

МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Экономика и управление»

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

*Методические рекомендации к лабораторным работам
для студентов направления подготовки
27.03.05 «Инноватика»
дневной формы обучения*



Могилев 2019



УДК 519.2
ББК 22.17
Т 33

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Экономика и управление» «26» февраля 2019 г.,
протокол № 7

Составитель ст. преподаватель Е. Г. Галкина

Рецензент канд. экон. наук, доц. Т. В. Романькова

Методические рекомендации к лабораторным работам предназначены для
студентов направления подготовки 27.03.05 «Инноватика», изучающих дисци-
плину «Теория вероятностей и математическая статистика».

Учебно-методическое издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Ответственный за выпуск	И. В. Ивановская
Технический редактор	С. Н. Красовская
Компьютерная верстка	Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 26 экз. Заказ №

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».

Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 7.03.2019.

Пр-т Мира, 43, 212022, Могилев.

© Белорусско-Российский
университет, 2019



Содержание

Содержание и выполнение лабораторных работ	4
Лабораторная работа № 1. Решение задач на вычисление классической вероятности	5
Лабораторная работа № 2. Вычисление вероятностей сложных событий с использованием формул сложения и умножения вероятностей	7
Лабораторная работа № 3. Решение задач на применение формулы полной вероятности, формулы Байеса	10
Лабораторная работа № 4. Применение встроенных функций EXCEL для вычисления числовых характеристик случайных величин	12
Лабораторная работа № 5. Исследование различных законов распределения. Закон редких событий (формула Пуассона)	13
Лабораторная работа № 6. Локальная и интегральная теоремы Муавра-Лапласа	15
Лабораторная работа № 7. Определение интегральной и дифференциальной функции распределения случайной величины	17
Лабораторная работа № 8. Группировка статистических данных. Построение гистограммы и полигона средствами EXCEL.....	20
Лабораторная работа № 9. Применение встроенных статистических функций EXCEL для определения числовых характеристик выборки	25
Лабораторная работа № 10. Знакомство с пакетом STATISTICA, изучение законов распределения, используемых в статистике.....	27
Лабораторная работа № 11. Вычисление теоретических частот в интервальном вариационном ряду	30
Лабораторная работа № 12. Применение критериев согласия к выборочным данным.....	32
Лабораторная работа № 13. Решение задач на проверку статистических гипотез	35
Лабораторная работа № 14. Проверка гипотез в EXCEL	37
Лабораторная работа № 15. Вычисление коэффициента корреляции Пирсона в EXCEL.....	43
Лабораторная работа № 16. Вычисление уравнения линейной регрессии в EXCEL	44
Список литературы	47



Содержание и выполнение лабораторных работ

В результате выполнения лабораторной работы студенту необходимо научиться обоснованно и результативно применять основные понятия вероятностного анализа, основ статистического описания, проверки статистических гипотез, а также основ анализа парных зависимостей.

Ход выполнения лабораторной работы

- 1 Изучить теоретический материал по теме выполняемой лабораторной работы.
- 2 Ознакомиться с основами применения методики вероятностных и статистических методов по изучаемой теме.
- 3 Выполнить задачи для самостоятельного решения. В случае отсутствия условия таких задач получить условия задач у преподавателя.
- 4 Сделать выводы по полученным результатам.
- 5 Составить отчет.

Форма отчета

Результаты работы оформить в виде листа Excel с соответствующими формулами для их расчета. Сохранить на диске S в каталоге группы в своей папке лист с именем Лабораторная работа № X (X – номер лабораторной работы).

К защите лабораторной работы допускаются только студенты, выполнившие работу и оформившие отчет. Защита проходит в форме устного и письменного собеседования, когда студент отвечает на вопросы преподавателя, которые приведены в данных методических рекомендациях в списке контрольных вопросов к каждой лабораторной работе, а также дополняет свои ответы письменно примерами формул и расчетов по ним.



Лабораторная работа № 1. Решение задач на вычисление классической вероятности

Цель работы: научиться вычислять вероятности различных случайных событий.

Методические указания

Классическое определение вероятности. Если исходы опыта *равновероятны*, то вероятностью события A называется отношение числа исходов, *благоприятствующих* данному событию, к числу всех возможных исходов опыта, т. е.

$$p(A) = \frac{m}{n}. \quad (1)$$

где m – число исходов опыта, благоприятствующих событию;

n – число всех возможных исходов.

Свойства вероятностей.

1 Вероятность любого события – есть число, заключенное между нулем и единицей, т. е. $0 \leq p(A) \leq 1$. Вероятность невозможного события равна 0, а вероятность достоверного события равна 1.

2 Если события A и B *несовместны*, то $p(A + B) = p(A) + p(B)$.

3 Вероятность любого события A в сумме с вероятностью противоположного события A равна единице: $p(A) + p(\bar{A}) = 1$.

Если вероятность интересующего нас события A по каким-либо причинам вычислить трудно, то можно попытаться вычислить вероятность противоположного события, а затем с помощью свойства 3 вычислить искомую вероятность события A .

Задача 1. Брошены две игральные кости. Найти вероятности следующих событий:

A – на обеих костях выпало одинаковое число очков;

B – сумма числа очков не меньше 11;

C – число очков на первой кости больше, чем на второй;

D – сумма очков четная;

E – сумма числа очков больше трех.

Решение

Число очков, благоприятствующих каждому из названных событий, легко подсчитать, если все возможные исходы опыта перечислить в виде таблицы 1. В каждой клетке таблицы первая цифра указывает число очков на первой кости, вторая – на второй кости.

Если кости симметричны и однородны, то все перечисленные исходы опыта равновероятны. Тогда



$p(A) = 6/36 = 1/6$ (благоприятствуют исходы: 11, 22, 33, 44, 55, 66);

$p(B) = 3/36 = 1/12$ (благоприятствуют три исхода: 56, 65, 66).

Таблица 1 – Возможные исходы опыта

Вариант выпадения очков на двух игральных костях					
11	12	13	14	15	16
21	22	23	24	25	26
31	32	33	34	35	36
41	42	43	44	45	46
51	52	53	54	55	56
61	62	63	64	65	66

Непосредственный подсчет числа благоприятствующих исходов дает:

$$p(C) = 15/36 = 5/12;$$

$$p(D) = 18/36 = 1/2;$$

$$p(E) = 33/36 = 11/12.$$

Контрольные вопросы

- 1 Дайте классическое определение вероятности. Как определяется классическая вероятность?
- 2 Какие события несовместны?
- 3 Какие события независимы?
- 4 Определите понятие «сочетание». Что называется «размещением»? Как вычислить число «перестановок»? Как связаны числа сочетаний, размещений и перестановок?
- 5 Дайте определение противоположного события и выведите формулу для его вероятности.
- 6 Укажите границы применения классической вероятности.
- 7 В каких пределах изменяется вероятность случайного события?
- 8 Дайте определение статистической вероятности и приведите примеры.
- 9 Что называется полной группой событий? Приведите пример событий, образующих полную группу.
- 10 Вероятность какого события равна нулю (единице)?



Лабораторная работа № 2. Вычисление вероятностей сложных событий с использованием формул сложения и умножения вероятностей

Цель работы: научиться вычислять вероятности сложных событий с использованием формул сложения и умножения вероятностей.

Методические указания

В теории вероятностей события рассматривают на фоне комплекса условий, которые их порождают.

Рассмотрим множество событий, которые можно наблюдать в эксперименте при фиксированном комплексе условий.

Суммой событий A и B называется событие, состоящее в появлении хотя бы одного из событий A или B . Сумму событий A и B обозначают через $A + B$.

Приведенное понятие можно проиллюстрировать следующим образом.

Пусть комплекс условий состоит в том, что внутри прямоугольника наугад бросают точку. Обозначим через A попадание точки внутрь левого круга, а через B – внутрь правого круга. Тогда события $A + B$, AB и \bar{A} состоят в попадании точки внутрь областей, закрашенных на соответствующей части рисунка 2.1.

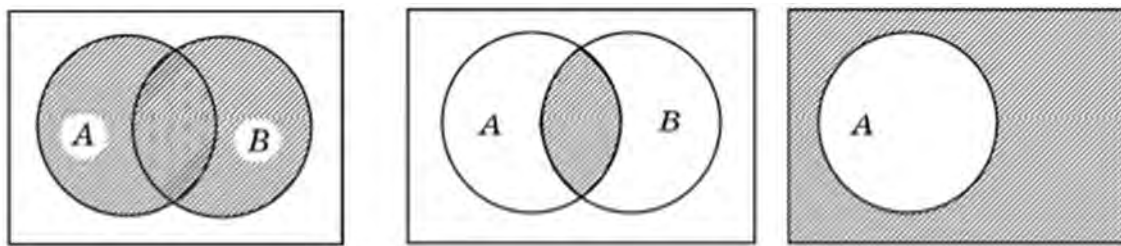


Рисунок 1 – Графическое изображение событий $A + B$, AB и \bar{A}

Произведением событий A и B называют событие, состоящее в появлении событий A и B в одном и том же опыте. Обозначают произведение событий A и B через AB .

