# Работа № 2. Выборки и их представление

##  Основные понятия

Напомним, что такое выборка, вариационный ряд, эмпирическое распределение, группирование, гистограмма, выборочные характеристики и др.

**Выборкой** *х1, ..., хn* объема *n* из совокупности, распределенной по *F(х),* называется *n* независимых наблюдений над случайной величиной *ξ* с функцией распределения *F(x).*

**Вариационным рядом** *х(1)≤ х(2)≤ ... ≤ х(n)* называется выборка, записанная в порядке возрастания ее элементов.

Каждому наблюдению из выборки присвоим вероятность, равную 1/*n*; получим распределение, которое называют **эмпирическим**; ему соответствует **функция эмпирического распределения**

*≡ = ,*

где *μn(х)* - число членов выборки, меньших *х*. Значение этой функции для статистики определяется тем, что при *n* → ∞

*→ F(x)*

(теорема Гливенко).

Выборки больших объемов труднообозримы; разобъем диапазон значений выборки на равные интервалы и подсчитаем для каждого интервала **частоту**- количество наблюдений, попавших в него; частоты, отнесенные к общему числу наблюдений *n*, называют **относительными частотами**; графическое представление распределения частот по интервалам − **гистограммой**; **накопленной частотой** для данного интервала называют сумму частот данного интервала и всех тех, что левее его.

Числовые характеристики эмпирического распределения называются **выборочными характеристиками**: выборочные **среднее (**математическое ожидание**), дисперсия:**

*= , s2=*

выборочный **момент** порядка *к*:

*mk = ;*

выборочные **квантили** ζp порядка *р* - корни уравнения

*F(ζp)=p*,

которыми являются члены вариационного ряда

*ζ(p)=ξ([np]+1*),

где *[nр]* означает целую часть *nр*; частным случаем (*p* = 0.5) является выборочная **медиана -** центральныйчлен вариационного ряда. Значение выборочных характеристик состоит в том, что при *n* → ∞ они стремятся к истинным значениям распределения *F(х).*

Приведем с помощью пакетов примеры. Исходные данные находятся в табл.1 ( *E(a)* в таблице означает показательное (экспоненциальное) распределение с математическим ожиданием, равным *a).*

 таблица1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ¹  | Закон | *n* | α | ¹  | Закон  | *n* | α |
| 1 | *R* [0, 2] | 50 | 0.03 | 14 | *N* (1,4) | 60 | 0.01 |
| 2 | *N*(2, 0.25) | 60 | 0.02 | 15 | *E* (5) | 70 | 0.03 |
| 3 | *E* (3) | 70 | 0.01 | 16 | *R* [0.3] | 80 | 0.1 |
| 4 | *R* [1, 3] | 80 | 0.02 | 17 | *N* (1,4) | 50 | 0.3 |
| 5 | *N* (1, 1) | 50 | 0.01 | 18 | *E* (1) | 60 | 0.2 |
| 6 | *E* (2) | 60 | 0.03 | 19 | *R* [1,3] | 70 | 0.03 |
| 7 | *R* [2, 3] | 70 | 0.01 | 20 | *N* (1,1) | 80 | 0.02 |
| 8 | N (0, 4) | 80 | 0.03 | 21 | *E* (2) | 50 | 0.01 |
| 9 | *E* (3) | 50 | 0.02 | 22 | *R* [2,3] | 60 | 0.02 |
| 10 | *R* [0, 2] | 60 | 0.03 | 23 | *N* (2,1) | 70 | 0.01 |
| 11 | *N* [2, 1] | 70 | 0.02 | 24 | *E* (3) | 80 | 0.03 |
| 12 | *E* (4) | 80 | 0.01 | 25 | *R* [1,2] | 50 | 0.01 |
| 13 | *R* [1, 2] | 50 | 0.02 |  |  |  |  |

