

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
Бийский технологический институт (филиал)
федерального государственного бюджетного образовательного
учреждения высшего профессионального образования
«Алтайский государственный технический университет
им. И.И. Ползунова»

А.В. Фролов

**ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА
И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ПРОГРАММЕ
STATISTICA**

Методические рекомендации к выполнению лабораторной работы
по дисциплинам «Статистические методы в управлении качеством»,
«Квалиметрия», «Управление процессами»
для студентов технических вузов
направления подготовки 27.03.02 «Управление качеством»
различных форм обучения

Бийск
Издательство Алтайского государственного технического
университета им. И.И. Ползунова
2015

УДК 004.9(076)

Рецензент: А.Ю. Козлюк, к. т. н., доцент кафедры ПБиУК
БТИ АлтГТУ

Фролов, А.В.

Законы распределения показателей качества и проверка статистических гипотез в программе Statistica: методические рекомендации к выполнению лабораторной работы по дисциплинам «Статистические методы в управлении качеством», «Квалиметрия», «Управление процессами» для студентов технических вузов направления подготовки 27.03.02 «Управление качеством» различных видов подготовки и форм обучения / А.В. Фролов; Алт. гос. техн. ун-т, БТИ. – Бийск: Изд-во Алт. гос. техн. ун-та, 2015. – 20 с.

Методические рекомендации содержат основные сведения о законах распределения показателей качества и проверке статистических гипотез в программе Statistica. Анализ гипотез является статистической процедурой для проверки обоснованности гипотезы, рассматривающей параметры качества одной или нескольких выборок с определенным уровнем доверия. Такая проверка дает возможность сделать вывод о верности гипотезы на основании рассмотрения параметров совокупности данных о качестве продукции с определенным уровнем доверия.

Методические рекомендации предназначены для студентов технических вузов направления подготовки бакалавров 27.03.02 (221400.62), изучающих курсы «Статистические методы в управлении качеством», «Квалиметрия», «Управление процессами».

Данные методические рекомендации написаны с учетом формирования у студентов общекультурных и профессиональных компетенций в процессе изучения дисциплины «Статистические методы в управлении качеством» в соответствии с требованиями Федерального государственного образовательного стандарта высшего профессионального образования.

УДК 004.9(076)

Рассмотрены и одобрены
на заседании кафедры ПБиУК.
Протокол №12/14 от 20.11.2014 г.

© Фролов А.В., 2015
© БТИ АлтГТУ, 2015

СОДЕРЖАНИЕ

1 Вычисление вероятностей и моделирование распределений случайных величин.....	4
1.1 Вычисление вероятностей для дискретных случайных величин	4
1.2. Вычисление вероятностей и квантилей для непрерывных случайных величин.....	6
1.3 Моделирование распределений случайных величин.....	7
2 Проверка статистических гипотез.....	9
3 Однофакторный дисперсионный анализ	14
4 Порядок выполнения работы (4 часа).....	17
5 Контрольные вопросы.....	18
Список использованных источников	19

1 ВЫЧИСЛЕНИЕ ВЕРОЯТНОСТЕЙ И МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН

1.1 Вычисление вероятностей для дискретных случайных величин

В пакете STATISTICA для вычисления вероятностей биномиального распределения используются две функции:

- **Binom** ($x; p; n$) – вероятность того, что случайная величина X , имеющая биномиальное распределение с параметрами n и p , примет значение x , т. е. $P[X = x]$;

- **IBinom** ($x; p; n$) – суммарная накопленная вероятность, $P[X \leq x]$ (аналог функции распределения в точке x).

Чтобы выполнить вычисления с помощью этих функций и ввести результаты вычислений в какую-либо переменную (например, Var1), нужно открыть окно спецификации Var1 (**Variable Specs...**) и в поле **long name** записать функцию **=Binom** ($x; p; n$) или **=IBinom** ($x; p; n$) и задать вместо x , p и n необходимые значения.

Пример 1.1. Пусть случайная величина X имеет биномиальное распределение с параметрами $n = 5$ и $p = 0,3$. Вычислим вероятности событий: $P[X = 2]$ и $P[X \leq 4]$.

Решение. Установите курсор на Var1 и нажмите правую кнопку мыши. В открывшемся меню выберите **Variable Specs...** (рисунок 1.1).