Событие, состоящее в *не появлении события A* , называется *противоположным* событием и обозначается через \bar{A} .

Если в каждом опыте два события A и B всегда либо оба происходят, либо оба не происходят, то такие события называют *равносильными* или *эквивалентными* и записывают: $A = B$.

Говорят, что события A_1, A_2, \dots, A_m образуют *полную группу* событий, если они попарно несовместимы и в каждом опыте непременно происходит одно и только одно из этих событий.

Словесные рассуждения можно перевести в символическую запись с помощью соответствий «или» \Leftrightarrow «+»; «и» \Leftrightarrow «·»; «не A » \Leftrightarrow \bar{A} ;

«равносильно» \Leftrightarrow «= \Rightarrow ».

Вероятность события A , вычисленная при условии, что событие B произошло, называется *условной вероятностью* события A и обозначается через $p(A/B)$.

Теорема умножения вероятностей. Вероятность произведения событий равна вероятности одного события, умноженной на вероятность другого события, вычисленную при условии, что первое событие произошло, т. е.

$$p(AB) = p(A) \cdot p(B/A) = p(B) \cdot p(A/B). \quad (2)$$

События называются *независимыми*, если появление одного из них не изменяет вероятности появления другого. Если события независимы, то

$$p(A/B) = p(A), \quad p(B/A) = p(B) \quad \text{и} \quad p(AB) = p(A) \cdot p(B). \quad (3)$$

Теорема сложения вероятностей. Вероятность суммы событий A и B равна сумме вероятностей этих событий без вероятности их совместного появления:

$$p(A + B) = p(A) + p(B) - p(AB). \quad (4)$$

События называются *несовместными*, если их появление в одном и том же опыте невозможно. Если события A и B несовместны, то

$$p(A + B) = p(A) + p(B). \quad (5)$$

Задача. В урне 10 белых и пять черных шаров. Из урны извлекают два шара. Найти вероятность того, что шары:

- 1) белые;
- 2) одного цвета;
- 3) разного цвета.

Решение

Пусть событие A означает извлечение белого шара, B – извлечение черного шара и пусть индекс есть номер извлечения.

Тогда в случае 1 искомое событие имеет вид A_1A_2 (первый шар белый и второй шар белый). Поскольку A_1 и A_2 зависимы, используем формулу вероятности произведения для произвольных событий

$$P(A_1A_2) = P(A_1)P_{A_1}(A_2).$$

$P(A_1)$ находим согласно формуле (1). Так как среди 15 шаров 10 белых, получаем $P(A_1) = 10/15$.

$P_{A_1}(A_2)$ есть условная вероятность события A_2 (второй шар белый) при условии, что A_1 (первый шар белый) произошло. Но если первым взят белый шар, то среди 14 оставшихся шаров белых 9, поэтому $P_{A_1}(A_2) = 9/14$.



Получаем

$$P(A_1A_2) = P(A_1)P_{A_1}(A_2) = \frac{10}{15} \cdot \frac{9}{14} = \frac{3}{7}.$$

В случае 2 искомое событие есть $A_1A_2 + B_1B_2$ (оба шара белые или оба шара черные), причем слагаемые несовместны, а сомножители зависимы. Тогда вероятность, согласно формулам сложения и умножения для несовместных и зависимых событий, равна

$$\begin{aligned} P(A_1A_2 + B_1B_2) &= P(A_1A_2) + P(B_1B_2) = P(A_1)P_{A_1}(A_2) + P(B_1)P_{B_1}(B_2) = \\ &= \frac{10}{15} \cdot \frac{9}{14} + \frac{5}{15} \cdot \frac{4}{14} = \frac{11}{21}. \end{aligned}$$

В случае 3 будем находить вероятность события $A_1A_2 + B_1B_2$ (первый шар белый, второй – черный или, наоборот, первый – черный, второй – белый).

Получим

$$\begin{aligned} P(A_1B_2 + B_1A_2) &= P(A_1B_2) + P(B_1A_2) = \\ &= P(A_1)P_{A_1}(B_2) + P(B_1)P_{B_1}(A_2) = \frac{10}{15} \cdot \frac{5}{14} + \frac{5}{15} \cdot \frac{10}{14} = \frac{10}{21}. \end{aligned}$$

Задача для самостоятельного решения

1 На железнодорожной станции имеется пять путей. Сколькими способами можно расставить на них три состава? Какова вероятность того, что составы случайно будут расставлены на путях в порядке возрастания их номеров?

2 Покупая карточку лотереи «Спортлото», игрок должен зачеркнуть 6 из 49 возможных чисел от 1 до 49. Если при розыгрыше тиража лотереи он угадает все 6 чисел, то имеет шанс выиграть значительную сумму денег. Сколько возможных комбинаций можно составить из 49 по 6, если порядок чисел безразличен? Чему равна вероятность угадать все шесть номеров?

Контрольные вопросы

1 Запишите формулу произведения вероятностей и приведите пример ее применения.

2 Запишите формулу сложения вероятностей и приведите пример ее применения.

3 Какие события называются совместными?

4 Какие события называются независимыми?

5 Какие события называются независимыми попарно, а какие независимыми в совокупности?



6 Что называется условной вероятностью события A ?

7 Чему равна сумма вероятностей противоположных событий?

Лабораторная работа № 3. Решение задач на применение формулы полной вероятности, формулы Байеса

Цель работы: научиться вычислять полную вероятность события.

Методические указания

Событие A может произойти только при выполнении одного из предположений (гипотез) H_1, H_2, \dots, H_n , которые образуют **полную группу несовместных событий**. Каждая из гипотез осуществляется случайным образом и представляет собой случайное событие, вероятности гипотез известны. Тогда вероятность события A вычисляется по формуле

$$P(A) = \sum_{i=1}^n P(H_i) \cdot P(A/H_i). \quad (6)$$

Опыт произведен, и произошло некоторое событие A . Тогда определить апостериорные (послеопытные) вероятности гипотез с учетом того, что произошло именно событие A $p(H_1/A), p(H_2/A), \dots, p(H_n/A)$, можно по формуле Байеса

$$P(H_i/A) = \frac{P(H_i \cdot A)}{P(A)} = \frac{P(H_i) \cdot P(A/H_i)}{P(A)} = \frac{P(H_i) \cdot P(A/H_i)}{\sum_{j=1}^n P(H_j) \cdot P(A/H_j)}. \quad (7)$$

Задача. Радиоприемное устройство имеет блок обработки сигналов, который позволяет отделить полезный сигнал от помехи без искажений. Если отношение уровня сигнала к уровню помехи менее 1,2, то вероятность выделить полезный сигнал без искажений равна 0,1, если отношение уровня сигнала к уровню помехи от 1,2 до двух, то вероятность – 0,8, а если превышает 2, то вероятность равна 1. Приемник принял сигнал, причем поступление сигнала с помехой любого уровня равновероятно. Найти вероятность того, что он будет обработан без искажений.

Решение

Определим событие A – приемник обработал сигнал без искажений. Выдвигаем гипотезы:

- 1) H_1 – приемник принял сигнал с отношением уровня сигнала к уровню помехи менее 1,2;
- 2) H_2 – приемник принял сигнал с отношением уровня сигнала к уровню помехи от 1,2 до двух;



ню помехи от 1,2 до двух;

3) H_3 – приемник принял сигнал с отношением уровня сигнала к уровню помехи более двух.

Вероятности гипотез (т. к. по условию они равновероятны):

$$p(H_1) = p(H_2) = p(H_3) = 1/3 .$$

Определим условные вероятности события A при каждой гипотезе:

$$p(A/H_1) = 0,1; \quad p(A/H_2) = 0,8; \quad p(A/H_3) = 1.$$

По формуле полной вероятности (6) найдем вероятность события A :

$$p(A) = 1/3 \cdot (0,1 + 0,8 + 1) = 0,633.$$

Задача для самостоятельного решения

1 Экспортно-импортная фирма собирается заключить контракт на поставку сельскохозяйственного оборудования в одну из развивающихся стран. Если основной конкурент фирмы не станет одновременно претендовать на заключение контракта, то вероятность получения контракта оценивается в 0,45; в противном случае – в 0,25. По оценкам экспертов компании, вероятность того, что конкурент выдвинет свои предложения по заключению контракта, равна 0,40. Чему равна вероятность заключения контракта?

