Забелина С. Б.
канд. пед. наук, доцент, доцент кафедры
высшей алгебры, математического анализа, геометрии
Государственный университет просвещения,
zabelina_sb@mail.ru
Шилова З. В.
канд. пед. наук, доцент, доцент
кафедры гуманитарных и естественных наук
Государственный университет «Дубна»
zoya@soi.su

ПРИМЕНЕНИЕ МЕТОДОВ КОРРЕЛЯЦИОННО-РЕГРЕССИОННОГО АНАЛИЗА ПРИ РЕШЕНИИ ПРОФЕССИОНАЛЬНО-ОРИЕНТИРОВАННЫХ ЗАДАЧ

Аннотация: в статье раскрывается один из аспектов применения методов математической статистики, а, именно корреляционно-регрессионного анализа при решении профессионально-ориентированных задач. Профессионально-педагогическая подготовка учителя математики предполагает применение профессионально ориентированных задач при обучении студентов математике, в том числе при изучении таких дисциплин, как «Математическая статистика», «Анализ данных» и др. В статье на примере практической задачи приведено решение с использованием методов корреляционного-регрессионного анализа.

Ключевые слова: корреляция, регрессия, математическая статистика, профессионально ориентированные задачи.

Zabelina Svetlana B.
Candidate Sciences (Educ.), Associate Professor, Associate Professor
of the Department of Higher Algebra, Mathematical Analysis, Geometry,
State University of Education,
zabelina_sb@mail.ru

Shilova Zoia V.
Candidate Sciences (Educ.), Associate Professor, Associate Professor

of the Department of Humanities and Natural Sciences,
State University «Dubna»,
zoya@soi.su

APPLICATION OF CORRELATION AND REGRESSION ANALYSIS METHODS WHEN SOLVING PROFESSIONALLY-ORIENTED TASKS

***Abstract*:** this article aims at describing one of the aspects of using methods of mathematical statistics while training the decision of the professionally oriented tasks. Professional and pedagogical training of a mathematics teacher involves the use of professionally oriented tasks in teaching students mathematics. Methods of mathematical statistics is considered in such subjects as “Mathematical Statistics”, “Data Analysis” etc. The article uses the example of a practical problem to provide a solution using the methods of correlation and regression analysis.

***Keywords*:** correlation, regression, mathematical statistics, professionally oriented tasks.

Математические знания являются неотъемлемой частью общечеловеческой культуры, математические знания необходимы учащимся при изучении других дисциплин, в быту и при формировании профессиональной направленности студентов. Проблема профессиональной направленности обучения математике изучена во многих аспектах. Рассмотрим один из них, посвященный использованию в учебном процессе профессионально ориентированных задач, которые, в свою очередь, органично соединяют в своей фабуле математические понятия, закономерности и профессионально значимое содержание.

Профессионально-педагогическая подготовка учителя математики предполагает применение профессионально ориентированных задач при обучении студентов, что, в свою очередь, позволяет наиболее эффективно реализовать профессиональную направленность обучения математике. Здесь под профессионально ориентированной задачей будем понимать сюжетные задачи, фабула которой заимствована из той или иной сферы профессиональной деятельности человека, а решение отыскивается математическими средствами. Фабула задачи может содержать термины, связанные с производством или профессией, ситуации, связанные с конкретными видами профессий или отраслями производства.

Рассмотрим аспекты профессионально-педагогической подготовки учителя математики при обучении математической статистике. Отметим, что методы математической статистики играют большую роль в анализе и обобщении полученных данных, что позволяет сформировать у обучающихся умение анализировать, обобщать, прогнозировать, выявлять тенденции в развитии явлений (экономических, статистических и др.). Между тем ни один из методов математической статистики, взятый сам по себе, не может претендовать на универсальность, на полную гарантию объективности получаемых данных. Поэтому каждому исследователю следует стремиться, с одной стороны к совершенствованию техники применения любого конкретного метода, а с другой – к комплексному, взаимоконтролирующему использованию разных методов для решения одной и той же задачи (проблемы). Поэтому владение всей системой методов дает возможность получить существенные теоретические и практические результаты по исследуемой задаче (проблеме).

В свою очередь, методы математической статистики дают возможность научить обучающихся:

1) компактно и информативно описывать статистические данные;

2) устанавливать степень достоверности сходства и различия исследуемых объектов на основании результатов измерений их показателей;

3) анализировать наличие или отсутствие зависимости между различными показателями (явлениями);

4) количественно описывать эти зависимости;

5) выявлять информативные показатели;

6) классифицировать изучаемые объекты и прогнозировать значения их показателей и характеристик и т.д.