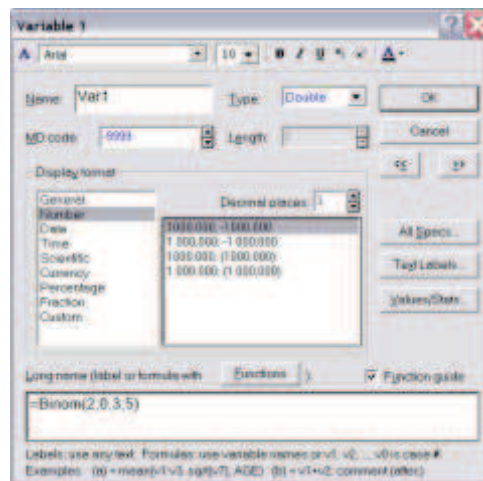


Рисунок 1.1 – Окно Variable Specs... для Var1

В поле **long name** введите функцию **=Binom (2; 0,3; 5)**, **ОК**. На экране появляется сообщение: Expression OK. Recalculate the Variable now? (Формула записана верно. Пересчитать переменную сейчас?), да. Поля переменной Var1 заполняются числами 0,309 = $P[X = 2]$. Аналогично, введя в поле **long name** функцию **=IBinom(4; 0,3; 5)**, получим $P[X < 4] = 0,998$.

В поле **long name** можно вставить соответствующую формулу, выбрав ее из списка после нажатия кнопки **FUNCTIONS** (функции) в окне **Variable Specs...** (см. рисунок 1.1) и нажав кнопку **Insert**. Список функций показан на рисунке 1.2.

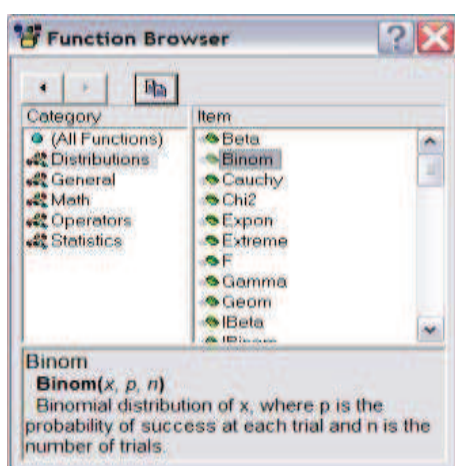


Рисунок 1.2 – Список функций

Если в окне списка функций (см. рисунок 1.2) нажать кнопку **SYNTAX**, то появится текст, поясняющий правила для записи формул и назначение функций. Функции, используемые в вероятностных расчетах, поясняются в дополнительном тексте, который появляется, если щелкнуть мышью по **Distribution functions and their integrals** (функции распределения и их интегралы).

Аналогично вычисляются вероятности и рассматриваются свойства распределения Пуассона (функция **Poisson (κ; λ)**)

$$\text{Poisson}(\kappa; \lambda) = \frac{\lambda^\kappa}{\kappa!} e^{-\lambda} = P[X = \kappa], \quad \kappa = 0, 1, 2, \dots$$
 и геометрического распределения (функция **Geom (x; p)**). $\text{Geom}(\kappa; p) = P[X = \kappa] = (1 - p)^{\kappa-1} p, \quad \kappa = 0, 1, 2, \dots$

С помощью функций **IPoisson**($x; \lambda$) и **IGeom**($x; p$) вычисляются соответствующие накопленные вероятности: $P[X \leq x]$.

1.2. Вычисление вероятностей и квантилей для непрерывных случайных величин

Для непрерывных распределений, например, нормального вычисляются три функции:

- **Normal** ($x; \mu; \sigma$) – значение функции плотности нормального распределения $N(\mu, \sigma)$ в точке x , где μ – математическое ожидание, σ^2 – дисперсия, а σ – среднеквадратическое (стандартное) отклонение;

- **Inormal** ($x; \mu; \sigma$) – функция распределения нормального распределения $N(\mu, \sigma)$ в точке x ;

- **Vnormal** ($x; \mu; \sigma$) – значение функции обратной к функции **Inormal** ($x; \mu; \sigma$) в точке x . Обратная функция к функции распределения используется для вычисления квантилей и моделирования распределений случайных величин.

