Специальный математический практикум

с.н.с. Уфимцев Михаил Валентинович
факультет ВМК МГУ

Введение

Два практикума, описанные в этой книге, разработаны в поддержку курса лекций "Математические аспекты обработки и интерпретации данных физического эксперимента", читаемого профессором Б.М. Щедриным для студентов физического факультета МГУ, специализирующихся в области физики полимеров и тонких органических плёнок. К сожалению, общая программа физфака предусматривает лишь довольно краткое знакомство с вероятностно-статистическими методами, а практических занятий, в том числе на компьютерах, по вычислению ошибок сложных функций, расчёту вероятностей, построению доверительных интервалов и проверке гипотез об экспериментальных данных вообще не предусмотрено.

Однако не подлежит сомнению, что в своей дальнейшей работе, в том числе при выполнении курсовых и дипломных работ, эти знания понадобятся. В ещё большей мере это будет злободневно при оформлении полученных результатов в публикациях, докладах на конференциях, да и для внутренней уверенности в правильности полученных выводов.

Автор стремится ознакомить читателей с основами применений теории вероятностей и статистики на примере (главным образом) физических и инженерных задач. Большей частью используемая математическая техника не сложная и требует, помимо тщательности выкладок на бумаге, вычисления на компьютере сравнительно простых величин – чаще всего выборочного среднего и выборочной дисперсии. Поэтому выполнение заданий практикума возможно на персональных компьютерах с весьма скромными возможностями: в принципе достаточно работать в MS DOS с готовыми программными модулями автора, а в качестве "интеллектуального калькулятора" использовать QBASIC.EXE, входящий в MS DOS.

Такой ригоризм облегчает студенту проникновение в *суть* рассматриваемых методов, так как нужно минимальное время для того, чтобы стало что-то получаться даже у студентов, для которых этот практикум – их первый опыт работы на персональной ЭВМ. У тех, кто уже не новичок в компьютерных средствах и имеет опыт работы в готовых "интеллектуальных" системах типа Excel, Mathematica, Macsyma, Marple или Matlab, конечно, есть более разнообразные возможности сделать требуемые задачи. Но, учитывая, что многие студенты не программировали на машинных языках (по крайней мере – помимо довольно отрывочных занятий по вычислительному практикуму) и не знакомы с особенностями упомянутых выше "интеллектуальных" систем, автор не счёл себя вправе подменить освоение вероятностно-статистических методов обучением программированию или работе, скажем, в Mathematica.

Изложение материала предусматривает знакомство с теорией методов и применение их в практических задачах. Теоретическое изложение ведётся на нескольких уровнях освоения: ознакомление с вопросом по рекомендуемой литературе (в пособии дана "выжимка" из этих книг), апелляция к материалу спецкурса Б.М. Щедрина, а также разъяснения автора, частично уточняющие и дополняющие первые два источника. Вы увидите, что первый и второй практикумы отличаются по стилю изложения из-за того, что материал первого практикума – более традиционный, знакомый физикам по общему физическому практикуму. Особенно непривычна для студентов тема проверки статистических гипотез.

Порядок выполнения заданий практикума и оформления результатов

Выполнение заданий потребует от Вас как умения проводить весьма сложные выкладки на бумаге, так и пользования компьютером. Необходимо также предварительное рассмотрение теоретического материала, работа с рекомендованной литературой. Допуск к работе даётся студенту только после беседы с преподавателем, свидетельствующей о подготовленности практиканта к задаче.

Задания практикума оформляются в журнале (тетради) лабораторных работ студента. Желательно всю возможную подготовительную работу (запись условий задач, чтение теории по вопросу, изложение метода решения задачи) выполнять накануне и соответственно готовить журнал. В журнале должны быть записаны условия задачи, значения входных параметров, выбираемых практикантом (если они есть), результаты моделирования данных (если это предусмотрено в задаче), используемый метод решения, узловые промежуточные вычисления, полученный результат и его обсуждение с окончательными выводами. Вспомогательные расчёты можно выполнять (или фиксировать их компьютерный результат) на черновике.

Средства для выполнения работ по вычислительному практикуму.

I. Выполнение промежуточных выкладок.

(а) Все компьютеры в компьютерном классе работают в оболочке Norton Commander и имеют в составе математического обеспечения интерпретатор QBASIC.EXE, в котором можно проводить расчёты. Если выполняющий задание находится в директории D: \TUTOR, то для вызова Бейсика нужно просто набрать

D: \ TUTOR > qbasic.

После загрузки интерпретатора языка нужно очистить экран, нажав клавишу <Esc>. Вычисления можно производить, набрав

print(< *требуемое выражение* >).

Если Вы не допустили ошибок (о них Вам сообщит интерпретатор), то ответ Вы можете получить, нажав клавишу <F5>. После того, как Вы зафиксировали его, Вы можете вернуться к экрану работы интерпретатора, нажав любую клавишу.

Пример:

```
x = 0.1
y = 3.14159
z = -2.
print(x*x+sin(y/8.)+exp(z))
```

<F5>

На экране выдач появится искомый результат <.5280184>.

Примечание: для повышения производительности работы можно сначала набрать несколько строк с нужными формулами, а затем нажать <F5>. При этом каждый результат печати будет отображаться в отдельной строке, например:

```
x = 2.
print(LOG(x))
y = 3.
print(SQR(y))
z = 3.14159
print(ABS(COS(z))
<F5>
```

На экране выдач появится искомый результат

```
0.6931472
1.732051
1
```

Из функций QBASIC Вам могут быть полезны:

ABS – абсолютное значение числа;

ATN – арктангенс;

COS – косинус;

EXP – экспонента;

LOG – натуральный логарифм;

MOD – остаток от деления;

RANDOM– генерация равномерно распределённых псевдослучайных чисел;

SGN – знак числа;

SIN – синус;

SQR – квадратный корень;

TAN – тангенс.

Дальнейшие детали по этим и по другим функциям Вы можете узнать, обратившись к Help (подпункт главного меню или “горячая клавиша” <Shift>+<F1>).

Прекратить работу в QBASIC и вернуться в DOS можно, нажав “горячую клавишу” <Alt>+F, X. Перед окончанием сеанса QBASIC спросит, нужно ли сохранить созданный в процессе Вашей работы файл. Вы должны ответить <No> (установите курсор на этом слове с помощью клавиши табуляции <Tab>).

(б) Вычисление простейших статистических характеристик, а также сравнение значений выборочных статистик с критическими значениями (уровнями значимости), само получение этих значений для наиболее ходовых распределений, вычисление функции вероятностей для дискретных распределений и функции плотности вероятностей для непрерывных распределений реализовано в следующих утилитах:

1) как для первого, так и для второго практикума – BINOM.EXE, ERFI.EXE, POISSON.EXE, T1.EXE;

2) для первого практикума – ERF.EXE, ERR_EST.EXE, MULNORM.EXE;

3) только для второго практикума – CHISQR.EXE, ERR1.EXE, ERR_EST1,EXE, GNORM.EXE, INT_POIS.EXE, ICHI.EXE, KOLMOGOR.EXE, SNEDECOR.EXE и TSTUD.EXE. Краткое описание этих программ см. в Приложении 1.

II. Графическая демонстрация результатов работы и печать на принтере осуществляется подпрограммой plt.exe (описание её возможностей можно посмотреть в файле plt.hlp). Вызов:

D: \ TUTOR > plt .

Просматриваемые данные должны быть записаны в файле нужного формата в директории пользователя. Можно (и нужно) распечатать график на матричном принтере и вставить в журнал лабораторных работ (<Alt> + P) .

III. Наряду с компьютерными программами Вы можете (а для решения некоторых задач это просто необходимо) воспользоваться статистическими таблицами из Приложения 2 (основной источник: Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке: Методы обработки данных. М.: Мир. 1980); но эти материалы понадобятся Вам, в основном, в следующем семестре при выполнении практикума "Построение доверительных интервалов и проверка гипотез". Необходимые разъяснения по этим материалам – в теоретическом обосновании практикума "Построение доверительных интервалов и проверка гипотез".

IV. Наконец, у Вас может быть возможность работы в более современной компьютерной среде, чем мы можем Вам предоставить. Это, например, работа в графической среде Windows 3.1 (или Windows 3.11), Windows -95 или Windows -NT с такими средствами, как электронная таблица Excel в Microsoft Office, графическая программа Microcal Origin, пакет Mathematica или же какой-либо статистический пакет. Или же Вы можете вообще работать не в MS-DOS, а в Unix. Всё это допускается при условии, что Вы подробно опишите Ваши вычисления и промежуточные результаты и, конечно, окончательный результат и Ваши выводы.

Вообще, только приветствуется разнообразие способов решения задачи, в зависимости от доступных средств. Помимо перечисленных выше возможностей, Вы можете выполнять промежуточные вычисления в заданиях 1) с помощью карандаша и бумаги; 2) на логарифмической линейке; 3) с помощью калькулятора. **Физик должен найти выход в любом положении !** Правда, при отказе от возможностей компьютера у Вас могут быть проблемы с визуализацией результатов (показ графиков, распечатка файлов и проч.).

Требования к оформлению работы

Как принято со времён классицизма в драматургии, правильно оформленная работа включает 3 части: вводную часть, основную часть, и заключительную часть. В первой части описывается постановка решаемой задачи и предполагаемый метод её решения. В основной части содержатся применяемые при решении задачи предположения и формулы, а также численные результаты применения этих моделей и формул. Итоговые результаты и выводы, а также, возможно, сравнение решений в других постановках или по другим формулам, приводятся в части III. В общем, для Вас эта схема знакома по общему физическому практикуму. Примерная схема –

Вводная часть:

1. Название работы.
2. Её назначение.
3. Условия задачи.
4. Используемые методы и основные формулы.

