МИНИСТЕРСТВО ОБРАЗОВАНИЯ УКРАИНЫ

ДОНЕЦКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

АВТОМОБИЛЬНО-ДОРОЖНЫЙ ИНСТИТУТ

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

АДИ

ДонГТУ

Горловка 1998

Министерство образования Украины

Донецкий государственный технический университет

Автомобильно-дорожный институт

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Утверждено на
заседании кафедры
"Высшая математика"
Протокол № 9
от 21.04.98 г.

Горловка 1998

УДК 512.21

Методические указания по математической статистике
/Сост.: Е.А.Королев, Л.П.Вовк, Р.Н.Корсунь - Горловка: АДИ ДонГТУ,
1998.- с. 38

Включает основные положения математической статистики и варианты индивидуальных заданий для выполнения контрольных работ.

Составители Е.А.Королев
Л.П.Вовк
Р.Н.Корсунь

Отв. за выпуск Е.А.Королев

Рецензент В.Г.Хребет

1. Общие замечания

При изучении курса высшей математики наибольшие трудности возникают у студентов при решении практических задач. Решение и анализ задач позволяют понять и запомнить основные теоремы и формулы данного курса. Умение решать задачи - лучший критерий оценки глубины изучения программного материала.

Данные методические указания относятся к теме "Математическая статистика". Для развития навыков самостоятельного решения практических задач предложены индивидуальные задания по основным разделам темы. Цель заданий - закрепить знание основных теорем и формул, освоить приемы решения задач.

Каждый студент выполняет пять задач, условия которых приведены в разделе 3 настоящей работы. Выбор номеров задач необходимо производить согласно варианту, соответствующему номеру зачетной книжки. Пример, если студент имеет номер зачетной книжки 160987183, то для выбора номеров задач следует принять во внимание лишь три последние цифры и расположить под ними три первые буквы русского алфавита, например,

1	8	3
а	б	в

Затем нужно обратиться к помещенной ниже табл. 1 и из каждого ее вертикального столбца, обозначенного внизу определенной буквой, взять одно число, стоящее в той горизонтальной строке, номер которой соответствует этой букве или разности |б-в|.

Таблица 1

Номер строки	Номера задач контрольных заданий				
1	2	12	25	35	43
2	0	13	21	33	44
3	3	11	26	36	45
4	8	18	24	34	41
5	6	17	30	37	42
6	7	19	22	32	50
7	9	14	27	38	46
8	5	20	28	31	49
9	1	15	23	39	47
0	4	16	29	40	48
	в	|б-в|	б	в	|б-в|

Например, номера задач, которые нужно решить для варианта 160987183, следующие:

3 17 28 36

2. Элементы математической статистики

2.1. Точечные оценки параметров распределения

Пусть для изучения количественного (дискретного или непрерывного) признака X из генеральной совокупности извлечена выборка $x_1, x_2, \dots, x_k$ объема n. Наблюдавшиеся значения x_i признака X называют ***вариантами***, а последовательность вариант, записанных в возрастающем порядке, - ***вариационным рядом***.

Статистическим распределением выборки называют перечень вариант x_i вариационного ряда и соответствующих им частот n_i(сумма всех частот равна объему выборки n) или относительных частот ω_i (сумма всех относительных частот равна единице).

Статистическое распределение выборки можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты интервала принимают сумму частот вариант, попавших в этот интервал).

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X<x$;

$$F^*(x) = \frac{n_x}{n}, \qquad (1)$$

где n_x—число вариант, меньших x; n —объем выборки.
Эмпирическая функция обладает следующими свойствами:

Свойство 1.

Значения эмпирической функции принадлежат отрезку [0;1].

Свойство 2.

$F^*(x)$—неубывающая функция.

Свойство 3.

Если x_1— наименьшая варианта, а x_k— наибольшая, то $F^*(x)=0$ при $x \le x_1$ и $F^*(x)=1$ при $x>x_k$.