2 Экономист-аналитик условно подразделяет экономическую ситуацию в стране на «хорошую», «посредственную» и «плохую» и оценивает их вероятности для данного момента времени в 0,15; 0,70 и 0,15 соответственно. Некоторый индекс экономического состояния возрастает с вероятностью 0,60, когда ситуация «хорошая»; с вероятностью 0,30, когда ситуация «посредственная»; с вероятностью 0,10, когда ситуация «плохая». Пусть в настоящий момент индекс экономического состояния возрос. Чему равна вероятность того, что экономика страны на подъеме?

Контрольные вопросы

- 1 Что такое условная вероятность?
- 2 Для чего применяется формула Байеса? Приведите примеры.
- 3 Для чего применяется формула Бернулли? Приведите примеры.
- 4 Что значит «полная вероятность»? Как ее посчитать? Приведите примеры.



Лабораторная работа № 4. Применение встроенных функций EXCEL для вычисления числовых характеристик случайных величин

Цель работы: научиться применять встроенные функции EXCEL для вычисления числовых характеристик случайных величин.

Задание 1

Вычисление математического ожидания.

Рассмотрим нахождение математического ожидания для ряда дискретных значений. На рисунке 2 представлен ряд дискретных значений, причём в левом столбце представлены значения, а в правом – их вероятности. Математическое ожидание представленного ряда значений вычисляется по следующей формуле: $=\text{СУММПРОИЗВ}(A2:A11;B2:B11)$. Результат вычислений математического ожидания помещён в ячейку B13.

B13		=СУММПРОИЗВ(A2:A11;B2:B11)				
	A	B	C	D	E	Строка формул
1	Значение	Вероятность				
2	1	0,0769				
3	2	0,1538				
4	3	0,0256				
5	4	0,1538				
6	5	0,0769				
7	6	0,0256				
8	7	0,1026				
9	8	0,1795				
10	9	0,0769				
11	10	0,1282				
12						
13		5,7436				

Рисунок 2 – Вычисление математического ожидания

Задание 2

Вычисление дисперсии.

Рассмотрим пример нахождения дисперсии случайной величины. Найдём математическое ожидание и возведём его квадрат (рисунок 3).

Результат вычислений представлен в ячейке B13.

Теперь найдём математическое ожидание квадрата случайной величины. Для этого возведём случайную величину в квадрат и запишем значения в столбец C. Затем найдём математическое ожидание для случайной величины из столбца C. Результат вычислений представлен в ячейке C13.

Вычисление математического ожидания квадрата случайной величины. После этого остается вычислить разность между ячейками C13 и B13, что и будет являться дисперсией случайной величины X .

B13		=СУММПРОИЗВ(A2:A11;B2:B11)*2					
	A	B	C	D	E	Строка формул	H
1	Значение	Вероятность					
2	1	0,0769					
3	2	0,1538					
4	3	0,0256					
5	4	0,1538					
6	5	0,0769					
7	6	0,0256					
8	7	0,1026					
9	8	0,1795					
10	9	0,0769					
11	10	0,1282					
12							
13		32,9888					
14							

Рисунок 3 – Вычисление математического ожидания

Задание 3

Вычисление среднего квадратического отклонения случайной величины.

После завершения работы с предыдущим заданием Вам необходимо вычислить квадратный корень из найденного значения дисперсии, что и будет являться средним квадратическим отклонением. В MS Excel квадратный корень вычисляется с использованием функции КОРЕНЬ(число). Поместите результат в ячейку B15.

Контрольные вопросы

- 1 Что называют дискретной случайной величиной?
- 2 Что называют законом распределения случайной величины?
- 3 Что такое математическое ожидание случайной величины? По какой формуле вычисляется данная характеристика?
- 4 Что такое дисперсия случайной величины? По какой формуле вычисляется данная характеристика?
- 5 Что такое среднее квадратическое отклонение случайной величины? По какой формуле вычисляется данная характеристика?

Лабораторная работа № 5. Исследование различных законов распределения. Закон редких событий (формула Пуассона)

Цель работы: изучить различные законы распределения, закон редких событий (формулу Пуассона).

Задача 1. Вероятность искажения одного символа при передаче сообщения равна 0,01. Какова вероятность, что сообщение, содержащее 200 символов, со-

держит не более двух искажений? Найти точное и приближенные значения вероятности, используя локальную формулу Муавра-Лапласа и формулу Пуассона.

Решение

Следует найти вероятность того, что в $n = 200$ испытаниях по схеме Бернулли с вероятностью успеха 0,01 число успехов будет не более двух. Для точного вычисления вероятности используем функцию БИНОМРАСП. Таким образом, для $n = 200$, $m = 2$, $p = 0,01$ находим

$$P_{200}(0 \leq m \leq 2) = \text{БИНОМРАСП}(2; 200; 0,01; 1) = 0,67.$$

Найдем приближенное значение вероятности, используя локальную формулу Муавра-Лапласа. Согласно этой формуле, вероятность $P_n(m)$ приближенно равна плотности нормального распределения в точке m со средним np и среднеквадратичным отклонением \sqrt{npq} . Таким образом,

$$\begin{aligned} P_{200}(0 \leq m \leq 2) &= P_{200}(0) + P_{200}(1) + P_{200}(2) \approx \\ &\approx \text{НОРМРАСП}(0; 2; \sqrt{1,98}; 0) + \text{НОРМРАСП}(1; 2; \sqrt{1,98}; 0) + \\ &+ \text{НОРМРАСП}(2; 2; \sqrt{1,98}; 0) = 0,607. \end{aligned}$$

Найдем приближенное значение той же вероятности, используя формулу Пуассона. Согласно этой формуле, вероятность $P_n(m)$ при малых p приближенно равна вероятности пуассоновского распределения в точке m со средним значением $\lambda = np$. Значения вероятностей для распределения Пуассона возвращает функция ПУАССОН. Таким образом,

$$P_{200}(0 \leq m \leq 2) \approx \text{ПУАССОН}(2; 200 \cdot 0,01; 1) = 0,676676.$$

Заметим, что погрешность, полученная при использовании формулы Пуассона, в данном случае на порядок ниже, чем при использовании локальной формулы Муавра-Лапласа. Смысла использовать интегральную формулу Муавра-Лапласа в данном случае нет, поскольку интервал значений m мал ($0 \leq m \leq 2$).

Найдем приближенное значение вероятности, используя теперь интегральную формулу Муавра-Лапласа. Вероятность

$$P_n(m_1 \leq m \leq m_2) \approx \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right),$$

т. е. приближенно равна вероятности попадания в интервал $(m_1; m_2)$ нормальной случайной величины со средним np и среднеквадратичным отклонением \sqrt{npq} .



Следовательно, если $F_{\xi}(x)$ – функция распределения нормальной случайной величины с параметрами $a = np$ и $\sigma^2 = npq$, то

$$P_n(m_1 \leq m \leq m_2) \approx F_{\xi}(m_2) - F_{\xi}(m_1).$$

Таким образом,

$$P_n(0 \leq m \leq 2) \approx \text{НОРМРАСП}(2; 2; \sqrt{1,98}; 1) - \text{НОРМРАСП}(0; 2; \sqrt{1,98}; 1) = 0,4223.$$

Задача для самостоятельного решения

1 Кандидат на выборах считает, что 20 % избирателей в определенной области поддерживают его избирательную платформу. Если 64 избирателя случайно отобраны из числа избирателей данной области, найдите вероятность того, что отобранная доля избирателей, поддерживающих кандидата, не будет отличаться по абсолютной величине от истинной доли более, чем на 0,07.

2 Авиакомпания знает, что в среднем 5 % людей, делающих предварительный заказ на определенный рейс, не будет его использовать. Если авиакомпания продала 160 билетов на самолет, в котором лишь 155 мест, чему равна вероятность того, что место будет доступно для любого пассажира, имеющего заказ и планирующего улететь?

Контрольные вопросы

- 1 Назовите и кратко охарактеризуйте законы распределения.
- 2 Охарактеризуйте условия применения закона редких событий (формула Пуассона).

Лабораторная работа № 6. Локальная и интегральная теоремы Муавра-Лапласа

Цель работы: изучить условия применения локальной и интегральной теоремы Муавра-Лапласа.

Методические указания

Пусть в каждом из n независимых испытаний событие A может произойти с вероятностью p , $q = 1 - p$ (условия схемы Бернулли). $P_n(k)$ – вероятность k появлений события A в n испытаниях. $P_n(k_1, k_2)$ – вероятность того, что число появлений события A находится между k_1 и k_2 .

