УДК 519.221

Лабораторные работы в курсе математической статистики

© Ветров Л.Г., А.Л. Сунчалина

МГТУ им. Н.Э. Баумана, Москва, 105005, Россия

Рассматриваются методические аспекты преподавания курса математической (прикладной) статистики в высшем учебном заведении. Обоснована необходимость включения в программу курса лабораторных работ. Приведен ориентировочный список тем лабораторных работ, содержащий методы статистического моделирования, оценки параметров распределений, проверку статистических гипотез, дисперсионный анализ, регрессионный анализ, непараметрические методы статистики и методы выборочного обследования. Для каждой лабораторной работы указан примерный список включаемых в нее задач. Частично указаны методы решения задач с использованием электронных таблиц Excel.

Ключевые слова: *математическая статистика, лабораторная работа, статистическое моделирование, оценки параметров, критерии проверки гипотез, дисперсионный и регрессионный анализ.*

Введение. При чтении курса теории вероятностей и математической статистики на семинарские занятия обычно отводится от 30 до 50 % времени. И при этом в теории вероятностей при проведении семинарских занятий не возникает никаких проблем: всегда можно подобрать такой набор задач, который закрепит теоретический материал, излагаемый на лекциях. Однако в зависимости от количества часов и контингента обучаемых студентов этот набор может изменяться от минимально необходимого до достаточно широкого. Обычно задачи по теории вероятностей не требуют сложных вычислений и больших затрат времени. В курсе математической статистики, к сожалению, все обстоит значительно сложнее. За редким исключением, любая задача математической статистики требует большого количества вычислений. Если ограничиваться решением задач с малыми выборками, то практически теряется смысл многих задач: доверительные интервалы большие, ошибки второго рода при проверке гипотез большие и т. д. При использовании выборок большого объема решение задач становится затруднительным по причине большого количества вычислений. Даже подсчет выборочного среднего и выборочной дисперсии может занять непозволительно большое время. Однако за последнее время вычислительная техника совершила такой скачок, что арифметические вычисления вообще не должны представлять каких-либо трудностей и занимать много учебного времени. При этом желательно для проведения вычислений использовать наиболее распространенные и всем доступные программные продукты (например, электронные таблицы Excel позволяют без больших усилий и затрат времени проводить статистическую обработку данных).

Таким образом, изменившиеся обстоятельства требуют изменения формы проведения учебного процесса. В курсе математической статистики — это может быть введение в курс обучения лабораторных работ. При сохранении семинарских занятий (в сокращенном виде) на них имеет смысл разбирать простейшие примеры с выяснением смысла используемых статистических процедур, а также учить студентов «читать» результаты обработки при использовании соответствующей статистической процедуры из «Пакета анализа» электронных таблиц Excel, выделять наиболее информативные показатели, выдаваемые программой, и делать соответствующие выводы.

Далее рассмотрим примерный список лабораторных работ, который может варьироваться в зависимости от объема часов, выделяемых на дисциплину.

1. Методы статистического моделирования, закон больших чисел и сравнение оценок параметров. Курс теории вероятностей обычно заканчивается предельными теоремами: закон больших чисел и центральная предельная теорема. Для выполнения усиленного закона больших чисел необходимо и достаточно существование конечного математического ожидания (обычно формулируется без доказательства). Но что будет, если это условие не выполняется, студент может прочувствовать, только увидев своими глазами, как ведет себя выборочное среднее в том и другом случае. Цель работы — научить студента пользоваться функциями Excel, моделировать выборки с заданным непрерывным законом распределения (метод обратной функции) и на примере двух распределений (Гаусса и Коши) посмотреть, как ведет себя выборочное среднее для этих распределений с ростом объема выборки. Для ряда распределений (равномерное, Гаусса, Бернулли, биномиальное, пуассоновское, дискретное) в «Пакете анализа» имеется программа «Генерация случайных чисел». Для других непрерывных распределений обычно используется метод обратной функции: если $\xi \approx R[0;1]$ (ξ равномерно распределена на отрезке [0,1]), то $\eta = F^{-1}(\xi)$ имеет функцию распределения $F(x)$, где $F(x)$ — произвольная непрерывная функция распределения.