##  Выполнение в пакете STATGRAPHICS

***Генерация выборки***

Работа начинается с главного меню пакета (панель *STATGRAPHICS Statistical Graphics System):*

 |STATGRAPHICS Statistical Graphics System|

DATA MANAGEMENT AND SYSTEM UTILITIES TIME SERIES PROCEDURES

 A. Data Management L. Forecasting

 B. System Environment M. Quality Control

 C. Report Writer and Graphics Replay N. Smoothing

 D. Graphics Attributes O. Time Series Analysis

PLOTTING AND DESCRIPTIVE STATISTICS ADVANCED PROCEDURES

 E. Plotting Functions P. Categorical Data Analysis

 F. Descriptive Methods Q. Multivariate Methods

 G. Estimation and Testing R. Nonparametric Methods

 H. Distribution Functions S. Sampling

 I. Exploratory Data Analysis T. Experimental Design

ANOVA AND REGRESSION ANALYSIS MATHEMATICAL AND USER

 PROCEDURES

 J. Analysis of Variance U. Mathematical Functions

 K. Regression Analysis V. Supplementary Operations

 **рис. 1. Главное меню**

Выполнение:

*H.Distribution functions* (законы распределения) — *5.Random Number* *Generation* (генерация случайных чисел) - из списка *Distributions available* (возможные распределения) выбираем нужное и его номер вводим в окно *Distribution number* - F6 (исполнение) — вводим параметры распределения и объем выборки *Number of samples*; исходное состояние датчика случайных чисел (окно *seed*) оставим без изменения (однако, оно не должно превышать 2147483646) - F6 — вводим имя файла, в котором будем хранить все данные этой работы (в виде различных переменных): *File: WORK* (например), вводим имя переменной, в которой будет находиться наша выборка: *Variable* (переменная): *x* - *F6.*

Выборка сгенерирована. Посмотрим полученную выборку:

*Ctrl + Break* (быстрый возврат в главное меню вместо многократного *Esc* или *F10*) — *A.Data Management* (управление данными) — *1.Display Data Directory* — выбираем нашу переменную *WORK.x* - *F6.*

Наблюдаем выборку. Выпишем значения выборки или выведем на печать (клавиша *F4*) или сохраним (*F3;* повторный вызов: *Report Writer & Graphics Replay* (составление отчетов и вызов графиков) *- Replay Texts & Graphic Files* (вызов текстов и графических файлов ) ).

Посмотрим выборку графически. После возврата в главное меню *(Ctrl + Break):*

 *E.Plotting Function* (графические функции) — *1.X-Y Line and Scatterplots* (*x-y* графики) — вводим данные для графика: по оси *x* должны быть целые числа от 1 до *n*: в строку *x* записываем оператор (для *n* = 50, например):

 *COUNT 50*

этот оператор создает массив целых чисел от 1 до 50; в строку *y* записываем *x*; в окне *Points: Yes* (точки нужны), в окне *Lines: Yes* (клавишей «пробел», линии нужны) - *F6.*

График выведем на печать (*F4*) или сохраним (*F3*).

***Построение вариационного ряда***

**1-й способ**

*A.2.File Operations* **—** вводим в окно *file name: WORK* (можно так: *Ctrl+F7* (список файлов) — выделить нужный - *ENTER*) *- Desired* operation: C (Edit - редактирование) - *F6* — выделяем переменную *x - ENTER -* *F6*-(наблюдаем выборку) - *F5* (опции) - *Sort in ascending order* (сортировка в порядке возрастания ) - *F6 - Save and exit* (запоминание и выход). Если бы требовалось не менять содержимое переменной *x*, следовало бы сначала скопировать ее в другую переменную (операцией *Update*).

**2-й способ**

Сначала загрузим оператор сортировки *SORTUP*, который относится к разряду загружаемых:

*V.Supplementary Operations* (дополнительные операции) — *1.Load Operation and Function* (загрузка операторов и функций) — *Mathematical function - Read* (после использования загружаемых операторов их желательно выгрузить опцией *Erase*, чтобы не занимать память).