Основная часть:

5. Применение методов и формул, описанных в п.4.

(а) Используемая модель.

(б) Аналитические выкладки.

(в) Численные расчёты, с описанием используемых программных средств, если таковые есть.

Итоговые результаты и выводы:

6. Окончательный ответ в задаче, желательно – с указанием его точности, уровня доверия и т.п.

7. Выводы.

8. Возможные пути развития (достоинства и недостатки использованного метода, сравнение с другими методиками, пути обобщения и т.д.).

Практикум

“Элементы теории погрешностей (абсолютная, относительная, вероятностная). Элементы теории вероятностей"

Теоретическое обоснование

Из [1]. “*Физический практикум. Механика и молекулярная физика*” // Под ред. В.И.Иверoновой. М.: Наука. 1967. (Вводная часть)

Где измерения – там и погрешности. Бывают систематические и случайные погрешности. Первые из них характеризуются постоянством появления при определенных условиях. Для них увеличение числа измерений N влияния этих ошибок не уменьшит. Случайные ошибки могут возникать, а могут и не возникать, их воздействие лишено определенного знака (как уменьшают, так и увеличивают действительное значение величины), усреднение при увеличении числа измерений взаимно компенсирует случайные ошибки (в $\sqrt{N}$ раз меньше).

Если $\Delta x_1, \Delta x_2, ..., \Delta x_k$ – ошибки измерений $x_1, x_2, ..., x_k$, то *средняя абсолютная ошибка* измерений равна

$$|\Delta x| = (|\Delta x_1| + |\Delta x_2| + ... + |\Delta x_k|) / k. \tag{1}$$

Отношения $|\Delta x_1| / |x_1|$, $|\Delta x_2| / |x_2|$, ... носят название *относительных ошибок* отдельных измерений, и, наконец, для

$$x = x_1 + x_2 + ... + x_k$$

величина

$$E = \pm \frac{|\Delta x|}{|x|} \tag{2}$$

определяет *среднюю относительную ошибку* измерений (обычно в %).

В линейном приближении абсолютная ошибка имеет такую же форму, как формула для линейного члена приращения функции в дифференциальном исчислении, например:

для степенной функции $x = t^n$, $t > 0$ получим $\Delta x = n\, t^{n-1} \Delta t$

$$\Rightarrow E = \pm \frac{|\Delta x|}{|x|} = n \frac{|\Delta t|}{|t|}. \tag{3}$$

Вообще если $x = f(t)$ – сложная функция, то

$$\Delta x = \pm \Delta t * \frac{df(t)}{dt}, \text{ а } E = \pm \frac{\Delta t}{f(t)} \cdot \frac{df(t)}{dt}. \tag{4}$$

Если данные имеют нормальное (Гауссово) распределение с параметрами μ (математическое ожидание) и σ (среднеквадратическое отклонение), то *среднее значение* при повторных измерениях (оценка для μ)

$$\bar{x} = \mathrm{mean}(x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (5)$$

будет иметь среднюю ошибку (среднеквадратическое отклонение)

$$\sigma(\mathrm{mean}(x_i)) = \sigma / \sqrt{n}\,. \qquad (6)$$

Вероятная ошибка (*срединное отклонение*) δ отдельного измерения x – это ошибка такой величины, что вероятность $\Pr\{x \in [\mu-\delta, \mu+\delta]\} = 1/2$. В *допущении нормального* (Гауссова) распределения эта величина равна

$$\delta = 0.6745 * \sigma\,. \qquad (7)$$

С вероятностью $\Pr\{x \in [\mu-\sigma, \mu+\sigma]\} = 0.6826$ в допущении нормального (Гауссова) распределения значение параметра μ лежит в интервале $[\mu-\sigma, \mu+\sigma]$, в то время как другие широко употребительные значения границ доверительных интервалов дают $\Pr\{x \in [\mu-2\sigma, \mu+2\sigma]\} = 0.9545$, а $\Pr\{x \in [\mu-3\sigma, \mu+3\sigma]\} = 0.99730$.

Максимальная, или предельная, ошибка измерений $z = x \pm y$ равна

$$\Delta z = |\Delta x| + |\Delta y|, \qquad (8)$$

а ее относительное значение

$$E = \Delta z \,/\, |\,z\,|\,. \qquad (9)$$

Из книги [2] Гришин В.К., Живописцев Ф.А., Иванов В.А. *Математическая обработка и интерпретация физического эксперимента*. М.: изд. МГУ. 1988 (Стр. 37–50).

Пусть дана нелинейная функция $F(y)$ одного случайного аргумента y, причём математическое ожидание y равно a, дисперсия y равна $D(y)$, а $\sigma(y)=\sqrt{\boldsymbol{D}\{\boldsymbol{y}\}}$ – среднеквадратическое отклонение. Обозначим через $E\{F(y)\}$ математическое ожидание $F(y)$, а через $D\{F(y)\}$ – дисперсию $F(y)$. Тогда имеем для $F(y)$ в линейном приближении

$$F(y) \approx F(a) + \frac{\partial F(a)}{\partial y} * (y - a), \qquad (10)$$

откуда

$$E\{F(y)\} \approx F(a), \qquad (11)$$

$$D\{F(y)\} \approx \left(\frac{\partial F}{\partial y}\right)^2_{y=a} \cdot D(y),\ \sigma\{F(y)\} \approx \left|\frac{\partial F}{\partial y}\right|\Bigg|_{y=a} \cdot \sigma(y). \qquad (12)$$

Если же задать $F(y)$ с точностью более высокого порядка, то

$$E\{F(y)\} \approx F(a) + \frac{1}{2}\frac{\partial^2 F}{\partial y^2}\Bigg|_{y=a} \cdot D(y)\ , \qquad (13)$$

а

$$D\{F(y)\} \approx \left(\frac{\partial F}{\partial a}\right)^2 \cdot D(y) + \frac{1}{4}\left(\frac{\partial^2 F}{\partial a^2}\right)^2 \cdot (\mu_4(y) - \sigma^4(y)) + \left(\frac{\partial F}{\partial a}\right) \cdot \left(\frac{\partial^2 F}{\partial a^2}\right) \cdot \mu_3(y), \qquad (14)$$

где $\frac{\partial F}{\partial a} \equiv \left(\frac{\partial F}{\partial y}\right)\Bigg|_{y\,=\,a}$ и т.д.; $\mu_3(y)$, $\mu_4(y)$ – центральные моменты 3-го и 4-го порядков для y. Для нормального распределения (14) запишется в виде

$$D\{F(y)\} \approx \left(\frac{\partial F}{\partial a}\right)^2 \cdot D(y) + \frac{1}{2}\left(\frac{\partial^2 F}{\partial a^2}\right)^2 \cdot \sigma^4(y)\ . \qquad (14')$$

Для нелинейной функции нескольких случайных аргументов $F(y_1, y_2, ..., y_k)$ (обозначения для математических ожиданий $E\{y_i\} = a_i$, дисперсия y_i равна $D(y_i)$, а коэффициент корреляции для y_i и y_j есть ρ_{ij}) имеем

$$F(y_1, y_2, ..., y_k) \approx F(a_1, a_2, ..., a_k) + \sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right) \cdot (y_i - a_i)\ , \qquad (15)$$

так что для приближенного значения мат. ожидания и дисперсии получаем

$$F(y_1, y_2, ..., y_k) \approx F(a_1, a_2, ..., a_k)\ , \qquad (16)$$

$$D\{F(y_1, y_2, ..., y_k)\} \approx$$
$$\sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right)^2 \cdot D(y_i) + 2\sum_{i<j}\left(\frac{\partial F}{\partial a_i}\right) \cdot \left(\frac{\partial F}{\partial a_j}\right) \cdot \rho_{ij} \cdot \sigma(y_i) \cdot \sigma(y_j), \qquad (17)$$

а когда $\rho_{ij} = 0$, (17) переходит в

$$D\{F(y_1, y_2, ..., y_k)\} \approx \sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right)^2 \cdot D(y_i). \qquad (17')$$

При более точном разложении F получим

$$F(y_1, y_2, ..., y_k) \approx F(a_1, a_2, ..., a_k) +$$
$$\frac{1}{2}\sum_{i=1}^{k}\left(\frac{\partial^2 F}{\partial a_i^{\,2}}\right)\cdot D(y_i) + \sum_{i<j}\left(\frac{\partial^2 F}{\partial a_i \partial a_j}\right)\cdot \rho_{ij}\cdot \sigma(y_i)\cdot\sigma(y_j); \qquad (18)$$

для независимых с.в. $y_1, y_2, ..., y_k$ имеем

$$D\{F(y_1, y_2, ..., y_k)\} \approx \sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right)^2\cdot D(y_i) + \frac{1}{4}\sum_{i=1}^{k}\left(\frac{\partial^2 F}{\partial a_i^{\,2}}\right)^2\cdot(\mu_4(y_i) - D^2(y_i)) +$$

$$+\sum_{i<j}\left(\frac{\partial^2 F}{\partial a_i \partial a_j}\right)^2 D(y_i)\cdot D(y_j) + \sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right)\cdot\left(\frac{\partial^2 F}{\partial a_i^{\,2}}\right)\cdot\mu_3(y_i), \qquad (19)$$

где

$$\frac{\partial F}{\partial a_i} = \left(\frac{\partial F}{\partial y_i}\right)\bigg|_{\{y_i\}=\{a_i\}}, \quad \frac{\partial^2 F}{\partial a_i \partial a_j} = \left(\frac{\partial^2 F}{\partial y_i \partial y_j}\right)\bigg|_{\{y_i\}=\{a_i\}}$$

, $\{y_i\} = y_1, ..., y_k$, $\{a_i\} = a_1, ..., a_k$.

