

Министерство образования и науки Украины
Севастопольский национальный технический университет



**«ВЫПОЛНЕНИЕ ПРОВЕРКИ
СТАТИСТИЧЕСКИХ ГИПОТЕЗ И
ДИСПЕРСИОННОГО АНАЛИЗА
ПРИ ПОМОЩИ EXCEL»**

МЕТОДИЧЕСКИЕ УКАЗАНИЯ

к лабораторной работе

по дисциплинам «Прикладная статистика»

для студентов специальностей:

для студентов специальностей:

7.050201 – «Менеджмент организаций»,

7.050104 – «Финансы»

7.050106 – «Учет и аудит»

7.050107 – «Экономика предприятия»

всех форм обучения

Севастополь
2007



УДК 681.5.015.:330.43

«Выполнение проверки статистических гипотез и дисперсионного анализа при помощи Excel» методические указания к лабораторной работе по дисциплине «Прикладная статистика» для студентов специальностей: 7.050201 – «Менеджмент организаций», 7.050104 – «Финансы», 7.050106 – «Учет и аудит», 7.050107 – «Экономика предприятия» всех форм обучения / Сост. А.А. Загорулько, А.Д. Горобец. – Севастополь: Изд-во СевНТУ, 2007г. – 16с.

Целью методических указаний является закрепление теоретических знаний и приобретение практических навыков по темам «Проверка статистических гипотез» и «Дисперсионный анализ» при решении ситуаций с помощью Excel. Методические указания предназначены для студентов экономических специальностей всех форм обучения.

Методические указания утверждены на заседании кафедры менеджмента и экономико-математических методов, (протокол № 7 от «16» марта 2007г.)

Допущено учебно-методическим центром СевНТУ в качестве методических указаний

Рецензент: Одинцова Т.М., канд. эконом. наук, доцент кафедры «Финансы и кредит»

СОДЕРЖАНИЕ

1. Цель работы.....	4
2. Теоретическая часть.....	4
2.1. Проверка статистических гипотез.....	4
2.2. Проверка статистических гипотез в Excel.....	5
2.3. Дисперсионный анализ.....	7
2.4. Выполнение дисперсионного анализа в Excel.....	8
3. Варианты заданий.....	12
3.1. Варианты заданий по теме «Проверка статистических гипотез».....	12
3.2. Варианты заданий по теме «Однофакторный дисперсионный анализ».....	13
4. Содержание отчета.....	14
5. Контрольные вопросы.....	14
Библиографический список.....	15

1. ЦЕЛЬ РАБОТЫ

Данные методические указания предназначены для закрепления теоретических знаний и получения практических навыков студентов при проверке статистических гипотез и выполнении дисперсионного анализа в Excel.

2. ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1. Проверка статистических гипотез

Статистическая гипотеза - некоторое предположение о законе распределения случайной величины или о параметрах этого закона в рамках данной выборки.

Пример статистической гипотезы: «генеральная совокупность распределена по нормальному закону», «различие между дисперсиями двух выборок незначимо» и т.д.

При аналитических расчетах часто необходимо выдвигать и проверять гипотезы. Проверка статистической гипотезы осуществляется с помощью статистического критерия в соответствии со следующим алгоритмом:

1) формулировка гипотезы. Гипотеза формулируется в терминах различия величин. Например, есть случайная величина x и константа a . Они не равны (арифметически), но нужно установить, значимо ли статистически между ними различие. Существует два типа критериев:

- а) двухсторонний критерий вида: $x \neq a$;
- б) односторонний критерий вида: $x < a$ или $x > a$.

Необходимо отметить, что знаки $>$, $<$, $=$ здесь используются не в арифметическом, а в «статистическом» смысле. Их необходимо читать «значимо больше», «значимо меньше», «различие незначимо».

2) Установка закона распределения. Далее необходимо установить или постулировать закон распределения. Существуют также критерии, которые не зависят от вида распределения - так называемые непараметрические критерии.

3) Вычисление тестовой статистики. Тестовая статистика - некоторая функция от рассматриваемых величин, закон распределения которой точно известен и ее можно сравнить с табличным значением.

4) Сравнение с табличным значением. Затем тестовая статистика сравнивается с табличным значением. Тестовая статистика всегда зависит от доверительной вероятности, и, в некоторых случаях, от дополнительных параметров. Так, в приведенном выше примере сравнения двух дисперсий тестовая статистика сравнивается с табличным значением критерия Фишера («критическим» значением), которое зависит от доверительной вероятности и числа степеней свободы дисперсий.