*Ctrl+F5* (быстрый выход в исполнительное окно) — *SORTUP x -* *ENTER* (наблюдаем вариационный ряд, при этом содержащие переменной *x* не изменилось).

 ***Построение графика функции эмпирического распределения***

*F.3.Frequency Histogram — Data: x - F6* — поправляем некоторые параметры графика: *No of classes* (число классов): 200 (или еще больше: — 500, чтобы на каждый интервал попало не больше одного наблюдения), *Cumulative: Yes*, (накопленные частоты, т.е. функция распределения), *Relative: Yes*, (относительные частоты) - *F6*.

Наблюдаем функцию эмпирического распределения. Выводим ее на печать или сохраняем.

***Группирование данных***

*F.2.Frequency Tabulation — Data: x - F6* — поправляем, если нужно параметры группирования: нижний (*Lower limit*) и верхний (*Upper Limit*) пределы (минимальное и максимальное значения выборки приведены ниже на экране), число интервалов группирования *No of classes: 10 - F6 — Display Table - ENTER.*

Наблюдаем таблицу группированных данных. Выводим ее на печать или сохраняем*.*

***Построение гистограммы частот***

*F.3.Frequency Tabulation — Data: x - F6* — поправляем параметры графика: *No of classes: 10 - F6.*

Наблюдаем гистограмму. Выводим ее на печать или сохраняем*.*

***Определение выборочных характеристик***

Определим выборочные среднее, дисперсию, cтандартное отклонение, медиану (сравним их с теоретическими значениями), минимальное и максимальное значения выборки, размах:

*F.1 - Summary Statistics - Data rectors:* *x* — в окне *Statistics* оставляем те буквы — коды, которым соответствуют нужные нам статистики *A, B, E, F, H, I, J - F6.*

Наблюдаем таблицу выборочных значений. Выписываем ее и сравниваем с теоретическими значениями.

***Проверка гипотезы о типе распределения***

Проверим выборку с помощью критерия Колмогорова - Смирнова .

*H.1. - Distribution Fitting Data vector: x*, вводим код распределения в окно *Distribution number - F6* — ââîäèì вместо оценок теоретические значения параметров - *F6 — Histogram - ENTER* — поправляем параметры графика: *No of classes 200, Cumulative: Yes - F6.*

Наблюдаем функции эмпирического и теоретического распределений; определим по графику значение статистики



— ìåðû различия между этими функциями. Величина *Dn,* конечно же, определяется пакетом:

*Esc - Esc* — вместо опции *Histogram* выбираем *K-S Test - ENTER.* Сообщается значение статистики « *...statistic DN = ....» и «...significance* *level* = ...» т.е. уровень значимости

.

Если эта вероятность мала (сотые доли или меньше), гипотезу о соответствии наблюдений теоретическому распределению следует отклонить. В противном случае признают, что наблюдения не противоречат гипо­тезе.

***Описание двумерных выборок***

**Пример.** В табл.2 приведены результаты химического анализа 32 образцов сланцевых пород на содержание двуокиси кремния (SiO2 – *x*) и двуокиси алюминия (Al2O3 – *y*).

Построим диаграмму рассеяния для этой выборки, определим выборочные характеристики: среднее, дисперсии, коэффициент корреляции и построим диаграмму рассеяния и двумерную гистограмму.

