



Научные записки НИЦ МКВК

№ 15

2022

Г.В. Стулина, Ш.М. Кенжабаев

Реализация программы R по исходным данным проекта WUFMAS



Научно-информационный центр
Межгосударственной координационной водохозяйственной комиссии
Центральной Азии

Г.В. Стулина, Ш.М. Кенжабаев

**Реализация программы R
по исходным данным проекта
WUFMAS**

Ташкент 2022

Цель работы – продемонстрировать возможности работы программы R с данными и получить результаты для дальнейшего анализа сельхозпроизводства.

Введение

В данной статье приводятся результаты обработки данных сельхозпроизводства с использованием программы R.

В качестве исходных использованы данные, полученные в проекте «Водопользование и сельскохозяйственное производство в республиках Центральной Азии» (WUFMAS), части проекта WARMAP (TACIS project, 1996-2000). Возврат к проекту WARMAP объясняется наличием уникальных данных, полученных в результате исследований в процессе выполнения указанных проектов. Уникальность данных заключается в методике их отбора.

В проекте была получена большая база данных, гарантировано достоверных, так как собирались экспертами при личном их участии в ходе ежедневных наблюдений и фиксации всех деталей сельскохозяйственного производства.

Объекты исследований

По проекту WUFMAS в 1996 было начато систематическое измерение использования факторов производства и выхода продукции на опытных полях в 36 опытных хозяйствах, которые располагались по всему бассейну Аральского моря, согласно распределению орошаемых земель.

Методика

Методика сбора данных

Данные по хозяйству в целом собираются раз в год и ежемесячно по отчётам хозяйств. На каждом опытном поле были намечены пять опытных участков, на которых проводятся агрономические измерения (включая собранный с этих участков урожай). Также записываются данные о

собранном урожае со всего поля по записям, которые ведутся в хозяйстве. Все эти данные записываются в специальный комплект таблиц с помощью кодов, которыми обозначаются использованные материалы, машины и механизмы, операции на культурах и собранная продукция.

На каждом поле были проведены почвенные изыскания и периодически отбирались образцы почвы, пробы оросительной, дренажной и грунтовой воды отправлялись с этих полей на анализ в лабораторию САНИИРИ¹. С ближайших к хозяйствам метеостанций собирались среднемесячные климатические данные, а данные по испарению с испарителя и по осадкам измерялись техниками непосредственно в хозяйствах.

Заполненные таблицы ежемесячно представлялись в центральный офис проекта для ввода данных в базу данных WUFMAS в среде MS Access. Эта база данных в настоящий момент имеет большой объем информации по сельхозпроизводству и является полезным инструментом для планирования на уровне хозяйств и на национальных уровнях, а также для развития частного коммерческого сектора. Ниже представлены обобщение и анализ некоторых из имеющихся данных.

Метод R

R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. Широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ.

Скрипт для корреляции агротехнических параметров культуры в программе R

1. Активизации нужных пакетов:

```
library(Hmisc)
```

```
library(corrplot)
```

```
library(PerformanceAnalytics)
```

2. Открытые данных в формате.csv:

```
wms <-read.csv("F:/R/exr6/WUFMAS.csv")
```

¹ Среднеазиатский научно-исследовательский институт ирригации им. В.Д. Журина

3. Общие сведения о данных в квантилях:

```
summary(wms)
```

4. Выборка матрицу данных

```
wt<- as.matrix(wms[1:58,2:12])
```

5. Получение коэффициента корреляции Пирсона:

```
cor<-cor(wt)
```

```
cor
```

6. Получение коэффициента корреляции и значения p :

```
res<-rcorr(wt)
```

```
res
```

7. График корреляции и значения p :

```
chart.Correlation(wt, histogram=TRUE, pch=10) # to plot cor. & p-value  
as mat.
```

8. График только корреляции:

```
corrplot(cor, type="upper", order="hclust", tl.col="black", tl.srt=45)
```

Корреляция – это мера степени и направления связи между значениями двух переменных, статистический показатель вероятности связи между двумя переменными, измеренными в количественной величине.

Коэффициент Пирсона – это мера корреляции, подходящая для двух непрерывных (метрических переменных), измеренных на одной и той же выборке корреляционного анализа. Проводится для проверки гипотезы о связях между переменными с использованием коэффициентов

Значение p – это вероятность, которая рассматривается как числовая мера объективной возможности осуществления некоторого события при

единичном испытании. Согласно классическому определению, вероятность события выражается отношением числа благоприятствующих осуществлению этого события исходов m к числу всех равновозможных и несовместных исходов n . Значение p – это, буквально, мера вероятности получения статистически значимого результата – статистическая значимость.