Квантиль распределения случайной величины X порядка p [$0 < p < 1$] будет обозначаться x_p . Для непрерывной случайной величины X , x_p есть решение уравнения: $F(x_p) = p$ или $P[X < x_p] = p$.

Таким образом, **Vnormal** ($p; 0; 1$) = x_p , где x_p – квантиль порядка p стандартного нормального распределения $N(0, 1)$ с математическим ожиданием $\mu = 0$ и среднеквадратическим отклонением $\sigma = 1$. Например, **Vnormal** ($0,95; 0; 1$) = 1,645.

Аналогичным образом задаются функции плотности, функции распределения и функции, обратные к функциям распределения для других непрерывных случайных величин. Приведем список обозначений функций плотности для распределений, которые будут использоваться в дальнейшем изложении.

Expon ($x; \lambda$) – плотность экспоненциального распределения с параметром λ ;

IExpon ($x; \lambda$) – функция распределения экспоненциального распределения с параметром λ ;

VExpon ($p; \lambda$) – значение функции обратной к функции **IExpon**;

Chi2 ($x; k$) – плотность распределения χ^2 с k степенями свободы;

IChi2 ($x; k$) – функция распределения χ^2 с k степенями свободы;

VChi2 ($p; k$) – значение функции, обратной к функции **IChi2**;

F ($x; k_1, k_2$) – плотность распределения Фишера с k_1 и k_2 степенями свободы;

IF ($x; k_1; k_2$) – функция распределения Фишера с k_1 и k_2 степенями свободы;

VF ($p; k_1; k_2$) – значение функции, обратной к функции **IF**;

Student ($x; k$) – плотность распределения Стьюдента с k степенями свободы;

IStudent ($x; k$) – функция распределения Стьюдента с k степенями свободы;

VStudent ($p; k$) – значение функции, обратной к функции **IStudent**.

1.3 Моделирование распределений случайных величин

В задачах статистического анализа сложных систем, например при разработке систем автоматического проектирования (САПР), исследовании систем массового обслуживания, широко используется метод моделирования выборки из генеральной совокупности с заданным законом распределения.

Пусть случайная величина X имеет функцию распределения $F(x)$. Как известно из теории вероятностей, случайная величина $Y = F(X)$ имеет равномерное распределение $R(0, 1)$.

Отсюда следует, что случайная величина X может быть получена из равномерно распределенной случайной величины Y по формуле $X = F^{-1}(Y)$, где F^{-1} – функция, обратная к F (заведомо существующая для случайных величин непрерывного типа).

Таким образом, для моделирования выборки из непрерывного распределения с функцией распределения $F(x)$ нужно сначала получить выборку из генеральной совокупности, имеющей равномерное распределение $R(0, 1)$, а затем использовать функцию, обратную к функции распределения $F(x)$, соответствующей случайной величины. В пакете **STATISTICA** распределение $R(0, 1)$ моделируется с помощью функции: **=rnd(1)**.

Для примера смоделируем выборку из равномерного распределения $R(0, 1)$ и запишем результат в переменной **Var4**. Курсор на поле **Var4** → щелчок правой кнопкой мыши → **Variable Specs...** → в поле **long name** введем формулу: **=rnd(1)** → **OK**.

После выполнения пересчета 200 значений переменной **Var4** будут заполнены числами, представляющими случайную выборку наблюдений из генеральной совокупности, имеющей равномерное распределение $R(0, 1)$. Чтобы получить выборку из нормального распределения $N(m = 7; \sigma^2 = 4)$ и записать ее в переменную **Var5**, нужно в поле **long name** переменной **Var5** записать формулу: **=Vnormal (V4; 7; 2)**.

Можно в качестве аргумента функции **Vnormal** сразу записать **rnd(1)**, тогда соответствующая формула будет: **=Vnormal (rnd(1); 7; 2)**.

Аналогично моделируются выборки для любого непрерывного распределения.

2 ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Критерий Стьюдента (*t-критерий*) является наиболее часто используемым методом проверки статистической гипотезы о равенстве средних двух выборок.