Локальная теорема Лапласа.

Если n – велико, а p – отлично от 0 и 1, то



$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi\left(\frac{k-np}{\sqrt{npq}}\right), \quad (8)$$

где $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ – функция Гаусса.

Задача 1. Найти вероятность того, что событие A (переключение передач) наступит 70 раз на 243-километровой трассе, если вероятность переключения на каждом километре равна 0,25.

Решение

$$n = 243, k = 70, p = 0,25, q = 0,75.$$

$$1) x = \frac{k-np}{\sqrt{npq}} = \frac{70-243 \cdot 0,25}{\sqrt{243 \cdot 0,25 \cdot 0,75}} = \frac{9,25}{6,75} = 1,37.$$

$$2) \varphi(1,37) = 0,1661.$$

$$3) P_{243}(70) = \frac{1}{1,75} \cdot 0,1561 = 0,0231.$$

Интегральная теорема Лапласа.

Если вероятность события A в каждом испытании постоянна и отлична от нуля и единицы, то вероятность $P(k_1, k_2)$ того, что событие A появится в n испытаниях от k_1 до k_2 раз, приближенно равна определенному интегралу:

$$P(k_1, k_2) \approx \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-\frac{z^2}{2}} dz, \quad (9)$$

$$\text{где } x' = \frac{k_1 - np}{\sqrt{npq}}, \quad x'' = \frac{k_2 - np}{\sqrt{npq}}.$$

Так как $\int_{x'}^{x''} e^{-\frac{z^2}{2}} dz$ не выражается через элементарные функции, то его значение находится по таблице значений функции Лапласа $\Phi(x) = \int_0^x e^{-\frac{z^2}{2}} dz$.

Функции Гаусса и Лапласа обладают **свойствами**, которые необходимо знать при использовании таблиц значений этих функций:

а) $\varphi(-x) = \varphi(x)$, $\Phi(-x) = -\Phi(x)$;

б) при больших x верно $\varphi(x) \approx 0$, $\Phi(x) \approx 0,5$.

Таким образом, $P(k_1, k_2) = \Phi(x'') - \Phi(x')$.

Задача 2. Вероятность выпуска нестандартной лампы $p = 0,1$. Чему равна вероятность того, что в партии из 2000 ламп число стандартных не менее 1790 шт.



Решение

$$p = 0,9, \quad q = 0,1, \quad n = 2000, \quad k_1 = 1790, \quad k_2 = 2000.$$

$$x'' = \frac{2000 - 2000 \cdot 0,9}{\sqrt{2000 \cdot 0,9 \cdot 0,1}} = 14,9,$$

$$x' = \frac{1790 - 2000 \cdot 0,9}{\sqrt{2000 \cdot 0,1 \cdot 0,9}} = -0,75.$$

$$P_{2000}(1790 \leq k \leq 2000) = \Phi(14,9) - \Phi(-0,75) = 0,5 + 0,273 = 0,773.$$

Теоремы Лапласа дают удовлетворительное приближение при $npq \geq 9$. При чем чем ближе значения q, p к 0,5, тем точнее данные формулы. При маленьких или больших значениях вероятности (близких к 0 или 1) формула дает большую погрешность (по сравнению с исходной формулой Бернулли).

Задача для самостоятельного решения

1 Для мастера определенной квалификации вероятность изготовить деталь отличного качества равна 0,75. За смену он изготовил 400 деталей. Найти вероятность того, что в их числе 280 деталей отличного качества.

2 В продукции некоторого производства брак составляет 15 %. Изделия отправляются потребителям (без проверки) в коробках по 100 штук. Найти вероятности событий В – «наудачу взятая коробка содержит 13 бракованных изделий».

Контрольные вопросы

- 1 Теорема Пуассона.
- 2 Что называется схемой Бернулли?
- 3 Локальная формула Муавра-Лапласа.
- 4 Интегральная формула Муавра-Лапласа.

Лабораторная работа № 7. Определение интегральной и дифференциальной функции распределения случайной величины

Цель работы: изучить метод построения интегральной и дифференциальной функции распределения случайной величины.

Методические указания

Случайная величина может быть определена путем задания значений самой величины и вероятностей этих значений.



Однако, такой метод применим далеко не всегда. Например, в случае непрерывной случайной величины, ее значения могут заполнять некоторый произвольный интервал. Очевидно, что в этом случае задать все значения случайной величины сложно.

Пусть x – действительное число. Вероятность события, состоящего в том, что X примет значение, меньшее x , т. е. $X < x$, обозначим через $F(x)$.

Функцией распределения называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x .

$$F(x) = P(X < x). \quad (10)$$

Функцию распределения также называют *интегральной функцией*.

Функция распределения существует как для непрерывных, так и для дискретных случайных величин. Она полностью характеризует случайную величину и является одной из форм закона распределения.

Для дискретной случайной величины функция распределения имеет вид:

$$F(x) = \sum_{x_i < x} P(X = x_i). \quad (11)$$

Знак неравенства под знаком суммы показывает, что суммирование распространяется на те возможные значения случайной величины, которые меньше аргумента x .

Функция распределения дискретной случайной величины X разрывна и возрастает скачками при переходе через каждое значение x_i .

Задача 1. Задана интегральная функция распределения:

$$F(x) = \begin{cases} 0 & \text{при } x \leq 1; \\ 0,3 & \text{при } 1 < x \leq 4; \\ 0,4 & \text{при } 4 < x \leq 8; \\ 1 & \text{при } x > 8. \end{cases}$$

Построить ее график.

Решение

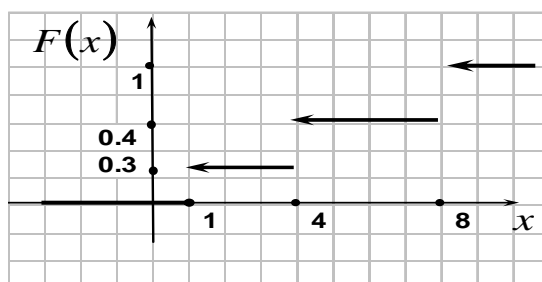


Рисунок 4 – График функции $F(x)$



Свойства функции распределения.

1 Значения функции распределения принадлежат отрезку $[0, 1]$:

$$0 \leq F(x) \leq 1.$$

2 $F(x)$ – неубывающая функция $F(x_2) \geq F(x_1)$ при $x_2 \geq x_1$.

3 Вероятность того, что случайная величина примет значение, заключенное в интервале (a, b) , равна приращению функции распределения на этом интервале

$$P(a \leq X < b) = F(b) - F(a).$$

4 На минус бесконечности функция распределения равна нулю, на плюс бесконечности функция распределения равна единице.

5 Вероятность того, что непрерывная случайная величина X примет одно определенное значение, равна нулю.

Таким образом, не имеет смысла говорить о каком-либо конкретном значении случайной величины. Интерес представляет только вероятность попадания случайной величины в какой-либо интервал, что соответствует большинству практических задач.

С помощью функции распределения можно дать более строгое определение непрерывной случайной величины.

Случайная величина X называется *непрерывной*, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме отдельных точек.

В отличие от дискретных случайных величин вероятность отдельного значения для непрерывной случайной величины равна нулю:

$$P(X = c) = 0, \quad \forall c \in R.$$

Поэтому для непрерывной случайной величины X имеем

$$P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = P(a \leq X \leq b) = F(b) - F(a). \quad (12)$$

Для непрерывных случайных величин, кроме функции распределения, существует еще один удобный способ задания закона распределения – *плотность вероятности*.

Пусть функция распределения $F(x)$ данной непрерывной случайной величины X непрерывна и дифференцируема всюду, кроме отдельных точек.

Плотностью распределения непрерывной случайной величины X (или плотностью вероятности, или просто плотностью) называется функция $\varphi(x) = F'(x)$.

Функцию $\varphi(x)$ называют также дифференциальной функцией распределения.

Плотность распределения любой непрерывной случайной величины неотрицательна, $\varphi(x) \geq 0$; обладает свойством нормированности:



$$\int_{-\infty}^{+\infty} \varphi(x) dx = 1; \quad \lim_{x \rightarrow \pm\infty} \varphi(x) = 0. \quad (13)$$

График функции $\varphi(x)$ называется кривой распределения.

Функция распределения $F(x)$ выражается через плотность распределения формулой

$$F(x) = \int_{-\infty}^x \varphi(x) dx. \quad (14)$$

Вероятность попадания непрерывной случайной величины X в промежуток $[a, b)$ определяется равенством

$$P(a \leq X < b) = \int_a^b \varphi(x) dx. \quad (15)$$

Контрольные вопросы

- 1 Что определяет функция распределения?
- 2 Назовите свойства интегральной функции распределения.
- 3 Назначение дифференциальной функции распределения случайной величины.