5) Вывод. На основании сравнения делается вывод о том, выполняется ли гипотеза (например, значимо ли различие и т.д.).

2.2. Проверка статистических гипотез в Excel

Рассмотрим пример.

Туристическая фирма в среднем реализует 21 путевку в страны ближнего зарубежья в день. Количество проданных путевок за последнюю неделю составило: 17, 19, 25, 32, 27, 30, 28. Полагая, что уровень значимости равен 5% определить:

- выполняет ли компания план по продаже путевок в страны ближнего зарубежья;
- увеличился ли среднего ежедневного объема продаж путевок за последнюю неделю.

Для выполнения проверки статистических гипотез в Excel необходимо произвести расчеты и вычислить значение критерия Стьюдента.

Решение.

На первом этапе выдвигаются нулевая и альтернативная гипотезы:

$$H_0: \mu = 21$$

$$H_1: \mu \neq 21.$$

После чего необходимо определить табличное значение критерия Стьюдента. Для этого в меню «Вставка» выбирается команда «Функция». Устанавливается категория «Статистические» и выбирается функция «СТЮДРАСПРОБР». Пример заполнения окна для определения табличного значения критерия Стьюдента представлен на рисунке 1.

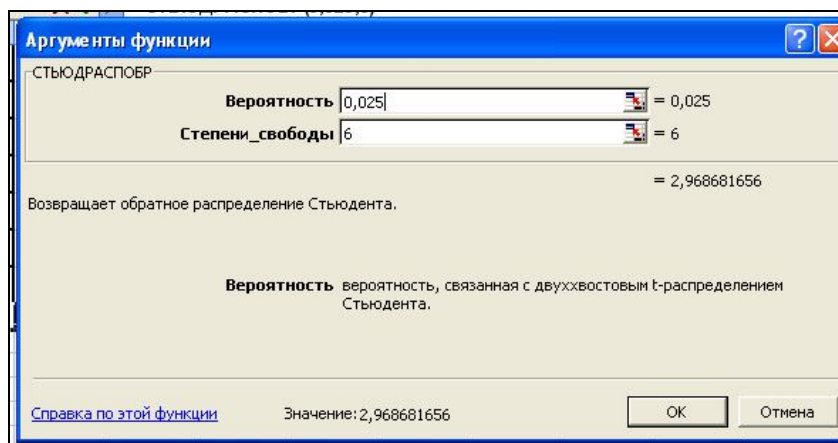


Рисунок 1 – Пример заполнения окна для определения табличного значения критерия Стьюдента

В поле «Вероятность» вводится заданный уровень значимости (α), а так как альтернативная гипотеза имеет следующий вид: $H_1: \mu \neq 21$, то необходимо применять двусторонний тест. При использовании двустороннего теста вместо значения α берется значение $\alpha/2$ (для рассматриваемого примера $\alpha/2=0,025$).

В поле «Степени свободы» вводится значение, равное $n-1$, где n - число элементов в выборке (для рассматриваемого примера $n-1=6$).

В результате мы определили, что табличное значение критерия Стьюдента для данной задачи равно 2,968.

Правило принятия (отвержения) гипотезы: если $TR < -2,968$ и $TR > 2,968$, то отвергается H_0 и принимается H_1 .

Далее необходимо выполнить следующие вычисления:

Вычисление t расчетного (TR) производится по следующей формуле:

$$TR = \frac{\bar{x} - \mu_{H_0}}{s / \sqrt{n}}, \quad (1)$$

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}. \quad (2)$$

Таблица 1 - Промежуточные вычисления

X	$X - \bar{X}$	$(X - \bar{X})^2$
17	-8,43	71,0649
19	-6,43	41,3449
25	-0,43	0,1849
32	6,57	43,1649
27	1,57	2,4649
30	4,57	20,8849
28	2,57	6,6049
		185,7143

Таким образом имеем:

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{185,7143}{6}} = 5,563$$

$$TR = \frac{\bar{x} - \mu_{H_0}}{s / \sqrt{n}} = \frac{25,43 - 21}{5,563 / \sqrt{7}} = 2,11$$

На заключительном этапе происходит принятие статистического решения.