Статистической гипотезой называется любое предположение о виде или параметрах некоторого закона распределения. Проверяемую гипотезу обычно называют нулевой и обозначают H_0 (например, H_0 : между строками и столбцами таблицы нет зависимости; коэффициент корреляции равен 0; средние некоторого показателя в двух выборках равны, закон распределения признака соответствует нормальному закону и т. д.). Наряду с нулевой гипотезой рассматривают альтернативную или конкурирующую гипотезу H_1 , являющуюся логическим отрицанием H_0 (например, H_1 : между строками и столбцами таблицы есть зависимость; коэффициент корреляции не равен 0; средние некоторого показателя в двух выборках не равны, закон распределения признака не соответствует нормальному закону и т. д.). Эти гипотезы представляют собой две возможности выбора в задачах статистической проверки гипотез.

Уровнем значимости критерия называется вероятность α совершить ошибку 1-го рода, т. е. отвергнуть гипотезу H_0 , когда она верна.

Мощностью критерия равной $(1 - \beta)$, называется вероятность не допустить ошибку 2-го рода. Вероятности α , β однозначно определяются выбором критической области. Очевидно, желательно сделать α , β как угодно малыми, однако это противоречивые требования. Лишь при увеличении объема выборки возможно одновременное уменьшение вероятностей α , β .

Уровень значимости p – это максимально приемлемая для исследователя вероятность ошибочно отклонить нулевую гипотезу, когда на самом деле она верна, т. е. допускаемая вероятность ошибки первого рода. Величина уровня значимости устанавливается исследователем произвольно, однако обычно принимается равным 0,05, либо 0,01, либо 0,001. В программе *STATISTICA* приемлемой границей статистической значимости приняты значения p , меньшие либо равные 0,05. Если p меньше либо равно 0,05, то результат считается статистически значимым, если p меньше либо равно 0,01, то результат считается статистически высоко значимым.

Теоретически *t*-критерий может применяться, даже если размеры выборок очень небольшие, переменные нормально распределены (внутри групп), а дисперсии наблюдений в группах не слишком различны. Предположение о нормальности можно проверить, исследуя распределение (например, визуально с помощью гистограммы) или применяя какой-либо критерий нормальности. Равенство дисперсий в двух группах можно проверить с помощью *F*-критерия. Уровень значимости *p* *t*-критерия равен вероятности ошибочно отвергнуть гипотезу о равенстве средних двух выборок, когда в действительности эта гипотеза имеет место.

Для запуска программы в верхнем меню **Statistics** надо выбрать команду **Basic Statistic Tables** (основные статистики/таблицы). Откроется меню команды, в котором *t*-критерий представлен четырьмя процедурами:

- **t-test, independent, by variables** (*t*-критерий для независимых выборок) применяется, если надо сравнить средние случайных величин, полученных по двум разным (независимым) выборкам;

- **t-test, independent, by groups** (*t*-критерий для независимых выборок с группирующей переменной) используется, если надо сравнить средние случайных величин двух независимых групп, полученных из одной выборки при помощи группирующей переменной;

- **t-test, dependent samples** (*t*-критерий для зависимых выборок) применяется, если надо сравнить средние случайных величин двух зависимых групп;

- **t-test, single samples** (*t*-критерий простые выборки).

В перечисленных процедурах в качестве нулевой гипотезы предполагается, что средние в группах равны.

Рассмотрим работу процедуры **t-test, dependent samples**, используя таблицу данных на рисунке 2.1, в которой приведены данные пробега пятнадцати японских и европейских автомобилей. В столбце 2 указан тип топлива: *P* – бензин; *G + P* – бензин и газ; *D* – дизель. В столбцах 3, 4, 5 приведены пробеги автомобилей до первой серьезной поломки, требующей ремонта на СТО – *Пробег1*; до капитального ремонта двигателя – *Пробег2*; после капитального ремонта – *Пробег3*.

Предположим, что величины пробега автомобилей в столбцах таблицы имеют нормальное распределение. На основании этих данных нужно определить, существенно ли отличаются средние величины пробега автомобилей до капитального ремонта и после капитального

ремонта двигателей; средние величины пробега в зависимости от типа используемого топлива и места производства.

