

УДК 519.25

**АНАЛИЗ ВЛИЯНИЯ РАЗЛИЧНЫХ ФАКТОРОВ НА РЕЗУЛЬТАТ
ВОЛЕЙБОЛЬНОЙ ИГРЫ**

Азарова Е.А., студент гр. СПмоз-241, I курс
Научный руководитель: Гоголин В.А., д. н., профессор
Кузбасский государственный технический университет
имени Т.Ф. Горбачева,
г. Кемерово

Волейбол – один из самых популярных и захватывающих видов спорта в России. Чтобы успешно играть в волейбол, требуется отличная выносливость, способность быстро перемещаться по площадке, выполнять сложные стратегические маневры, а также демонстрировать силу и ловкость. Для этого спортсмены годами тренируются и совершенствуют свои навыки.

Основная цель волейбола – направить мяч на сторону соперника так, чтобы он приземлился на площадке противника или чтобы игрок защищающейся команды допустил ошибку. Исследования, представленные в данной статье, помогут понять, какие факторы действительно влияют на исход игры, а какие не оказывают существенного воздействия.

В 2023 году состоялся чемпионат России по волейболу, в котором участвовали 15 команд. В этой статье будут проанализированы средние показатели команд, оказавшие влияние на результаты игр. Вследствие чего, мы найдем один наиболее значимый фактор. Среди таких данных можно выделить:

- количество победных матчей
- средний рост игроков;
- средний возраст игроков;
- среднее количество подач;
- среднее количество нападающих ударов;
- среднее количество принятых подач.

Такой анализ можно провести с использованием метода математической статистики, которые позволят определить, существует ли зависимость между этими факторами и результатами игры.

Статистические данные по играм сборных приведены в таблице.

Таблица – Средние статистические данные команд за год

Команда	Y	X1	X2	X3	X4	X5
	Количество победных матчей	Рост	Возраст	Количество сделанных подач	Нападающие удары	Количество принятых подач
Зенит-	14	201	27	471	486	468

Казань						
Локомотив	14	199	29	466	475	456
Белогорье	13	200	30	464	469	455
Динамо (М)	12	187	26	454	460	445
Зенит (СПб)	10	200	28	445	458	439
Кузбасс	9	196	28	433	439	425
Факел	9	200	26	432	437	424
Енисей	9	199	29	438	449	415
Динамо-ЛО	8	197	30	425	436	412
Нова	7	199	29	421	425	409
Газпром-Юрга	5	199	29	400	408	388
Урал	7	201	29	396	405	369
Нефтяник	6	198	29	389	399	378
АСК	3	196	30	368	345	356
Юрга-Самотрол	4	200	27	375	380	348

В данном исследовании имеем 5 факторов. Предположим, что они оказывают влияние на результат игры. Необходимо найти уравнение множественной линейной регрессии, проверить его значимость и значимость факторов. Для этого используем анализ данных и получим вывод итогов (см. рис. 1).

ВЫВОД ИТОГОВ						
Регрессионная статистика						
Множественный R	0,971035007					
R-квадрат	0,942908984					
Нормированный R-квадрат	0,911191753					
Стандартная ошибка	1,036415403					
Наблюдения	15					
Дисперсионный анализ						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	5	159,6659213	31,93318427	29,72860353	0,016473	
Остаток	9	9,667411997	1,074156889			
Итого	14	169,3333333				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верх</i>
У-пересечение	-35,97080344	18,46512662	-1,948039901	0,083228274	-77,7418219	5,80
Переменная X 1	0,007374865	0,090377731	0,081600464	0,936750327	-0,197073767	0,21
Переменная X 2	-0,072189365	0,227319708	-0,31756756	0,758056999	-0,586422269	0,4
Переменная X 3	0,141840329	0,086060235	1,648151772	0,13372409	-0,052841448	0,31
Переменная X 4	-0,010675779	0,043604122	-0,244834168	0,812075809	-0,109315157	0,08
Переменная X 5	-0,025381637	0,049714973	-0,510543116	0,621957059	-0,137844719	0,08

Рис. 1 – Вывод итогов

Зададим уровень значимости (вероятность ошибки) $\alpha = 0,05$.