Так как $-2,968 < TR < 2,968$, тогда с $\alpha = 0,05$ гипотеза H_0 о равенстве среднего значения количества проданных путевок за день в размере 21 штуки принимается. Следовательно, туристическая компания выполняет план по продаже путевок в страны ближнего зарубежья.

При ответе на второй вопрос задания нулевая и альтернативная гипотеза будут иметь вид:

$$H_0: \mu = 21$$

$$H_1: \mu > 21.$$

Так как альтернативная гипотеза имеет следующий вид: $H_1: \mu > 21$, то необходимо применять односторонний тест. При использовании одностороннего теста табличное значение критерия Стьюдента определяется таким же образом, как и для двустороннего теста, но в поле «Вероятность» указывается значение α (для рассматриваемого примера $\alpha=0,05$). Таким образом, табличное значение критерия Стьюдента для одностороннего теста равно 2,446.

$|TR| < 2,446$, то принимается гипотеза H_0 о равенстве среднего значения количества проданных путевок в размере 21 штуки в день при $\alpha=0,05$. Следовательно, нельзя говорить об увеличении среднего ежедневного объема продаж путевок за последнюю неделю.

2.3 Дисперсионный анализ

Дисперсионным анализом называют статистический метод анализа результатов, зависящих от действия количественных или качественных факторов. Включает в себя проверку гипотез, связанных с оценкой выборочной дисперсии. Может показаться странным, что процедура сравнения средних называется дисперсионным анализом. В действительности, это связано с тем, что при исследовании статистической значимости различия между средними двух (или нескольких) групп, сравниваются (анализируются) выборочные дисперсии. Фундаментальная концепция дисперсионного анализа предложена Фишером в 1920 году

Дисперсионный анализ может быть использован для выявления совместного влияния экономических факторов, неподдающихся количественному измерению, на изучаемый показатель. Суть метода в том, что общая вариация результирующего показателя расчленяется на части, соответствующие раздельному и совместному влиянию различных качественных факторов, и остаточную вариацию, аккумулирующую влияние всех неучтенных факторов. Статистическое изучение этих частей позволяет делать выводы о том, действительно ли оказывает влияние на результирующий показатель тот или иной фактор.

Можно выделить три основных вида гипотез:

1) $S_1^2 = S_2^2$ (значимо ли различие между двумя дисперсиями?);

2) $S_1^2 > S_2^2$ (одна дисперсия значимо больше другой?);

3) $S_1^2 = S_2^2 = \dots = S_m^2$ (значимо ли различие между несколькими дисперсиями?).

Дисперсия вычисляется из случайных величин, и поэтому сама также является случайной величиной. Дисперсии, в отличие от средних, подчиняются распределению χ^2 .

Первые две гипотезы дисперсионного анализа проверяются с помощью критерия Фишера. Причем первая гипотеза - с помощью двустороннего критерия, а вторая - с помощью одностороннего. Строго говоря, эти критерии не равны, но в общем случае разницей можно пренебречь. Проверка производится по следующей формуле:

$$F = \frac{S_1^2}{S_2^2} \quad (3)$$

С третьей гипотезой (сравнение нескольких дисперсий или проверка однородности дисперсий) дело обстоит сложнее. Попарно сравнивать дисперсии некорректно. В дисперсионном анализе для сравнения нескольких дисперсий существует два критерия: критерий Бартлетта и критерий Кохрена.

2.4 Выполнение дисперсионного анализа в Excel

Рассмотрим дисперсионный анализ на следующем примере: за месяц известны данные о выработке рабочего за время работы в первую и во вторую смены.

Таблица 2 - Исходные данные

Смена	Выработка рабочего, нормо-час
1	12,1; 11,1; 12,6; 12,9; 11,6; 13,1; 12,6; 12,4; 11,6; 17,3; 12,9; 11,6; 12,4
2	9,9; 11,4; 13,4; 10,4; 12,9; 12,6; 13,9; 13,4; 12,4; 9,9; 10,2; 11,2; 9,7

Можно ли считать, что расхождение между уровнями выработки рабочего в первую и во вторую смены незначительно, т.е. можно ли считать, что генеральные средние в двух подгруппах одинаковы и, следовательно, выработка рабочего может быть охарактеризована общей средней.

Решение.