Результаты

Расчет по программе R

Были проведены расчеты для основных культур: пшеницы, риса, хлопчатника.

Получены два показателя, матрица значения корреляции и матрица значения P , по результатам построены графики корреляции и значения P для каждой из трех культур.

А. Результаты по пшенице

Матрица корреляции

	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area	1.00	0.09	0.32	-0.18	-0.09	0.00	-0.16	-0.75	-0.16	0.10	-0.17
mech	0.09	1.00	0.20	0.07	0.03	0.00	-0.07	0.02	0.09	0.08	-0.20
chem	0.32	0.20	1.00	-0.19	-0.29	0.01	-0.05	-0.44	-0.33	0.12	-0.09
work	-0.18	0.07	-0.19	1.00	0.18	0.32	-0.06	0.29	0.49	-0.30	0.16
nitr	-0.09	0.03	-0.29	0.18	1.00	-0.11	-0.06	0.37	0.31	0.17	-0.10
phos	0.00	0.00	0.01	0.32	-0.11	1.00	-0.05	0.06	0.02	0.06	0.10
pots	-0.16	-0.07	-0.05	-0.06	-0.06	-0.05	1.00	0.29	-0.15	0.00	-0.19
npkt	-0.75	0.02	-0.44	0.29	0.37	0.06	0.29	1.00	0.22	-0.12	0.06
watr	-0.16	0.09	-0.33	0.49	0.31	0.02	-0.15	0.22	1.00	-0.10	0.11
seed	0.10	0.08	0.12	-0.30	0.17	0.06	0.00	-0.12	-0.10	1.00	0.06
yild	-0.17	-0.20	-0.09	0.16	-0.10	0.10	-0.19	0.06	0.11	0.06	1.00

Матрица значения r

	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area		0.4860	0.0152	0.1674	0.5089	0.9963	0.2406	0.0000	0.2303	0.4780	0.2092
mech	0.4860		0.1227	0.5895	0.8341	0.9925	0.5986	0.9078	0.5201	0.5301	0.1368
chem	0.0152	0.1227		0.1483	0.0249	0.9504	0.6868	0.0005	0.0110	0.3612	0.4811
work	0.1674	0.5895	0.1483		0.1887	0.0138	0.6377	0.0279	0.0001	0.0211	0.2322
nitr	0.5089	0.8341	0.0249	0.1887		0.4261	0.6308	0.0042	0.0174	0.2029	0.4591
phos	0.9963	0.9925	0.9504	0.0138	0.4261		0.7059	0.6715	0.8821	0.6506	0.4583
pots	0.2406	0.5986	0.6868	0.6377	0.6308	0.7059		0.0301	0.2774	0.9924	0.1503
npkt	0.0000	0.9078	0.0005	0.0279	0.0042	0.6715	0.0301		0.1000	0.3510	0.6716
watr	0.2303	0.5201	0.0110	0.0001	0.0174	0.8821	0.2774	0.1000		0.4581	0.4284
seed	0.4780	0.5301	0.3612	0.0211	0.2029	0.6506	0.9924	0.3510	0.4581		0.6344
yild	0.2092	0.1368	0.4811	0.2322	0.4591	0.4583	0.1503	0.6716	0.4284	0.6344	

$n = 58$ – количество использованных данных в программе

Здесь:

area – площадь, га

mech – использование механизмов, час/га

chem – использование агрохимикатов, кг/га

work – использование труда, час/га

nitr – использование азота, кг/га

phos – использование фосфора, кг/га

pots – использование калия, кг/га

npkt – использование удобрения, кг/га

watr – использование воды, тыс.м³/га

seed – использование семян, кг/га

yild – основная продукция на поле (урожайность), т/га

Наибольший интерес представляет собой корреляция между полученной урожайностью и факторами производства. Анализировалось 58 полей.

Величина урожая более отзывчива на использовании удобрений, агрохимикатов и особенно на количество семян. В меньшей степени на затраты техники, что вполне понятно, так как подготовка полей ведется одновременно в хозяйстве и ручного труда, который в производстве пшеницы минимальный (рис. 1).