На рис. 1 и 2 представлены графики, описывающие поведение выборочного среднего и выборочной медианы соответственно для двух распределений: Гаусса и Коши.

Анализ этих графиков позволяет студенту сделать вывод, что для распределения Коши (математическое ожидание не существует) выборочное среднее является неприемлемой оценкой центра распределения. Далее, для этих двух распределений можно сравнить поведение выборочных средних и выборочных медиан и сделать соответствующие выводы. В результате этой работы студент должен уяснить, что если генеральная совокупность имеет распределение с «тяжелыми хвостами»

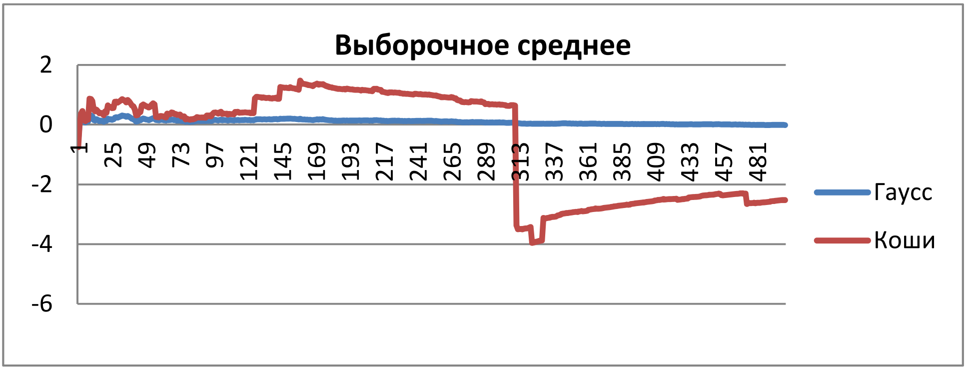

Рис. 1. График зависимости выборочного среднего от объема выборки

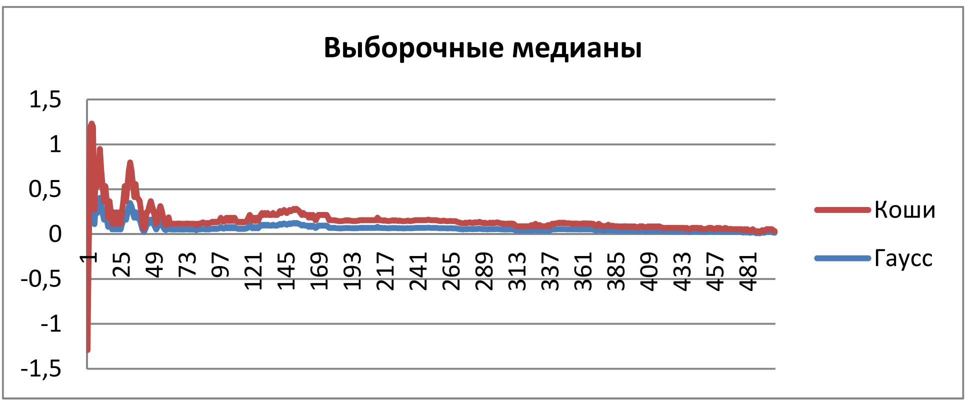

Рис. 2. График зависимости выборочной медианы от объема выборки

(например, зарплата жителей региона), то оценка среднего генеральной совокупности с помощью выборочного среднего может оказаться неэффективной.

Вторая задача лабораторной работы демонстрирует возможности методов статистического моделирования при сравнении эффективности различных несмещенных оценок параметров. Известно, что эффективность несмещенных оценок характеризуется их дисперсией: чем меньше дисперсия, тем оценка лучше.