Таблица 2

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ¹  | *X* | *Y* | ¹ | *X* | *Y* | ¹ | *X* | *Y* | ¹ | *X* | *Y* |
| 1 | 57.8 | 17.2 | 9 | 53.9 | 16.1 | 17 | 53.8 | 16.3 | 25 | 50.9 | 14.7 |
| 2 | 54.6 | 17.9 | 10 | 60 | 14.8 | 18 | 53.1 | 17.2 | 26 | 49.6 | 16.1 |
| 3 | 54.8 | 18.8 | 11 | 56.2 | 17 | 19 | 51.5 | 15.8 | 27 | 52.2 | 19.5 |
| 4 | 51.7 | 19.9 | 12 | 55.2 | 17.8 | 20 | 54 | 15 | 28 | 50.5 | 15.6 |
| 5 | 61.1 | 16 | 13 | 53.3 | 19.9 | 21 | 50.4 | 14.4 | 29 | 51.1 | 18.1 |
| 6 | 62.3 | 17.8 | 14 | 57.9 | 17.1 | 22 | 53 | 15.3 | 30 | 52.2 | 19,5 |
| 7 | 52.2 | 18.8 | 15 | 54 | 15.5 | 23 | 53.3 | 16.6 | 31 | 49.2 | 15.7 |
| 8 | 49.2 | 19.3 | 16 | 52.6 | 17.6 | 24 | 51.6 | 14.9 | 32 | 49.3 | 13.2 |

**а) Ввод данных:**

*А.2.File Operations — file name: WORK, Desired operation: С* (операция *Edit* (редактирование) в списке *Operations) − F6 − F6* − В окне *Add additional column* (добавить дополнительные столбцы) вводим имена новых переменных: *Name: x1, Type: N* (тип вещественный), *Width: 13* (или меньше; − число десятичных разрядов) *− F6 − Name: y1 − F6 − ESC −* вводим данные в колонки *x1* и *y1. − F6 − Save and Exit* (запомнить и выйти) − *ENTER* (появляется описание созданных переменных).

**Замечание.** При вводе удобнее вводить числа без десятичной точки, а затем операцией *J* и опцией *A (ASSIGNMENT* − назначение) разделить на 10.

**б) Представление выборки диаграммой рассеяния:**

*E.1.X-Y Line and* *Scatterplots* − вносим в строку *x: x1,* в строку *у: y1 − F6.* Появляется диаграмма рассеяния, которую можно отредактировать *(F5)*: изменять надписи, диапазоны величин по осям, частоту делений и т.д. Диаграмму выводим на печать *(F4).*

**в) Определение выборочных характеристик:**

*F.1.Summary Statistics − Data vectors: x1,* во второй строке: *y1,* *Statistics: A, B, F, H, I, − F6.* Появляется таблица выборочных характеристик для двух переменных *x1* и *x2.*

Определение выборочной ковариационной матрицы: *Q.2.Covariance Analysis* (анализ ковариаций) − *Data vectors: x1*, во второй строке: *y1 − F6.* Появляется матрица ковариаций (в данном случае, 2×2).

выборочная корреляционная матрица определяется процедурой *Q.1.Correlation Analysis.*

**г) Построение двумерной гистограммы:**

*F.7.Three − Dimensional Histogram − Sample 1: x1, Sample 2: y1 − F6 −* Поправляем, если необходимо, параметры графика − *F6.*

 Появляется трехмерный график. Выводим его на печать или сохраняем.

##  Выполнение в пакете STATISTICA

***Генерация выборки***

Сгенерируем, например, выборку объема *n* =50 с показательным распределением со средним значением 5.

Создадим новый файл:

*File - New Data -* укажем имя файла в окне *File Name : descript* (например) - *OK.* На экране сетка-таблица; в ее заголовке указаны название и размеры : *10v \* 10c* - ( 10 переменных ( variables ) - столбцов по 10 наблюдений ( cases ) - строк.

Преобразуем таблицу к размерам 1×50:

кнопка *Vars* (на экране) - *Delete;* окно *Delete Variables:* укажем какие переменные- столбцы убрать : *From variable : var 2, To variable : var 10 - OK* - Кнопка *Cases* - *Add* ( добавление ) - окно *Add Cases:* укажем, сколько строк добавить и куда : *Number of Cases to Add* : *40, Insert after Case : 1* ( например ) - OK.