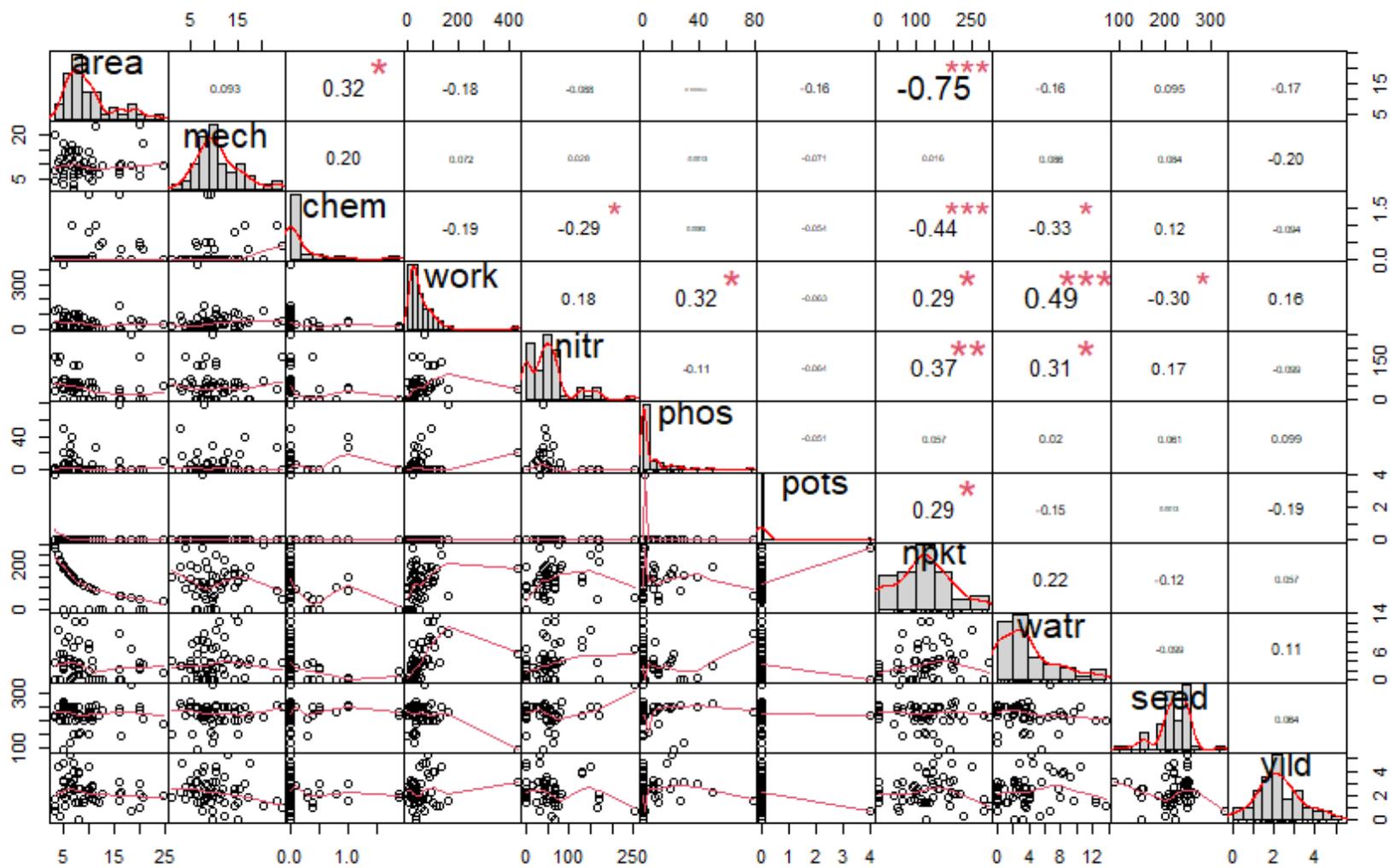


Рис. 1. График корреляции и значения p основных параметров пшеницы

Урожай ц/га в меньшей степени зависит от площади самого поля.

Соответственно, рис. 2 демонстрирует, на что влияет каждый параметр. Так затраты ручной работы определяются затратами воды на орошение, внесение удобрений. Использование удобрений связано с количеством поливов, так как после сева внесение удобрений проводится совместно с поливом.