Для модели равномерного распределения $R[0;\theta]$ предлагается сравнить три несмещенные оценки: $\theta_1 = 2\overline{X}$ (метод моментов), $\theta_2 = \frac{n+1}{n} X_{(n)}$ (скорректированная оценка метода максимального правдоподобия) и $\theta_3 = 2X_{(n/2)}$ (удвоенная выборочная медиана).

Для θ_1 дисперсия считается легко, с несколько большими трудностями можно посчитать дисперсию θ_2, а вот для θ_3 это сделать значительно сложнее. Смоделировав выборку достаточно большого размера с распределением $R[0;1]$ (функция «Случайное число») и вычислив значения этих оценок, можно сделать вывод об их эффективности по близости значений оценок к 1. На рис. 3 представлена типичная картина зависимости этих трех оценок от объема выборки (объем выборки порядка 2000).

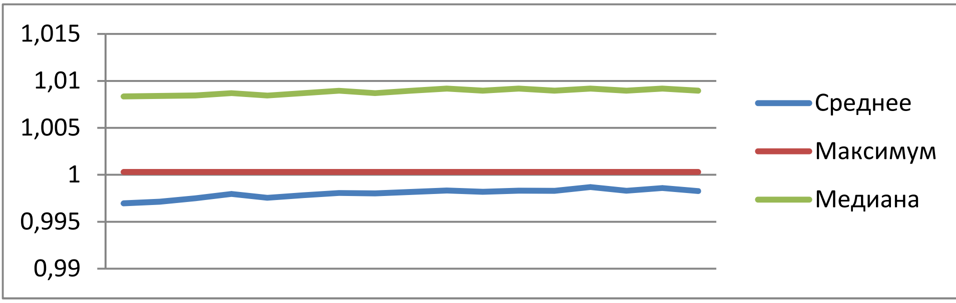

Рис. 3. Сравнение оценок

2. Построение доверительных интервалов и проверка статистических гипотез. При построении доверительных интервалов для параметров основных параметрических моделей обычно достаточно рассчитать выборочные характеристики $\left(\bar{X};\, S^2\right)$ и квантили соответствующих распределений статистик. Все это содержится в наборе функций Excel, и задача построения доверительных интервалов сводится к очень простым вычислениям. Единственное, на что следует обратить внимание студентов, — это то, что среди функций Excel имеется две функции дисперсии: «ДИСПР» — обычная выборочная дисперсия $S^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 / n$ и «ДИСП» — «исправленная» выборочная дисперсия $\bar{S}^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 / (n-1)$.

В «Пакете анализа» для проверки гипотез есть ряд критериев, но студент должен учитывать два момента: а) все эти критерии имеют отношение к нормальному (гауссов) распределению; б) это критерии однородности, т. е. сравниваются средние значения или дисперсии двух выборок. Поэтому при использовании этих критериев для негауссовых распределений они не гарантируют заданную вероятность ошибки первого рода, а для проверки простой гипотезы о среднем или дисперсии границы критической области следует вычислять с помощью функций Excel. Еще одна особенность программ проверки гипотез в «Пакете анализа» — это альтернативные гипотезы: односторонние критические значения можно использовать при проверке основной гипотезы $H_0 : a_1 = a_2$ при альтернативе $H_1 : a_1 > a_2$. Для того чтобы разобраться в этом, достаточно в лабораторной работе взять две выборки с заведомо различающимися средними значениями и применить критерий проверки гипотезы о равенстве средних, поменяв местами выборки. Из табл. 1 видно, что критерий работает адекватно, если в качестве первой выборки взята выборка с большим средним.