	Пробег до кап. ремонта, после кап. ремонта				
	1	2	3	4	5
	Пробег	Тип топлива	Пробег1	Пробег2	Пробег3
Opel Astra	Europe	P	65	240	230
Skoda Fabia 1.2	Europe	P	70	250	220
Mitsubishi Pajero	Japan	G+P	110	300	280
Skoda Ambiente 1.6	Europe	P	60	230	230
Nissan Almera 1.5	Japan	G+P	90	280	260
Nissan Maxima 2.0 GX	Japan	G+P	100	300	280
Audi A4 2.0 MultiTronic	Europe	P	80	250	230
Nissan Maxima 3.0 SE	Japan	P	110	310	310
Mitsubishi Pajero III	Japan	G+P	95	320	280
Toyota Corolla	Japan	G+P	100	300	300
Toyota Carina	Japan	D	110	310	300
VW Passat 1.8T	Europe	D	70	275	250
VW Bora 1.6	Europe	D	80	260	230
Subaru Legacy	Japan	D	105	315	350
VW Golf 1.6	Europe	D	75	250	240

Рисунок 2.1 – Исходные данные

Так как сравниваемые группы основываются на одной и той же совокупности наблюдений (случаев), тестирувавшихся три раза, необходимо использовать для анализа *t-критерий* для зависимых выборок.

После выбора команды **t-test, dependent samples** в открывшемся окне процедуры (рисунок 2.2) нажмите кнопку **Variables**. Выберите две сравниваемые между собой переменные: *Пробег2*, *Пробег3*, нажмите ОК. Программа вернется в диалоговое окно модуля, в котором надо нажать на кнопку **Summary: t-test**.

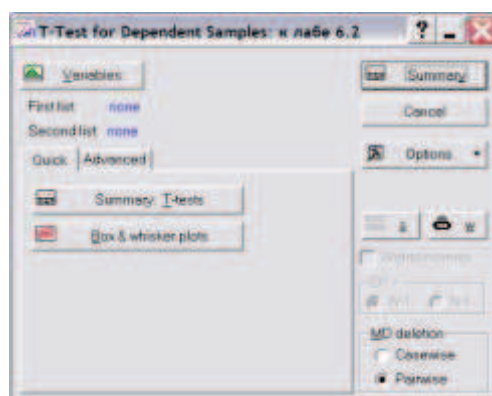


Рисунок 2.2 – Окно процедуры t-test, dependent samples

Появится таблица с результатами анализа (рисунок 2.3), в которой будут приведены значения следующих статистик:

- *Mean* (средние значения переменных);
- *Std. Dev.* (стандартные отклонения значений переменных);
- *N* (число наблюдений в группе);
- *Diff* (разница между средними);
- *Std. Dev. Diff* (стандартное отклонение от среднего значения);
- *t* (значение критерия Стьюдента);
- *F* (значение критерия Фишера);
- *df* (число степеней свободы);
- *p* (*p*-уровень значимости *t*-критерия), вероятность отклонения гипотезы.

Для нашего примера *p* меньше 0,05, поэтому гипотезу о равенстве средних отвергаем. Таким образом, средний пробег всех автомобилей до капитального ремонта значительно превышает средний пробег после капитального ремонта.

Variable	T-test for Dependent Samples (к лабе 6.2)							
	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Пробег2	279,3333	30,52322						
Пробег3	266,0000	37,94733	15	13,33333	17,89520	2,885678	14	0,011976

Рисунок 2.3 – Результаты анализа t-test, dependent samples

STATISTICA позволяет применять *t*-критерий для независимых выборок, используя одну независимую (группирующую) переменную (например, место производства) и одну зависимую переменную (например, пробег автомобиля).

Допустим, необходимо проверить, равны ли средние пробега автомобилей до первой серьезной поломки, до капитального ремонта двигателя и после капитального ремонта для автомобилей европейского и японского производства. В меню процедуры **Basic Statistic Tables** выберите команду **t-test, independent, by groups**. Откроется рабочее окно процедуры. Нажмите кнопку **Variables** и определите группирующую переменную *Произв.* и зависимые переменные *Пробег1*, *Пробег2*, *Пробег3*.

При необходимости ввода текстовой информации имеется кнопка **text label** (текстовые значения) для вызова *Диспетчера текстовых значений*. Диспетчер текстовых значений работает следующим образом. В графу «Текст» следует занести текстовые *варианты* переменной, а в графу «Число» – для них *коды* (произвольные сочетания цифр). При формировании таблицы с исходными данными с клавиатуры следует печатать не текстовые символы, а коды, тогда на экране автоматически будут появляться текстовые значения переменной.