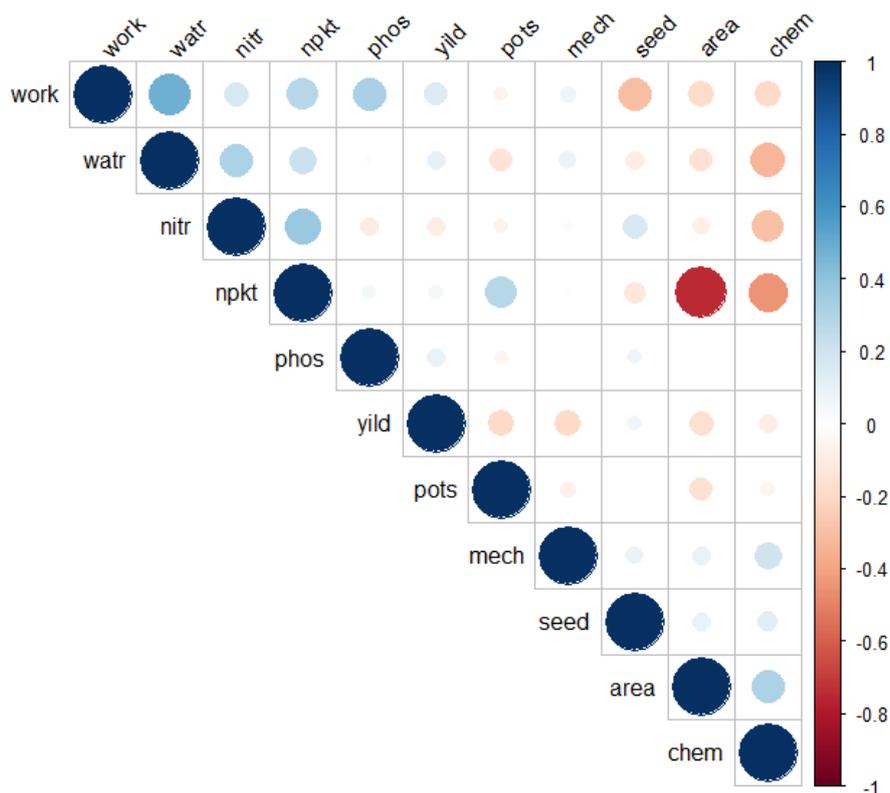


Рис. 2. График корреляции основных параметров пшеницы

Б. Результаты по рису

Матрица корреляции

	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area	1.00	-0.10	-0.40	-0.34	-0.51	-0.41	-0.31	-0.89	0.13	0.26	0.16
mech	-0.10	1.00	-0.33	0.59	-0.13	0.17	-0.19	-0.05	0.78	-0.07	0.67
chem	-0.40	-0.33	1.00	-0.13	0.36	0.19	0.34	0.53	-0.42	-0.17	-0.45
work	-0.34	0.59	-0.13	1.00	-0.09	0.31	0.16	0.27	0.14	-0.31	0.17
nitr	-0.51	-0.13	0.36	-0.09	1.00	0.27	-0.16	0.58	-0.12	-0.04	-0.02
phos	-0.41	0.17	0.19	0.31	0.27	1.00	-0.06	0.41	-0.03	-0.48	-0.04
pots	-0.31	-0.19	0.34	0.16	-0.16	-0.06	1.00	0.46	-0.25	-0.15	-0.44
npkt	-0.89	-0.05	0.53	0.27	0.58	0.41	0.46	1.00	-0.31	-0.32	-0.35
watr	0.13	0.78	-0.42	0.14	-0.12	-0.03	-0.25	-0.31	1.00	0.21	0.87
seed	0.26	-0.07	-0.17	-0.31	-0.04	-0.48	-0.15	-0.32	0.21	1.00	0.22
yild	0.16	0.67	-0.45	0.17	-0.02	-0.04	-0.44	-0.35	0.87	0.22	1.00

Матрица значения p

P	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area		0.6356	0.0501	0.0961	0.0086	0.0400	0.1328	0.0000	0.5224	0.2132	0.4550
mech	0.6356		0.1068	0.0019	0.5338	0.4214	0.3758	0.8271	0.0000	0.7533	0.0002
chem	0.0501	0.1068		0.5314	0.0770	0.3514	0.0921	0.0059	0.0368	0.4127	0.0236
work	0.0961	0.0019	0.5314		0.6665	0.1256	0.4378	0.1933	0.5137	0.1354	0.4272
nitr	0.0086	0.5338	0.0770	0.6665		0.1888	0.4348	0.0022	0.5726	0.8431	0.9397
phos	0.0400	0.4214	0.3514	0.1256	0.1888		0.7688	0.0428	0.8842	0.0158	0.8470
pots	0.1328	0.3758	0.0921	0.4378	0.4348	0.7688		0.0203	0.2200	0.4687	0.0264
npkt	0.0000	0.8271	0.0059	0.1933	0.0022	0.0428	0.0203		0.1376	0.1236	0.0877
watr	0.5224	0.0000	0.0368	0.5137	0.5726	0.8842	0.2200	0.1376		0.3073	0.0000
seed	0.2132	0.7533	0.4127	0.1354	0.8431	0.0158	0.4687	0.1236	0.3073		0.2813
yild	0.4550	0.0002	0.0236	0.4272	0.9397	0.8470	0.0264	0.0877	0.0000	0.2813	