Лабораторная работа № 8. Группировка статистических данных. Построение гистограммы и полигона средствами EXCEL

Цель работы: изучение методов графического изображения статистических данных.

Методические указания

Рассмотренные ранее распределения вероятностей случайной величины (СВ) опираются на знание закона распределения СВ. Для практических задач такое знание – редкость. Здесь закон распределения обычно неизвестен, или известен с точностью до некоторых неизвестных параметров. В частности, невозможно рассчитать точное значение соответствующих вероятностей, т. к. нельзя определить количество общих и благоприятных исходов. Поэтому вводится статистическое определение вероятности. По этому определению вероятность равна отношению числа испытаний, в которых событие произошло, к общему числу произведенных испытаний. Такая вероятность называется статистической частотой.



Статистическим распределением выборки называют перечень вариантов и соответствующих им частот или относительных частот.

Связь между эмпирической функцией распределения и функцией распределения (теоретической функцией распределения) такая же, как связь между частотой события и его вероятностью.

Для построения выборочной функции распределения весь диапазон изменения случайной величины X (выборки) разбивают на ряд интервалов (карманов) одинаковой ширины. Число интервалов обычно выбирают не менее 3 и не более 15. Затем определяют число значений случайной величины X , попавших в каждый интервал (абсолютная частота, частота интервалов).

Частота интервалов – число, показывающее, сколько раз значения, относящиеся к каждому интервалу группировки, встречаются в выборке. Поделив эти числа на общее количество наблюдений n , находят относительную частоту (частость) попадания случайной величины X в заданные интервалы.

По найденным относительным частотам строят гистограммы выборочных функций распределения.

Гистограмма распределения частот – это графическое представление выборки, где по оси абсцисс (Ox) отложены величины интервалов, а по оси ординат (Oy) – величины частот, попадающих в данный классовый интервал. При увеличении до бесконечности размера выборки выборочные функции распределения превращаются в теоретические: гистограмма превращается в график плотности распределения.

Сумма всех абсолютных частот называют объемом выборки (т. е. количество всех объектов выборки). Например, на рисунке 5 в гистограмме по оси ординат отложены n_i/h (h – длина интервала), следовательно, объем выборки равен $2 \cdot (1 + 2 + 4 + 5) = 24$.

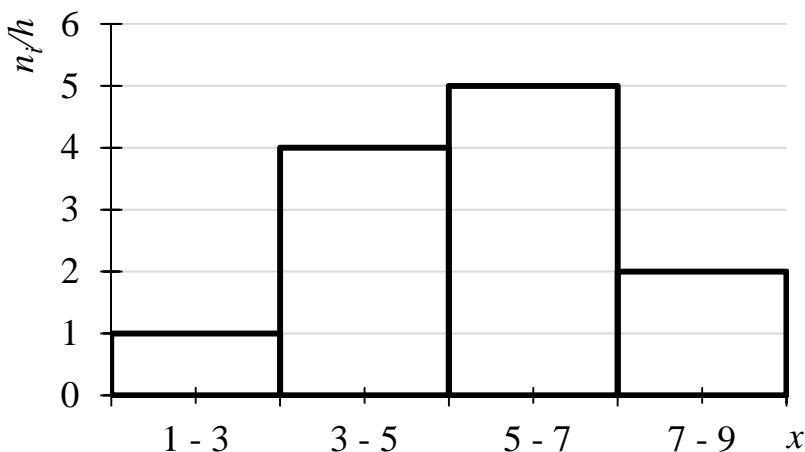


Рисунок 5 – Гистограмма распределения частот

Накопленная частота интервалов – это число, полученное последовательным суммированием частот в направлении от первого интервала к последнему, до того интервала включительно, для которого определяется накопленная частота.



В Excel для построения выборочных функций распределения используются специальная функция **ЧАСТОТА** и процедура **Гистограмма** из пакета анализа.

Функция **ЧАСТОТА** (*массив_данных, двоичный_массив*) вычисляет частоты появления случайной величины в интервалах значений и выводит их как массив цифр, где:

– *массив_данных* — это массив или ссылка на множество данных, для которых вычисляются частоты;

– *двоичный_массив* — это массив интервалов, по которым группируются значения выборки.

Процедура **Гистограмма** из **Пакета анализа** выводит результаты выборочного распределения в виде таблицы и графика. Параметры диалогового окна **Гистограмма**:

– **Входной диапазон** – диапазон исследуемых данных (выборка);

– **Интервал карманов** – диапазон ячеек или набор граничных значений, определяющих выбранные интервалы (карманы). Эти значения должны быть введены в возрастающем порядке. Если диапазон карманов не был введен, то набор интервалов, равномерно распределенных между минимальным и максимальным значениями данных, будет создан автоматически.

– **Выходной диапазон** предназначен для ввода ссылки на левую верхнюю ячейку выходного диапазона;

– переключатель **Интегральный процент** позволяет установить режим включения в гистограмму графика интегральных процентов;

– переключатель **Вывод графика** позволяет установить режим автоматического создания встроенной диаграммы на листе, содержащем выходной диапазон.

Задание 1

Построить эмпирическое распределение веса студентов в килограммах для следующей выборки: 64, 57, 63, 62, 58, 61, 63, 70, 60, 61, 65, 62, 62, 40, 64, 61, 59, 59, 63, 61.

Решение

1 В ячейку A1 вводим слово Наблюдения, а в диапазон A2:A21 – значения веса студентов.

2 В ячейку B1 вводим названия интервалов *Вес, кг*. В диапазон B2:B8 вводим граничные значения интервалов (40, 45, 50, 55, 60, 65, 70). Результаты представлены на рисунке 6.

3 Вводим заголовки создаваемой таблицы: в ячейки C1 – **Абсолютные частоты**, в ячейки D1 – **Относительные частоты**, в ячейки E1 – **Накопленные частоты** (см. рисунок 6).

4 С помощью функции **Частота** заполняем столбец абсолютных частот, для этого выделите блок ячеек C2:C8. Вызовите **Мастер функций** (кнопка *fx*). В появившемся диалоговом окне выбираем категорию **Статистические** и функцию **ЧАСТОТА**, после чего нажимаем кнопку ОК. Получаемое диалоговое окно представлено на рисунке 7.



	A	B	C	D	E
1	Наблюдения	Вес, кг			
2	64	40			
3	57	45			
4	63	50			
5	62	55			
6	58	60			
7	61	65			
8	63	70			
9	70				
10	60				
11	61				
12	65				
13	62				
14	62				
15	40				
16	64				
17	61				
18	59				
19	59				
20	63				
21	61				

Рисунок 6 – Вид электронной таблицы после ввода данных

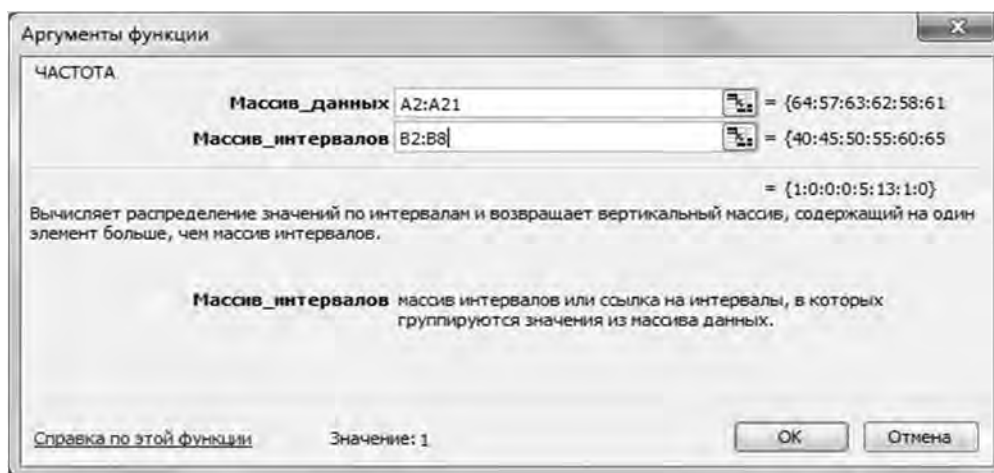


Рисунок 7 – Вид диалогового окна для ввода параметров функции ЧАСТОТА

Указателем мыши в рабочее поле **Массив_данных** вводим диапазон данных наблюдений (A2:A21). В рабочее поле **Двоичный_массив** мышью вводим диапазон интервалов (B2:B8). Слева на клавиатуре последовательно нажимаем комбинацию клавиш Ctrl + Shift + Enter. В столбце C должен появиться массив абсолютных частот (рисунок 8).