Значимость $F = 0,0165$, что меньше $\alpha = 0,05$. Отсюда следует, что уравнение регрессии значимо. Однако P -значения всех коэффициентов превышают 0,05. Это значит, что ни один из факторов нельзя использовать в уравнении, так как его не существует.

Данное явление называется мультиколлинеарностью. Его необходимо исключить, найдя один из взаимосвязанных факторов, который оказывает большее влияние на игру.

Для начала найдем связь между факторами, используя диаграмму рассеивания и величину R^2 . Этими факторами могут быть: $X1$ и $X2$, $X1$ и $X3$, $X1$ и $X4$, $X1$ и $X5$, $X2$ и $X3$, $X2$ и $X4$, $X2$ и $X5$, $X3$ и $X4$, $X3$ и $X5$, $X4$ и $X5$ (см. рис. 2 – рис. \ 11).

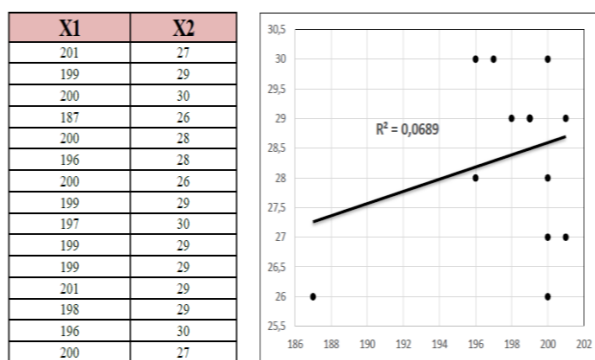


Рис. 2 – взаимосвязь между факторами $X1$ и $X2$

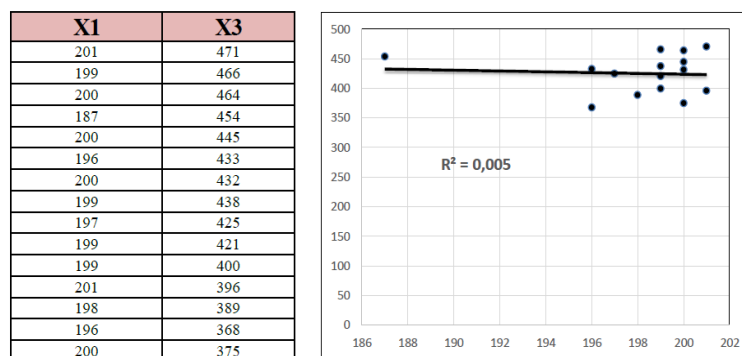


Рис. 3 – взаимосвязь между факторами $X1$ и $X3$