Таблица 1

Двухвыборочный t-тест с одинаковыми дисперсиями

	X	*Y*		*Y*	*X*
Среднее	**3**	**6,2**	Среднее	**6,2**	**3**
Дисперсия	2,5	7,7	Дисперсия	7,7	2,5
Наблюдения	5	5	Наблюдения	5	5
Объединенная дисперсия	5,1		Объединенная дисперсия	5,1	
df	8		df	8	
t-статистика	**–2,24045**		t-статистика	**2,240448**	
t критическое одностороннее	**1,859548**		t критическое одностороннее	**1,859548**	
t критическое двухтороннее	**2,306004**		t критическое двухстороннее	**2,306004**	

Приведем пример еще одной задачи, решение которой на обычных семинарах невозможно. Задача состоит в нахождении необходимого объема выборки для построения критерия проверки двух простых гипотез $H_0:\sigma^2=\sigma_1^2$, $H_1:\sigma^2=\sigma_2^2$ о дисперсии нормального распределения с заданными вероятностями ошибок первого и второго рода α и β. Теория говорит, что

$$n^*=\min\left\{n:\frac{\chi_{\alpha,n}}{\chi_{1-\beta,n}}>\frac{\sigma_1^2}{\sigma_2^2}\right\}\quad(\sigma_1<\sigma_2),$$

где $\chi_{\alpha,n}$ — квантиль распределения «хи-квадрат» уровня α с n степенями свободы.

Но в любой книге [1–7] по математической статистике приводятся таблицы квантилей распределения χ_n^2 только для ограниченного числа значений α и n. Поэтому по этим таблицам найти n^* невозможно. Однако использование таблиц Excel и встроенных в них функций позволяет решить эту задачу для любых значений σ_1,σ_2, α и β. Рассмотрим пример. Пусть проверяются две простые гипотезы $H_0:\sigma^2=1$, $H_1:\sigma^2=1,4$. Надо построить критерий с вероятностями ошибок первого и второго рода соответственно $\alpha=0,01$, $\beta=0,05$. Какой минимальный объем выборки для этого необходим?

В табл. 2 приведен фрагмент листа Excel с решением данной задачи, из которого следует, что минимальный объем выборки равен 286. При этом если в формулах для квантилей распределения χ^2_n значения α и β берутся из фиксированных ячеек, то для решения задачи с новыми α и β достаточно изменить содержание только этих двух ячеек.

Таблица 2

n	$\chi_{\alpha,n}$	$\chi_{1-\beta,n}$	$\chi_{\alpha,n} / \chi_{1-\beta,n}$	σ_1^2 / σ_2^2
283	230,61	323,24	0,713	0,714
284	231,51	324,31	0,714	0,714
285	232,42	325,37	0,714	0,714
286	233,32	326,44	0,715	0,714

Эта же задача при проверке гипотез о средних значениях вполне может быть и должна быть решена на семинаре или на лекции, так как ее решение сводится к вычислению значений функции Лапласа, для которой в любой книге по математической статистике приведены таблицы. Аналогичная задача может быть решена для проверки двух простых гипотез для экспоненциального распределения. В этом случае задача также сводится к сравнению квантилей распределения χ^2_{2n}.

В современном курсе математической статистики особое место должно быть отведено непараметрическим методам [8, 9], так как применение методов математической статистики ушло далеко за пределы приложений в технике, где использование нормального распределения и других параметрических моделей вполне оправданно. Например, в приложениях в экономике возникают распределения, для которых подобрать соответствующую параметрическую модель весьма затруднительно.

В работе имеет смысл проверить гипотезу об однородности двух выборок с использованием рангового критерия Манна — Утни — Уилкоксона. Можно, конечно, воспользоваться каким-либо серьезным статистическим пакетом (типа «Статистика»), но в этом случае «пакет» выдаст только результат — значения статистики критерия, но для студента останется «черным ящиком». Кроме того, подсчет статистики Манна — Утни — Уилкоксона очень просто организуется с использованием таблиц Excel, и при этом достаточно просто и наглядно можно объяснить студенту, какие значения статистики (большие или малые) говорят в пользу той или иной альтернативы. Еще одно преимущество таблиц Excel состоит в том, что студент, заготовив на странице шаблон, имеет возможность использовать его для новых данных — достаточно в ячейки с данными внести их новые значения. Тогда на выходе получим для этих данных значения статистики критерия.