Сгенерируем выборку:

выделим столбец - переменную *Var1* ( щелчком мыши по ее заглавию) - нажмем правую клавишу - в открывшемся меню выберем *Variable specs* ( спецификации переменной ) - в появившемся окне *Variable 1* введем *Name x* ( например ) , в нижнем поле *Long name* вводится выражение, определяющее переменную. Ввод можно сделать набором на клавиатуре или с помощью клавиши *Functions*, выбирая в меню *Kategory* и *Name* требуемую функцию и вставляя клавишей *Insert*. Для задания закона распределения следует ввести, например,

 *=rnd(2)* для *R*[0, 2],

 *=Vnormal(rnd(1); 2; 0.5 )* для *N(2, σ2=0.52),*

 *=VExpon(rnd(1); 0.2 )* для *E(5)* со средним 1/0.2=5; (для нашего примера вместо значения параметра *λ=*0.2 можно набрать выражение 1/5).

Такая форма задания определяется способом генерации: с помощью функции, обратной (буква *V*) к функции распределения и генератора случайных чисел *R*[0, 1] *( rnd(1))*.

Распечатаем выборку командой *Print* меню *File*.

Посмотрим выборку графически:

*Graphs - Custom Graphs* (настраиваемые графики) - *2D graphs -* в открывшемся окне все можно оставить по умолчанию - .*OK*. Наблюдаемый график (рис.2) распечатаем.

**Рис. 2. Наблюдения, распределенные по показательному закону со средним 5 (n = 50).**

***Построение вариационного ряда***

Первый способ:

выделим требуемую переменную (столбец) - нажмем правую клавишу мыши - выберем *Quiq Stats Graphs* (быстрые статистики и графики) - *Values / Stats of Vars* (значения и статистики ) - наблюдаем вариационный ряд и выборочное среднее (*mean*) и стандартное отклонение ( *SD* ).

Второй способ:

войдем в модуль *Data Menagement* (двойной щелчек левой клавишей мыши на чистом поле и выбор модуля в окне *Module Switcher*; если модуль уже загружен, то *Alt+Tab* до появления модуля) - *Analysis Sort* - устанавливаем имя переменной, тип сортировки: *Ascen* (по возрастанию ) или *Desc* ( по убыванию) - *OK*.

***Функция эмпирического распределения***

Первый способ:

*Graphs - Stats 2D Graphs - Histogram -* в появившемся окне установим: *Graph Type : Regular, Cumulative Counts* (накопленные частоты), *Fit Type* (подбираемый тип) : *Exponential* (для нашего примера) или *off* (без подбора)*, Variablles: x,* *Categories* (число интервалов группирования) : 250 - *OK*.

Наблюдаем график функции эмпирического распределения (рис. 3). График можно отредактировать: изменить линии, точки, фон, шкалы, надписи; для этого необходимо подвести стрелку в нужное иесто и дважды щелкнуть левой клавишей мыши. Выведем его на печать или сохраним.

**Рис.3. Функция эмпирического распределения**

Второй способ:

упорядочим по возрастанию нашу выборку (см. ***Построение вариационного ряда***);

образуем новую переменную *F* для значений функции:

клавиша *Var - Add - ...* ( см. ***Генерация выборки***) - выделим новую переменную *NEWVAR -* правая клавиша мыши - *Variable Specs ... - Name: F - Long name:*

*= V0/50*

(оператор *V0* создает массив целых чисел) ; построим график:

*Graphs - Custom Graphs - 2D Graph -* в новом окне установим: в поле *X: x,* в поле *Y: F, Step Plot* (ступеньки, но не *Line Plot -* линии) *- OK*.

Наблюдаем функцию эмпирического распределения (с точностью до мелкого группирования с 250 интервалами).