После нажатия на ОК программа вернется в рабочее окно модуля; укажите в нем коды группирующей переменной *Europe* и *Japan*. Щелкните кнопкой **Summary: t-tests**, откроется таблица с результатами анализа (рисунок 2.4). По данным таблицы можно сделать вывод, что средние отличаются существенно. Вывод статистически достоверен, так как верна гипотеза о равенстве дисперсий (*p Variances* значительно больше, чем 0,05).

T-tests, Grouping, Пробега (к лабе 6.2)									
Group 1: Europe									
Group 2: Japan									
Variable	Mean Europe	Mean Japan	t-value	df	p	Std Dev Europe	Std Dev Japan	F-ratio Variances	p
Пробег1	71.4296	102.5000	-7.98944	13	0.000002	7.48013	7.55929	1.021277	0.995853
Пробег2	250.7143	304.3750	-7.80851	13	0.000003	14.28785	12.37437	1.329446	0.711739
Пробег3	232.9571	295.0000	-5.71278	13	0.000071	9.51180	27.25541	8.210526	0.000368

Рисунок 2.4 – Результаты анализа t-test, independent, by groups

Если предположить, что значения величин пробега в столбцах *Пробег2* и *Пробег3* получены по разным выборкам (тестировались различные группы автомобилей), для сравнения средних можно применить процедуру **t-test, independent, by variables**. После выбора этой команды откроется рабочее окно модуля. Укажите имена анализируемых переменных и щелкните по ОК. По данным таблицы результатов (рисунок 2.5) можно сделать вывод, что верна гипотеза о равенстве средних, при этом также верна гипотеза о равенстве дисперсий.

T-test for Independent Samples (к лабе 6.2)									
Note: Variables were treated as independent samples.									
Group 1 vs. Group 2	Mean Group 1	Mean Group 2	t-value	p	Std Dev Group 1	Std Dev Group 2	F-ratio Variances	p	
Пробег2 vs. Пробег3	279.3333	266.0000	1.060370	0.298038	30.52322	37.94733	1.545617	0.425377	

Рисунок 2.5 – Результаты анализа t-test, independent, by variables

3 ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

Для сравнения средних в более чем двух группах необходимо воспользоваться модулем дисперсионного анализа *ANOVA*.

Модуль **Breakdown & one-way ANOVA** (группировка и однофакторный дисперсионный анализ ANOVA) определяет внутригрупповые описательные статистики и корреляции для зависимых переменных в каждой из нескольких групп, определенных одной или большим числом группирующих (независимых) переменных. Сравнивает средние и определяет, в каких именно группах средние отличаются между собой. В качестве нулевой гипотезы предполагается, что средние в группах равны.

Работу данного модуля проследим на примере данных из таблицы, приведенной на рисунке 2.1. Выделите в модуле **Basic statistic Tables** команду **Breakdown & one-way ANOVA**. Откроется рабочее окно команды. Нажмите кнопку **Variables 1** и выберите **Grouping variables** (группирующие переменные) *Произв.* и *Tun. топл.* и **Dependent variables** (зависимые переменные) *Пробег1*, *Пробег2*. Нажмите ОК и, вернувшись в исходное окно, щелкните кнопкой **Codes for grouping variables** (коды для группирующих переменных). Выберите коды для группирующих переменных в диалоговом окне **Select codes for indep. vars (factors)** (коды для независимых факторов).

Чтобы выбрать все коды переменной, можно либо ввести номера кодов в соответствующем поле ввода, либо нажать на кнопку **All** (все), либо поставить * на соответствующем поле ввода. Нажатие ОК без задания каких-либо значений эквивалентно определению всех значений для всех переменных. Если перед выбором кодов необходимо посмотреть значения переменной, нажмите кнопку **Zoom** (информация), которая откроет окно **Value/Stats** (значение/статистики). В этом окне будет выведен отсортированный список значений переменной (при этом будут отображаться все значения независимо от условий выбора наблюдений).