При нормальном распределении (19) упрощается:

$$D\{F(y_1, y_2, ..., y_k)\} \approx \sum_{i=1}^{k}\left(\frac{\partial F}{\partial a_i}\right)^2\cdot D(y_i) + \frac{1}{2}\sum_{i=1}^{k}\left(\frac{\partial^2 F}{\partial a_i^{\,2}}\right)^2\cdot D^2(y_i) +$$
$$+\sum_{i<j}\left(\frac{\partial^2 F}{\partial a_i \partial a_j}\right)^2 D(y_i)\cdot D(y_j). \qquad (19')$$

Из книги [3] Крамер Г. *Математические методы статистики*. М.: Мир. 1975. Гл.27.

Пусть оценка дисперсии $D(x)$ по выборке $x_1, ..., x_n$ есть

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 . \qquad (20)$$

– выборочный центральный момент второго порядка. Тогда $E\{m_2\} =$

$$\frac{n-1}{n}\mu_2 = \frac{n-1}{n}D(x), \;\; E(m_2^{\,2}) = \mu_2^{\,2} + \frac{\mu_4 - 3\mu_2^{\,2}}{n} - \frac{2\mu_4 - 5\mu_2^{\,2}}{n^2} + \frac{\mu_4 - 3\mu_2^{\,2}}{n^3}, \qquad (21)$$

где μ_ν – центральный момент порядка ν, следовательно,

$$D(m_2) = \frac{\mu_4 - \mu_2^{\,2}}{n} - \frac{2(\mu_4 - 2\mu_2^{\,2})}{n^2} + \frac{\mu_4 - 3\mu_2^{\,2}}{n^3} . \qquad (22)$$

Ковариация $\bar{x}$ и m_2 равна

$$\mathrm{cov}(\bar{x}, m_2) = \mu_{11}(\bar{x}, m_2) = \frac{n-1}{n}\mu_3 \quad . \tag{23}$$

Для нормальной совокупности

$$E\{m_2\} = \frac{n-1}{n}\sigma^2, \; D(m_2) = \frac{2(n-1)}{n^2}\sigma^4 \quad . \tag{24}$$

Для моментов высших порядков

$$E\{m_3\} = \frac{(n-1)(n-2)}{n^2}\mu_3, \; E\{m_\nu) = \mu_\nu + O(n^{-1}) \quad , \tag{25}$$

а

$$D\{m_\nu\} = \frac{\mu_{2\nu} - 2\nu\cdot\mu_{\nu-1}\cdot\mu_{\nu+1} + \nu^2\mu_{\nu-1}^2\cdot\mu_2}{n} + O(n^{-2}) \quad , \tag{26}$$

$$\mu_{11}(m_\nu, m_\rho) = \frac{\mu_{\nu+\rho} - \nu\cdot\mu_{\nu-1}\cdot\mu_{\rho+1} - \rho\cdot\mu_{\rho-1}\cdot\mu_{\nu+1} - \mu_\rho\cdot\mu_\nu + \rho\cdot\nu\mu_{\rho-1}\cdot\mu_{\nu-1}}{n} + O(n^{-2})$$

.

Несмещённая оценка дисперсии есть

$$s^2 = \frac{n}{n-1}m_2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad . \tag{20'}$$

Для неё

$$E\{\, s^2 \,\} = D(x)\,, \tag{21'}$$

для нормальной совокупности

$$D\{\, s^2 \,\} = \frac{2}{n-1}\sigma^4 \tag{24'}$$

Теорема о функциях от моментов. Пусть задана функция от выборочных центральных моментов $H(m_\nu, m_\rho)$ такая, что:

1) В некоторой окрестности точки $m_\nu = \mu_\nu$, , $m_\rho = \mu_\rho$ H непрерывна и имеет непрерывные первые и вторые производные по аргументам m_ν и m_ρ.

2) Для всевозможных значений величин x_i выполняется неравенство $|H| < C{\cdot}n^p$, где C и p – неотрицательные постоянные.

Тогда, если H_0, H_1 и H_2 – значения $H(m_\nu, m_\rho)$ и ее первых частных производных в точке $m_\nu = \mu_\nu$, $m_\rho = \mu_\rho$, то среднее значение (математическое ожидание) и дисперсия с.в. $H(m_\nu, m_\rho)$ будут определяться формулами:

$$E(H) = H_0 + O(n^{-1}), \tag{27}$$

$$D(H) = \mu_2(m_\nu)\cdot H_1^2 + 2\,\mu_{11}(m_\nu, m_\rho)\cdot H_1\cdot H_2 + \mu_2(m_\rho)\cdot H_2^2 + O(n^{-3/2})\,. \tag{28}$$

3) Если объём выборки $n \to \infty$, то $H(m_\nu, m_\rho) \sim N(E(H), D(H))$.

Некоторые используемые вероятностные распределения

Биномиальное $Bi(n, \varepsilon)$, $n >= 1$, $0 <= \varepsilon <= 1$; функция вероятностей равна

$$p_k = P\{\xi = k\} = C_n^k \varepsilon^k \cdot (1-\varepsilon)^{n-k},\ k = 0, \ldots, n, \tag{29}$$

причём $E\{\xi\} = np$, $D\{\xi\} = np\cdot(1 - p)$.

Пуассона $Po(\theta)$, $0 <= \theta$, с функцией вероятностей

$$p_k = P\{\xi = k\} = e^{-\theta}\frac{\theta^k}{k!},\ k = 0,1, \ldots \tag{30}$$

причём $E\{\xi\} = D\{\xi\} = \theta$.

Дискретное равномерное с n значениями, с функцией вероятностей

$$p_k = P\{\xi = k\} = 1/n,\quad k = 1,2, \ldots, n. \tag{31}$$

Непрерывное равномерное, $R(\mu, \omega)$, с функцией плотности вероятностей (ф.п.в.)

$$f(x) = \begin{cases} \dfrac{1}{\omega}, & x \in [\mu - \dfrac{\omega}{2}, \mu + \dfrac{\omega}{2}]; \\ 0, & x \notin [\mu - \dfrac{\omega}{2}, \mu + \dfrac{\omega}{2}] \end{cases}. \tag{32}$$

причём $E\{\xi\} = \mu$, $D\{\xi\} = \omega^2/12$.

Одномерное нормальное $N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma > 0$, с ф.п.в.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\cdot exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\ -\infty < x < \infty. \tag{33}$$

Стандартное нормальное $N(0, 1)$ с ф.п.в.

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\cdot exp\left(-\frac{x^2}{2}\right),\ -\infty < x < \infty; \tag{33'}$$

интеграл от неё

$$\Phi(x) = \int_{-\infty}^{x}\varphi(t)dt = P\{\xi \le x\} \tag{34}$$

называется *интегралом (функцией) ошибок*.

Распределение χ_k^2 (Хи-квадрат с k степенями свободы) для с.в.

$$\xi = \zeta_1^{\,2} + \zeta_2^{\,2} + \ldots + \zeta_k^{\,2}, \tag{35}$$

где статистически независимые с.в. $\zeta_1, \ldots, \zeta_k$ имеют стандартное нормальное $N(0, 1)$ распределение, имеет ф.п.в.

$$f_k(x) = \begin{cases} \dfrac{\left(\dfrac{x}{2}\right)^{k/2-1} exp\left\{-\dfrac{x}{2}\right\}}{2 \cdot \Gamma(k/2)}, & x \geq 0 \;; \\ 0, & x < 0 \end{cases} \tag{36}$$

функция распределения также будет обозначаться как $\chi_k^{\,2}$.

Распределение **Стьюдента** с k степенями свободы St_k определено для с.в.

$$\tau = \frac{\xi}{\sqrt{\dfrac{\zeta_k}{k}}}, \tag{37}$$

с.в. $\xi \sim N(0, 1)$, $\zeta_k \sim \chi_k^{\,2}$, причём с.в. ξ и ζ_k статистически независимы; во втором нашем практикуме функция распределения Стьюдента с k степенями свободы обозначается как t_k, а ф.п.в. для τ есть

$$f_k(t) = \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{k}{2}\right) \cdot \sqrt{\pi k}} \cdot \left(1 + \frac{t^2}{k}\right)^{-\frac{(k+1)}{2}}, \; -\infty < t < +\infty. \tag{38}$$

Наконец, распределение **Снедекора** (-Фишера), или F–распределение, определено для отношения независимых с.в.

$$F = \frac{\zeta_n / n}{\zeta_m / m} = \frac{m \cdot \zeta_n}{n \cdot \zeta_m}, \tag{39}$$

где $\zeta_n \sim \chi_n^{\,2}$, $\zeta_m \sim \chi_m^{\,2}$; его функция распределения обозначается $F_{n,m}$. Для случайной величины F ф.п.в. равна

$$f(x) = \frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right) \cdot \Gamma\left(\dfrac{n}{2}\right)} \cdot \left(\frac{n}{m}\right)^{\frac{n}{2}} \cdot x^{\frac{n}{2}-1} \left[1 + \frac{n \cdot x}{m}\right]^{-\frac{n+m}{2}}, \; x \geq 0. \tag{40}$$

Задания по практикуму

Работа № 1. Различные виды погрешностей (абсолютная, относительная, вероятностная).

Цель работы – знакомство с различными видами ошибок. Прежде всего, это задачи, выполняемые с бумагой и карандашом (или ручкой), которые будут распределены по выбору преподавателя:

1. (а) – (и)
2. (й) – (с)
3. 27 – 29[1]

Они такого же типа, как рассматривавшиеся на 1-ом курсе в общем физическом практикуме, и послужат вам для напоминания, “для разгона”, а также для ознакомления с предоставленными вам вычислительными средствами.

На этом занятии вы также смоделируете на компьютере совокупности данных с помощью готовых программ:

Программа Т1.

