶]		0,030

[¶]This coefficient is set to zero because it is redundant.

Coefficients
Target: Neonolum

Model Term	Coefficient	Std. Error	t	Sig.
Intercept	91,597	67,675	1,353	,177
Hosp_transformed	-3,330	0,304	-10,955	,000
AST_transformed	348,117	55,325	6,292	,000
Silver_transformed	3,895	0,710	5,487	,000
Ca_transformed	-28,402	5,522	-5,144	,000
Venozseb_transformed=0	-83,518	19,646	-4,251	,000
Venozseb_transformed=1	0,000 [¶]			
KIM_transformed	-11,616	2,788	-4,167	,000
Sons_transformed=0	-16,781	4,114	-4,079	,000
Sons_transformed=1	0,000 [¶]			
dinam_transformed=0	-15,056	3,723	-4,044	,000
dinam_transformed=1	0,000 [¶]			
Qansizma_transformed=0	-16,486	4,278	-3,853	,000
Qansizma_transformed=1	-13,044	4,033	-3,234	,001
Qansizma_transformed=2	0,000 [¶]			
Mermer_transformed=0	-49,205	14,487	-3,397	,001

[¶]This coefficient is set to zero because it is redundant.

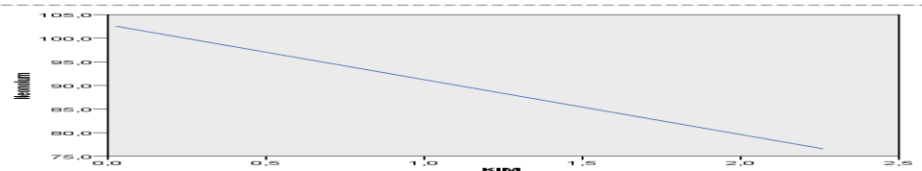
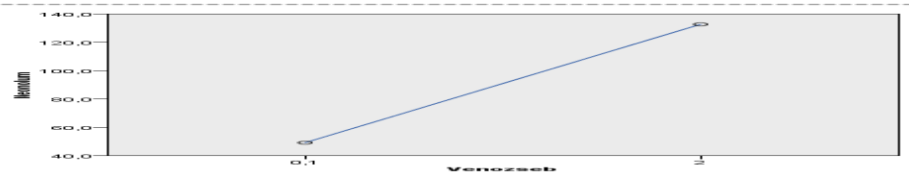
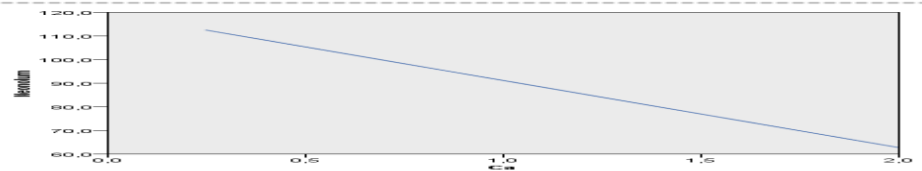
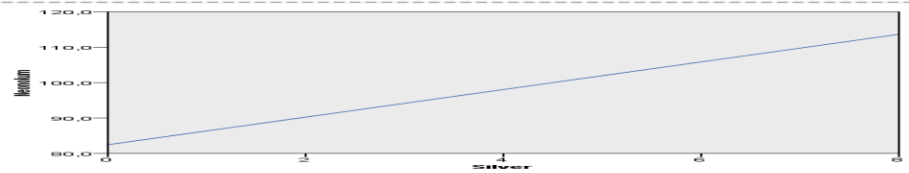
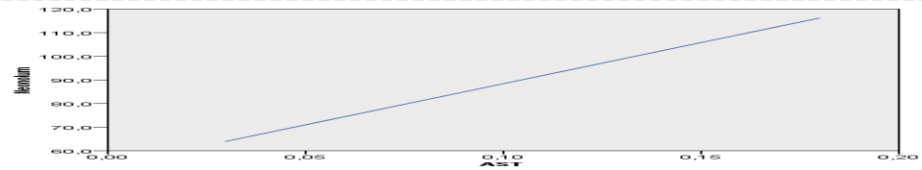
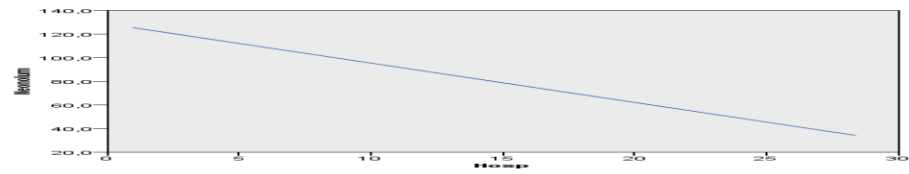
Coefficients
Target: Neonolum