n= 25 – количество использованных данных

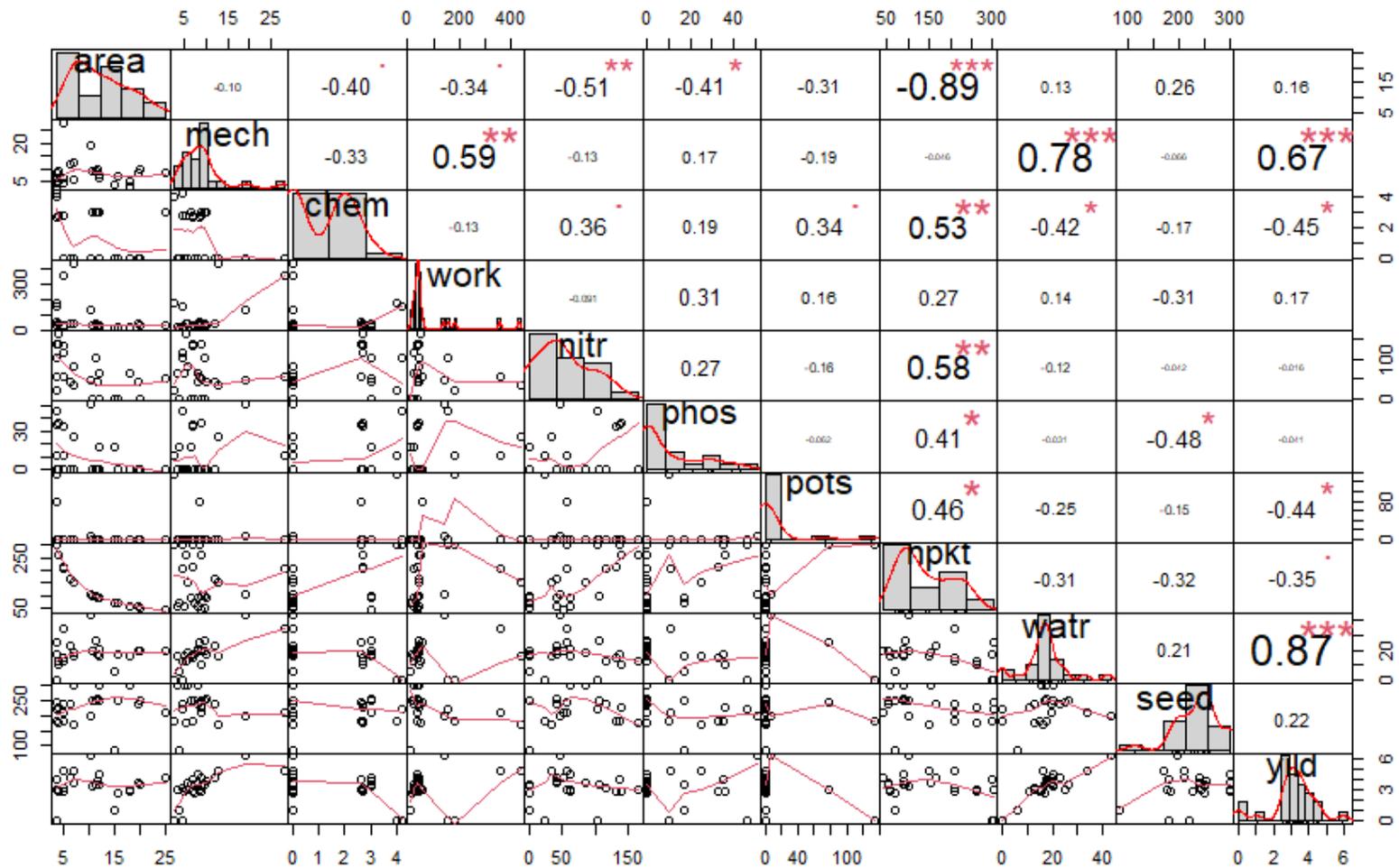


Рис. 3. График корреляции и значения p основных параметров риса

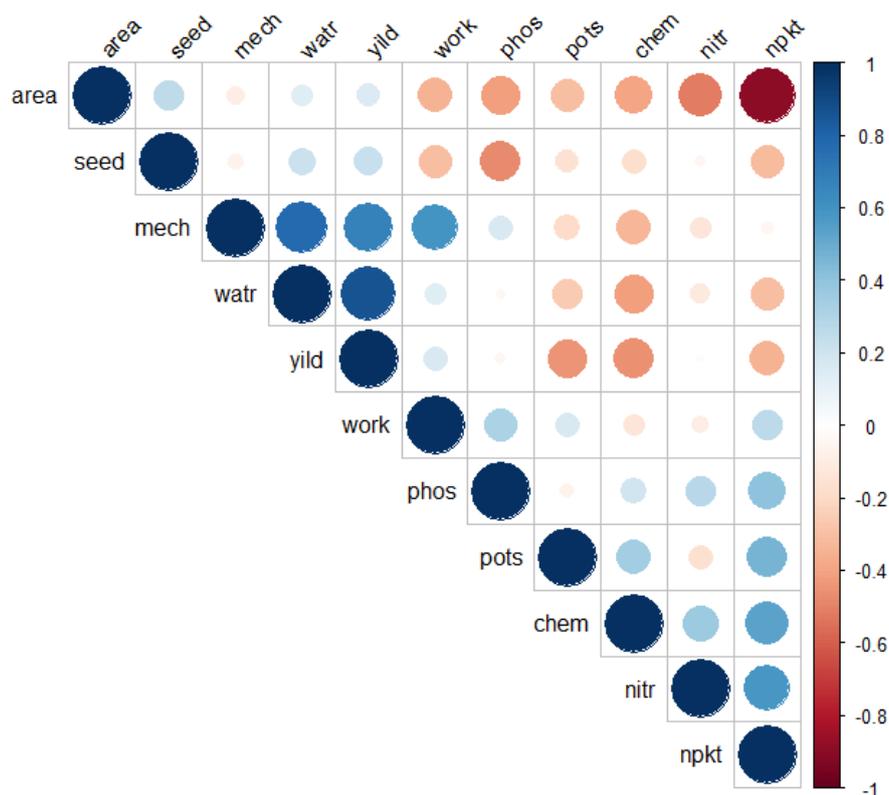


Рис. 4. График корреляции основных параметров риса

Здесь:

area – площадь, га

mech – использование механизмов, час/га

chem – использование агрохимикатов, кг/га

work – использование труда, час/га

nitr – использование азота, кг/га

phos – использование фосфора, кг/га

pots – использование калия, кг/га

npkt – использование удобрений, кг/га

watr – использование воды, тыс.м³/га

seed – использование семян, кг/га

yild – основная продукция на поле (урожайность), т/га

Было проанализировано

При возделывании риса основное воздействие на продуктивность и урожай культур оказывают масштабы использования механизмов в час/га (матрица корреляции), а также количества поданной воды.

Внесение химических веществ и урожайность имеют отрицательную корреляцию при большой статистической значимостью, т.е. с большой степенью вероятности. Показатели указывают на необходимость пересмотреть норму внесения химикатов.