	A	B	C	D	E	F	G
	Наблюдения	Вес, кг	Абсолютные частоты	Относительные частоты	Накопленные частоты		
1							
2	64	40	1				
3	57	45	0				
4	63	50	0				
5	62	55	0				
6	58	60	5				
7	61	65	13				
8	63	70	1				
9	70						

Рисунок 8 – Фрагмент электронной таблицы после вычисления абсолютных частот



5 В ячейке С9 найдем общее количество наблюдений. Активизируем ячейку С9, на вкладке **Главная** нажимаем кнопку **Сумма**. Убедившись, что диапазон суммирования указан правильно, нажимаем клавишу Enter (рисунок 9).

	A	B	C	D	E	F	G
	Наблюдения	Вес, кг	Абсолютные частоты	Относительные частоты	Накопленные частоты		
1							
2	64	40	1				
3	57	45	0				
4	63	50	0				
5	62	55	0				
6	58	60	5				
7	61	65	13				
8	63	70	1				
9	70		20				
10	60						

Рисунок 9 – Фрагмент электронной таблицы после суммирования диапазона С2:С8

6 Заполняем столбец относительных частот. В ячейку вводим формулу для вычисления относительной частоты: $= C2/\$C\9 . Нажимаем клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) копируем введенную формулу в диапазон и получим массив относительных частот. Результат представлен на рисунке 10.

	A	B	C	D
	Наблюдения	Вес, кг	Абсолютные частоты	Относительные частоты
1				
2	64	40	1	0,05
3	57	45	0	0
4	63	50	0	0
5	62	55	0	0
6	58	60	5	0,25
7	61	65	13	0,65
8	63	70	1	0,05
9	70		20	
10	60			

Рисунок 10 – Фрагмент электронной таблицы после вычисления относительных частот

7 Заполним столбец накопленных частот. В ячейку D2 скопируем значение относительной частоты из ячейки E2. В ячейку D3 вводим формулу: $= E2 + D3$. Нажимаем клавишу Enter. Протягиванием (за правый нижний угол при нажатой левой кнопке мыши) копируем введенную формулу в диапазон D3:D8. Получим массив накопленных частот. Результат вычислений представлен на рисунке 11.

8 Построим диаграмму относительных и накопленных частот. Щелчком указателя мыши по кнопке на панели инструментов вызовем на вкладке **Вставка** команду **Диаграммы**. В появившемся диалоговом окне выберем тип диаграммы **Гистограмма**.

	A	B	C	D	E	F	G
	Наблюдения	Вес, кг	Абсолютные частоты	Относительные частоты	Накопленные частоты		
1							
2	64	40	1	0,05	0,05		
3	57	45	0	0	0,05		
4	63	50	0	0	0,05		
5	62	55	0	0	0,05		
6	58	60	5	0,25	0,3		
7	61	65	13	0,65	0,95		
8	63	70	1	0,05	1		
9	70		20				
10	60						
11	61						

Рисунок 11 – Результат вычислений

Задание для самостоятельной работы

1 Построить выборочные функции распределения (относительные и накопленные частоты) для роста 20 студентов (сантиметры): 181, 169, 178, 178, 171, 179, 172, 181, 179, 168, 174, 167, 169, 171, 179, 181, 181, 183, 172, 176.

2 Найти распределение по абсолютным частотам для следующих результатов тестирования в баллах: 79, 85, 78, 85, 83, 81, 95, 88, 97, 85 (используйте границы интервалов 70, 80, 90).

Контрольные вопросы

- 1 Опишите этапы построения гистограммы распределения частот.
- 2 Что показывает частота интервалов?
- 3 Как определяется объем выборки?

Лабораторная работа № 9. Применение встроенных статистических функций EXCEL для определения числовых характеристик выборки

Цель работы: научиться применять встроенные статистические функции EXCEL для определения числовых характеристик выборки.

Методические указания

Раздел математики, посвященный методам сбора, анализа и обработки статистических данных для научных и практических целей, называется математической статистикой.

Данный раздел математики имеет дело с массовыми явлениями и тесно связан с теорией вероятностей, т. к. базируется на ее математическом аппарате.

Цели статистического исследования – исследование соотношений между статистическими данными (описательная статистика) и использование результа-



тов данных исследований для прогнозирования и принятия решений (аналитическая статистика).

Статистические данные представляют собой данные, полученные в результате обследования большого числа объектов или явлений.

По охвату статистической совокупности исследование может быть сплошное или не сплошное. При сплошном исследовании группа формируется из всех единиц изучаемого явления (генеральная совокупность), а при не сплошном – только группа этих единиц (выборка).

Конечной же целью изучения выборочной совокупности (выборки) является получение информации не о ней самой как таковой, а о генеральной совокупности. Поэтому обычно стремятся сделать так, чтобы выборка наилучшим образом представляла генеральную совокупность, была репрезентативной и представительной. Например, если мы хотим получить данные о поступающих во все вузы города, то абитуриенты данного университета есть выборка из более широкой генеральной совокупности – всех абитуриентов вузов города, и тем не менее эта выборка не обязательно будет являться представительной. В тех же случаях, когда генеральная совокупность недостаточно известна, обычно не удается предложить более лучшего способа, чем случайный выбор элементов для выборки.

Возможности MS Excel для анализа данных.

В мастере функций Excel имеется ряд специальных функций, предназначенных для вычисления выборочных характеристик.

Функция СРЗНАЧ вычисляет среднее арифметическое из нескольких массивов (аргументов) чисел. Аргументы число1, число2, ... – это от 1 до 30 массивов, для которых вычисляется среднее.

Функция МЕДИАНА позволяет получать медиану заданной выборки.

Функция МОДА вычисляет наиболее часто встречающееся значение в выборке.

Функция ДИСП позволяет оценить дисперсию по выборочным данным.

Функция СТАНДОТКЛОН вычисляет стандартное отклонение.

Пример 1 – Провести статистический анализ методом описательной статистики показателей в регионе 1 (выборка 1) и регионе 2 (выборка 2). Результат расчета представлен на рисунке 12.

Задание для самостоятельной работы

1 Найти среднее значение, медиану, моду, стандартное отклонение результатов бега на дистанцию 100 м у группы студентов (секунды): 12,8; 13,2; 13,0; 12,9; 13,5; 13,1.

2 Определить основные статистические характеристики для данных измерений роста групп студенток (сантиметры): 164, 160, 157, 166, 162, 160, 161, 159, 160, 163, 170, 171.

3 Найти наиболее популярный туристический маршрут из четырех, реализуемых фирмой, если за неделю последовательно были реализованы следующие маршруты: 1, 3, 3, 2, 1, 1, 4, 4, 2, 4, 1, 3, 2, 4, 1, 4, 4, 3, 1, 2, 3, 4, 1, 1, 3.



Основные статистические характеристики		
	Выборка 1	Выборка 2
1		
2		
3	1	49
4	1	51
5	1	49
6	1	51
7	1	49
8	1	51
9	1	49
10	1	51
11	1	49
12	491	51
13	сумма	500
14	среднее	50
15	дисперсия	24010
16	станд. отклонение	1,11111111
17	медиана	1
18	мода	49

Рисунок 12 – Пример расчета основных статистических характеристик

Контрольные вопросы

- 1 Чем отличается генеральная и выборочная совокупности?
- 2 Что такое выборочная медиана?
- 3 Что такое выборочная мода?
- 4 Что называют дисперсией выборки?
- 5 Что такое стандартное отклонение?
- 6 Какие основные функции пакета MS Excel применяются для определения основных статистических характеристик?

Лабораторная работа № 10. Знакомство с пакетом STATISTICA, изучение законов распределения, используемых в статистике

Цель работы: ознакомиться с основными возможностями пакета STATISTICA, изучить законы распределения, используемые в статистике.

Методические указания

Statistica – пакет прикладных программ, программный пакет, для статистического анализа, разработанный компанией StatSoft, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов.

Программный пакет Statistica широко применяется в финансовой, промышленной, медицинской, экономической, страховой, коммерческой и некоммерческой сферах деятельности людей. Продукт удобен в использовании, легок в осво-

ении, оснащен интуитивно понятным интерфейсом. При помощи нескольких несложных действий пользователь сможет выбрать и применить аналитические инструменты, которые требуются для статистического анализа.

Данные в системе Statistica организованы в виде электронных таблиц, как в привычной для пользователей программе Excel. Файл содержит наблюдения и переменные. Наблюдения можно рассматривать как эквивалент записей в базах данных (или строк электронной таблицы), а переменные – как эквивалент полей (столбцов электронной таблицы). Каждое наблюдение состоит из набора значений переменной.