Среди методов математической статистики отдельно выделим методы корреляционного анализа и регрессионного анализа, они рассматриваются при изучении таких дисциплин, как: «Математическая статистика», «Анализ данных» и др.

Корреляционно-регрессионный анализ предполагает решение, в основном, двух типов задач: 1) выяснить, насколько сильная связь между признаками; 2) построить уравнение регрессии и выполнить прогнозирование.

Рассмотрим задачу, позволяющую проанализировать тенденции общего роста внутреннего рынка легковых автомобилей (данные взяты с сайта http://www.e-stat.ru/), применяя методы корреляционно-регрессионного анализа.

Выделим временной тренд продаж автомобилей за 24 месяца.

год	месяц	месяц	продажа
1 год	январь	1	75505
	февраль	2	75608
	март	3	82615
	апрель	4	87571
	май	5	90814
	июнь	6	94539
	июль	7	98282
	август	8	101599
	сентябрь	9	99492
	октябрь	10	97094
	ноябрь	11	95576
	декабрь	12	92638
2 год	январь	13	93570
	февраль	14	95569
	март	15	100289
	апрель	16	101867
	май	17	109334
	июнь	18	112349
	июль	19	117285
	август	20	120948
	сентябрь	21	115827
	октябрь	22	113370
	ноябрь	23	113337
	декабрь	24	111933

Выполним построение регрессии. Для регрессии вида:

$$Y = \beta X + \alpha$$

найдем коэффициенты по формулам:

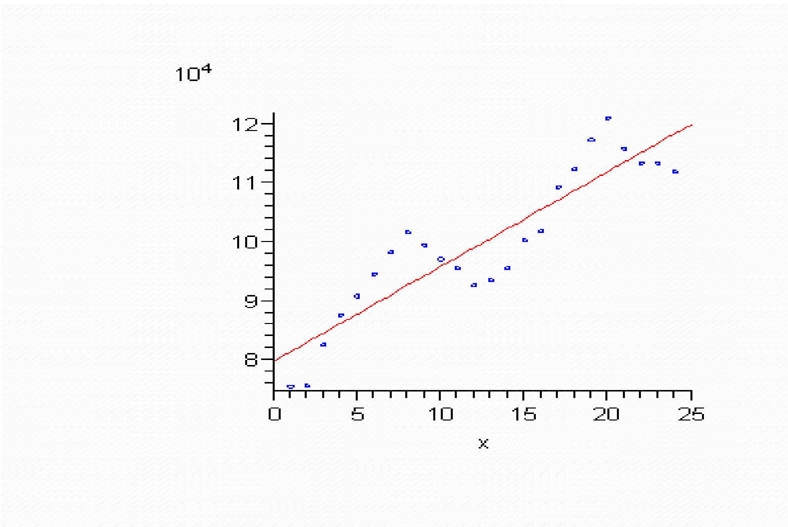
$$\beta = \frac{24\left(\sum_{n=1}^{24} X_n Y_n\right) - \left(\sum_{n=1}^{24} X_n\right)\left(\sum_{n=1}^{24} Y_n\right)}{24\left(\sum_{n=1}^{24} X_n^2\right) - \left(\sum_{n=1}^{24} X_n\right)^2}$$

$$\alpha = \frac{1}{24}\left(\sum_{n=1}^{24} Y_n\right) - \frac{1}{24}\,\beta\left(\sum_{n=1}^{24} X_n\right)$$

Вычислим коэффициента регрессии: β =1608,88, α =79764,46.

Тогда линейная регрессия будет иметь вид: Y = 1608,88·X + 79764,46.

Коэффициент β позволяет сформировать представление взаимосвязи между двумя переменными, является мерой чувствительности одной переменной к изменениям другой. Он показывает, насколько изменяется значение одной переменной при изменении другой, то есть если мы изменим значение X на 1 единицу, то Y изменится на 1608,88 единиц. Это позволяет нам прогнозировать и анализировать влияние изменений одной переменной на другую. Построим точечную диаграмму и линию тренда.



Рассмотрим аспект статистической значимости построенной модели, используя коэффициенты корреляции и детерминации, критерии Фишера и Стьюдента.

Коэффициент корреляции:

$$corr(X, Y) = \frac{E(X\,Y)\ \mathrm{K}\ E(X)\,E(Y)}{\sqrt{Var(X)\,Var(Y)}}$$

где

$$E(XY) = \frac{1}{24}\left(\sum_{k=1}^{24} X_k Y_k\right) - \frac{1}{576}\left(\sum_{k=1}^{24} X_k\right)\left(\sum_{k=1}^{24} Y_k\right)$$

$$Var(X) = \frac{1}{24}\left(\sum_{k=1}^{24} X_k^2\right) - E(X)^2$$

$$Var(Y) = \frac{1}{24}\left(\sum_{k=1}^{24} Y_k^2\right) - E(Y)^2$$

$$corr(X, Y) := 0.9037200$$

Коэффициент корреляции показывает, что связь прямая и сильная (линейная).