Непараметрический критерий Манна — Утни — Уилкоксона применяется для проверки гипотезы о совпадении распределений против альтернативы правого или левого сдвига. В качестве статистики критерия выступает статистика

$$U = \sum_{i=1}^{n}\sum_{j=1}^{m} u_{i,j},$$

где

$$u_{i,j} = \begin{cases} 1, & x_i < y_j, \\ 1/2, & x_i = y_j, \\ 0, & x_i > y_j. \end{cases}$$

Вычисление этой статистики «вручную» при больших n и m является весьма затруднительной задачей.

Если расположить данные для двух выборок так, как это изображено в табл. 3, и в ячейку B2 занести формулу

=ЕСЛИ($A2=B$1;0,5;ЕСЛИ($A2<B$1;1;0)),

то «растянув» ячейку B2 на весь прямоугольный диапазон и просуммировав содержимое ячеек этого диапазона, получим значение статистики Манна — Утни — Уилкоксона.

Таблица 3

	A	B	…	
1		Y_1	…	Y_m
2	X_1			
…	…			
$n + 1$	X_n			

При этом изменение данных о выборках (первая строка и первый столбец) автоматически изменяет значение статистики критерия.

Можно также на лабораторных работах (если время позволяет) отработать проверку гипотезы согласия Колмогорова и критерия однородности Колмогорова — Смирнова. «Вручную» вычислять значения статистик этих критериев не очень просто (особенно для выборок большого объема). Использование же таблиц Excel позволяет организовать вычисление статистик достаточно легко, при этом главное состоит в том, что студент осознает, «что он вычисляет», а не просто пользуется пакетом как «черным ящиком».

3. Дисперсионный анализ. В приложениях математической статистики очень часто возникает задача выяснения влияния на результат качественного фактора, например влияние вида протектора на износоустойчивость шины (в технике), тип методики обучения на результаты обучения (в образовании), пол работающего на уровень средней зарплаты (в социологии) и т.д. Все это задачи дисперсионного анализа. Для ответа на вопрос, влияет ли данный фактор на результат, можно использовать однофакторный и двухфакторный дисперсионный анализ из «Пакета анализа».

В результате этой лабораторной работы студент должен научиться понимать результаты работы данной программы: где находится информация и как по этой информации сделать правильные выводы. Кроме того, он должен понимать специфику таблиц: при использовании двухфакторного анализа без повторений для любого набора двух значений факторов задано ровно одно наблюдение. Это специфика двухмерных таблиц. При наличии более двух факторов информация хранится в многомерных таблицах, что требует использования многомерных массивов и специальных вычислительных процедур (программирование в среде типа Pascal и т. п.). Но с точки зрения практического применения однофакторного и двухфакторного анализа вполне достаточно.

4. Регрессионный анализ. Эта работа имеет, пожалуй, самое большое значение в курсе математической статистики, так как нет более важной задачи, чем получить достаточно точную связь отклика Y с влияющими на него факторами $X_1 ... X_n$: $Y = \varphi(X_1, ..., X_n) + \varepsilon$. В «Пакете анализа» имеется алгоритм нахождения линейной связи $Y = a_0 + a_1 X_1 + ... + a_n X_n + \varepsilon$, использующий метод наименьших квадратов. Студент должен понимать, что если в линейной модели ошибки ε имеют нормальное распределение и для различных наблюдений линейно не зависимы и одинаково распределены $\varepsilon_i \approx N(0; \sigma^2)$, то метод наименьших квадратов дает оптимальные коэффициенты для параметров модели линейной регрессии. В то же время он должен осознавать, что если распределение ошибок не является гауссовым, то гарантия оптимальности оценок отсутствует. Кроме того, метод наименьших квадратов проявляет неустойчивость по отношению к «загрязнению выборки» (появлению больших выбросов). В лабораторной работе по регрессионному анализу может быть предложен ряд задач.

