

Сыдыкова Асель Жекшенбековна,
ст. препод.,
кафедра «Прикладная математика и информатика»
ИИТ КГТУ им. И. Раззакова
Аблабекова Чынара Азисовна,
к.ф.-м.н., доцент
кафедры программной инженерии
Международного университета Кыргызской Республики
Досболова Фарида Кубанычбековна,
преподаватель
кафедры программной инженерии
Международного университета Кыргызской Республики

Сыдыкова Асель Жекшенбековна,
улук окутуучу И. Раззакова а. КМТУ МТИнун
Колдонмо математика жана информатика кафедрасы
Аблабекова Чынара Азисовна,
физика - математика илимдеринин кандидаты, доцент
Кыргыз республикасынын эл аралык университети программалык
инженерия кафедрасы
Досболова Фарида Кубанычбековна,
окутуучу Кыргыз республикасынын эл аралык
университети программалык инженерия кафедрасы

Ablabekova Ch. A,
Ph.D., Associate Professor
Department of Software Engineering from the
International University of the Kyrgyz Republic
Sydykova A. Zh,
Senior Lecturer, Department of Applied Mathematics
and Computer Science IIT KSTU named after I.
Razzakov. **Dosbolova F.K,**
Lecturer
Departments of Software Engineering The
International University of the Kyrgyz Republic

**ИСПОЛЬЗОВАНИЯ ЦИФРОВЫХ ТЕХНОЛОГИЙ
ПРИ ИЗУЧЕНИЕ КОРРЕЛЯЦИОННОГО
АНАЛИЗА**

**ЦИФРАЛЫК ТЕХНОЛОГИЯЛАРДЫ КОЛДОНУП
КОРРЕЛЯЦИЯЛЫК АНАЛИЗДИ ОКУУ**

**USING DIGITAL TECHNOLOGIES
WHEN STUDYING CORRELATION ANALYSIS**

Аннотация. В данной статье описан опыт внедрения современных компьютерных технологий в области математического моделирования, как более интерактивное средство в обучении дисциплины «Стохастическое моделирование». Целью является ознакомление с возможностями и методикой выполнения корреляционного анализа путем использования цифровых инструментальных вычисления коэффициента парной корреляции и построение многофакторной модели планирования производства. В статье показываются возможности применения прикладного пакета программирования MS Excel и высокоуровневого языка программирования Python на примере корреляционного анализа производства. На конкретном примере построена многофакторная модель, осуществляется корреляционный анализ основных показателей производства, реализовано на основе модели с использованием цифровых технологий.

Ключевые слова: цифровых технологий, математическая модель, корреляционный анализ, коэффициент корреляции и т.д.

Аннотациясы. Бул макалада «Стохастикалык моделдөө» дисциплинасын окутууда кыйла интерактивдүү курал катары математикалык моделдөө чөйрөсүндө заманбап компьютердик технологияларды киргизүү тажрыйбасы баяндалат. Максаты жуптук корреляция коэффициентин эсептөө жана өндүрүштү пландаштыруунун көп факторлуу моделин түзүү үчүн санариптик каражаттарды колдонуу менен корреляциялык анализ жүргүзүүнүн мүмкүнчүлүктөрү жана ыкмалары менен тааныштыруу. Макала өндүрүштүн корреляциялык анализинин мисалында MS Excel тиркеме программалоо пакетин жана жогорку деңгээлдеги Python программалоо тилин колдонуу мүмкүнчүлүктөрүн көрсөтөт. Көп факторлуу модель конкреттүү мисалга курулат, негизги өндүрүштүк көрсөткүчтөрдүн корреляциялык анализи жүргүзүлөт, санариптик технологияларды колдонуу менен моделдин негизинде ишке ашырылат.

Негизги сөздөр: санариптик технологиялар, математикалык модель, корреляциялык анализ, корреляция коэффициенти ж.б.