***Группирование данных***

*Analysis Frequency Tables -* в окне *Frequency Tables* зададим *No of exact intervals: 10* (10 интервалов группирования; или *Step size: 2, starting at: 0*), в поле *Display options* отметим *Cumulative frequences* ( накопленные частоты ), *Percentages* (проценты - относительные частоты), *Cumulative Percentages* (накопленные частоты ) - *OK.*

Наблюдаем таблицу группированных данных. Выведем ее на печать или сохраним.

***Построение гистограммы частот***

*Graphs - Stats 2D Graphs - Histograms -* в появившемся окнеустанавливаем: имя переменной, *Graph Type: Regular, Fit Type; off* ( без подбора ) или нужный тип, число интервалов группирования *Categories:* или *Auto* (автоматический выбор числа интервалов) - *OK.*

 Наблюдаем гистограмму (рис. 4). Отредактируем график, если необходимо. Выведем на печать или сохраним.

**Рис. 4. Гистограмма.**

***Выборочные характеристики***

первый способ: на заголовке столбца с выборкой щелкнем правой клавишей мыши - *Quick Basic Stats... - Descriptives of var -* получаем таблицу с характеристиками: *mean* (среднее), *Confid 95%*  ( доверительные границы нижняя и верхняя с уровнем доверия 0.95 ), *Sum* ( сумма ), *Minimum, Maximum, Range* ( размах ), *Variance* ( дисперсия ), *Std. Dev.* ( стандартное отклонение ) и др. Сравним выборочное среднее, медиану и стандартное отклонение с соответствующими теоретическими значениями. Это же можно сделать через меню: *Anflisis - Quick Basic Stats ...*

Второй способ: на заголовке столбца с выборкой щелкнем правой клавишей мыши - *Block Stats / Columns* (блок статистик по колонкам ) - выделим необходимое или *All.*

***Описание двумерных выборок***

Ввод данных: зададим новую таблицу 2×32, назовем столбцы *X* и *Y*. Заполним таблицу вручную заданными в табл.2 значениями.

Диаграмма рассеяния:

*Graphs - Stats* 2*D Graphs... - Scatterplots... -* вводим значения по осям *X* и *Y* (нажав на кнопку *Variables* и выбрав переменные ) - *OK.*

Распечатаем диаграмму (рис. 5) или сохраним.

**Рис. 5. Диаграмма рассеяния**

Выборочные характеристики.

Выделим те переменные, по которым требуются выборочные характеристики - щелкнем правой клавишей мыши - *Quick Basic Stats - Descriptivs of VARS...* Наблюдаем таблицу выборочных характеристик (тех же, что иыше). Отпечатаем таблицу или сохраним.

 Выборочные характеристики можно внести в таблицу данных, в конец соответствующих столбцов. Выделим нужные столбцы, далее см. вторую часть п. ***Выборочные характеристики.***.

Определим корреляционную матрицу:

*Analysis - Correlation matrices - Two lists - First list: All - Second list: All - OK - Cancel* (отмена предложения на новую матрицу).

Матрицу отпечатаем или сохраним.

двумерная гистограмма (рис. 6).

*Graphs - Stat 3D Sequential Graphs - Bivariate Gistogram -* установим по осям *X* и *Y* требуемые переменные ( кнопкой *Variables* ), зададим число интервалов по каждой оси - *OK.*

 Распечатаем гистограмму.

**Рис. 6. Двумерная гистограмма.**

## Выполнение в пакете SPSS

Предварительно отметим:

1) диалоги заканчиваются нажатием кнопок *ОК, Define* или *Continue* для исполнения или *Cancel* для отмены;

1. кнопка со стрелкой (треугольником) означает перемещение выбранного элемента из одного списка в другой в направлении стрелки.

***Генерация выборки***

Сгенерируем 2 выборки с заданными законами распределения, например, выборки объема *n* = 50 c нормальным законом распределения со средним 5 и стандартным отклонением 1 и показательным (экспоненциальным) законом со средним 5.