Model Term	95% Confidence Interval		Importance
	Lower	Upper	
Intercept	-41,824	225,018	
Hosp_transformed	-3,930	-2,731	0,281
AST_transformed	239,044	457,189	0,093
Silver_transformed	2,496	5,295	0,071
Ca_transformed	-39,288	-17,516	0,062
Venozseb_transformed=0	-122,250	-44,787	0,042
Venozseb_transformed=1			0,042
KIM_transformed	-17,113	-6,120	0,041
Sons_transformed=0	-24,891	-8,871	0,039
Sons_transformed=1			0,039
dinam_transformed=0	-22,396	-7,717	0,038
dinam_transformed=1			0,038
Qansizma_transformed=0	-24,920	-8,051	0,035
Qansizma_transformed=1	-20,995	-5,093	0,035
Qansizma_transformed=2			0,035
Mermer_transformed=0	-77,766	-20,645	0,030

[¶]This coefficient is set to zero because it is redundant.

Регрессионный анализ

Estimated Means
Target: Neocelum
Estimated means charts for the top ten significant effects ($p < .05$) are displayed.



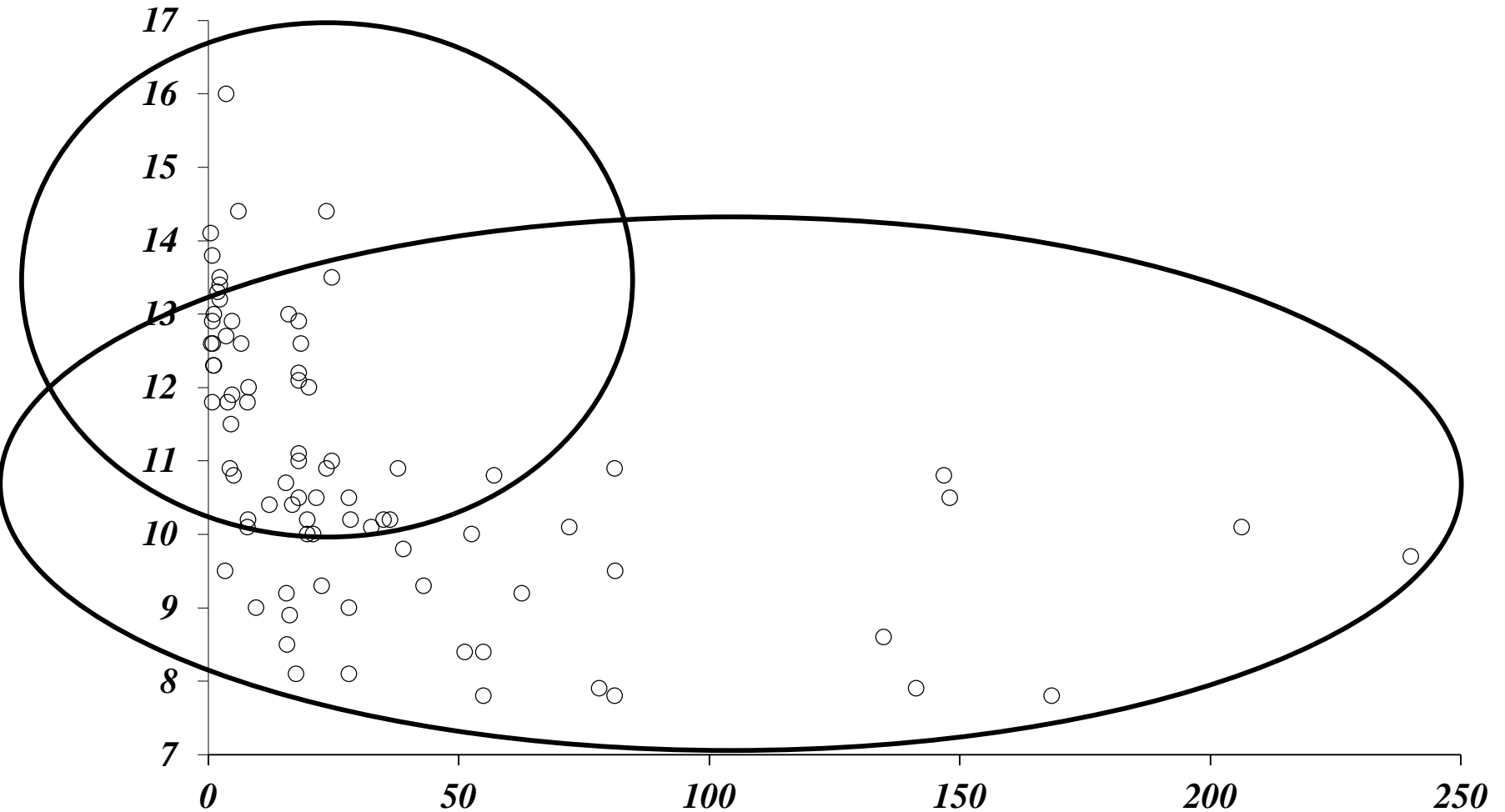
Регрессионный анализ

$$\text{Результат} = A_0 + A_1 * \text{Фактор}_1 + A_2 * \text{Фактор}_2 + \dots + A_k * \text{Фактор}_k$$

Далее полученная формула проходит «экзамен», оценивается специфичность, чувствительность и т.д. и естественно оценивается статистическая значимость результатов

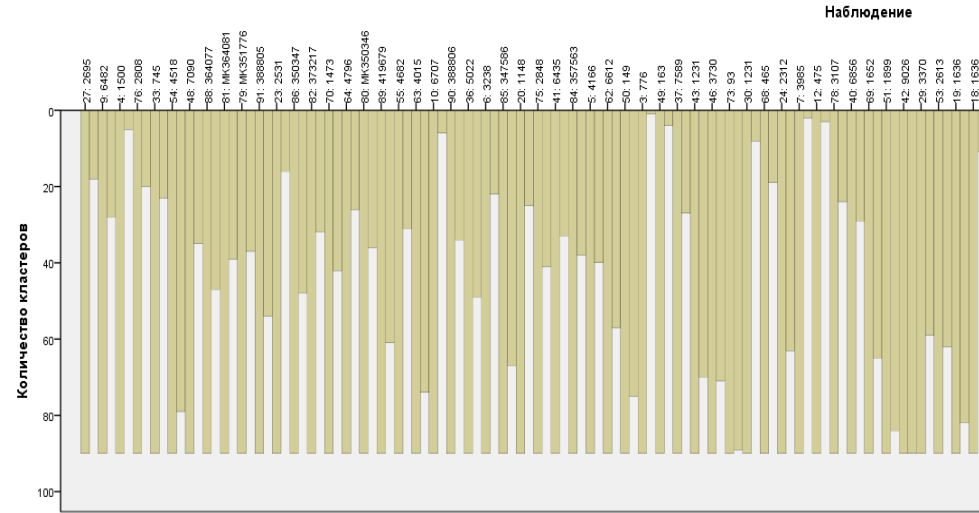
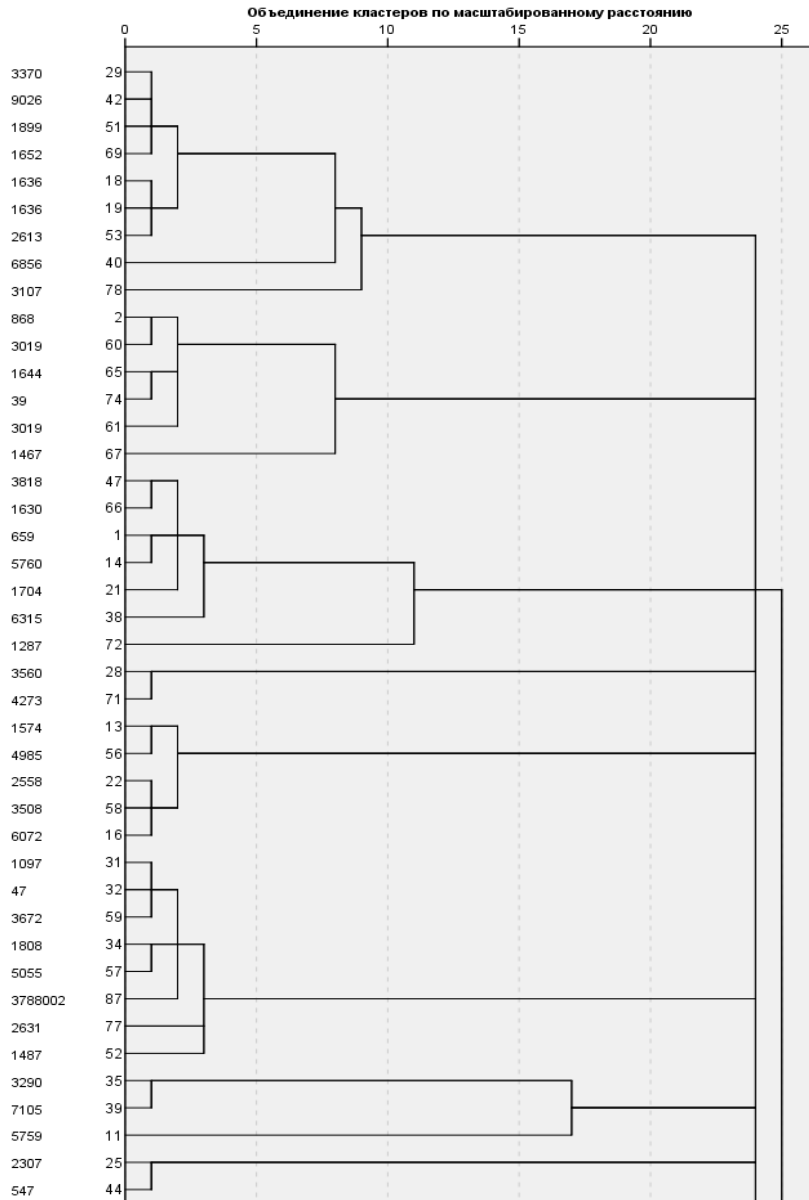