Нажмите ОК в диалоговом окне **Statistics by Groups (Breakdown)**. Откроется новое диалоговое окно (рисунок 3.1) **Statistics by Groups-Results** (внутригрупповые, описательные статистики – результаты), которое предоставляет различные процедуры и настройки для анализа данных внутри групп. Цель такого анализа – лучшее понимание различий между группами. Информационная часть окна сообщает, что зависимых – две переменные: *Пробег1*, *Пробег2*; группирующих – две переменные: *Произв.* с двумя кодами (*Europe, Japan*) и *Tun. топл.* с тремя кодами» (*P, G + P, D*). На рисунке активизирована вкладка **Quick**. На ней находятся следующие кнопки:

- **Summary: Table of statistics** – итоговая таблица средних;
- **Detailed two-way tables** – подробные двухходовые таблицы;
- **Analysis of Variance** – дисперсионный анализ;
- **Interaction plot** – графики взаимодействий;
- **Categorized box & whisker plot** – категоризованные диаграммы размаха. Щелкните кнопкой **Summary: Table of statistics**, появится таблица результатов (рисунок 3.2). В приведенной таблице имеются описательные статистики для выбранных переменных, разбитых на группы. Так, в столбцах 3 и 5 показаны средние (*means*) переменных *Пробег1* и *Пробег2*, в столбцах 4 и 7 – количество автомобилей, в столбцах 5, 8 – среднеквадратические отклонения (*Std.Dev*).

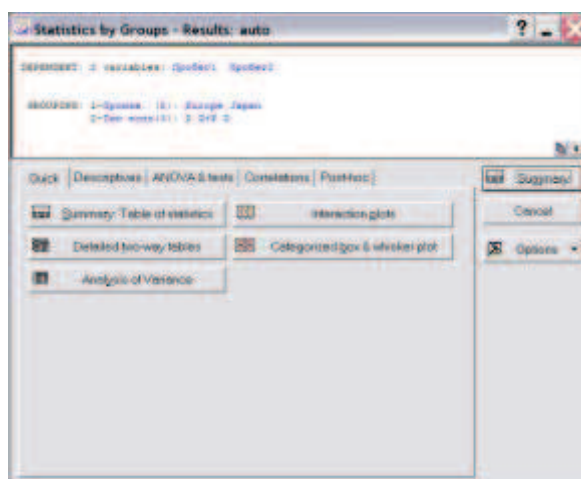


Рисунок 3.1 – Диалоговое окно

Breakdown Table of Descriptive Statistics (auto)							
N=15 (No missing data in dep. var. list)							
Против.	Тип топл.	Пробег1 Means	Пробег1 N	Пробег1 Std Dev	Пробег2 Means	Пробег2 N	Пробег2 Std Dev
Europe	P	68,7500	4	8,53913	242,5000	4	9,57427
Europe	G+P		0			0	
Europe	D	75,0000	3	5,00000	261,6667	3	12,58306
Japan	P	110,0000	1	0,00000	310,0000	1	0,00000
Japan	G+P	99,0000	5	7,41620	300,0000	5	14,14214
Japan	D	107,5000	2	3,53553	312,5000	2	3,53553
All Groups		88,0000	15	17,60682	279,3333	15	30,52322

Рисунок 3.2 – Таблица результатов

Для проверки значимости различий в средних указанных групп надо использовать процедуру **Analysis of Variance** (анализ дисперсий). Щелкните кнопкой **Analysis of Variance** на вкладке **ANOVA & tests**. Откроется таблица результатов Analysis of Variance (рисунок 3.3). Из таблицы видно, что можно отвергнуть гипотезу о равенстве средних переменных *Пробег1*, *Пробег2* в группах. Так как число групп более двух, то из таблицы не видно, какие группы вызвали значительное отличие средних. Процедура **Post-hoc** (апостериорные сравнения средних) позволяет устранить этот недостаток.

Analysis of Variance (auto)								
Marked effects are significant at p < .05000								
Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
Пробег1	3638.75	4	909.688	501.250	10	50.1250	19.14509	0.000111
Пробег2	11639.17	4	2909.792	1464.167	10	146.4167	20.72255	0.000679

Рисунок 3.3 – Таблица результатов Analysis of Variance

Следует выделить вкладку **Post-hoc**, на которой представлены различные апостериорные процедуры. Для проверки можно назначить *p*-level for highlighting (*p*-уровень значимости для выделения). Далее используем процедуру **LSD test or planned comparison** (критерий наименьшей значимости (НЗР)). Появится таблица, состоящая из вероятностей (рисунок 3.4).