Заготовим таблицу с 2 столбцами и *n* = 50 строками:

на экране таблица с пустыми клетками; прокрутим ее до 50-й строки и выделим клетку во 2-м столбце - введем любой символ, например, точку - *Enter.* Таблица 50 × 2 образована.

Создадим соответствующий файл на диске в *D:\TMP:*

*File - New - Data* - на вопрос *save* ...? отвечаем *Yes* - в окне *Save As Data File : Name : D:\TMP WORK. SAV* (например) - *ОК.*

Присвоим переменным удобные имена *х1* и *х2:*

выделим первый столбец, кликнув мышью по заголовку - *Data - Define Variable*...(определение переменной) - *Variable Name: x1 - OK*.

Аналогично - второй столбец.

Сгенерируем выборку с нормальным распределением:

*Transform* (преобразование) - *Compute* (вычислить) - в поле *Target Variable* (выходная переменная - столбец) введем имя переменной, в которую будет занесен результат: *х1*; в списке *Functions*  выделим *NORMAL (stddev)* (*standart deviation* - стандартное отклонение), перенесем в поле *Numeric Expression* (вычисляющее выражение): *NORMAL (1) + 5 - OK - Change...? - OK*.

Сгенерируем в *х2* выборку с показательным распределением со средним 5: действия аналогичны предыдущим, однако, *Numeric Expression*:

*- 5\* LN (UNIFORM (1)),*

поскольку случайная величина - *a lnξ,,* где ξ ~ *R* [0, 1], имеет показательное распределение со средним *а*; оператор *UNIFORM (x)* генерирует равномерно на [0, *x*] распределенные случайные числа.

Посмотрим выборку графически:

*Graphs - Line* - выберем *Simple* (простой), в поле *Data in Chart Are* (данные для графика) выберем *Values of individual cases* (значения отдельных наблюдений) - *Define* - â появившемся окне *Define Simple* из левого списка переместим кнопкой - стрелкой *х1* в поле *Line Represent - OK*.

Наблюдаем график; его можно отредактировать (кнопка *Edit*); сохраним его:

*File - Save As - Name: Fig1.cht*. (например) - *ОК*

или распечатаем: *File -Print...*

Посмотрим выборку *х2*: действия аналогичны.

Сохраненные графики можно посмотреть еще раз:

*File - Open - Chart* ... - в поле *Files:* выделим *Fig1 - OK.*

***Построение вариационного ряда***

*Data - Sort Cases...* - в поле *Sort by: x1* (переносом из левого списка), в поле Sort Orden: Ascending (возрастание, в отличие от *Descending* - сортировка по убыванию) - *ОК.* Сортировка проводится по указанной переменной - столбцу *х1*, но сразу для всех столбцов.

***Построение графика функции эмпирического распределения***

Сначала построим график для выборки *х1:*

*Statistics - Summarize - Frequencies*...- в поле *Variable(s): x1*, отметим *Display frequency tables* (показ таблицы частот) - *ОК*. В окне *Output* (выход - окно результатов и протокола работы) появляется таблица, первый столбец которой *Value* (значение) - вариационный ряд, пятый - *Cum. Percent* (накопленные частоты в процентах) - соответствующие значения функции эмпирического распределения в процентах. Переносим столбец *Value* в таблицу *WORK:*

выделяем столбец *Value* (если в столбце *Value* имеются пустые клетки, следует выделить соответствующие строки и их удалить) - *Edit - Copy* (копирование в буфер) - выделяем в таблице *WORK* первый справа свободный столбец: *Var - Edit - Past* (вставить) - получаем новый столбец с вариационным рядом. Для удобства присвоим ему имя *х1v* (например) (выполнение см. выше).