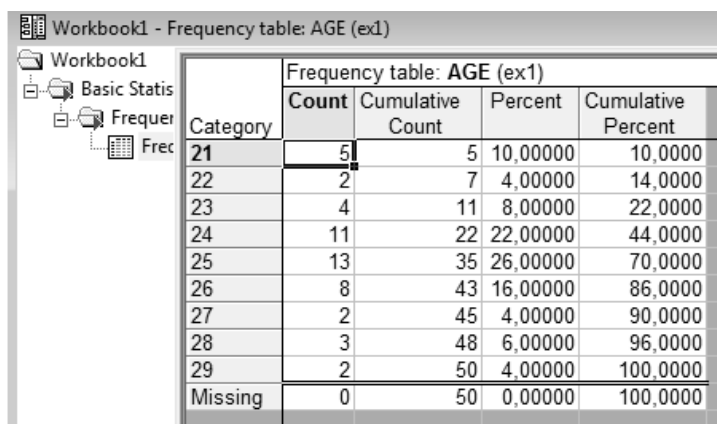
Пример применения пакета Statistica.

Проведено исследование 50 человек, где фиксировались их предпочтения в просмотре по телевидению соревнований по футболу, бейсболу и баскетболу, а также возраст опрашиваемых. Частота просмотра соревнований фиксировалась в шкале «всегда», «как правило», «иногда» и «никогда». Результаты опроса получить у преподавателя.

Требуется занести данные обследования в файл данных.

Ход решения задания.

Внесем результаты опроса в программный пакет Statistica. Построим дискретный ряд распределения переменной Age – Возраст. Таблица частот (ряд распределения), полученная для переменной **Age – Возраст** с помощью опции Alldistinctvalues, представлена на рисунке 13.



Workbook1 - Frequency table: AGE (ex1)				
Category	Frequency table: AGE (ex1)			
	Count	Cumulative Count	Percent	Cumulative Percent
21	5	5	10,00000	10,0000
22	2	7	4,00000	14,0000
23	4	11	8,00000	22,0000
24	11	22	22,00000	44,0000
25	13	35	26,00000	70,0000
26	8	43	16,00000	86,0000
27	2	45	4,00000	90,0000
28	3	48	6,00000	96,0000
29	2	50	4,00000	100,0000
Missing	0	50	0,00000	100,0000

Рисунок 13 – Дискретный ряд распределения переменной Age – Возраст

Построим атрибутивный ряд распределения переменной **Football – Футбол**. Для этого для переменной **Football – Футбол** выберем опцию **Alldistinctvalues** – Все различающиеся значения и **withtextvalues** – С текстовыми значениями (рисунок 14).

Построим ряд распределения с заданным количеством интервалов, используя опцию **No.ofexactintervals** – Число равных интервалов (рисунок 15).

Построим двухвходовую таблицу частот, воспользовавшись группой кнопок **Stub-and-bannertable** (рисунок 16).

Построим по таблице сопряженности график **CategorizedHistograms** – Категоризованных. Результат построения графика представлен на рисунке 17.

Frequency table: FOOTBALL (ex1)				
Category	Count	Cumulative Count	Percent	Cumulative Percent
always	24	24	48,00000	48,0000
usually	4	28	8,00000	56,0000
sometime	11	39	22,00000	78,0000
never	11	50	22,00000	100,0000
Missing	0	50	0,00000	100,0000

Рисунок 14 – Атрибутивный ряд переменной **Football – Футбол**

Frequency table: AGE (ex1)					
From	To	Count	Cumulative Count	Percent	Cumulative Percent
20,20000	<x<=21,80000	5	5	10,00000	10,0000
21,80000	<x<=23,40000	6	11	12,00000	22,0000
23,40000	<x<=25,00000	24	35	48,00000	70,0000
25,00000	<x<=26,60000	8	43	16,00000	86,0000
26,60000	<x<=28,20000	5	48	10,00000	96,0000
28,20000	<x<=29,80000	2	50	4,00000	100,0000
Missing		0	50	0,00000	100,0000

Рисунок 15 – Ряд распределения с заданным количеством интервалов

Summary Frequency Table (ex1)
 Marked cells have counts > 10
 (Marginal summaries are not marked)

FOOTBALL	BASEBALL always	BASEBALL usually	BASEBALL sometime	BASEBALL never	Row Totals
always	13	5	4	2	24
usually	2	1	0	1	4
sometime	0	2	8	1	11
never	1	2	2	6	11
All Grps	16	10	14	10	50

Рисунок 16 – Двухвходовая таблица сопряженности частот переменных **FOOTBALL** и **BASEBALL**

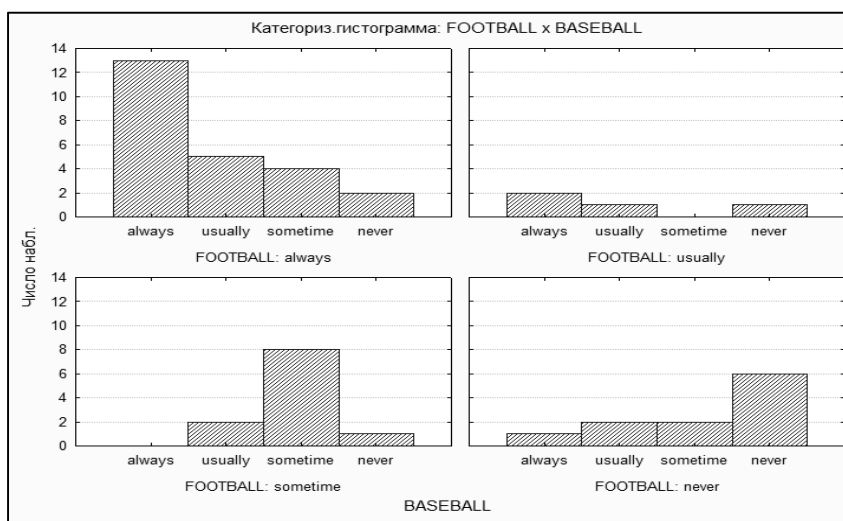


Рисунок 17 – Категоризованная гистограмма, построенная по таблице сопряженности



Контрольные вопросы

- 1 Опишите назначение пакета Statistica.
- 2 Опишите последовательность импорта данных из Excel в Statistica.

Лабораторная работа № 11. Вычисление теоретических частот в интервальном вариационном ряду

Цель работы: освоить методику вычисления теоретических частот в интервальном вариационном ряду.

Эмпирические частоты получают в результате опыта (наблюдения). *Теоретические частоты*, обозначаемые далее через v_i , находят с помощью равенства

$$v_i = m \cdot P_i, \quad (16)$$

где m – количество испытаний;

$P_i = P(z_{i-1} < X < z_i)$ – теоретическая вероятность попадания значений случайной величины в i -й промежуток.

$$z = \frac{x_i - \bar{x}}{s}. \quad (17)$$

Например, вероятность попадания в соответствующий интервал при нормальном распределении:

$$P_i = \Phi(z_{1i}) - \Phi(z_{2i}),$$

где $\Phi(z)$ – функция Лапласа.

Теоретические вероятности вычисляются в условиях выдвинутой гипотезы о законах распределения изучаемой случайной величины.

Задача 1. Даны результаты испытания стойкости 200 удлиненных сверл диаметра 4 мм (таблица 2).

Таблица 2 – Результаты испытания стойкости сверл диаметра 4 мм

Стойкость сверла x_i , ч	3...3,2	3,2...3,4	3,4...3,6	3,6...3,8	3,8...4
Частота m_i	16	50	70	44	20

Найти теоретические частоты нормального закона распределения.



Решение

Вычислим теоретические частоты. Для этого пронормируем x , т. е. перейдем к случайной величине z . Теоретические частоты нормального закона распределения вычислим по формуле (15). Составим расчетную таблицу 3.

Таблица 3 – Расчетная таблица теоретических частот

Интервал	3...3,2	3,2...3,4	3,4...3,6	3,6...3,8	3,8...4	Итого
z_{1i}	$-\infty$	-1,384	-0,468	0,449	1,366	–
z_{2i}	-1,384	-0,468	0,449	1,366	$+\infty$	–
Φ_{1i}	-0,5	-0,417	-0,18	0,173	0,414	–
Φ_{2i}	-0,417	-0,18	0,173	0,414	0,5	–
P_i	0,083	0,237	0,353	0,241	0,086	1
v_i	16,627	47,384	70,66	48,133	17,196	200

Задача для самостоятельного решения

Дан интервальный статистический ряд распределения частот экспериментальных значений случайной величины X (таблицы 4 и 5).

Требуется найти теоретические частоты нормального закона распределения.