Первая задача — сравнение метода наименьших квадратов (параметрический метод) с непараметрическим методом Тейла (угловой коэффициент — медиана всех угловых коэффициентов прямых, проведенных через все пары различных точек) для простой линейной регрессии $Y = aX + b + \varepsilon$. Например, можно по данным о расстояниях до галактик и их скоростям удаления численными методами проверить выполнимость закона Хаббла: «скорость удаления галактики прямо пропорциональна расстоянию до нее». По реальным данным [5] рассчитать линию регрессии по методу наименьших квадратов и по методу Тейла [8, 9]. Оба метода дают приблизительно одинаковые модели. Добавив одно «загрязняющее» (резко выпадающее) наблюдение, построить обе модели линейной регрессии и сравнить результаты. Метод наименьших квадратов содержится в «Пакете анализа», метод Тейла требует вычислений «вручную», однако использование таблиц Excel позволяет достаточно легко провести расчеты коэффициентов линейной модели по обоим методам. Для подсчета коэффициентов непараметрической модели линейной регрессии по методу Тейла данные о выборках необходимо расположить так, как показано в табл. 4, и в ячейку D3 необходимо занести формулу =(D\$2–\$B3)/(\$D1–A\$3). Затем скопировать содержимое этой ячейки во все ячейки, находящиеся над диагональю квадратной таблицы правее столбца D (можно растянуть на всю квадратную таблицу, но в этом случае на диагонали будут стоять значения деления 0/0).

Таблица 4

	A	B	C	D	T	F
1			X_1	X_2	…	X_n
2			Y_1	Y_2	…	Y_n
3	X_1	Y_1				
4	X_2	Y_2				
	…	…				
	X_n	Y_n				

Еще одна особенность таблиц Excel заключается в том, что аргументом любой функции является прямоугольный диапазон таблицы. Поэтому, чтобы найти медиану совокупности чисел, стоящих над диагональю квадратной матрицы, необходимо предварительно, используя процедуру «Вырезать», превратить эту совокупность в столбец или строку.

Вторая задача — отбор наиболее информативных параметров. Обычно при применении модели линейной регрессии исследователь в качестве факторов использует всю доступную ему информацию. Однако среди факторов могут оказаться такие, которые на отклик никак не влияют. Построив такую модель, студент должен научиться выделять те факторы, которые на отклик не влияют. Он должен понимать, что построение модели линейной регрессии ничто иное, как ортогональное проектирование вектора отклика Y на линейное пространство, порожденное факторами $X_1,...,X_n$, и добавление к $X_1,...,X_n$ еще одного фактора X_{n+1}, ортогонального Y, проекции на новое линейное пространство не изменит.

В пакете на выходе программы линейной регрессии имеется параметр P-значение, который как раз и говорит о степени влияния данного фактора на отклик. Если P-значение для параметра a_i больше 0,05, то имеет смысл попробовать исключить данный фактор из модели.

О качестве модели говорит такой показатель, как множественный коэффициент корреляции. Это, по сути, коэффициент корреляции между реальными значениями отклика Y и прогнозируемыми значениями по модели линейной регрессии. Другим параметром, характеризующим качество модели, является коэффициент «нормированный R-кадрат», который учитывает число используемых факторов. Чем ближе этот коэффициент к 1, тем более адекватной является подобранная модель. Если исключение фактора из модели приводит к возрастанию коэффициента этого фактора из модели «нормированный R-квадрат», то этот фактор использовать в модели не имеет смысла.

В табл. 5 приведен фрагмент итогов обработки модели линейной регрессии зависимости отклика Y от факторов X_1, X_2 и X_3. Анализ столбца, содержащего P-значение, говорит о том, что, возможно, второй фактор в этой модели является лишним.