Для того чтобы ответить на поставленные вопросы, рассчитаем среднюю выработку рабочих в каждой смене. Величина выработки в первую и вторую смены различна. Теперь возникает вопрос о том, насколько существенны эти расхождения, нужно проверить предположение о возможном влиянии сменности на выработку рабочих. Результаты расчетов сведены в таблицу 3.

Таблица 3 – Промежуточные расчеты для проведения дисперсионного анализа

Смена	Средняя выработка, нормо-часы $\bar{y}_j = \frac{\sum_{j=1}^n y_j}{n}$	Число смен в месяце n_j	Сумма квадратов отклонений вариантов от групповой средней $\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$	Квадраты отклонений групповых средних от общей средней $(\bar{y}_j - \bar{y})^2$
1	2	3	4	5
1	12.63	13	28.0877	0.2461
2	11.63	13	28.0707	0.2461
Итого	$\bar{y} = \frac{12.63 + 11.63}{2} = 12.13$	26	56.1584	

Используя данные графы 4 и графы 5, рассчитаем S_1^2 и S_2^2 .

Число степеней свободы для расчета внутригрупповой дисперсии равно 24 (26-2), а для расчета межгрупповой дисперсии число степеней свободы равно 1 (2-1).

$$S_1^2 = \frac{\sum_{j=1}^k (\bar{y}_j - \bar{y})^2 * n_j}{k - 1} = \frac{0.2461 * 13 + 0.2461 * 13}{1} = 6.4$$

$$S_2^2 = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2}{n - k} = \frac{56.1584}{24} = 2.339$$

Рассчитаем значение критерия Фишера по следующей формуле:

$$F = \frac{S_1^2}{S_2^2} \quad (4)$$

$$F = \frac{6.4}{2.339} = 2.736$$

В соответствии с числом степеней свободы для расчета внутригрупповой и межгрупповой дисперсий (24 и 1) в таблице F-распределения для $\alpha=5\%$ находим $F_{\text{табл}} = 4.26$.

При этом выдвигается две гипотезы. Нулевая гипотеза гласит о том, что различия выработки рабочего в первую и вторую смены несущественны.

Альтернативная гипотеза: существуют существенные различия в значении выработки рабочего в первую и во вторую смены.

Так как расчетное значение критерия Фишера значительно меньше табличного значения критерия Фишера, то гипотеза о несущественности различия выработки рабочего в первую и вторую смены не опровергается, т.е. сменность не оказывает влияния на уровень выработки рабочего.

Выполнение однофакторного дисперсионного анализа в Excel.

Для того, чтобы провести дисперсионный анализ в Excel, необходимо активировать команду «Анализ данных». Для этого проходится следующий путь: Сервис -> Надстройки -> Пакет анализа. После этого в меню «Сервис» появляется команда «Анализ данных» и выбирается команда «Однофакторный дисперсионный анализ».

Далее необходимо заполнить окно «Анализ данных»:

«Входной интервал» - вводится ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.

«Группирование» - установите переключатель в положение. По столбцам или По строкам в зависимости от расположения данных во входном диапазоне.

«Метки в первой строке/Метки в первом столбце» - если первая строка исходного диапазона содержит названия столбцов, установите переключатель в положение Метки в первой строке. Если названия строк находятся в первом столбце входного диапазона, установите переключатель в положение Метки в первом столбце. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

«Альфа» - введите уровень значимости, необходимый для оценки критических параметров F-статистики. Уровень альфа связан с вероятностью возникновения ошибки типа I (опровержение верной гипотезы).

«Выходной диапазон» - введите ссылку на левую верхнюю ячейку выходного диапазона. Размеры выходной области будут рассчитаны автоматически, и соответствующее сообщение появится на экране в том случае, если выходной диапазон занимает место существующих данных или его размеры превышают размеры листа.

«Новый лист» - установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

«Новая книга» - установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку A1 на первом листе в этой книге.

Пример заполнения окна «Однофакторный дисперсионный анализ» представлен на рисунке 2.