Полигоном частот называют ломаную, отрезки которой соединяют точки $(x_1,n_1),(x_2,n_2),\dots,(x_k,n_k)$, где x_i —варианты выборки и n_i—соответствующие им частоты.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки $(x_1,\omega_1),(x_2,\omega_2),\dots,(x_k,\omega_k)$, где x_i —варианты выборки и ω_i—соответствующие им относительные частоты.

При непрерывном распределении признака весь интервал, в котором заключены все наблюденные значения признака, разбивают на ряд частичных интервалов длины h и находят n_i — сумму частот вариант, попавших в интервал.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников основаниями которых служат частичные интервалы длины h, а высоты равны отношению $\frac{n_i}{h}$ (плотность частоты). Площадь частичного i-го прямоугольника равна $h \cdot \frac{n_i}{h} = n_i$ — сумме частот вариант, попавших в i-й интервал. Площадь гисограммы частот равна сумме всех частот, т.е. объему выборки n.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников основаниями которых служат частичные интервалы длины h, а высоты равны отношению $\frac{\omega_i}{h}$ (плотность относительной частоты). Площадь частичного i-го прямоугольника равна $h \cdot \frac{\omega_i}{h} = \omega_i$ — относительной частоте вариант, попавших в i-й интервал. Площадь гисограммы относительных частот равна сумме всех относительных частот, т.е. единице.

Точечной называют статистическую оценку, которая определяется одним числом.

Несмещенной называют точечную оценку, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки.

Смещенной называют точечную оценку, математическое ожидание которой равно оцениваемому параметру.

Несмещеной оценкой генеральной средней (математического ожидания) служит выборочная средняя

$$\overline{x}_B = \frac{\sum_{i=1}^{k} n_i x_i}{n} \qquad (2)$$

где x_i —варианта выборки, n_i — частота варианты x_i,

$n = \sum_{i=1}^{k} n_i$ — объем выборки.

Если первоначальные варианты x_i — большие числа, то для упрощения расчета целесообразно вычесть из каждой варианты одно и то же число C, т.е. перейти к условным вариантам $u_i = x_i - C$

(в качестве C выгодно принять число, близкое к выборочной средней; поскольку выборочная средняя неизвестна, число C выбирают "на глаз"). Тогда

$$\overline{x}_B = C + \frac{\sum n_i u_i}{n}. \qquad (3)$$

Смещенной оценкой генеральной дисперсии служит выборочная средняя

$$D_B = \frac{\sum\limits_{i=1}^{k} n_i (x_i - \overline{x}_B)^2}{n}; \qquad (4)$$

эта оценка является смещенной, так как

$$M[D_B] = \frac{n-1}{n} D_\Gamma. \qquad (5)$$

Более удобна формула

$$D_B = \overline{x}^2 - [\overline{x}]^2 = \frac{\sum n_i x_i^2}{n} - \left[\frac{\sum n_i x_i}{n}\right]^2. \qquad (6)$$

Если первоначальные варианты x_i — большие числа, то целесообразно вычесть из всех вариант одно и то же число C, равное выборочной средней или близкое к ней, т.е. перейти к условным вариантам $u_i = x_i - C$ (дисперсия при этом не изменится). Тогда

$$D_B(X) = D_B(u) = \overline{u}^2 - [\overline{u}]^2 = \frac{\sum n_i u_i^2}{n} - \left[\frac{\sum n_i u_i}{n}\right]^2. \qquad (7)$$

Если первоначальные варианты являются десятичными дробями с k десятичными знаками после запятой, то чтобы избежать действий с дробями, умножают первоначальные варианты на постоянное число $C = 10^k$, т.е. переходят к условным вариантам $u_i = Cx_i$. При этом дисперсия увеличивается в C^2 раз. Поэтому, найдя дисперсию в условных вариантах, надо разделить ее на C^2:

$$D_B(X) = \frac{D_B(u)}{C^2}. \qquad (8)$$

Несмещенной оценкой генеральной дисперсии исправленная выборочная дисперсия

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{\sum n_i (x_i - x_B)^2}{n-1}. \qquad (9)$$