Таблица 5

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	**0,9959**
R-квадрат	0,9919
Нормированный R-квадрат	**0,9912**
Стандартная ошибка	1,1581
Наблюдения	40

	Коэффиц.	*Станд. ошибка*	*t-статистика*	*P-значение*
Y-пересечение	0,597	0,569	1,049	**0,301111846**
X_1	2,047	0,032	63,800	**1,24059E-38**
X_2	–0,498	0,668	–0,746	**0,4606597**
X_3	5,424	0,534	10,153	**4,13493E-12**

Дальнейший анализ, после исключения второго параметра из модели, показывает (табл. 6), что предположение имеет основание: хотя множественный коэффициент корреляции несколько уменьшился, коэффициент «нормированный R-квадрат» при этом вырос.

Таблица 6

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	**0,9956**
R-квадрат	0,9917
Нормированный R-квадрат	**0,9913**
Стандартная ошибка	1,1512
Наблюдения	40

	Коэффиц.	*Станд. ошибка*	*t-статистика*	*P-значение*
Y-пересечение	0,322	0,431	0,747	**1,23831E-5**
X_1	2,049	0,032	64,712	**1,12573E-39**
X_3	5,349	0,522	10,257	**2,28937E-12**

Аналогично можно решать задачу подбора степенной модели $Y = a_0 + a_1X + a_2X^2 + \ldots + a_nX^n + \varepsilon$ наименьшей степени, адекватно описывающей влияние фактора X на отклик Y.

В результате студент должен научиться при построении линейной модели линейной регрессии отбирать наиболее значимые факторы и строить линейную модель, если она действительно описывает влияние заданных факторов на отклик.

Третья задача — сведение нелинейной модели к линейной. Часто первичный визуальный анализ зависимости Y от X (X — одномерный параметр) показывает, что надеяться на линейную зависимость не приходится. В этом случае можно подбирать либо экспериментальную степенную модель $Y = a_0 + a_1X + a_2X^2 + \ldots + a_lX^l$, либо использовать некоторые функциональные зависимости типа $Y = \dfrac{a}{bX + c}$ (гиперболические), либо зависимости, возникающие как результат теоретических исследований (закон Архимеда или закон Арениуса в физике, закон Кобба — Дугласа в экономике и др.). Очень часто эти модели с помощью соответствующих преобразований можно свести к линейным моделям и после этого для поиска параметров модели воспользоваться моделью линейной регрессии.

В данной лабораторной работе можно предложить студентам среди заданных 5–6 параметрических нелинейных моделей по заданным значениям фактора X и отклика Y подобрать модель, наиболее адекватно описывающую значения отклика Y по значению фактора X. Основным показателем качества модели является значение коэффициента R-квадрат при построении соответствующей линейной модели. Ясно, что даже при выборках небольшого объема за одно семинарское занятие невозможно обработать результаты подбора модели линейной регрессии для шести моделей, в то время как на лабораторной работе для этого достаточно 30…45 мин, если студент умеет в Excel заносить в ячейку заданную формулу, копировать содержимое ячейки в заданный диапазон и использовать опцию «Регрессия» из «Пакета анализа».

5. Элементы теории выборочных обследований. В современном курсе прикладной статистики некоторая часть курса должна быть посвящена методам выборочного обследования (если в программе не предполагается отдельного курса «Методы выборочного обследования»). Выборочные обследования применяются к генеральным совокупностям ограниченного объема, и зачастую генеральная совокупность бывает существенно неоднородной. Эти особенности генеральных совокупностей требуют своих методов оценки среднего. Для этой части курса имеет смысл провести лабораторную работу, содержащую решение следующих задач:

а) сравнение оценок простого случайного отбора, пропорционального отбора и оптимального отбора (при этом исходные данные могут быть как смоделированными с помощью генератора случайных чисел, так и почерпнуты из литературы [10, 11];

б) анализ оценок по регрессии в случае наличия двух коррелированных характеристик объектов, информация об одной из которых известна полностью.