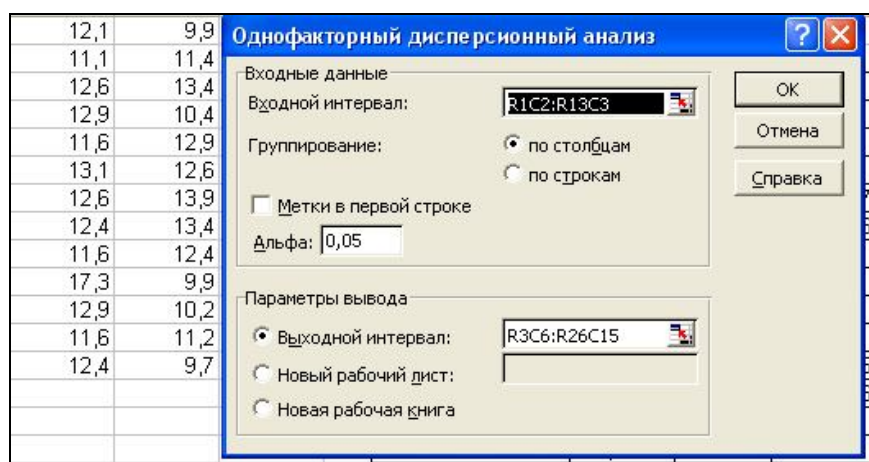


Рисунок 2 – Пример заполнения окна «Однофакторный дисперсионный анализ»

Результаты расчетов однофакторного дисперсионного анализа представлены на рисунке 3.

Однофакторный дисперсионный анализ						
ИТОГИ						
<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
Столбец 1	13	164,2	12,63077	2,34064103		
Столбец 2	13	151,3	11,63846	2,33923077		
Дисперсионный анализ						
<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
Между группами	6,400385	1	6,400385	2,73528203	0,111176312	4,259675279
Внутри групп	56,15846	24	2,339936			
Итого	62,55885	25				

Рисунок 3 – Результаты расчетов по однофакторному дисперсионному анализу

Интерпретация результатов:

«Группы» - данные по выработке в первую и вторую смены.

«Счет» - количество наблюдений в каждой из групп.

«Сумма» - сумма элементов каждой из групп.

«Среднее» - средняя выработка в каждой из групп.

«Дисперсия» - рассчитывается дисперсия по каждой из групп;

SS - сумма квадратов;

df - число степеней свободы;

MS – средний квадрат;

F – расчетное значение отношения Фишера;

P - уровень значимости для вычисленного F;

F критическое – табличное значение отношения Фишера.

Результаты расчетов аналогичны результатам, полученным при расчетах вручную.

3. ВАРИАНТЫ ЗАДАНИЙ

В ходе выполнения работы студенту необходимо:

- выполнить индивидуальное задание по теме «Проверка статистических гипотез». Задание выполняется по варианту, выбранному в разделе 3.1. Расчеты проводятся в Excel (пример приведен в разделе 2.2.). В завершении расчетов сделать выводы;

- выполнить индивидуальное задание по теме «Однофакторный дисперсионный анализ». Задание выполняется по варианту, выбранному в разделе 3.2. Расчеты проводятся в Excel (пример приведен в разделе 2.4.). В завершении расчетов сделать выводы;

3.1. Варианты заданий по теме «Проверка статистических гипотез»

Имеются данные о количестве выданных пластиковых кредитных карт за 30 дней. Начальник отдела кредитования хочет проверить предположение о том, что отделением в среднем выдается 350 карт в день. Полагая, что уровень значимости $\alpha = 0.05$, определите: какой вывод сделает отдел маркетинга на основе статистической информации. Данные представлены в таблице 4.

Таблица 4 – Исходные данные

День	Вариант						
	1	2	3	4	5	6	7
1	300	280	340	350	293	345	298
2	290	270	342	370	311	367	330
3	295	265	335	360	290	372	321
4	280	256	338	352	295	354	309
5	275	279	341	347	280	354	298
6	305	254	330	350	325	347	295
7	310	302	327	320	343	348	278
8	315	307	325	324	354	370	290
9	290	304	343	329	360	380	327
10	295	300	354	326	339	381	325
11	293	298	360	332	337	352	343
12	311	330	339	343	360	324	354
13	314	321	337	345	352	356	360
14	322	309	312	360	347	345	356

Продолжение таблицы 4

День	Вариант						
	1	2	3	4	5	6	7
15	320	298	330	356	314	367	368
16	317	295	325	368	322	387	371
17	305	278	333	371	320	369	358
18	345	290	340	358	317	345	363
19	312	295	314	363	305	352	325
20	320	293	322	345	345	376	343
21	332	311	320	354	298	342	354
22	330	290	317	356	330	346	360
23	310	295	339	327	321	368	339
24	325	280	335	325	309	367	315
25	333	330	327	343	298	365	290
26	340	321	345	354	322	387	295
27	342	309	348	360	335	354	293
28	345	310	346	339	337	387	311
29	330	312	342	337	348	387	314
30	335	333	335	339	350	383	323

3.2. Варианты заданий по теме «Однофакторный дисперсионный анализ»

Имеются данные о доле своевременно погашенных кредитов в различных отделениях банка «Импакс» за год (таблица 5). Можно ли считать, что расхождение между уровнями возвратов кредитов в различных филиалах банка несущественно.