Кластерный анализ (метод ближайшего соседа)



Кластерный анализ (метод ближайшего соседа)

Дендрограмма с использованием метода ближайшего соседа.





Примерная структура научной статьи

- ✓ *Актуальность*
- ✓ *Материал исследования*
- ✓ *Методы исследований*
- ✓ *Основные результаты*
- ✓ *Обсуждение результатов*
- ✓ *Выводы*
- ✓ *Список литературы*
- ✓ *Резюме*
- ✓ *Ключевые слова*

Примерный объем и структура диссертационных работ

<i>Часть</i>	<i>диссертация доктора философии</i>	<i>диссертация доктора наук</i>
<i>Введение</i>	<i>7-10 стр.</i>	<i>15-20 стр.</i>
<i>Обзор литературы</i>	<i>≤ 30 стр.</i>	<i>≤ 60 стр.</i>
<i>Материал и методы</i>	<i>12-15 стр.</i>	<i>20-30 стр.</i>
<i>Основная часть</i>	<i>≥ 50 стр.</i>	<i>≥ 100 стр.</i>
<i>Заключение</i>	<i>15-20 стр.</i>	<i>30-40 стр.</i>
<i>Выводы</i>	<i>5-6 выводов</i>	<i>10-12 выводов</i>
<i>Практические рекомендации</i>	<i>3-5 пр. рекомен.</i>	<i>6-10 пр. рекомен.</i>
<i>Список литературы</i>	<i>130-180 источников (> 50% - 10л ; > 25% - 5 л)</i>	<i>250-350 источников (> 50% - 10л ; > 25% - 5 л)</i>
<i>Приложения</i>	<i>По требованию</i>	<i>По требованию</i>
<i>Иллюстрации</i>	<i>30-40 табл. + рис.</i>	<i>60-80 табл. + рис.</i>
<i>Итого</i>	<i>130-180 стр. текст: 100-150 стр.</i>	<i>250-350 стр. текст: 200-250 стр.</i>



Презентация окончена)

Спасибо за внимание!