Более удобна формула

$$s_X^2 = \frac{\sum n_i x_i^2 - \frac{\left[\sum n_i x_i\right]^2}{n}}{n-1}. \qquad (10)$$

В условных вариантах она имеет вид

$$s_u^2 = \frac{\sum n_i u_i^2 - \frac{\left[\sum n_i u_i\right]^2}{n}}{n-1}. \qquad (11)$$

причем, если $u_i = x_i - C$, то $s_X^2 = s_u^2$; если $u_i = Cx_i$, то $s_X^2 = \frac{s_u^2}{C^2}$.

ЗАДАЧА №1.

Найти оценки числовых характеристик генеральной совокупности по данным выборки объема n=50.

x_i	1	2	3	4
n_i	20	15	10	5

Решение. Находим среднюю выборочную по формуле (2)

$$\bar{x}_B = \frac{1\cdot 20 + 2\cdot 15 + 3\cdot 10 + 4\cdot 5}{50} = 2.$$

Смещенная оценка дисперсии генеральной по (3)

$$D_B = 5 - 4 = 1$$

Исправленная дисперсия по (9) равна

$$S^2 = \frac{50-1}{50}\left(5 - 2^2\right) = 0.98$$

Несмещенная оценка дисперсии генеральной

$$S^2 = 0.98.$$

Несмещенная оценка математического ожидания генеральной совокупности.

$$\bar{x}_B = 2.$$

$$\bar{y}_x - 35{,}60 = 0{,}76\frac{10{,}2}{5{,}35}(x - 31{,}70),$$

или окончательно

$$\bar{y}_x = 1{,}45x - 10{,}36.$$

2.4. Статистическая проверка статистических гипотез

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений.

Нулевой (основой) называют выдвинутую гипотезу $\mathbf{H_0}$.

Конкурирующей (альтернативной) называют гипотезу $\mathbf{H_1}$, которая противоречит нулевой.

В итоге проверки гипотезы могут быть допущены ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза. Вероятность ошибки первого рода называют ***уровнем значимости*** и обозначают через α

Ошибка второго рода состоит в том, что будет принята неправильная гипотеза. Вероятность ошибки второго рода обозначают через β.

Статистическим критерием (или просто ***критерием***) называют случайную величину $\mathbf{K}$, которая служит для проверки гипотезы.

Наблюдаемым (эмпирическим) значением $\mathbf{K_{набл}}$ называют то значение критерия, которое вычислено по выборкам.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых нулевую гипотезу принимают.

Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия принадлежит критической области, то нулевую гипотезу отвергают ; если наблюдаемое значение критерия принадлежит области принятия гипотезы, то гипотезу принимают.

Критическими точками *(*границами*)* $\mathbf{k_{кр}}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Правосторонней называют критическую область, определяемую неравенством $\mathbf{K} > \mathbf{k_{кр}}$, где $\mathbf{k_{кр}}$ - положительное число.

Левосторонней называют критическую область, определяемую неравенством $\mathbf{K} > \mathbf{k_{кр}}$, где $\mathbf{k_{кр}}$ - отрицательное число.

Односторонней называют правостороннюю или левостороннюю критическую область.

Двусторонней называют критическую область , определяемую неравенствами $K<k_1$, $K>k_2$, где $k_2>k_1$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенствами (в предположении, что $k_{кр}>0$)

$$K<-k_{kp}, \quad K>k_{kp},$$

или равносильным неравенством

$$|K|>-k_{kp}.$$

Для отыскания критической области задаются уровнем значимости α и ищут критические точки, исходя из следующих соотношений:

а) для правосторонней критической области

$$P\left(K>k_{kp}\right)=\alpha \quad \left(k_{kp}>0\right);$$

б) для левосторонней критической области

$$P\left(K<k_{kp}\right)=\alpha \quad \left(k_{kp}<0\right);$$

в) для двусторонней симметричной области

$$P\left(K>k_{kp}\right)=\frac{\alpha}{2} \quad \left(k_{kp}>0\right),$$

$$P\left(K<-k_{kp}\right)=\frac{\alpha}{2}.$$

Обозначим через n и m объемы больших ($n>30$, $m>30$) независимых выборок, по которым найдены соответствующие выборочные средние $\overline{X}$ и $\overline{Y}$.