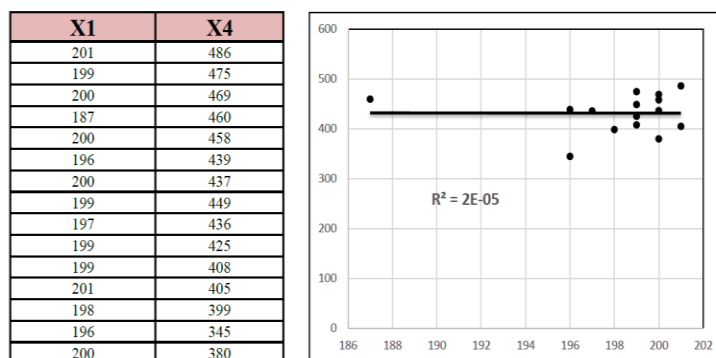


Рис. 4 – взаимосвязь между факторами X1 и X4

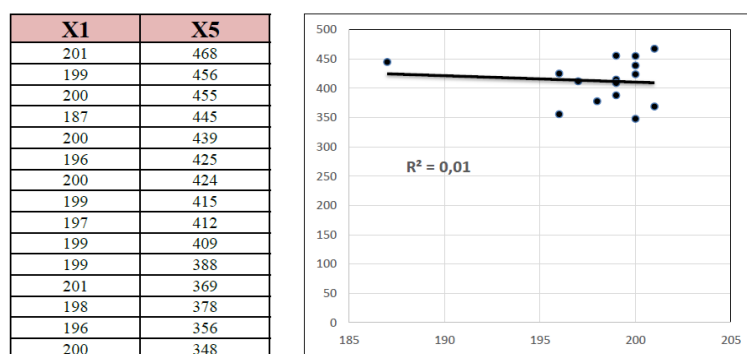


Рис. 5 – взаимосвязь между факторами X1 и X5

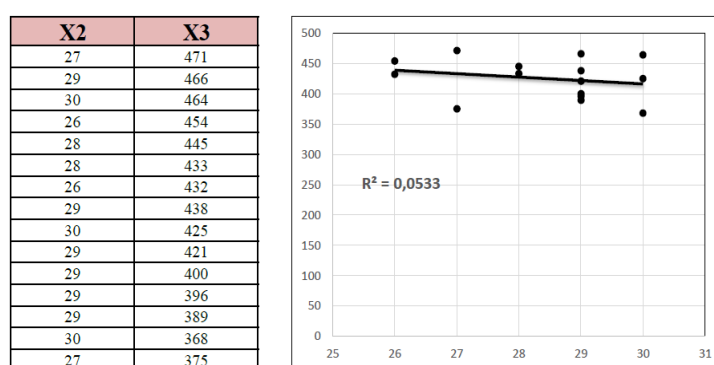


Рис. 6 – взаимосвязь между факторами X2 и X3

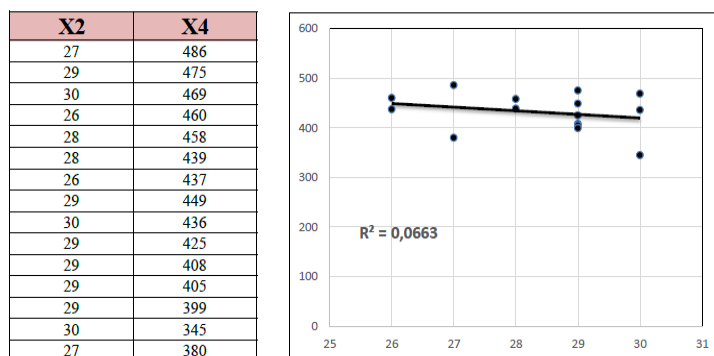


Рис. 7 – взаимосвязь между факторами X2 и X4

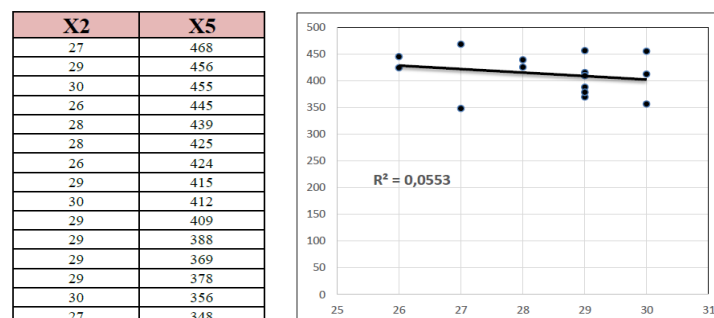


Рис. 8 – взаимосвязь между факторами X2 и X5

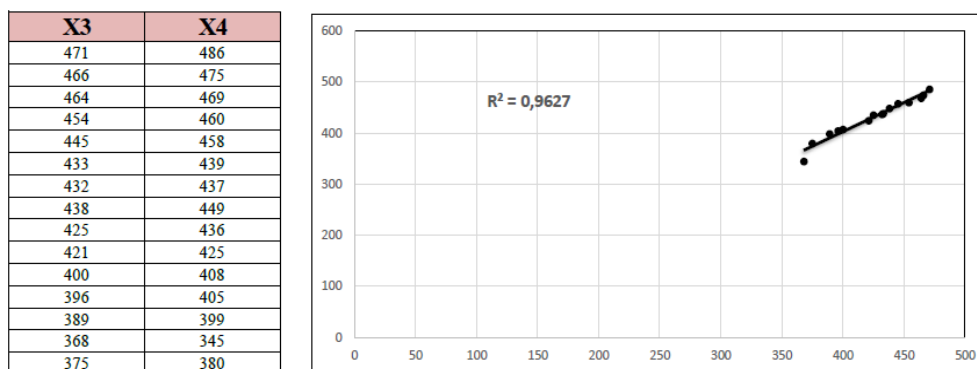


Рис. 9 – взаимосвязь между факторами X3 и X4

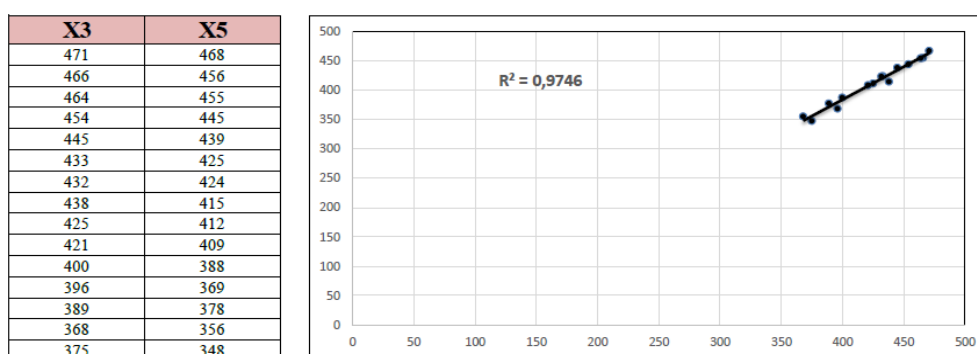


Рис. 10 – взаимосвязь между факторами X3 и X5

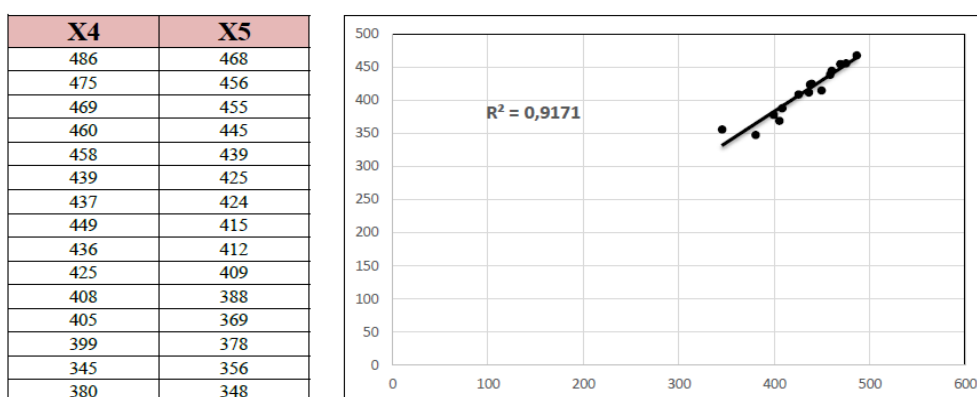


Рис. 11 – взаимосвязь между факторами X4 и X5

На рис. 2 – 8 можно увидеть, что взаимосвязи между факторами практически нет, так как значение R^2 близко к нулю. Отсюда можно сделать вывод, что такие факторы как рост и возраст не взаимосвязаны с другими факторами и не оказывают влияния на победу команды.

Поэтому рассмотрим рис. 9 – 11, на которых значение коэффициента корреляции наиболее приближено к единице. Это значит, что факторы X3, X4, X5 могут быть тесно связаны между собой.

Наибольший $R^2=0,9746$ (см. рис. 10), следовательно, можем предположить, что факторы X3 и X5 имеют наибольшую взаимосвязь между собой. Однако, использовать их в уравнении одновременно нельзя, поэтому оставим только один.

Выбор одного из двух взаимосвязанных факторов может осуществляться по различным методикам. Воспользуемся самой простой для получения наилучшей модели. Для этого необходим независимый фактор, пусть это будет X_1 , и один из взаимосвязанных факторов (X_3 или X_5).

Рассмотрим две модели (см. рис. 12 и рис. 13).