Аналогично переносим столбец *Cum. Percent* и назовем его *F* (например). Строим график: *Graphs - Scatter...- Simple - Define - Yaxis: F, X* *Axis: x1v - OK - Edit* - кнопка \* в окне *Markers* (метки) выберем точку , *Apply All* - закроем окно *Markers* - кнопка и линии (в виде зигзага) - выберем *Left step* (левые ступеньки), *Apply All - Close.*

Наблюдаем функцию эмпирического распределения; сохраняем график или распечатываем.

Аналогично строим функцию эмпирического распределения для выборки *х2*. Сравниваем эти две функции.

***Построение гистограммы частот***

Построим гистограмму для выборки *х1: Graphs - Histogram...- Variable:* *x1 - OK*. Наблюдаем гистограмму; сохраним ее или распечатаем.

Аналогично - для *х2*.

***Определение выборочных характеристик***

*Statistics - Summarize - Descriptives...- Variable(s): x1, x2,* убираем выделения внизу *- Options* - отмечаем нужное: *Mean, Sum, Std. Deviation* (стандартное отклонение), *Range* (размах), *Minimum, Maximum - Continue - OK.*

Наблюдаем таблицу, в которой показаны отмеченные характеристики для обеих выборок. Выделяем таблицу и сохраняем ее:

*File - Save As - Name: Descr. Lst* (например) - *ОК*. Сравниваем выборочные средние и стандартные отклонения с теоретическими.

***Проверка гипотезы о типе распределения***

Проверим обе наши выборки с помощью критерия Колмогорова - Смирнова на нормальность распределения и равномерность:

*Statistics - Nonparametric Tests - 1 Sample K - S* - в поле *Test Variable List:* *x1, x2* (переносом из списка слева), в поле *Test Destribution* отметим *Normal,* *Uniform - OK.*

В окне *Output* даются результаты тестирования двух выборок по двум гипотезам: итого 4 сообщения. Например, результат тестирования *х1* на нормальность (*Test distribution - Normal*): приводятся параметры гипотетического распределения (оценки) *Mean* è *Standart Deviation*; статистика *Dn* Колмогорова (*Most estreme differences Absolute*), *z = Dn  (K – S Z)* и уровень значимости *2 – Tailed P*; если последний порядка сотых долей или меньше, гипотезу следует отклонить.

Выписываем упомянутые значения и делаем выводы.

Заметим, что такой способ проверки при отклонении гипотезы можно считать корректным, а при принятии - это не совсем так (см. более подробные руководства по статистике).

***Описание двумерных выборок***

а) Ввод данных: в свободные два столбца введем с клавиатуры данные из табл. 2; назовем их *x* и *y.*

б) Диаграмма рассеяния:

*Graphs - Scatter...- Simple - Define - X Axis: x, Y Axis: y - OK*.

 Наблюдаем диаграмму; сохраним ее или распечатаем.

в) Выборочные характеристики.

Некоторые характеристики см выше.

Определение корреляционной матрицы:

*Statistics - Correlate - Bivariate* - в поле *Variables: x, y* (переносом из левого списка), *Correlation Coefficients*: *Pearson Test of Significance: Two - tailed* (двусторонний тест Пирсона на значимость отличия от нуля) - *ОК*.

В окне *Output* имеем таблицу 2 × 2 коэффициентов корреляции и уровней значимости *Р*; если *Р* порядка сотых долей или меньше, гипотезу о нулевом значении коэффициента следует отклонить. Если *Р* порядка 0.1 или более, коэффициент корреляции следует считать нулевым. Матрицу выделяем и сохраняем ее или распечатываем.

г) Трехмерная диаграмма.

Для примера образуем третью переменную (столбец) *z*, равную *x + y.*

Построим диаграмму:

*Graphs - Scatter...- 3D - Define - X Axis: x, Y Axis: y, Z* *Axis: z - OK*.

Наблюдаем трехмерную диаграмму. Будем изменять точку обозрения: *Edit - Spin* - вращаем трехмерную совокупность с помощью 6 кнопок, находим удачную точку - *End Spin*.

Сохраняем рисунок или распечатываем.