Может быть предложен еще рад задач для лабораторных работ, все определяется временем, отводимым на изучение того или иного раздела.

Заключение. Число лабораторных работ, так же как и число решаемых на них задач, естественно, зависит от объема часов, выделяемых на изучаемую дисциплину, и от контингента обучаемых студентов. Но в любом случае изменение формы обучения студентов курса математической (прикладной) статистики с использованием вычислительной техники в процессе практических занятий должно принести пользу. Естественно, для обработки статистических данных существуют специальные пакеты программ (Statistica, Statgraphyics, SPSS, SAS и др.), содержащие значительно более мощные средства статистического анализа, но их использование в процессе обучения требует выполнения лицензионных соглашений. Однако в наше время статистическая информация обычно хранится в таблицах Excel и поэтому естественно, по крайней мере, первичную статистическую обработку данных проводить в этой среде. Опыт чтения курса прикладной статистики на факультете ИБМ МГТУ им. Н.Э. Баумана выявил еще одну особенность излагаемого подхода. Студенты экономических специальностей относятся к выполнению лабораторных работ весьма ответственно, так как кроме новых знаний по прикладной статистике они попутно учатся работать с таблицами Excel, что для экономиста просто необходимо, а в основной программе изучение Excel не предусмотрено.

ЛИТЕРАТУРА

[1] Горяинов В.Б., Павлов И.В. и др. *Математическая статистика*. Москва, Изд-во МГТУ им. Н.Э. Баумана, 2001, 423 с.

[2] Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Москва, Финансы и статистика, 1983, 472 с.

[3] Орлов А.И. *Прикладная статистика*. Москва, Изд-во «Экзамен», 2006. 672 с.

[4] Орлов А. И. *Эконометрика*. 3-е изд. Москва, Изд-во «Экзамен», 2004, 576 с.

[5] Лагутин М.Б. *Наглядная математическая статистика*. Москва, Бином. Лаборатория знаний, 2007, 472 с.

[6] Гмурман Г.Е. *Теория вероятностей и математическая статистика*. Москва, Высшая школа, 2003, 479 с.

[7] Кремер Н.Ш. Теория вероятностей и математическая статистика. Москва, ЮНИТИ-ДАНА, 2004, 573 с.

[8] Холлендер М., Вульф Д. *Непараметрические методы математической статистики*. Москва, Финансы и статистика, 1983.

[9] Дьячков А.Г. *Учебные материалы по курсу математической статистики для психологического факультета МГУ*. PDF: http://stat-msu.narod.ru

[10] Кокрен У. *Методы выборочного исследования*. Москва, Статистика, 1976, 440 с.

[11] Дженсен Р. *Методы статистических обследований*. Москва, Финансы и статистика, 1985, 479 с.

Статья поступила в редакцию 28.06.2013

Ссылку на эту статью просим оформлять следующим образом:

Ветров Л.Г., Сунчалина А.Л. Лабораторные работы в курсе математической статистики. *Инженерный журнал: наука и инновации*, 2013, вып. 5. URL: http://engjournal.ru/catalog/pedagogika/hidden/737.html

Ветров Леонид Георгиевич окончил механико-математический факультет МГУ им. М.В. Ломоносова, отделение математики в 1976 г., аспирантуру МГУ им. М.В. Ломоносова, канд. физ.-мат. наук, доц. кафедры «Высшая математика». Область научных интересов: теория вероятностей, математическая статистика, теория надежности. e-mail: lvetrov@mail.ru

Сунчалина Анна Леонидовна окончила МГТУ им. Н. Э. Баумана, факультет фундаментальных наук, кафедра прикладной математики в 2004 г.; аспирантуру МГТУ им. Н.Э. Баумана, факультет фундаментальных наук в 2007 г., канд. физ.-мат. наук, доц. кафедры «Высшая математика». Область интересов: теория вероятностей, математическая статистика, теория надежности, автострахование. e-mail: sunchalina@mail.ru