Назначение: Запись выборки данных в файл и чтение записанных данных из файла. Вычисление простейших выборочных характеристик: AVE - среднего, SUM - суммы отклонений от среднего, VAR - выборочной дисперсии, SIGMA - среднего квадратического отклонения. В случае размера выборки N = 1 две последние не вычисляются.

Описание:

При запуске программа запрашивает N – объём выборки данных. Затем она пытается открыть файл данных SAMPLE.DAT. Если он существует, то данные считываются из этого файла, и длина файла (число строк MM) сравнивается с введённым значением N. Если MM > N, то впоследствии печатается предупреждение IERR = 1 (IERR - диагностический параметр), а обработке подвергаются только первые N строк файла. Если MM = N, то обрабатывается весь файл SAMPLE.DAT. Наконец, если N > MM или же если файла SAMPLE.DAT не существует, то открывается файл SAMPLE.DAT со статусом ‘NEW’ (заново), и производится ввод данных в файл с клавиатуры <или же, владея некоторыми начальными навыками программирования на Фортране, Вы можете заменить строки

```
read (*, *) tmp
D(I) = TMP
```

на обращение к генератору случайных чисел (ГСЧ):

```
D(I) = PTGRR(ISEED)
```

, образовать объектный модуль t1.obj (командами компиляции

```
D:\TUTOR> for1 t1.for ;
D:\TUTOR> for2   )
```

, а затем утилиту t1.exe командой редактору связей

```
link t1.obj + as183.obj;
```
> .

Затем происходит обработка введённых (или считанных из файла SAMPLE.DAT) данных, причём в первом случае данные записываются в файл

[1] Текстовые задачи имеют номера 1–168, согласно их месту в общем списке задач. Последний пополнялся постепенно, и нередко рядом расположены задачи, относящиеся к разным темам.

SAMPLE.DAT. На экране печатается величина IERR (см. выше) и значения выборочных характеристик. Затем работа программы прекращается.

Программа ERR_EST
Назначение: Моделирование данных со случайными и систематическими ошибками, а также коррелированных данных. Оценка характеристик ошибок.
Описание:

Рассматривается выборка (совокупность независимых одинаково распределённых величин) объема (размера) N. Эту величину в диапазоне 64 <= N <= 512 назначает студент. Моделируются измерения вида

$$x_i = m + \varepsilon_i , \quad i = 0,1, \dots, N,$$

причём истинное значение m рассматриваемой величины равно 0, а ошибки измерений ε_i порождаются генератором случайных чисел с равномерным на интервале [0, 1] распределением.

Для запуска ГСЧ пользователь должен задать затравочное число ISEED – целое, нечётное, в интервале [1, 30001]. Начальное число $D = x_0$ печатается на экране, чтобы выполняющий задачу мог зафиксировать его в журнале лабораторной работы и сравнить его с результатами обработки выборки – выборочным средним AVE и выборочной дисперсией VAR. Полученные N чисел записываются в файл SAMPLE.DAT, и должны быть распечатаны студентом по окончании выполнения задачи.

Затем вычисляются выборочные характеристики: среднее значение, сумма квадратов отклонений от среднего, среднеквадратическое отклонение, выборочная дисперсия. Выполняющий задание должен сравнить найденные значения с теоретическими оценками характеристик для равномерного [1/2, 1] распределения.

На следующем этапе введённое Вами значение объёма выборки запоминается: N1 := N, а значение N принудительно полагается N := 64. Иными словами, дальше будут рассматриваться первые 64 значения $x_1 , \dots, x_{64}$, образующие вектор $\vec{x} = (x_1 , \dots, x_{64})^T$. Вычисляется матрица Гильберта 64-го порядка с элементами

$$H_{ij} = \left\{ \frac{1}{i + j - 1} \right\} , i, j = 1, \dots, N = 64.$$

Из книг по вычислительной линейной алгебре известно, что матрица H имеет почти линейно зависимые строки и столбцы. Производится нормировка H на сумму элементов столбца

$$H_{ij}^* = H_{ij} / t_i ,$$

где

$$t_i = \sum_{j=1}^{N} H_{ij} .$$

Рассматриваемый вектор $\vec{x}$ умножается на H^*: $\vec{y} = H^* \cdot \vec{x}$.

Нетрудно видеть, что математические ожидания $\vec{x}$ и $\vec{y}$ равны (докажите это !), но так как компоненты $\vec{y}$ сильно коррелированны, то в выдаче значений вектора $\vec{y}$ на экран (перепишите их в журнал ! или же при повторном вызове программы ERR_EST.EXE с <u>теми же значениями</u> N и ISEED организуйте выдачу данных не на экран, а в файл (в MS DOS это именуется переадресацией)

D:\TUTOR> ERR_EST.EXE >TMP.DAT

а затем распечатайте файл TMP.DAT на принтере) будут стоять очень близкие числа (их знак и величина зависят от выбранного значения ISEED). Для $\vec{y}$, как прежде для $\vec{x}$, вычисляются выборочные характеристики. Перепишите их в журнал, сравните с полученными для $\vec{x}$, и поразмыслите о причинах различий.

Наконец, рассматривается эффект систематических ошибок. Значение N восстанавливается: N := N1, и с помощью ГСЧ моделируется вектор с компонентами

$$x_i = m + \varepsilon_i - 0.1, \quad i = 0, 1, \ldots, N.$$

Он записывается в файл SAMPLE1.DAT, и для него вычисляются выборочные характеристики. Сравните их с результатами для вектора $\vec{x}$. Объясните сходства и различия.

Задания № 2, 3. <u>Применение метода линеаризации для оценивания погрешностей</u>

Требуется аккуратно дифференцировать, возводить в квадрат и извлекать корни, получая формульные выражения. В задачах с числовыми данными можно пользоваться QBASIC как калькулятором. В журнале работ должны быть подробные выкладки.

1. Лёгкие задачи – 67, 68.
2. Работа главным образом с карандашом и бумагой – задачи 101 - 106, 108 - 114.
3. Задачи повышенной сложности: 54, 55, 115 - 116.

Задание № 4. <u>Избранные дискретные распределения</u>.

1. Для выполнения некоторых задач этого и следующего задания потребуется программа графики PLT.EXE. Освойте её возможности.
2. Задачи 5, 7 – с компьютером и с PLT.
3. Задачи 4, 8 - 11, 13 - 16 – с компьютером и с бумагой.

Задание № 5. <u>Избранные непрерывные распределения</u>.

1. Задача 118 – с PLT.
2. Задачи 84, 85, 87, 88, 20, 21, 23 – характерные значения вероятностей для нормального распределения.
3. Задачи 28 - 29 имеют примерно ту же сложность.
4. Несколько сложнее задачи 25 и 86.
5. Задача 22 - творческого характера.
6. Задачи 119 - 122 – получение формул, работа с бумагой и карандашом.

Задание № 6. <u>Подведение итогов, завершение недоконченного, выставление оценок</u>.

Задачи по практикуму

“Элементы теории погрешностей (абсолютная, относительная, вероятностная. Элементы теории вероятностей"

(А) *Различные виды погрешностей и их применение*

Задана абсолютная ошибка Δx. Определить абсолютную и относительную ошибку для

(а) $f(x) = A \cdot \exp\{\alpha \cdot x\}$.

(б) $f(x) = A \cdot \ln\{\alpha \cdot x\}$.

(в) $f(x) = \sqrt[n]{x}$, $x > 0$.

(г) $f(x) = A \cdot x^n$, $x > 0$.

(д) $f(x) = A \cdot tg(x)$.

(е) $f(x) = A \cdot ctg(x)$.

(ё) $f(x) = A \cdot \arcsin(x)$.

(ж) $f(x) = A \cdot \arccos(x)$.

(з) $f(x) = A \cdot \arctan(x)$.

(и) $f(x) = A \cdot arcctg(x)$.

Для величин x и y среднеквадратичные отклонения известны и равны, соответственно, σx и σy. Величины x и y независимы. Для f(x, y) найти среднеквадратическую ошибку (абсолютную и относительную). Что будет в случае зависимых x и y ? Укажите границы применимости формул.

(й) $f(x, y) = A \cdot x + B \cdot y$.

(к) $f(x, y) = A \cdot x - B \cdot y$.

(л) $f(x, y) = x \cdot y$.

(м) $f(x, y) = x / y$.

Найдите максимальную абсолютную и относительную ошибки для f(x, y), если абсолютные ошибки для x и y известны и равны, соответственно, Δx и Δy. Укажите границы применимости формул.

(н) $f(x, y) = A \cdot x + B \cdot y$.

(о) $f(x, y) = A \cdot x - B \cdot y$.

(п) $f(x, y) = x \cdot y$.

(р) $f(x, y, z) = x \cdot y \cdot z$.

(с) $f(x, y) = x / y$.

27. В некотором химическом процессе стандартная жидкость из трех бутылей сливается в большой бак. Если среднее квадратическое отклонение объема жидкости в каждой бутыли равно 0.007 литра, то каково среднее квадратическое отклонение объема жидкости в большом баке ?

28. Две части A и B соединяются в одну трубу. Средние квадратические отклонения длин A и B равны 0.22 и 0.45 мм соответственно. Длина, добавляемая при соединении, имеет среднее квадратическое отклонение σ мм. Получите формулу для среднего квадратического отклонения длины C, предполагая, что A и B взаимно независимы.

29. При сборке два изделия C (см. упражнение 28) должны отличаться по длине менее чем на 1.25 мм. Предполагая нормальное распределение, найдите выражение для вероятности того, что два случайно взятых изделия C окажутся совместимыми. Как мала должна быть величина σ, чтобы эта вероятность составляла не менее 90 % ?