Генеральные дисперсии $D(X)$ и $D(Y)$ известны.

Для того чтобы, при заданном уровне значимости α, проверить нулевую гипотезу H_0: $M(X)=M(Y)$ *о равенстве математических ожиданий (генеральных средних) двух нормальных генеральных совокупностей с известными дисперсиями (в случае больших выборок), при конкурирующей гипотезе* H_1: $M(X) \neq M(Y)$, *надо вычислить наблюдаемое значение критерия*

$$Z_{набл}=\frac{\overline{x}-\overline{y}}{\sqrt{\frac{D(X)}{n}+\frac{D(Y)}{m}}} \tag{27}$$

и по таблице функции Лапласа найти критическую точку $z_{кр}$ *из равенства*

$$\Phi(z_{kp})=\frac{1-\alpha}{2}.$$

Если $|Z_{набл}|<z_{кр}$ *- нет оснований отвергнуть нулевую гипотезу.*

Если $|Z_{набл}|>z_{кр}$ *- нулевую гипотезу отвергают.*

При конкурирующей гипотезе H_1: $M(X) > M(Y)$ *находят критическую точку* $z_{кр}$ *по таблице функции Лапласа из равенства*

$$\Phi(z_{кр}) = \frac{1-2\alpha}{2}.$$

Если $Z_{набл} < z_{кр}$ *- нет оснований отвергнуть нулевую гипотезу.*

Если $Z_{набл} > z_{кр}$ *- нулевую гипотезу отвергают.*

При конкурирующей гипотезе H_1: $M(X) < M(Y)$ *находят вспомогательную точку* $z_{кр}$.

Если $Z_{набл} > -z_{кр}$ *- нет основания отвергнуть нулевую гипотезу.*

Если $Z_{набл} < -z_{кр}$ *- нулевую гипотезу отвергают.*

ЗАДАЧА №9 б

По двум независимым выборкам, объем которых $n=40$ и $m=50$, извлеченных из нормальных генеральных совокупностей найдены выборочные средние $x=130$ и $y=140$. Генеральные дисперсии известны $D(X)=80$, $D(Y)=100$. Требуется, при уровне значимости $0{,}01$, проверить нулевую гипотезу H_0:

$M(X)=M(Y)$, при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Решение. Найдем наблюдаемое значение критерия

$$Z_{набл} = \frac{x-y}{\sqrt{\frac{D(X)}{n}+\frac{D(Y)}{m}}} = \frac{130-140}{\sqrt{\frac{80}{40}+\frac{100}{50}}} = -5.$$

По условию конкурирующая гипотеза имеет вид $M(X) \neq M(Y)$, поэтому критическая область - двусторонняя.

Найдем правую критическую точку из равенства

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0{,}01}{2} = 0{,}495.$$

По таблице функции Лапласа (приложение 1) находим $z_{кр}=2{,}58$.

Так как $|Z_{набл}| > z_{кр}$, то, в соответствии с правилом, нулевую гипотезу отвергаем. Другими словами, ввыборочные средние различаются значимо.

ЗАДАЧА №10

По двум независимым выборкам, объемы которых $n=45$ и $m=40$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\bar{x} = 120$; $\bar{y} = 123{,}8$. Генеральные дисперсии известны: $D(X) = 90$; $D(Y) = 80$. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0: M(X) = M(Y)$ при конкурирующей гипотезе $H_1: M(X) \neq M(Y)$.