При возделывании риса используется и ручной труд для поливов на рисовом поле, вслед за вспашкой следуют нарезка чеков и часто ручная посадка риса, урожай находится в тесной связи от затрат техники и объема орошения (рис. 3, 4).

С. Результаты по хлопчатнику

Матрица корреляции

	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area	1.00	-0.22	0.32	-0.35	-0.14	0.01	-0.17	-0.86	-0.04	-0.36	-0.20
mech	-0.22	1.00	0.01	0.13	0.37	0.12	0.16	0.27	0.18	0.44	0.45
chem	0.32	0.01	1.00	-0.24	-0.03	0.08	0.06	-0.26	-0.07	-0.23	0.17
work	-0.35	0.13	-0.24	1.00	-0.01	-0.17	0.08	0.43	0.44	0.56	0.30
nitr	-0.14	0.37	-0.03	-0.01	1.00	-0.12	-0.03	0.15	-0.11	0.04	0.17
phos	0.01	0.12	0.08	-0.17	-0.12	1.00	0.32	-0.03	-0.23	-0.04	0.05
pots	-0.17	0.16	0.06	0.08	-0.03	0.32	1.00	0.14	-0.03	0.18	0.28
npkt	-0.86	0.27	-0.26	0.43	0.15	-0.03	0.14	1.00	0.11	0.39	0.25
watr	-0.04	0.18	-0.07	0.44	-0.11	-0.23	-0.03	0.11	1.00	0.48	0.27
seed	-0.36	0.44	-0.23	0.56	0.04	-0.04	0.18	0.39	0.48	1.00	0.32
yild	-0.20	0.45	0.17	0.30	0.17	0.05	0.28	0.25	0.27	0.32	1.00