(Б) *Применение метода линеаризации для оценивания погрешностей*

54. Данные образуют выборку y_1 , …, y_n из нормального $N(\mu, \sigma^2)$ распределения. Показать, что для выборочного стандартного отклонения

$$s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

, где $\bar{y}$ - выборочное среднее, в линейном приближении имеет место соотношение

$$\sigma\{ s_y \} = \frac{\sigma}{\sqrt{2\cdot(n-1)}} ,$$

где $\sigma\{z\}$ - стандартное отклонение для с.в. z.

55. Данные образуют выборку y_1 , …, y_n из нормального $N(\mu, \sigma^2)$ распределения. Показать, что для выборочной дисперсии

$$s_y{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 ,$$

где $\bar{y}$ - выборочное среднее, для оценки точности $s_y{}^2$ в линейном приближении имеют место формулы (здесь D{z} - дисперсия, а $\sigma\{z\}$ - стандартное отклонение для с.в. z)

$$D\{ s_y{}^2\} = \frac{2\sigma^4}{n-1} , \sigma\{ s_y{}^2\} = \sqrt{\frac{2}{n-1}}\cdot\sigma^2 .$$

Сравнить ее с точным вторым моментом для выборочной дисперсии в общем случае и для нормальной выборки в ([3], гл. 27).

66. Для определения точности регистрирующего прибора было произведено 5 независимых измерений. Получены результаты: y(1) = 2781, y(2) = 2836, y(3) = 2807, y(4) = 2763, y(5) = 2858. Определить приближенное значение дисперсии и среднего квадратического отклонения, характеризующих точность прибора, если значение измеряемой величины а) известно и равно Ave(y) = 2800; б) неизвестно. Оценить точность.

67. Найти среднюю квадратичную ошибку функции F = ln(n1/n2). Указать границы применимости формулы.

68. Найти формулу для оценки дисперсии D логарифма интенсивности излучения y ln(N / t).

101. Математическое ожидание числа бракованных аппаратов при проверке их на безотказность действия определяется формулой

$$T = N\cdot [1 - (1 - p / (\Omega\cdot N))^m] ,$$

где p - вероятность того, что испытание одного из аппаратов будет признано зачётным; Ω - среднее число зачётных испытаний до получения отказа в действии аппарата; N - число аппаратов, участвующих в проверке; m - число испытаний (зачётных и незачётных), приходящихся на один аппарат.

Пользуясь методом линеаризации, определить зависимость математического ожидания и дисперсии с.в. T от m, если N, p и Ω - независимые с.в., математические ожидания и дисперсии которых соответственно равны: $E[N] = 5$, $E[p] = 0.8$, $E[\Omega] = 4$, $D[N] = 1$, $D[p] = 0.1$, $D[\Omega] = 0.2$.

102. Максимальная высота полета спутника определяется формулой

$$Y = y_0 + (R + y_0)\cdot [\, (1 + l) / (1 - \lambda) / 2 - 1\,]\,,$$

где $\lambda = v^2 / (2\, g\, R)\, [1 + y_0\, / R\,]$, $l = \sqrt{1 - 4\lambda(1-\lambda)\cos^2\theta}$, y_0 - высота активного участка траектории, g - ускорение силы тяжести на поверхности Земли, R - радиус Земли.

Функция Y в области практически возможных значений случайных аргументов линеаризуется. Начальная скорость v и угол бросания θ - нормальные с.в. с функцией плотности вероятностей (ф.п.в.)

$$f(v,\theta) =$$

$$= \frac{1}{2\pi\sigma_v\sigma_\theta\sqrt{1-r^2}} exp\left\{-\frac{1}{2(1-r^2)}\left[\left(\frac{v-\bar{v}}{\sigma_v}\right)^2 + \left(\frac{\theta-\bar{\theta}}{\sigma_\theta}\right)^2 - 2r\cdot\left(\frac{v-\bar{v}}{\sigma_v}\right)\cdot\left(\frac{\theta-\bar{\theta}}{\sigma_\theta}\right)\right]\right\}$$

,где через $\bar{v}, \bar{\theta}$ обозначены математические ожидания с.в. v и θ, соответственно, σ_v, σ_θ - их стандартные отклонения, r – коэффициент корреляции. Найти приближенное значение дисперсии для максимальной высоты полета спутника.

103. Найти срединное отклонение с.в., имеющей ф.п.в.

$$f(x) = 0.5\cdot \exp\{-\,|\,x\,|\,\}.$$

104. Ф.п.в. случайной величины x имеет вид

$$f(x) = \begin{cases} \frac{1}{2l}, |x-a| \le l \\ 0, |x-a| > l. \end{cases}$$

Определить a) $E\,\{x\,\}$; b) $D\,\{x\,\}$; c) найти связь между средним квадратическим и срединным отклонениями с.в. x .

105. Определить срединную ошибку прибора, если систематических ошибок он не имеет, а случайные распределены по нормальному закону и с вероятностью 0.8 не выходят за пределы $\pm$ 20 м.

106. Количество тепла Q в калориях, выделяемое в проводнике с сопротивлением R при прохождении тока I в течение времени T, определяется формулой

$$Q = 0.24\, I^2 RT\,.$$

Ошибки измерения величин I, R, T являются независимыми нормальными с.в. с математическими ожиданиями $E\{I\} = 10a$, $E\{R\} = 30$ ом, $E\{T\} = 10$ мин и срединными отклонениями $E_I = 0.1$ *а*, $E_R = 0.2$ ом, $E_T = 0.5$ сек. Найти приближенное значение срединного отклонения с.в. Q .

107. Частота основного тона струны определяется формулой

$$\mathbf{\Omega} = 1/2\sqrt{\frac{P}{ML}} \quad ,$$

где P - сила натяжения, M - масса струны, L - длина струны.

Известны м.о. μ_P, μ_M, μ_L и средние квадратические отклонения σ_P, σ_M и σ_L . Определить рассеивание частоты основного тона струны из-за разброса силы натяжения, массы и длины струны, если соответствующие коэффициенты корреляции равны r_{PL}, r_{PM}, r_{ML} .

108. Сопротивление участка электрической цепи определяется формулой

$$Z = \sqrt{R^2 + \left(\mathbf{\Omega} L - \frac{1}{\mathbf{\Omega} C}\right)^2} \quad ,$$

где R - омическое сопротивление, L - индуктивность проводника тока, C - его емкость, Ω - частота тока.

Определить срединную ошибку в величине сопротивления из-за ошибок при независимых измерениях R , L , C и Ω, если заданы μ_R, μ_L, μ_C, μ_Ω и срединные отклонения E_R , E_L , E_C, E_Ω.

109. При параллельном соединении элементов сила тока в цепи определяется формулой

$$I = E / (R + W/n) ,$$

где E - ЭДС элемента, W - его внутреннее сопротивление, n - число элементов, R - сопротивление внешней части цепи.

Пользуясь методом линеаризации, определить м.о. и дисперсию силы тока, если с.в. E , R и W независимы, а μ_E, μ_R, μ_W и σ_E, σ_R, σ_W заданы.

110. Найти приближенное значение средних квадратических отклонений прямоугольных координат случайной точки

$$X = H \, ctg \, \varepsilon \, \cos \beta ,$$
$$Y = H \, ctg \, \varepsilon \, \sin \beta ,$$
$$Z = H ,$$

если с.в. H , ε и β независимы, а м.о. и средние квадратические отклонения их соответственно равны: $\mu_H = 6200$ м, $\mu_\varepsilon = 45°$, $\mu_\beta = 30°$, $\sigma_H = 25$ м, $\sigma_\beta = \sigma_\varepsilon = 0.001$ рад.

111. Переход от сферических координат к декартовым производится по формулам

$$X = R \sin \theta \, \cos \varphi ,$$
$$Y = R \sin \theta \, \sin \varphi ,$$

$$Z = R \cos \theta \ .$$

Ошибки в определении R , θ и φ независимы со средними квадратическими отклонениями σ_R = 10 м, $\sigma_\theta = \sigma_\varphi =$ 0.001 рад. Определить приближённое значение средних квадратических ошибок прямоугольных координат, если $\mu_\theta = \mu_\varphi = 45°$, $\mu_R = 10000$ м.

112. Пусть $Z = sin\ X\ Y$, где X и Y - независимые с.в. Найти приближённое значение σ_Z, если $\mu_X = \mu_Y = 0$, $\sigma_X = \sigma_Y = 0.001$.

113. Используя метод линеаризации, определить приближённое значение дисперсии с.в. $Z = \sqrt{kX^2 + Y^2}$, где $X = \sin V$, $Y = \cos V$, с.в. V распределена равномерно на отрезке (0., $\pi/2$) , а k - известная постоянная.

114. Для измерения объема конуса измерены: а) диаметр основания и высота; б) диаметр основания и длина образующей . В каком из этих двух случаев дисперсия ошибки определения объема конуса меньше, если м.о. высоты конуса $\mu_h = 8$ дм, диаметра основания $\mu_d = 12$ дм, длины образующей $\mu_l = 10$ дм, а $\sigma_h = \sigma_d = \sigma_l = 0.1$ дм ?

115. Случайная величина X подчиняется нормальному закону с ф.п.в.

$$f(x) = \frac{1}{10\sqrt{2\pi}} exp\left\{-\frac{(x+5)^2}{200}\right\} \quad .$$

Определить приближенное значение м.о. и дисперсии с.в. $Y = 1 / X$, учитывая первые два и три члена разложения в ряд Тейлора.

116. Учитывая три первых члена разложения функции $Y = \varphi(X)$ в ряд Тейлора, определить м.о. и дисперсию с.в. Y , если X нормальна.

117. Высота горной вершины H определяется по наклонной дальности D и углу места ε:

$$H = D \sin \varepsilon .$$

Найти приближённое значение срединной ошибки определения высоты, если $E_D = 80$ м, $E_\varepsilon = 0.001$, а наивероятнейшие значения соответственно равны μ_d = 12300 м и $\mu_\varepsilon = 31.2°$ (с.в. D и ε независимы и нормальны).