Решение. Объемы выборок большие (m>30), дисперсии генеральные известны, совокупности распределены нормально. Для проверки нулевой гипотезы используем критерий Z, имеющий нормальное распределение.

Найдем наблюдаемое значение критерия

$$Z_H = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{D(X)}{n} + \frac{D(Y)}{m}}} = \frac{120 - 123{,}8}{\sqrt{\frac{50}{25} + \frac{80}{40}}} = -1{,}9.$$

По условию задания конкурирующая гипотеза имеет вид

$H_1: M(X) \neq M(Y)$,

поэтому критическая область - двусторонняя.

Найдем правую критическую точку из равенства

$$\Phi(Z_K) = \frac{1 - \alpha}{2} = \frac{1 - 0{,}01}{2} = 0{,}495.$$

По таблице функции Лапласа (приложение 1) находим критическое значение критерия $Z_K = \cancel{2{,}85}$ 2,58

Так как $|Z_H| < Z_K$ - нет основания отвергать нулевую гипотезу. Другими словами, выборочные средние различаются незначимо.

2.5 Проверка гипотезы о нормальном законе распределения генеральной совокупности по критерию Пирсона

По данным выборки проверить гипотезу о том, что генеральная совокупность распределена по нормальному закону, т.е. выдвигаем H_0: генеральная совокупность распределена по нормальному закону.

Проверку H_0 проводим в следующем порядке:

1. Весь интервал наблюдаемых значений признака разбиваем на S интервалов равной длины точками x_i, i=0,1,2,...,S.

2. За варианты берем середины интервалов $x_i^* = \dfrac{x_i + x_{i+1}}{2}$.

По ним вычисляем среднее выборочное $\bar{X}_B$ и выборочное среднее квадратическое отклонение σ_B.

3. Вычислим вероятность попадания вариант в каждый интервал

по формуле

$$P_i = P(x_i < X < x_{i+1}) = \Phi\left(\frac{x_{i+1} - \bar{x}_B}{\sigma_B}\right) - \Phi\left(\frac{x_i - \bar{x}_B}{\sigma_B}\right).$$

4. Зная теоретические вероятности P_i, определим теоретические частоты $n_i = nP_i$, где n - объем выборки, $n = \sum n_i$.

5. Вычисляем наблюдаемое значение критерия Пирсона

$$\chi^2_H = \sum \frac{(n_i - n'_i)^2}{n'_i} \qquad (28)$$

6. По таблице (приложение 3) критических точек распределения χ^2 по уровню значимости α и числу степеней свободы $f = S - 3$ находим критическую точку правосторонней критической области χ^2_K.

7. При $\chi^2_H < \chi^2_K$ - нет оснований отвергать нулевую гипотезу о нормальном законе распределения генеральной совокупности, т.е. эмпирические и теоретические частоты различаются незначимо.

Если $\chi^2_H > \chi^2_K$ - отвергаем гипотезу о нормальном распределении. Это означает, что данные наблюдений не согласуются с теоретическими расчетами, т.е. различие имперических и теоретических частот значимо.

ЗАДАЧА №11

Используя критерий Пирсона, при уровне значимости 0,01, установить, случайно или значимо расхождение между эмпирическими частотами n_i и теоретическими частотами n_i', которые вычислены, исходя из гипотезы о нормальном распределении генеральной совокупности X:

n_i 8 16 40 72 36 18 10

n_i' 6 18 36 76 39 18 7

Решение. Найдем наблюдаемое значение критерия Пирсона

$$\chi^2_H = \sum \frac{\left(n_i - n_i'\right)^2}{n_i'}.$$

Составим расчетную таблицу 7.