LSD Test: Variable: Пробег1 (auto)							
Marked differences are significant at p < .05000							
Произв	Тип топли	(1)	(2)	(3)	(4)	(5)	(6)
		M=68.750	M=0.0000	M=75.000	M=110.00	M=99.000	M=107.50
Europe P (1)				0.274618	0.000395	0.000081	0.000087
Europe G+P (2)							
Europe D (3)		0.274618			0.001908	0.000019	0.000515
Japan P (4)		0.000395		0.001688		0.186500	0.778993
Japan G+P (5)		0.000081		0.000015	0.186500		0.181823
Japan D (6)		0.000087		0.000515	0.778993	0.181823	

Рисунок 3.4 – Таблица вероятностей

Если вероятность, стоящая в таблице на пересечении строки и столбца с соответствующими номерами групп, больше чем 0,05, то гипотезу о равенстве средних этих групп принимаем, в противном случае – отвергаем. Из таблицы видно, что верна гипотеза о равенстве средних в группах: 1,2; 1,3; 2,3; 4,5. Не верна гипотеза о равенстве средних в группах: 4,1; 4,2; 4,3; 5,1; 5,2; 5,3.

4 ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ (4 ЧАСА)

Цель работы: освоение методик вычисления вероятностей и моделирования распределений дискретных и непрерывных случайных величин, являющихся параметрами качества процессов, и методик проверки статистических гипотез и дисперсионного анализа в программе Statistica.

Выполнение работы:

- изучить технологии работ по расчету вероятностей случайных величин и моделированию распределений случайных величин;
- изучить технологии работ по методикам проверки статистических гипотез и дисперсионного анализа;
- каждой подгруппе получить у преподавателя исходные данные для выполнения заданий;
- выполнить задания в электронном виде, используя ПО «Statistica»;
- ответить на контрольные вопросы.

Задание 1. Рассчитать значения вероятностей для заданных исходных данных.

Задание 2. Сформировать массив случайных величин, используя моделирование распределений дискретных и непрерывных случайных величин.

Задание 3. Построить таблицу квантилей, используя функции распределений.

Задание 4. Проверить статистические гипотезы о равенстве средних значений двух нормальных распределений для зависимых выборок, независимых выборок и для одной выборки с группирующей переменной.

Задание 5. Провести однофакторный дисперсионный анализ.

5 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Перечислите основные законы распределения дискретных случайных величин и назовите соответствующие им функции.
2. Перечислите основные законы распределения непрерывных случайных величин и назовите соответствующие им функции.
3. Каким образом найти квантиль распределения?
4. Каким образом можно смоделировать ряд случайных величин, подчиняющихся определенным распределениям?
5. Какие типы статистических гипотез позволяет проверять программа Statistica?
6. Перечислите основные критерии, связанные с проверкой статистических гипотез и их роль в проверке.
7. В чем заключаются ошибки первого и второго рода?
8. В чем заключается задача дисперсионного анализа?

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Халафян, А.А. STATISTICA 6. Статистический анализ данных: учебник / А.А. Халафян. – М.: ООО «Бином-Пресс», 2008 – 512 с.
2. Вуколов, Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel: учебное пособие / Э.А. Вуколов. – М.: ФОРУМ: ИНФРА-М, 2004. – 464 с.
3. Клячкин, В.Н. Статистические методы в управлении качеством: компьютерные технологии: учебное пособие / В.Н. Клячкин. – М.: Финансы и статистика, 2007. – 304 с.

Учебное издание

Фролов Александр Валериевич

**ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА
И ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ПРОГРАММЕ
STATISTICA**

Методические рекомендации к выполнению лабораторной работы
по дисциплинам «Статистические методы в управлении качеством»,
«Квалиметрия», «Управление процессами»
для студентов технических вузов
направления подготовки 27.03.02 «Управление качеством»
различных форм обучения

Редактор Малыгина И.В.
Технический редактор Сидоренко О.В.
Подписано в печать 17.02.15. Формат 60×84 1/16.
Усл. печ. л. 1,16. Тираж 3 экз. Заказ 2015-25.
Печать – ризография, множительно-копировальный
аппарат «RISO EZ300».

Издательство Алтайского государственного
технического университета им. И.И. Ползунова.
656038, г. Барнаул, пр-т Ленина, 46.

Оригинал-макет подготовлен ИИО БТИ АлтГТУ.
Отпечатано в ИИО БТИ АлтГТУ.
659305, г. Бийск, ул. Трофимова, 27.