(В) *Избранные дискретные распределения*

4. В некотором процессе получается 2 % брака. Берется выборка из 10 независимых образцов. Какова вероятность того, что в выборке не окажется дефектов ?

5. Посмотреть графики биномиального распределения для $n = 8$,

a) $p = 0.1$ c) $p = 0.5$
b) $p = 0.3$ d) $p = 0.9$.

6. В окрестностях г. Кливленда (США) среднегодовое число погибших в автокатастрофах за период 1945-1951 гг. составило 0.5. В 1952 г. там погибло 3 человека. Не указывает это на реальный рост транспортной опасности ?

7. Посмотреть графики распределения Пуассона $Po(m)$ для

a) $m = 0.5$ b) $m = 1.0$ c) $m = 6.0$.

8. При выборе случайных цифр с одинаковой вероятностью может быть взята каждая цифра 0, 1, 2, 3, ..., 9. В выборке объема 3 все три отобранные цифры превосходят 6. Можно ли считать это неожиданным событием ?

9. Вероятность попадания в цель равна 0.90.

(а) Найдите математическое ожидание числа попаданий в 10 независимых попытках;

(б) какова вероятность более чем 5 попаданий ?

(в) какова вероятность двух или менее попаданий ?

10. Как много независимых попыток нужно сделать в условиях упражнения 9, чтобы вероятность попасть хотя бы один раз была не менее 0.99?

11. По заявлению некоторой фирмы вероятность того, что данное покрытие устойчиво против коррозии (что определяется путем стандартных испытаний) , равна 0.95. Было выбрано 20 независимых образцов.

(а) Если эти покрытия столь хороши, как утверждается, то как много брака можно ожидать ?

(б) Какова вероятность того, что будет обнаружен более чем один случай брака ?

12. Процент выхода, характеризующий эффективность химического процесса, распределен симметрично около среднего 87 %. В каждом из 6 независимых лабораторных испытаний выход оказался менее 87 %.

(а) Какова вероятность такого события ?

(б) Какова вероятность того, что эффективность окажется > 87 % (для этого процесса) в 8 случаях из 10 ?

13. Максимальное сопротивление резиновой ленты должно составлять 2.0 Ом*см. Ниже приведены числа дефектов (X), полученные в 1000 выборках по 10 образцов.

X	f		X	f		X	f
0	353		4	11		8	0
1	382		5	2		9	0
2	196		6	1		10	0
3	55		7	0			

(а) Сравните эти результаты с биномиальным распределением при $p = 0.1$ и $n = 10$.

(б) Сравните эти результаты с распределением Пуассона при $m = 1.0$.

(в) В каком случае данные аппроксимируются лучше ? Почему ?

14. Пусть вероятность получения качественного листа при нанесении на него покрытия равна 0.90.

а) Сколько бракованных листов следует ожидать в выборке объема 50 ?

б) Какова вероятность того, что в выборке объема 50 окажется более 5 плохих листов ?

в) Какова вероятность того, что окажется менее трех плохих листов ?

15. Какова вероятность двух или более дефектных образцов в выборке объема 20, если средняя доля дефектных образцов равна 0.12 ?

16. Фирма провела полевые испытания шести образцов одинакового оборудования и не обнаружила среди них неисправных. Как часто можно ожидать, что в выборке объема 6 будут присутствовать неисправные образцы, если

а) ожидаемая доля неисправных образцов равна 5 % ?

б) ожидаемая доля неисправных образцов равна 10 % ?

17. Фирма провела кампанию по уменьшению числа несчастных случаев. Среднее число таких случаев за предыдущие годы было 20. В год поведения указанной кампании их число оказалось равным 16. Можете ли вы сказать, что это произошло под влиянием неслучайных причин ? Оцените вероятность 16 и менее несчастных случаев, полагая, что условия не изменились.

18. Случайная величина X имеет распределение Пуассона со средним значением μ. Покажите, что математическое ожидание A**X равно

$$\exp\{\mu \cdot (A - 1)\}.$$

Каково среднее значение этой величины, если X имеет биномиальное распределение с параметрами n и p ?

97. Радиоаппаратура состоит из 1000 элементов. Вероятность отказа одного элемента в течение одного года работы равна 0.001 и не зависит от состояния других элементов. Какова вероятность отказа двух и не менее двух электронных элементов в год ?

98. Математическое ожидание числа отказов радиоаппаратуры за 10000 часов работы равно 10. Определить вероятность отказа радиоаппаратуры за 100 часов работы.

99. Аппаратура содержит 2000 одинаково надежных элементов, вероятность отказа для каждого из которых равна p = 0.0005. Какова вероятность отказа аппаратуры, если он наступает при отказе хотя бы одного из элементов ?

100. В аппаратурный отсек космической ракеты за время ее полета с вероятностью

$$p(r,\lambda) = e^{-\lambda}\frac{\lambda^r}{r!}$$

попадает r элементарных частиц. Условная вероятность для каждой из них попасть при этом в уязвимый блок равна p . Найти вероятность попадания в блок : а) ровно k частиц; б) хотя бы одной частицы.

(Г) *Некоторые непрерывные распределения*

119. Вычислить математическое ожидание и дисперсию случайной величины ξ , имеющей равномерное распределение $R(\mu, \omega)$ с функцией плотности вероятностей

$$f(x) = \begin{cases} \dfrac{1}{\omega}, & x \in [\mu - \dfrac{\omega}{2}, \mu + \dfrac{\omega}{2}]; \\ 0, & x \notin [\mu - \dfrac{\omega}{2}, \mu + \dfrac{\omega}{2}] \end{cases} .$$

120. Пусть $\xi \sim R(\mu, \omega)$. Показать, что с.в.

$$\eta = (\xi - \mu + \omega/2) / \omega \sim R(1/2,\ 1).$$

121. Пусть $\xi \sim R(1/2,\ 1)$. Доказать, что с.в. $\eta = -2\ln\xi$ имеет распределение $\chi_2^{\ 2}$ (Хи-квадрат с двумя степенями свободы) с функцией плотности вероятностей

$$f(y) = \frac{1}{2} exp\{-y/2\} .$$

23. Рычаг собирается из пяти секций. Исследование отдельных секций показало, что средняя длина крайних секций составляет 1.001 дюйма, а трех средних 1.999 дюйма. Средние квадратические отклонения длин всех секций равны 0.004 дюйма. Если осуществить случайную сборку и отдельные секции предполагаются распределенными нормально, то

а) Какова средняя длина сборки ?

б) Каково среднее квадратическое отклонение длины сборки ?

в) Какова вероятность того, что длина рычага превышает 8.002 дюйма ?

24. В выборку включены 224 последовательные бобины медной проволоки. Измеряется толщина пластикового покрытия проволоки. Можно ли на основе приведенных ниже данных считать, что толщина покрытия имеет нормальное распределение ? Сравните наблюдаемые и теоретические частоты.

Толщина X	Частота f
146	3
147	3
148	7
149	11
150	25
151	33
152	34
153	37
154	25
155	23
156	11
157	9
158	2
159	0
160	1

26. Найдите функцию плотности вероятностей суммы Z = X + Y, если функция плотностей совместного распределения (X, Y) есть P(x, y) = 1 для 0 <= x <= 1; 0 <= y <= 1 и P(x, y) = 0 вне единичного квадрата.

31. В таблице, приведенной ниже, X - твердость по Шору сложного сплава. Постройте гистограмму и сравните ее с нормальным распределением при $\mu = 73.5$, $\sigma = 3$.

X	f	X	f
57-58	1	73-74	38
59-60	3	75-76	22
61-62	5	77-78	18
63-64	8	79-80	15
65-66	10	81-82	11
67-68	13	83-84	6
69-70	20	85-86	3
71-72	25	87-88	2

(Г1) *Свойства нормального распределения*

118. Построить графики $f(x)$ – функции плотности нормального $N(\mu ,\sigma^2)$ распределения со значениями параметров: (а) $\mu = 0.$, $\sigma = 1.$; (б) μ = -5., σ = 1.; (в) μ = +5., σ = 1.; (г) μ = 2.5, $\sigma = 0.3$; (д) μ = -2.5, $\sigma = 3$. Интерпретируйте смысл параметров нормального распределения, объясните особенности поведения $f(x)$.

122. Показать, что если ξ – с.в., имеющая нормальное распределение $N(\mu, \sigma^2)$, то

$$E\{ \mid \xi - \mu \mid \} = \sqrt{\frac{2}{\pi}} \cdot \sigma \quad .$$

19. Предполагается, что внешний диаметр (ВД) годных для сборки стальных стержней распределен приблизительно по нормальному закону со средним 2.30 см и средним квадратическим отклонением 0.06 см, т.е. N(2.30, 0.06). Пределы допуска 2.31$\pm$ 0.10 см. Изделие с ВД ниже нижнего предела допуска считается ломом, тогда как при превышении ВД верхнего предела возможна доработка. Нужно получить ответы на следующие вопросы:

а) Каков получающийся процент лома ?

б) Сколько процентов продукции нуждается в доработке ?

в) Каков будет ожидаемый процент лома и изделий, требующих доработки, если изменить средний ВД до 2.31 см ?

20. Наружные диаметры втулок распределены нормально. Среднее этого распределения равно 2.000 дюйма, а среднее квадратическое отклонение составляет 0.003 дюйма. Определите вероятность того, что наружный диаметр втулки

а) равен или больше, чем 2.009 дюйма ;

б) меньше, чем 1.994 дюйма ;

в) заключен между 1.997 и 2.003 дюйма ;

г) заключен между 1.994 и 2.006 дюйма.