Таблица 5 – Исходные данные

Месяц	Номер отделения						
	1	2	3	4	5	6	7
Январь	87	64	76	92	67	57	90
Февраль	77	58	78	95	66	60	85
Март	67	55	75	93	68	62	72
Апрель	83	67	80	89	71	65	75
Май	91	70	82	90	75	68	80
Июнь	90	75	65	91	69	70	82
Июль	85	73	74	90	72	71	93
Август	72	69	79	92	65	73	89
Сентябрь	78	73	82	91	74	69	90
Октябрь	91	59	77	88	79	70	91
Ноябрь	85	64	81	79	75	72	91
Декабрь	77	67	85	92	72	68	93

Выберите номера отделений в соответствии с вариантом. Варианты заданий представлены в таблице 6.

Таблица 6 – Варианты заданий

Вариант задания	Номера отделений	Вариант задания	Номера отделений
1	1,2	7	3,4
2	1,3	8	5,6
3	1,4	9	5,7
4	1,5	10	2,5
5	2,3	11	3,5
6	2,4	12	1,6

4. СОДЕРЖАНИЕ ОТЧЕТА

- 1) Цель работы;
- 2) постановка задачи;
- 3) результаты расчетов, проведенных в Excel, по проверке статистических гипотез и однофакторному дисперсионному анализу;
- 4) выводы по выполнению работы.

5. КОНТРОЛЬНЫЕ ВОПРОСЫ

- 1) Сформулируйте правило принятия гипотез при помощи критерия Стьюдента?
- 2) От чего зависит табличное значение критерия Стьюдента?
- 3) Что такое дисперсия?
- 4) Как определяется число степеней свободы для расчета внутригрупповой и межгрупповой дисперсии?
- 5) Для чего проводится дисперсионный анализ?
- 6) Как рассчитывается среднееквадратическое отклонение?
- 7) Приведите примеры использования дисперсионного анализа?

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- 1) Болч Б., Многомерные статистические методы для экономики/ Б. Болч, К. Дж. Хуань. - М.: Статистика, 1979. - 317с.
- 2) Венецкий И.Г., Венецкая В.И. Основные математико-статистические понятия и формулы в экономическом анализе: Справочник / И.Г. Венецкий, В.И. Венецкая. - М.: Статистика, 1979. - 447с.
- 3) Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента / Н. Джонсон, Ф. Лион. – М.: Мир, 1981.-520с.
- 4) Елисеева Е.И. Статистика / И.И. Елисеева, И.И. Егорова. – М.: ТК Велби, Изд-во Проспект, 2003. – 448с.
- 5) Елисеева И.И. Общая теория статистики / И.И. Елисеева, М.М. Юзбашев. – М.: Финансы и статистика, 1999. – 275с.
- 6) Ефимова М.Р., Петрова Е.В., Румянцева В.Н. Общая теория статистики: Учебник. - М.: ИФРА-М, 1996 - 416с.
- 7) Назаров М.Г. курс социально-экономической статистики / М.Г. Назаров. – М.: Финстатинформ, ЮНИТИ-ДАНА, 2000. – 427с.
- 8) Салин В.Н. Макроэкономическая статистика / В.Н. Салин, В.Г. Медведев, С.И. Кудряшова и др. – М.: Дело, 2000. – 323с.
- 9) Салин В.Н. Социально-экономическая статистика / В.Н. Салин, Е.П. Шпаков. – М.: Гарданика Юрист, 2000. – 354с.
- 10) Шеффе Г. Дисперсионный анализ / Г. Шеффе. – М.: Наука. 1980.- 512с.

Заказ № _____ от « _____ » _____ 2007. Тираж _____ экз.

Изд-во СевНТУ