Annotation. This article describes the experience of implementing modern computer technologies in the field of mathematical modeling, as a more interactive tool in teaching the discipline "Stochastic Modeling". The goal is to familiarize with the possibilities and methods of performing correlation analysis by using digital tools for calculating the pair correlation coefficient and building a multi-factor model of production planning. The article shows the possibilities of using the MS Excel application programming package and the high-level Python programming language using the example of correlation analysis of production. A multi-factor model is built on a specific example, a correlation analysis of the main production indicators is carried out, implemented on the basis of the model using digital technologies.

Key words: digital technologies, mathematical model, correlation analysis, correlation coefficient, etc.

Введение

Корреляционная связь – это согласованное изменение признаков, отражающее тот факт, что изменчивость одного признака находится в соответствии с изменчивостью другого. Парная корреляция изучает взаимосвязи между двумя случайными величинами, множественная – между большим числом величин. Основная задача корреляционного анализа – выявление и

оценка связи между случайными величинами, основная задача регрессионного анализа – установление формы и изучение зависимости между случайными величинами [6].

Несмотря на широкое применение, корреляционный анализ имеет свои ограничения. Важно помнить, что корреляция не означает причинно-следственную связь. Многие исследователи подчеркивают важность критического подхода к интерпретации корреляционных данных, чтобы избежать неверных выводов и обеспечить точность исследований.

Преимуществом корреляционного анализа для решения задач внутреннего контроля является обеспечение более гибкого анализа производственных процессов, так как в контрольную выборку можно поместить любые производственные показатели и при их определенном наборе видеть общую картину состояния производства либо при другом, более локальном наборе показателей, видеть более точные данные по какому-либо одному участку производства.

Корреляционный анализ используется в экономике, социологии и психологии, медицине, управления качеством, биометрии и других сферах.

Однако корреляционный анализ имеет свою специфику и методику. Очень важно использование этого метода только при соблюдении предпосылок расчета того, или иного, коэффициента корреляции

Постановка задачи

Рассмотрим компанию по производству бытовой техники в Кыргызстане. Компания «М» в Кыргызстане, производит товары для дома, такие как сушилка для белья, гладильная доска, кронштейны и антенны для телевизоров, эко урны, эко мангал, товары для дезинфекции, распределительные щиты и другие товары.

Наша задача заключается в анализе и оптимизации производственного плана, учитывая шесть параметров:

Определим тесноту линейной взаимосвязи между объёмом производства (y) и факторными признаками: количеством занятых в отрасли человек (x1); производительностью труда (x2); производственные мощности (x3); количество ресурсов (x4); объемом используемых основных фондов (x5); фондоотдачей (x6).

Чтобы определить тесноту связи между переменными вычислим множественные коэффициенты корреляции, которые являются обобщением парных коэффициентов корреляции. Предварительно вычислим по формуле матрицу парной корреляции [2].

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} \quad (1)$$

Нужно определить тесноту связи между объёмом производства и данными параметрами. Для этого необходимо провести корреляционный анализ, чтобы определить, какие из этих параметров имеют наибольшее влияние на объем производства [4].

Анализ проведем как зависят величины, зависят ли вообще, нам дано измерения в виде данной таблицы, шесть величин x. В этой задаче x рассматриваем как факторы, а y как следствие. Выявим взаимосвязь между переменными и силу связи, понять на сколько сильно они связаны, отберем переменные x, которые будем использовать в нашей модели для расчета y.