21. Пусть установлено, что выход в граммах красителя стандартного цвета со специальным оттенком распределен нормально со средним 1550 и средним квадратическим отклонением 50. В скольким из 100 проверок вы ожидаете, что выход в среднем будет

а) ниже 1550 ;

б) выше 1650 ;

в) между 1525 и 1575 ;

г) выше 1470 ?

22. Изделия типа А должны подгоняться к изделиям типа В. Предполагается, что критический внешний размер А распределен нормально со средним 4.30 см и средним квадратическим отклонением 0.04 см. Изделия В имеют критический внутренний размер, который считается распределенным нормально со средним 4.36 см и средним квадратическим отклонением 0.04 см.

а) Какова ожидаемая доля случаев, когда изделия А и В, выбранные независимо и случайно, окажутся непригодными друг для друга ?

б) Надо ли считать необычным два случая несовместимости из 20 ?

84. Измеряемая с.в. x подчиняется закону распределения $N(10, 25)$. Найти симметричный относительно математического ожидания μ_x интервал, в который с вероятностью p попадет измеренное значение. Рассмотреть следующие числовые значения: а) p =0.9974; б) p = 0.9544; в) p = 0.50.

85. Случайная величина $x \sim N(\mu, \sigma^2)$. Пользуясь программой вычисления функции стандартного нормального распределения ERF.EXE, вычислить вероятность p_k того, что отклонение с.в. x от ее математического ожидания не превзойдет величины $k{\cdot}\sigma$ (ответ получить для трех значений k = 1, 2, 3).

25. Величина X распределена нормально и 1) Pr[X > 7] = 0.05; 2) Pr[X < 5] = 0.50. Найдите Pr[4 < X < 6].

86. В нормально распределенной совокупности 15% значений x меньше 12 и 40% значений x больше 16.2. Найти математическое ожидание и стандартное отклонение данного распределения.

22. Изделия типа А должны подгоняться к изделиям типа В. Предполагается, что критический внешний размер А распределен нормально со средним 4.30 см и средним квадратическим отклонением 0.04 см. Изделия В имеют критический внутренний размер, который считается распределенным нормально со средним 4.36 см и средним квадратическим отклонением 0.04 см.

а) Какова ожидаемая доля случаев, когда изделия А и В, выбранные независимо и случайно, окажутся непригодными друг для друга ?

б) Надо ли считать необычным два случая несовместимости из 20 ?

87. Деталь, изготовленная автоматом, считается годной, если отклонение x от номинала не превышает 10 мм. Точность изготовления деталей характеризуется стандартным отклонением σ. Считая, что для данной технологии σ = 5 и x распределено нормально, выяснить, сколько процентов годных деталей изготовляет автомат.

88. В условиях предыдущей задачи выяснить, какой должна быть точность изготовления, чтобы процент годных деталей повысился до 98 ?

Практикум

“Построение доверительных интервалов и проверка гипотез”

Теоретическое обоснование

Из книги [4] Джонсон Н., Лион Ф. *Статистика и планирование эксперимента в технике и науке: Методы обработки данных*. М.: Мир. 1980

Пояснения к табл.8.5: рассматривается либо одна выборка x_1 , ..., x_n объема n из совокупности $N(\mu , \sigma^2)$, имеющая выборочное среднее

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{41}$$

и выборочную дисперсию

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ , \tag{42}$$

либо две выборки объемов n_1 и n_2 , x_1 , ..., x_{n_1} и x_1 , ..., x_{n_2} , соответственно, из генеральных совокупностей $N(\mu_1 , \sigma_1^{\ 2})$ и $N(\mu_2 , \sigma_2^{\ 2})$, имеющие выборочные средние

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i \quad , \qquad \bar{x}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} x_i \tag{43}$$

и выборочные дисперсии s_1^2, s_2^2 . Рассматриваемые распределения статистик критериев: стандартное нормальное $N(0,\ 1)$, распределение χ_k^2 - χ^2 с k степенями свободы, St_k - распределение Стьюдента с k степенями свободы, $F_{k,\ m}$ - распределение Снедекора с $k,\ m$ степенями свободы. Критические точки обозначаются как u_{α} , $\chi^2_{k,\ \alpha}$, $t_{k,\ \alpha}$ и $F_{k,\ m,\ \alpha}$ – нижние критические точки; $u_{1-\alpha}$, $\chi^2_{k,\ 1-\alpha}$, $t_{k,\ 1-\alpha}$ и $F_{k,\ m,\ 1-\alpha}$ – верхние критические точки, соответственно. В последней из статистик критериев $(\sigma_1^2 \neq \sigma_2^2)$ число степеней свободы

$$\nu = \frac{(s_1^2 / n_1 + s_2^2 / n_1)^2}{\dfrac{(s_1^2 / n_1)^2}{n_1 - 1} + \dfrac{(s_2^2 / n_2)^2}{n_2 - 1}} \ . \tag{44}$$

<u>**Проверка гипотез.**</u>

Используемые в задачах критерии и статистики:

1.Критерий согласия χ^2 Пирсона.

Предположения : x_1 , ..., x_n - случайная выборка независимых одинаково распределенных величин.

Гипотеза H : x_1 , ..., x_n имеет заданное распределение $F(x)$ (или $F(x,\ \theta_1 , ...,\ \theta_m)$, значения параметров θ_1 , ..., θ_m неизвестны).

Альтернатива K : x_1 , ..., x_n имеет какое-то другое распределение $F_1(x)$.

Уровень значимости : α.

Статистика критерия : X^2, где

$$X^2 = \sum_{j=1}^{k} \frac{(r_j - np_j)^2}{np_j} \ , \qquad (45)$$

где $E_1, ..., E_j, ..., E_k$ - интервалы группировки значений с.в. $\xi \sim F(x); E_j \cap E_l = \varnothing$, $E_1 \cup .... \cup E_k = E$ - достоверному событию, $Pr\{\xi \in E_j\} = p_j$, r_j - частота (число x_i из $x_1, ..., x_n$,попавших в E_j).

Решение : если $X^2 < \chi^2_{k-1, 1-\alpha}$, то H принимается; в противном случае H отклоняется. Если рассматривается H с $F(x, \theta_1, ..., \theta_m)$, значения параметров $\theta_1, ..., \theta_m$ неизвестны, то если $X^2 < \chi^2_{k-1-m, 1-\alpha}$, то H принимается; в противном случае H отклоняется (здесь X^2 вычисляется при подставленных значениях оценок параметров $\theta_1, ..., \theta_m$) .

Маленькое замечание: во многих задачах на критерий Хи-квадрат не известны параметры проверяемого распределения, и их нужно предварительно оценить по выборке $x_1, ..., x_n$. Оценки для нормального распределения суть выражения (41), (42); для параметра ε биномиального распределения $Bi(n, \varepsilon)$ и для параметра θ распределения Пуассона $Po(\theta)$ оптимальной оценкой также служит выборочное среднее (41).

2. Сравнение дисперсий в задаче о k выборках из нормальных совокупностей (критерий Бартлетта).

Предположения : имеется k независимых случайных выборок из $N(\mu_j, \sigma_j^2)$, $j = 1, ..., k$.

Гипотеза H : $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma^2$.

Альтернатива K : $\exists j, l : \sigma_j^2 \neq \sigma_l^2$.

Уровень значимости : α.

Статистика критерия : X^{2*}, где

$$X^{2*} = \frac{1}{1 + \frac{\sum_j (1/\nu_j) - \frac{1}{\nu}}{3(k-1)}} \left[\nu \cdot \ln s_p^2 - \sum_j \nu_j \cdot \ln s_j^2 \right] \ , \qquad (46)$$

где s_j^2 - выборочная дисперсия (42) для j-ой выборки объема n_j, $\nu_j = n_j - 1$ - число степеней свободы для j-ой выборки,

$$\nu = \sum_j \nu_j \ , \quad s_p^2 = \nu^{-1} \sum_j \nu_j s_j^2 \ . \qquad (47)$$

Решение : Если $X^{2*} < \chi^2_{k-1, 1-\alpha}$, то H принимается; в противном случае H отклоняется. (*Замечание :* при вычислении X^{2*} корректирующий множитель $1./3 \cdot (k-1)^{-1} [\sum \nu_j^{-1} - \nu^{-1}]$ часто можно опустить).

3. Проверка независимости компонент в выборке двумерных нормальных величин.

Предположения : имеется n независимых реализаций

$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ двумерной нормальной случайной величины (с.в.) $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \sim N_2\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$.

Гипотеза H : ξ и η независимы, т.е. $\rho = 0$.

Альтернатива K ; $\rho \neq 0$.

Уровень значимости : α.

Статистика критерия : $t = r\sqrt{\dfrac{n-2}{1-r^2}}$, где

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad - \qquad (48)$$

выборочный коэффициент корреляции.

Решение : Если $|t| < t_{n-2,\,1-\alpha/2}$, то H принимается ; в противном случае H отвергается.

4. Сравнение более двух средних из нормальных совокупностей.

Предположим, что требуется проверить гипотезу о равенстве k м.о., т.е. H: $\mu_1 = \mu_2 = \dots = \mu_k$, когда извлечено k выборок $\{x_{ji}\}$, $i = 1, \dots, n_j$; $j = 1, \dots, k$, причем n_j - объем j-ой выборки.

Предположения : Имеется k независимых выборок $\{x_{ji}\}$, $i = 1, \dots, n_j$; $j = 1, \dots, k$, из нормальных популяций $N(\mu_j, \sigma^2)$ с общей дисперсией σ^2.

Гипотеза H : $\mu_1 = \mu_2 = \dots = \mu_k = \mu$.

Альтернатива K : $\exists j, l : \mu_j \neq \mu_l$.

Уровень значимости : α.