Из таблицы 7 находим наблюдаемое значение критерия:

$$\chi^2_H = 3{,}068.$$

По таблице критических точек распределения χ^2 (приложение 3) по уровню значимости 0,01 и числу степеней свободы **k=s-3=7-3=4** находим критическую точку правосторонней критической области

$$\chi^2_{kp}(0{,}01;7) = 1{,}3.$$

Так как $\chi^2_H < \chi^2_{kp}$ -нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности . Другими словами, расхождения между эмпирическими и теоретическими частотами незначимо (случайно).

Таблица 7.

i	n_i	n_i'	n_i-n_i'	$(n_i-n_i')^2$	$\frac{(n_i-n_i')}{n_i'}$
1	8	6	2	4	0.667
2	16	18	-2	4	0.224
3	40	36	4	16	0.448
4	72	76	-4	16	0.208
5	36	39	-3	9	0.234
6	18	18	-	-	-
7	10	7	3	9	1.287
e	n=200				$\chi^2_H = 3{,}068.$

№30 ✓

Y \ X	10	15	20	25	30	35	n_y
20	5	1	-	-	-	-	6
30	-	6	2	-	-	-	8
40	-	-	5	40	5	-	50
50	-	-	2	8	7	-	17
60	-	-	-	4	7	8	19
n_x	5	7	9	52	19	8	n=100

31-40. Заданы выборочные средние $\bar{x}$ и $\bar{y}$, найденные по выборкам объемов n=60 и m=50, извлеченных из нормальных генеральных совокупностей X и Y с известными дисперсиями D(X) и D(Y). Требуется при уровне значимости α=0.05 проверить нулевую гипотезу H_0:M(X)=M(Y) при конкурирующей гипотезе H_1: M(X)$\neq$M(Y), т.е. требуется установить, значимо или незначимо различаются выборочные средние.

№31	$\bar{x}$=638,	$\bar{y}$=625,	D(X)=30,	D(Y)=25.
№32	$\bar{x}$=130,	$\bar{y}$=125,	D(X)=60,	D(Y)=150.
№33	$\bar{x}$=260,	$\bar{y}$=250,	D(X)=90,	D(Y)=125.
№34	$\bar{x}$=390,	$\bar{y}$=375,	D(X)=180,	D(Y)=50.
№35	$\bar{x}$=520,	$\bar{y}$=500,	D(X)=72,	D(Y)=140.
№36	$\bar{x}$=650,	$\bar{y}$=640,	D(X)=78,	D(Y)=135.
№37	$\bar{x}$=780,	$\bar{y}$=785,	D(X)=120	D(Y)=100.
№38	$\bar{x}$=910,	$\bar{y}$=885,	D(X)=180	D(Y)=50.
№39	$\bar{x}$=1000,	$\bar{y}$=990,	D(X)=84,	D(Y)=130.
№40	$\bar{x}$=13.8,	$\bar{y}$=17.1,	D(X)=96,	D(Y)=120.

41-50. Требуется при уровне значимости α=0.05 порверить по критерию согласия Пирсона гипотезу о нормальном распределении генеральной совокупности, если известны эмпирические частоты n_i и теоретические частоты n_i'.

№41✓

n_i	6	10	20	27	19	12	6
n_i'	5	14	16	25	21	12	7

√№42

n_i	5	12	19	27	20	10	7
n_i'	6	13	14	27	22	12	6

№43

n_i	8	10	18	27	17	11	9
n_i'	4	15	16	25	21	12	7

№44

n_i	5	11	22	25	21	11	5
n_i'	5	13	17	25	21	12	7

№45

n_i	7	12	16	25	21	11	7
n_i'	5	12	18	29	20	10	6

№46

n_i	8	12	16	27	19	12	6
n_i'	5	17	13	25	21	12	7

№47

n_i	6	15	16	26	19	12	6
n_i'	5	17	13	25	21	12	7

№48

n_i	6	13	21	23	19	12	6
n_i'	5	14	16	25	21	13	6

№49

n_i	5	11	20	27	19	12	6
n_i'	6	13	16	25	20	13	7

№50

n_i	6	9	21	27	20	11	6
n_i'	5	13	17	25	21	12	7