X1	X3	Y	ВЫВОД ИТОГОВ						
201	471	14	Регрессионная статистика						
199	466	14							
200	464	13							
187	454	12							
			Множественный R	0,969942562					
			R-квадрат	0,940788573					
200	445	10	Нормированный R-квадрат	0,930920002					
196	433	9	ошибка	0,914078234					
200	432	9	Наблюдения	15					
199	438	9	Дисперсионный анализ						
197	425	8							
199	421	7							
199	400	5							
			Регрессия	df	SS	MS	F	Значимость F	
201	396	7	Остаток	12	10,02646822	0,835539018		4,30956E-08	
198	389	6	Итого	14	169,3333333				
196	368	3							
200	375	4		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	
			Y-пересечение	-34,85818942	14,58348023	-2,390251768	0,034119616	-66,63286325	
			Переменная X 1	0,001191856	0,070755	0,016844827	0,039868372	-0,152970047	
			Переменная X 2	0,101823843	0,007391999	13,77487177	1,02548E-08	0,08571806	

Рис. 12 – Первая модель с факторами X_1 и X_3

X1	X5	Y	ВЫВОД ИТОГОВ					
201	468	14	Регрессионная статистика					
199	456	14						
200	455	13						
187	445	12						
200	439	10	Множественный R	0,95222985				
			R-квадрат	0.906741687				
196	425	9	Нормированный R-квадрат	0.891198635				
			Стандартная ошибка	1,147161025				
200	424	9	Наблюдения	15				
199	415	9	Дисперсионный анализ					
197	412	8						
199	409	7						
199	388	5						
201	369	7	Регрессия	df	SS	MS	F	Значимость F
198	378	6	Остаток	12	15,791741	1,315978417		6,57848E-07
196	356	3	Итого	14	169,3333333			
200	348	4						
				Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%
			Y-пересечение	-33,32627084	18,28938559	-1,822164593	0,093431439	-73,1754188
			Переменная X 1	0,028524113	0,089023562	0,320410823	0,045416756	-0,165441567
			Переменная X 2	0,088107386	0,00817729	10,7746442	1,59019E-07	0,070290603

Рис. 13 – Вторая модель с факторами X_1 и X_5

Значимость F для первой модели равна $4,31 \cdot 10^{-8}$, для второй $6,57 \cdot 10^{-7}$. Обе величины меньше 0,05, однако, выберем ту, значение которой наименьшее, так как вероятность ошибки будет меньше. Отсюда следует, что наилучшим уравнением является первое, в него входят факторы X_1 и X_3 .

Далее проверим P -значения, они должны быть меньше 0,05. Для X_1 ($0,045 < 0,05$), для X_2 ($1,59 \cdot 10^{-7} < 0,05$), поэтому получаем уравнение:

$$y = -33,32 + 0,029X_1 + 0,088X_2$$

Однако необходимо выяснить какой же из этих факторов имеет наибольшее влияние на результат игры. Для этого вычислим

стандартизированные коэффициенты регрессии и выберем наибольший из них.

Воспользуемся формулами (см. рис. 13):

$$\beta_1 = b_1 \cdot \frac{S_{x1}}{S_y}; \beta_3 = b_3 \cdot \frac{S_{x3}}{S_y}$$

где S_{x1} , S_{x3} – средние квадратические отклонения факторов $X1$ и $X3$,

S_y - среднее квадратическое отклонение Y

Найдем средние квадратические отклонения (см. рис. 14):

Стандартизированные коэффициенты регрессии:

$$\beta_1 = 0,029 \cdot \frac{3,344}{3,359} = 0,0288$$

$$\beta_3 = 0,088 \cdot \frac{32,01}{3,359} = 0,838$$

Так как $\beta_3 > \beta_1$, то фактор $X3$ (количество сделанных подач) оказывает наибольшее влияние на Y .

Подведем итоги:

1. Наименее значимыми факторами являются рост, возраст, так как взаимосвязь этих факторов с другими самая маленькая.

2. Значимыми факторами являются количество сделанных подач и нападающих ударов, так как взаимосвязь между этими факторами и другими высокая, а значение R^2 приближенно к единице.

3. Самым значимым фактором является количество сделанных подач, так как взаимосвязь между этими факторами и другими самая высокая, а значение R^2 наиболее приближенно к единице.

Таким образом, на основании проведенного исследования и подведенных итогов, можно сделать вывод, что такой фактор, как количество сделанных подач значительно влияет на ход игры и успех команды, что непременно зависит от мастерства и опыта игроков.

Список литературы

1. Горелова, Г.В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel: учеб. пособие для вузов / Г. В. Горелова, И. А. Кацко //Ростов н/Д: Феникс, 2005. – 480 с.

2. Лига ставок: сайт – URL: <https://soccer365.ru/clubs/449/> (дата обращения 03.05.2023г.)