Таблица 1 –Значения результативного и факторных признаков

Отчетный период	Объем производства в сопоставимых ценах (млрд)	Факторные признаки					
		Труд		Оборудования		Капитал	
		Экстенсивный	Интенсивный	Экстенсивный	Интенсивный	Экстенсивный	Интенсивный
		Количество занятых в отрасли (чел.)	Производительность труда (млрд.чел.)	Производственные мощности	Количество ресурсов	Объем используемых основных фондов (млрд)	Фондоотдача (процент %)
		x1	x2	x3	x4	x5	x6
2010	3142855	416	340	1180	1430	1403232	32
2011	2518666	442	326	1180	1200	1215863	28
2012	3140108	456	338	1200	1304	1435038	31
2013	3228636	430	403	1220	1329	1445628	34
2014	3114684	398	345	1220	1310	1331032	33
2015	3482757	422	429	1220	1356	1482419	35
2016	3487215	412	438	1200	1358	1456646	35
2017	2551891	368	348	1200	1270	937359	27
2018	2494900	340	340	1200	1290	622575	29
2019	2526515	344	352	1200	1210	1226931	34

Используемые цифровые технологии

Используем встроенные инструменты, как анализ данных, исследуем однородность переменной, отличается ли значения переменной [7].

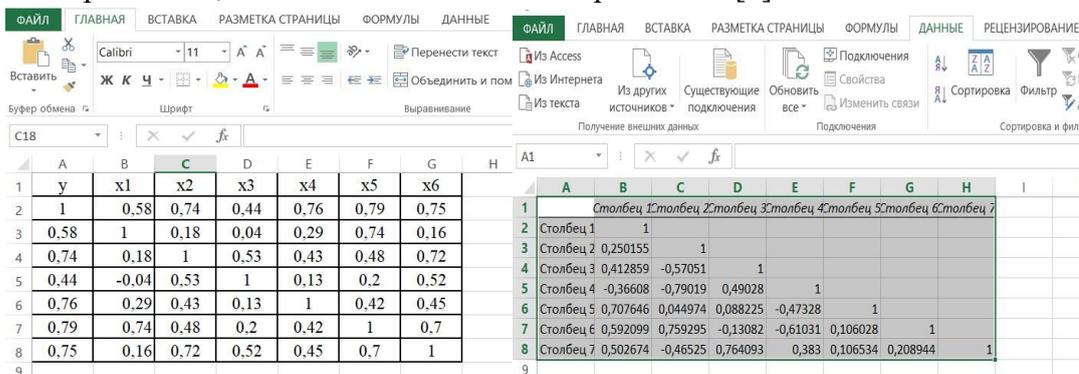


Рис. 1 - Интерпретация взаимосвязи между показателями Рис. 2 - Коэффициенты корреляции

Если коэффициенты парной корреляции имеют большие значения ($r_{y/x} > 0,7$), то согласно шкале Чеддока, между ними есть высокая корреляционная зависимость [3].

Коэффициенты парной корреляции результативного признака – объемов производства (y) с четырьмя факторными признаками: производительностью труда (x2); количество ресурсов (x4); объемом используемых основных фондов (x5); фондоотдачей (x6), имеют довольно большие значения ($> 0,7$), это говорит о достаточно высокой корреляционной зависимости результативного признака с перечисленными факторными признаками.