Коэффициент детерминации:

$$R^2 = corr(X, Y)^2$$

$$R^2 = 0.8167098$$

Коэффициент детерминации показывает, что регрессия объясняется на 81,67 % вариации признака.

Убедимся в значимости модели посредством статистики Фишера:

$$F = \frac{(N-2)R^2}{1 - R^2}$$

$$F = 98.02824$$

$$F_{crit} = F_{1,\ 22,\ 0.05} = 4{,}3$$

Эмпирическое значение статистического критерия больше критического значения, следовательно, регрессия статистически значима.

Проверим статистическую значимость коэффициента корреляции с помощью критерия Стьюдента:

$$t = \frac{corr(X, Y)\sqrt{N - 2}}{\sqrt{1 - corr(X, Y)^2}}$$

$$t = 9.900924$$

$t_{crit} = t_{22,\,0.05} = 2{,}07$

Эмпирическое значение статистического критерия больше критического значения, поэтому выборочный коэффициент корреляции значимо отличается от нуля, то есть статистически значим.

Найдем среднюю ошибку аппроксимации

$$A = \frac{1}{24}\left(\sum_{n=1}^{24}\left|\frac{Y_n - 1608.880\,X_n - 79764.46}{Y_n}\right|\right)$$

$A = 0.0477677$

Коэффициент аппроксимации составляет 4,78%, что существенно ниже максимально допустимой величины 8-10%, что говорит о хорошем качестве модели.

Рассмотрим статистику Дарбина-Уотсона

$$d = \frac{\sum_{k=2}^{24}(e_k - e_{k-1})^2}{\sum_{k=1}^{24} e_k^2}$$

$d = 0.378751$

Попали в зону положительной автокорреляции.

Выполним прогноз: 1) точечный прогноз для $x_0 = 25$, $y(x_0) = 1{,}199865 \cdot 105$; 2) интервальный прогноз с надежностью 95%:

$$\left[25\,\beta + \alpha - 2.07\,s\sqrt{1 + \frac{1}{N} + \frac{(25 - E(X))^2}{\sum_{n=1}^{24}(X_n - E(X))^2}},\ 25\,\beta + \alpha + 2.07\,s\sqrt{1 + \frac{1}{N} + \frac{(25 - E(X))^2}{\sum_{n=1}^{24}(X_n - E(X))^2}}\right]$$

или

$$[1.076087\,10^5,\ 1.323643\,10^5]$$

Точечный прогноз для $x_0 = 26$, $y(x_0) = 1{,}215953 \cdot 10^5$.

Выполним интервальный прогноз с надежностью 95%:

$$\left[26\,\beta+\alpha-2.07\,s\,\sqrt{1+\frac{1}{N}+\frac{(26-E(X))^2}{\sum_{n=1}^{24}\left(X_n-E(X)\right)^2}},\ 26\,\beta+\alpha+2.07\,s\,\sqrt{1+\frac{1}{N}+\frac{(26-E(X))^2}{\sum_{n=1}^{24}\left(X_n-E(X)\right)^2}}\right]$$

или

$$[1.090993\,10^5,\ 1.340913\,10^5]$$

Отметим, что при построении прогноза необходимо учитывать показатель сезонности, так как может оказать значительное влияние на точность прогноза. Между тем для эффективного изучения вероятного спроса на автомобили необходимо учитывать не только фактор времени, но и сезонность, а также зависимость от других параметров. Кроме того, на спрос на автомобили могут влиять и внешние факторы, такие как экономические показатели, доверие потребителей и наличие других вариантов передвижения.

Таким образом, использование в учебном процессе профессионально-ориентированных задач способствует повышению у студентов положительной мотивации обучения, формированию у них прочных базовых знаний, столь необходимых для профессиональной деятельности. Кроме того, это позволяет студентам систематизировать предметный материал, глубже осознать взаимосвязи изучаемых понятий, материал по дисциплинам «Математика», «Математическая статистика», «Анализ данных» и т. д., а также рационально решать целый ряд профессионально-ориентированных задач.

Литература

1. *Алибеков И.Ю.* Теория вероятностей и математическая статистика в среде MATLAB: учебное пособие. М.: Лань, 2019. 184 с.

2. *Гмурман В.Е.* Теория вероятностей и математическая статистика: учебник. М.: Юрайт, 2018. 480 с.

3. *Шилова З.В.* Теория вероятностей и математическая статистика: учебное пособие / З.В. Шилова, О.И. Шилов. Саратов: Ай Пи ар Букс, 2015. 158 с.