Статистика критерия : одна из статистик , используемых для решения поставленной задачи (в случае выборок равного объёма $n_1 = n_2 = \dots = n_k = n$) – стьюдентизированный размах

$$Q = \frac{R_{\bar{x}}}{s_p / \sqrt{n}} \quad , \qquad (49)$$

где $\boldsymbol{R_{\bar{x}}} = \max_j \bar{x}_j - \min_j \bar{x}_j$ - размах средних, $s_p^2 = \dfrac{\sum_j (n_j - 1)\cdot s_j^2}{\sum_j (n_j - 1)} = \dfrac{\sum_j \sum_i (x_{ji} - \bar{x}_j)^2}{\sum_j (n_j - 1)}$ - несмещённая оценка σ^2.

Решение : Если $Q < Q_{n,\,\alpha}$, то H принимается ; в противном случае H отвергается. Критические значения Q см. в табл. И.

5. Равенство средних значений совокупностей при парных или коррелированных выборках.

Во многих задачах две выборки взаимосвязаны в силу особенностей постановки эксперимента, или потому, что избежать этой связи невозможно. Например, это имеет место тогда, когда признаки А и В измеряются на одном и том же объекте. Чтобы учесть взаимосвязь выборок, берут соответствующие выборочные значения пáрами (x_1 из А и y_1 из В, x_2 из А и y_2 из В и т.д.), и исследуют их разности $D_i = x_i - y_i$.

Предположения : с.в. D_1 , D_2 , ..., D_n выбраны случайным образом из совокупности с распределением $N(\mu_D, \sigma_D^2)$.

Гипотеза : $\mu_D = 0$.

Альтернатива : $\mu_D \neq 0$.

Уровень значимости : α.

Статистика критерия : $T = \dfrac{\overline{D} - 0}{s_{\overline{D}}}$, (50)

где $s_{\overline{D}}$ - среднеквадратичное отклонение для выборочного среднего

$$\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i \quad .$$

Решение : Если $| T | < t_{n-1,\,1-\alpha/2}$, то H принимается ; в противном случае H отклоняется.

Рассматриваемая процедура применима, когда модель записана в виде :

$$x_i = X_i + Z_i' ,$$
$$y_i = X_i + Z_i'' + \delta. \qquad (51)$$

Здесь X_i могут быть как константами (параметрами), как и с. в. ; δ – константа ; Z_1', ..., Z_n' - взаимно независимые с.в., так же как и Z_1'', ..., Z_n'' . Величина Z_i' может быть связана с Z_i'', но она должна быть независима от Z_j'' при $j \neq i$. Тогда $D_i = x_i - y_i = -\delta + (Z_i' - Z_i'') = -\delta + Z_i$, и при условии, что дисперсия Var(Z_i) не зависит от i (это так, если каждая из величин Var(Z_i ‘) , Var(Z_i “) , Cov(Z_i ‘,Z_i “) постоянна), можно использовать описанный метод для проверки гипотезы $\delta = 0$.

6. Сравнение значения p_0 параметра совокупности и его выборочной оценки - случай биномиального распределения.

Предположения : С.в. $X \sim Bi(n, p)$.

Гипотеза H: $p = p_0$.

Альтернатива : $p > p_0$.

Уровень значимости : $\leq \alpha$ (см. последующие замечания).

Статистика критерия : X / n .

Решение : Представляется естественным принимать H, если $X / n \leq K$ и отклонять H в противном случае. Не всегда удается найти такое значение K , чтобы Pr { $X / n \leq K \mid H$ } = α. Поэтому K выбирается так, чтобы уровень значимости имел наибольшее возможное значение, $\leq \alpha$.

Если альтернативой H является $p \neq p_0$, то H принимается, если $K_1 \leq X / n \leq K_2$ и отклоняют, если “нет”. Один из подходов использует приближённые 100(1 - α) % -ные доверительные интервалы для p . Если p_0 попадает в этот интервал, H принимают, если нет – отклоняют. Доверительный интервал имеет вид $p_1 \le P \le p_2$, где p_1 – наибольшее из p , удовлетворяющих уравнению

$$\Pr\{ X \ge x_0 \} = \sum_{x=x_0}^{n} \binom{n}{x} \cdot p^x (1-p)^{n-x} \le \alpha / 2 \ , \qquad (52)$$

а x_0 - наблюдаемое значение X; p_2 - наименьшее из значений p , удовлетворяющих уравнению

$$\Pr\{ X \le x_0 \} = \sum_{x=0}^{x_0} \binom{n}{x} \cdot p^x (1-p)^{n-x} \le \alpha / 2 \ . \qquad (53)$$

Указанные p_1 и p_2 могут быть взяты из таблиц [5] Большев Л.Н., Смирнов Н.В. *Таблицы математической статистики.* М.: Наука. 1965. См. также табл. 1.

7. Сравнение двух долей – случай биномиального распределения.

Предположения : Есть две независимые с.в. X_1 и X_2 , $X_j \sim Bi(n_j, p_j), \quad j = 1, 2$.

Гипотеза H: $p_1 = p_2 = p$, или H: $p_1 - p_2 = 0$.

Альтернатива : $p_1 \ne p_2$.

Уровень значимости : α.

Статистика критерия :

$$U = \frac{\hat{p}_1 - \hat{p}_2}{S_{\hat{p}_1 - \hat{p}_2}} = \frac{X_1 / n_1 - X_2 / n_2}{S_{\hat{p}_1 - \hat{p}_2}} \ , \qquad (54)$$

где

$$S_{\hat{p}_1 - \hat{p}_2} = \sqrt{\left(\frac{X_1 + X_2}{n_1 + n_2}\right) \cdot \left(1 - \frac{X_1 + X_2}{n_1 + n_2}\right) \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \ . \qquad (55)$$

Если верна H , то оценкой p является

$$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2} \ . \qquad (56)$$

При достаточно больших n_1 , n_2 используется нормальная аппроксимация биномиального распределения ; при $p_1 = p_2$ асимптотически $U \sim N(0, 1)$.

Решение : Если $| U | < u_{1 - \alpha / 2}$, то H принимается ; в противном случае H отвергается.

8. Обнаружение выбросов в нормально распределенных данных.

Часто при обработке данных одно или несколько значений имеет подозрительно малую или же слишком большую величину. Такие аномальные наблюдения называются *выбросами* (outliers), и задача статистического анализа – подтвердить, что он (они) несовместимы с предписанной вероятностной моделью.

Предположения : Имеются n независимых с.в. $x_1, ..., x_n$.

Гипотеза H : $x_1, ..., x_n$ – выборка из совокупности с нормальным $N(\mu, \sigma^2)$ распределением; значения параметров μ и σ^2 не известны.

Альтернатива K : Одно из значений $x_l \sim N(\mu', \sigma^2)$ имеет м.о. $\mu' = E\{ x_l \} = \mu + \sigma \cdot \Delta$, $| \Delta | > 0$.

Уровень значимости : α.

Статистика критерия :

$$D = \max_i \frac{|x_i - \bar{x}|}{s} \qquad (57)$$

(максимальное стьюдентизированное абсолютное отклонение).

Решение : Если $D < D_{1-\alpha}$, то H принимается ; в противном случае H отвергается. Критические значения $D_{1-\alpha}$ даны в табл. 3. *Замечание :* в частном простейшем случае (45) сводится к хорошо известному практикам "правилу трех сигм": H отвергается, если $D > 3\sigma$ (или 3 s) / правило Шовене/.

<u>*Впрок*</u> : Можно использовать и другие критерии выявления выбросов, например, статистики Диксона

$$r_{pq,\,st} = \frac{x_{(p)} - x_{(q)}}{x_{(s)} - x_{(t)}} \ , \qquad (58)$$

частного вида $R_{r-q\ \ n-s} \equiv r_{pq,\ sq}$, где $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ – члены вариационного ряда, построенного по исходным данным x_1 , ..., x_n путем их упорядочивания по величине (порядковые статистики) . H отвергается, если $r_{p-1\ \ n-s}$ превышает соответствующее критическое значение. Дальнейшие детали ясны из табл. К и из книги [4].

9. Проверка гипотезы о значении параметра распределения Пуассона.

Предположения : выборка объема n независимых величин x_1, ..., x_n взята из распределения Пуассона Po(θ).

Гипотеза H : $\theta^* = \theta_0$.

Альтернатива K : $\theta^* = \theta_1 \neq \theta_0$.

Уровень значимости : α.

Статистика критерия : $X_0 = \sum_{i=1}^{n} x_i$. (59)

Решение : Если $X_{1,\alpha} \le X_0 \le X_{2,\alpha}$, то H принимается ; в противном случае H отклоняется. Здесь $\theta = n\,\theta_0$,

$$X_{1,\alpha}:\ Pr\{X \le X_{1,\alpha}\} = \sum_{x=0}^{X_{1,\alpha}} e^{-\theta}\frac{\theta^x}{x!} \le \alpha/2\ , \qquad (60')$$

$$X_{2,\alpha}:\ Pr\{X \ge X_{2,\alpha}\} = \sum_{x=X_{2,\alpha}}^{\infty} e^{-\theta}\frac{\theta^x}{x!} = 1 - \sum_{x=0}^{X_{2,\alpha}-1} e^{-\theta}\frac{\theta^x}{x!} \le \alpha/2\ . \qquad (60'')$$

Замечание : Приближенно границы 100·γ-%-ного доверительного интервала для θ получаются при решении следующих уравнений относительно $n\,\theta^*$:

$$Pr\{X \le X_0\} = \sum_{x=0}^{X_0} e^{-n\theta_2}\frac{(n\theta_2)^x}{x!} \le \frac{1-\gamma}{2}\ , \qquad (61')$$

$$Pr\{X \ge X_0\} = \sum_{x=X_0}^{\infty} e^{-n\theta_1}\frac{(n\theta_1)^x}{x!} \le \frac{1-\gamma}{2}\ . \qquad (61'')$$

(см. Таблицу Б и программу INT_POIS.EXE).