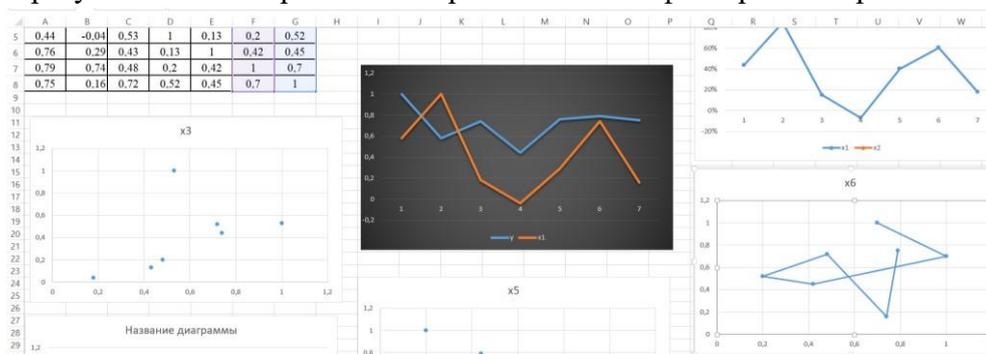


Рис. 3 - Графическая интерпретация
Таблица 2 –Матрица парной корреляции

Результативные и факторные признаки		Объём производства	Количество занятых в отрасли человек	Производительность труда	Производственные мощности	Количество ресурсов	Объём используемых основных фондов	Фондоотдачи
	y	x_1	x_2	x_3	x_4	x_5	x_6	
Объём производства	y	1	0,58	0,74	0,44	0,76	0,79	0,75
Количество занятых в отрасли человек	x_1	0,58	1	0,18	0,04	0,29	0,74	0,16
Производительность труда	x_2	0,74	0,18	1	0,53	0,43	0,48	0,72
Производственные мощности	x_3	0,44	-0,04	0,53	1	0,13	0,20	0,52
Количество ресурсов	x_4	0,76	0,29	0,43	0,13	1	0,42	0,45
Объём используемых основных фондов	x_5	0,79	0,74	0,48	0,20	0,42	1	0,70
Фондоотдачи	x_6	0,75	0,16	0,72	0,52	0,45	0,70	1

Корреляционный анализ дает основание исключить из модели экстенсивные переменные - количество занятых в отрасли человека (x_1) и (x_3).

Описание программы

Чтобы записывать компьютерные программы существует формальный язык программирования. Он состоит из набора правил действий, дающих команду, выполняющему ПК. Специалисты этого профиля владеют несколькими языками программирования, представляющего собой синтаксис, лексику и семантику условий, и выводов [1].

В нашей работе используем разработку программы для исследования корреляционной зависимости.

Python — язык программирования с высоким уровнем общего назначения, имеет динамическую строгий тип для автоматического управления памятью, ориентированного на повышение производительности разработчика, читаемости кода и его качества, имеет обеспечение переносимости написанного на нём программ. Также язык популярен тем, что интерпретируется и используется для написания скриптов [9].

Рис 5 – Фрагмент кода для регрессионной модели

```
def reg_m(y, x):
    ones = np.ones(len(x[0]))
    X = sm.add_constant(np.column_stack((x[0], ones)))
    for ele in x[1:]:
        X = sm.add_constant(np.column_stack((ele, X)))
    results = sm.OLS(y, X).fit()
    return results
print (reg_m(y, x).summary())
```

С помощью функция corrcoef опишем программный код для парных коэффициентов корреляции.

Рис 6 – Фрагмент кода парной корреляции

```
## Парная корреляция
corr=np.corrcoef(y, x)
print ("\nПарные коэффициенты корреляции \n", corr)
```

Опишем программный код для частного коэффициента корреляции.

Рис 7 – Фрагмент кода частного коэффициента корреляции

```
def ch_k_k(x):
    r=x
    private_corr_coef=[[0]*7]*7
    for i in range(1,7):
        for j in range(1, 7):
            if (i != j):
                r[i][j] = round((x[i][j] - (x[i][0] * x[j][0])) / math.sqrt((1 - math.pow(x[i][0], 2)) * (1 - math.pow(x[j][0], 2))),3)
                private_corr_coef[i][j]=r[i][j]

    print('\n')

    r[1:].clear()
    private_corr_coef=np.array(r)
    private_corr_coef.transpose()
    for i in range(7):
        private_corr_coef[0][i]=0
        private_corr_coef[i][0]=0
    print ("Частной коэффициент корреляции \n", private_corr_coef)
```

Далее представим программный код для построения графиков.

```
fig, ax = plt.subplots()
ax.plot(x0, y, label='Объем производства (y) построенных на основе эмпирических данных')
ax.plot(x0, y0, label='Объем производства (y) построенных на основании уравнение регрессии')
ax.legend(loc='upper left')
ax.set_xlabel(r'Годы')
ax.set_ylabel(r'Объем производства y')
plt.show()
```

Рис 8 – Фрагмент кода построения графика

Как видно, что в этой модели существует мультиколлинеарность. Для того чтобы избавиться от мультиколлинеарности необходимо исключить одну переменную из коллинеарных пар. Для определения какую переменную исключить из коллинеарных пар, вычислим парный и частный коэффициенты корреляции [5].