Матрица значения p

	area	mech	chem	work	nitr	phos	pots	npkt	watr	seed	yild
area		0.0290	0.0013	0.0005	0.1652	0.9223	0.0935	0.0000	0.7167	0.0003	0.0520
mech	0.0290		0.8915	0.2077	0.0002	0.2417	0.1212	0.0084	0.0829	0.0000	0.0000
chem	0.0013	0.8915		0.0169	0.7604	0.4424	0.5402	0.0091	0.5134	0.0213	0.0938
work	0.0005	0.2077	0.0169		0.8842	0.0901	0.4097	0.0000	0.0000	0.0000	0.0028
nitr	0.1652	0.0002	0.7604	0.8842		0.2267	0.7639	0.1500	0.2793	0.6655	0.0957
phos	0.9223	0.2417	0.4424	0.0901	0.2267		0.0013	0.7921	0.0229	0.6997	0.6086
pots	0.0935	0.1212	0.5402	0.4097	0.7639	0.0013		0.1863	0.7565	0.0815	0.0056
npkt	0.0000	0.0084	0.0091	0.0000	0.1500	0.7921	0.1863		0.2807	0.0000	0.0125
watr	0.7167	0.0829	0.5134	0.0000	0.2793	0.0229	0.7565	0.2807		0.0000	0.0069
seed	0.0003	0.0000	0.0213	0.0000	0.6655	0.6997	0.0815	0.0000	0.0000		0.0015
yild	0.0520	0.0000	0.0938	0.0028	0.0957	0.6086	0.0056	0.0125	0.0069	0.0015	

n= 97 – количество использованных данных

Здесь:

area – площадь, га

mech – использование механизмов, час/га

chem – использование агрохимикатов, кг/га

work – использование труда, час/га

nitr – использование азота, кг/га

phos – использование фосфора, кг/га

pots – использование калия, кг/га

npkt – использование удобрения, кг/га

watr – использование воды, тыс.м³/га

seed – использование семян, кг/га

yild – основная продукция на поле (урожайность), т/га

Для хлопчатника имеется следующая взаимосвязь: для получения высокого урожая хлопчатника необходимы: затраты техники, ручного труда, затем внесения удобрений одновременно с поливом (рис. 5, 6). Урожай хлопчатника в равной степени зависит от многих факторов внесения всех видов удобрения, чувствительный к объему затрат механизмов. Ручной труд используется при севе и поливах культуры.