После того как мы получили результаты парных и частных коэффициентов корреляции, сравнивая их значения между собой, определяем какие переменные исключить из модели. В нашем случае составленная модель прогнозирования производственных затрат при планировании производства поможет предприятию существенно планировать все затраты и брать во внимание параметры и факторы, влияющие на производство.

Результаты

Рассмотрен алгоритм построения корреляционного анализа. Алгоритм многофакторного модели был реализован в среде разработки IDLE, с использованием языка программирование Python.

В статье подтверждены применения методов корреляционного анализа наличия и значимости связи неслучайных входных параметров и случайных выходных результатов и предложена зависимость для расчета коэффициентов их линейной корреляции. Определены величины и значимость коэффициентов корреляции при планировании производства.

```

import pandas as pd

# Данные
data = {
    'Y': [3142855, 2518666, 3140108, 3228636, 3114684, 3482757, 3487215, 2551891, 2494900],
    'X1': [416, 442, 456, 430, 398, 422, 412, 368, 340, 344],
    'X2': [340, 326, 338, 403, 345, 429, 438, 348, 340, 352],
    'X3': [1180, 1180, 1200, 1220, 1220, 1220, 1200, 1200, 1200, 1200],
    'X4': [1430, 1200, 1304, 1329, 1310, 1356, 1358, 1270, 1290, 1210],
    'X5': [1403232, 1215863, 1435038, 1445628, 1331032, 1482419, 1456646, 937359, 622575],
    'X6': [32, 28, 31, 34, 33, 35, 35, 27, 29, 34]
}

# Создание DataFrame
df = pd.DataFrame(data)

# Расчет корреляционной матрицы
correlation_matrix = df.corr()

correlation_matrix

```

Рис 9 – Код парной корреляции

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Аблабекова Ч.А., Жээналиева Н.М., Досболова Ф.К. Использование цифровых технологий для транспортной задачи. // Вестник Международного университета Кыргызстана, — №4(56)2024. — с.3-9
https://drive.google.com/file/d/1JqUKT_7tbzo3pI2TcVSnGOdPjCVnJ5Br/view
2. Буровский, Е. А., Гришунина, Ю. Б. Задачи математической статистики и их решение с использованием языка программирования Python: учеб. Пособие - М.: Изд. дом Высшей школы экономики, 2022. — 64 с. [Электронный ресурс] –Режим доступа <https://publications.hse.ru/pubs/share/direct/776603038.pdf>
3. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учебное пособие - М.: Высшее образование, 2007. - 404 с. [Электронный ресурс] –Режим доступа. <https://elenagavrile.narod.ru/ms/gmurman.pdf>
4. Гришунина Ю. Б. Задачи математической статистики и их решение с использованием приложения Microsoft Excel. Учеб. пособие по дисциплине «Теория вероятностей и математическая статистика» - М.: РИО МИЭМ НИУ ВШЭ, 2023. [Электронный ресурс] –Режим доступа <https://publications.hse.ru/pubs/share/folder/x7b5b5bwln/77277022.pdf>
5. Степанова, М. Д. Проверка статистических гипотез: учебно-метод. пособие по курсу «Статистические основы индуктивного вывода» для студентов специальности «Искусственный интеллект» – Мн.: БГУИР, 2020. – 36 с.
6. Шихалев А. Корреляционный анализ. Непараметрические методы – Казан. ун-т, 2015. — 58с.
7. https://e.vyatsu.ru/pluginfile.php/462619/mod_resource/content/3B7.pdf
8. https://www.orgma.ru/sveden/education/Metod_Osnovy_biostatistiki_Pediatrica-003.pdf