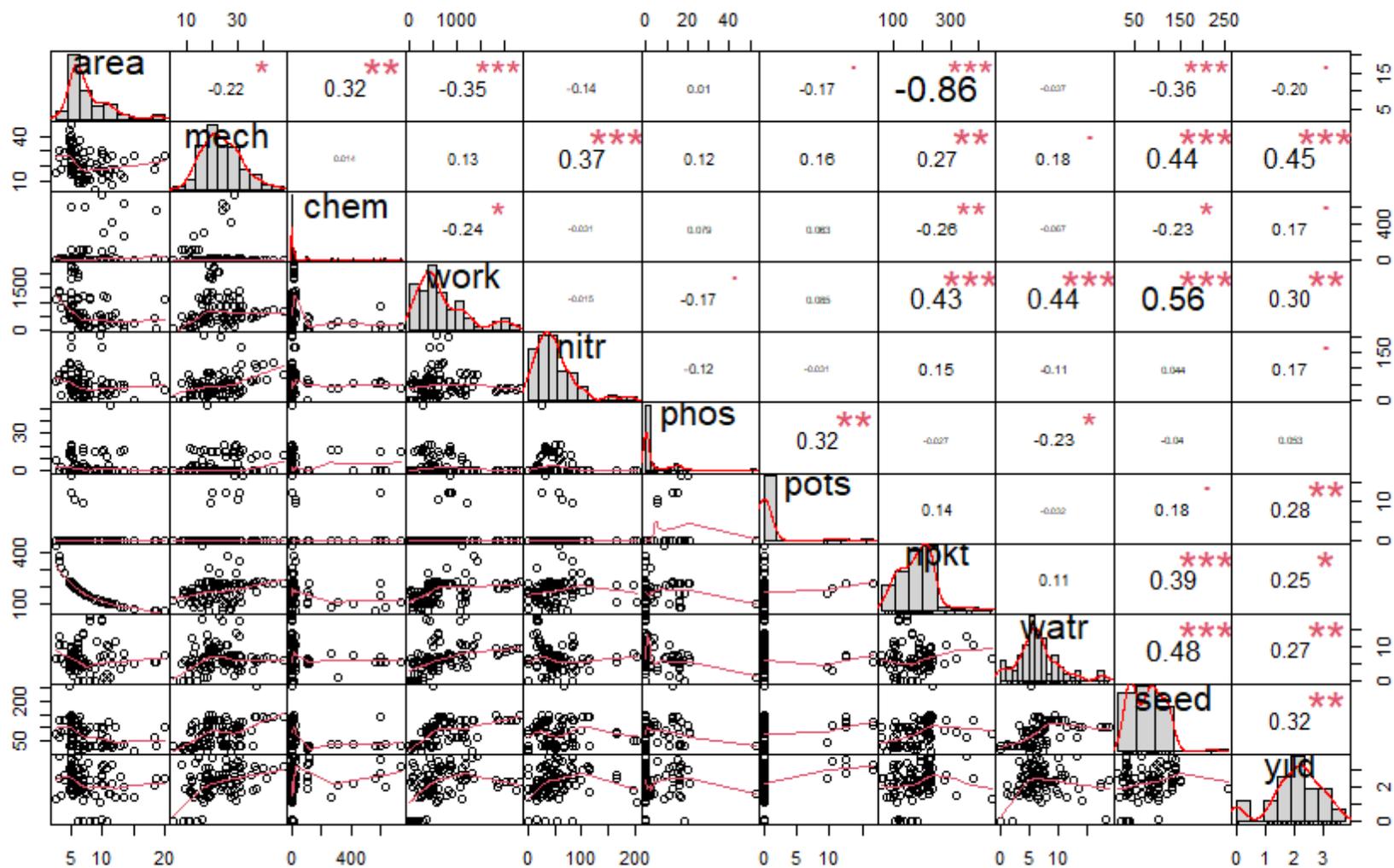


Рис. 5. График корреляции и значения r основных параметров хлопчатника

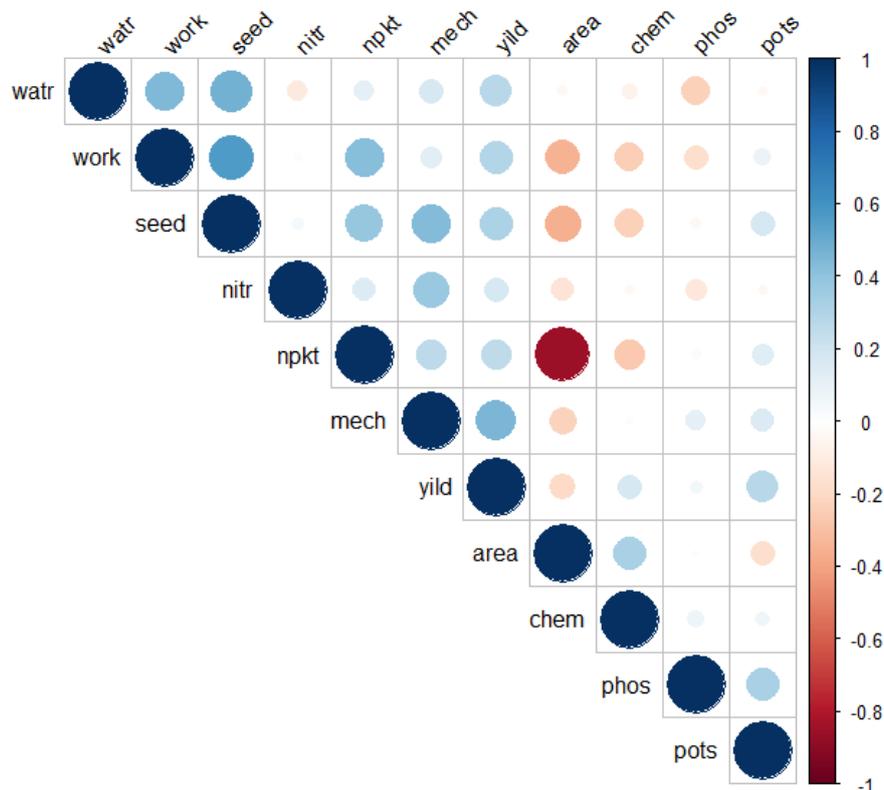


Рис. 6. График корреляции основных параметров хлопчатника

Выводы

Статистический анализ, выполненный по программе R, возможно использовать для оценки корреляции между переменными, а также для статистической значимости этой корреляции между факторами производства в сельском хозяйстве. Полученные данные возможно использовать для оценочной характеристики вклада каждого параметра – фактора производства на конечный результат.

Следующим этапом работы с программой R будет определение состояния сельскохозяйственных полей по космическим снимкам, анализ и увязка состояния с деталями производственного цикла выращивания сельхозкультур и с продуктивностью сельскохозяйственных культур.

Верстка: Беглов И.

Подготовлено к печати
в Научно-информационном центре МКВК

Республика Узбекистан, 100 187,
г. Ташкент, м-в Карасу-4, д. 11А

sic.icwc-aral.uz