Таблица 4 – Результаты исследования 100 напыленных образцов на прочность напыленного слоя

Прочность x_i , кг/мм ²	2,0...2,2	2,2...2,4	2,4...2,6	2,6...2,8	2,8...3,0
Частота m_i	7	22	38	23	10

Таблица 5 – Результаты исследования на разрыв 100 образцов дюралюминия

Предел прочности x_i , кг/мм ²	42...43	43...44	44...45	45...46	46...47
Частота m_i	8	25	36	22	9

Контрольные вопросы

1 Чем отличаются эмпирические частоты от теоретических в интервальном вариационном ряду?

2 Как вычислить вероятность попадания в соответствующий интервал при нормальном распределении случайной величины?

Лабораторная работа № 12. Применение критериев согласия к выборочным данным

Цель работы: изучить применение критериев согласия к выборочным данным.

При обработке статистических данных большого объема часто возникает ситуация, когда закон распределения генеральной совокупности не известен заранее. Однако сравнение гистограммы с известными кривыми функций плотностей позволяет выдвинуть гипотезу о виде распределения генеральной совокупности. Так, например, если гистограмма имеет один явно выраженный пик (рисунок 18, *a*), то можно предположить, что исследуемая генеральная совокупность распределена по нормальному закону $N(a, \sigma)$, т. е. имеет плотность

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (18)$$

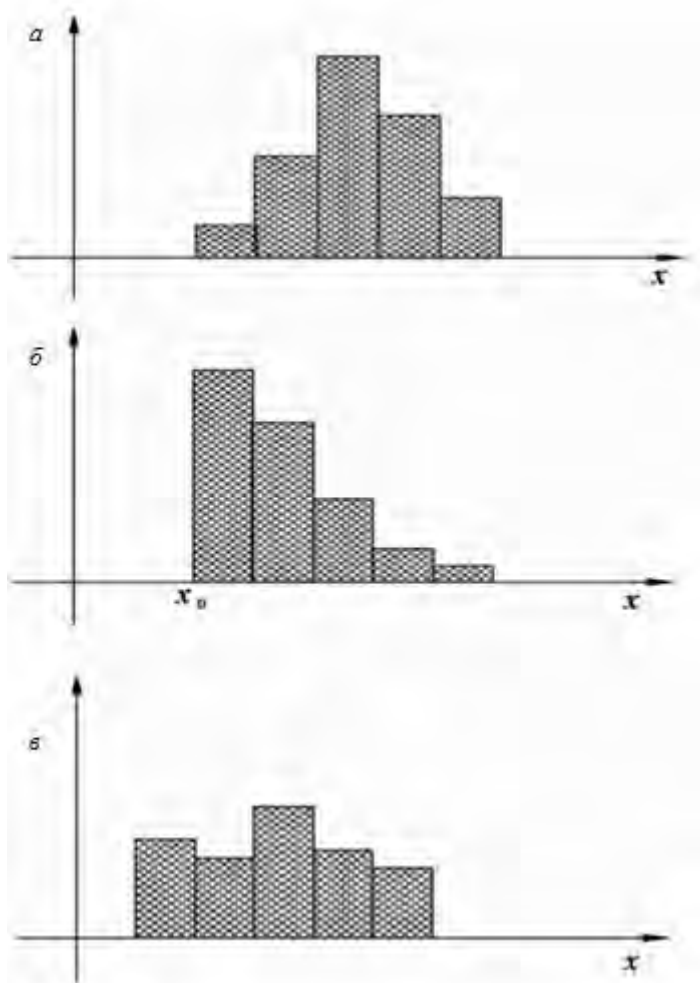


Рисунок 18 – К проверке гипотезы о законе распределения

Если гистограмма представляет собой «убывающие ступеньки прямоугольников» (рисунок 18, б), то генеральная совокупность может быть распределена по показательному закону:

$$p(x) = \begin{cases} 0, & x < x_0; \\ \lambda e^{-\lambda(x-x_0)}, & x \geq x_0. \end{cases} \quad (19)$$

Для гистограммы, представленной на рисунок 18, в, естественно выдвинуть гипотезу о равномерном распределении генеральной совокупности.

Возникает вопрос о критерии проверки по выборочным данным гипотезы о том, что случайная величина X подчиняется распределению с плотностью $y = p(x)$. Такие критерии называются *критериями согласия*.

Таким образом, критерием согласия называют критерий, который позволяет установить, является ли расхождение эмпирического и теоретического распределений случайным или значимым, т. е. согласуются ли данные наблюдений с выдвинутой статистической гипотезой или не согласуются.

Имеются несколько критериев согласия: критерий χ^2 (Пирсона), критерий Колмогорова, критерий Романовского и др.

Рассмотрим лишь один критерий согласия, использующий χ^2 -распределение и получивший название критерия согласия Пирсона (или критерия χ^2). Выдвигая гипотезу о виде распределения генеральной совокупности, необходимо различать два случая. В первом из них вид функции плотности определен в гипотезе полностью. Например, выдвигается гипотеза о том, что генеральная совокупность распределена по нормальному закону с параметрами $a = 0$ и $\sigma = 1$. Такие гипотезы называются *простыми*. Если же гипотеза состоит лишь в том, что функция плотности $p(x)$ принадлежит к некоторому семейству функций, то такая гипотеза называется *сложной*. Например, можно выдвинуть гипотезу о том, что генеральные совокупности распределены по показательному закону, не оговаривая значений параметров λ и x_0 . Такая гипотеза будет сложной.

Задача 1. По выборке объема $n = 14$ составлен группированный статистический ряд (таблица 6).

Таблица 6 – Группированный статистический ряд:

X	0...1	1...2	2...3	3...4	4...5	5...6	6...7	7...8
m_j	16	17	19	16	24	19	17	16

Проверить на уровне значимости $\alpha = 0,05$ гипотезу о равномерности распределения генеральной совокупности на отрезке $[0,8]$.

Решение

Нулевая гипотеза имеет вид:

$$H_0 : p_X(x) = p(x) = \begin{cases} \frac{1}{8-0}, & 0 \leq x \leq 8; \\ 0, & \text{для остальных } x. \end{cases}$$

Вычислим вероятность попадания случайной величины X в каждый интервал:

$$p_i = \int_{i-1}^i \frac{1}{8} dx = \frac{1}{8}(i - i + 1) = \frac{1}{8}, \quad i = 1, 2, \dots, 8.$$

Поэтому $np_i = 1/8 \cdot 144 = 18$ при любом i . Так как $np_i \geq 10$, то нет необходимости объединять несколько интервалов. Результаты дальнейших вычислений сведены в таблице 7.

Таблица 7 – Результаты вычислений

Номер интервала	m_i	np_i	$m_i - np_i$	$\frac{(m_i - np_i)^2}{np_i}$
1	16	18	-2	0,22
2	17	18	-1	0,06
3	19	18	1	0,06
4	16	18	-2	0,22
5	24	18	6	2,00
6	19	18	1	0,06
7	17	18	-1	0,06
8	16	18	-2	0,22
Σ	144	144	0	2,9

Таким образом, числовое значение $K_{наб} = 2,9$. Для заданного уровня значимости $\alpha = 0,05$ находим $\gamma = 1 - \alpha = 0,95$, $\chi^2 = (0,95; 7) = 14,1$. Так как $K_{наб} < \chi_{np, \alpha}$ то гипотеза H_0 принимается.

Контрольные вопросы

- 1 Что называется критерием согласия?
- 2 Как проверить гипотезу о виде распределения генеральной совокупности?



Лабораторная работа № 13. Решение задач на проверку статистических гипотез

Цель работы: изучить методику проверки статистических гипотез.

Под *статистической гипотезой* понимается всякое высказывание о генеральной совокупности (случайной величине X), проверяемое по выборочной совокупности (по результатам наблюдений). Не располагая сведениями о всей генеральной совокупности, высказанную гипотезу сопоставляют по определенным правилам с выборочными данными и делают вывод о том, можно принять гипотезу или нет. Эта процедура сопоставления называется *проверкой гипотезы*.

Рассмотрим этапы проверки гипотезы и используемые при этом понятия.

Этап 1. Располагая выборочными данными и руководствуясь конкретными условиями рассматриваемой задачи, формулируют гипотезу H_0 , которую называют *основной* или *нулевой*, и гипотезу H_1 , *конкурирующую* с гипотезой H_0 . Гипотезу H_1 называют также *альтернативной*.

Термин «конкурирующая» означает, что являются взаимоисключающими следующие два события:

- по выборке принимается решение о справедливости для генеральной совокупности гипотезы H_0 ;
- по выборке принимается решение о справедливости для генеральной совокупности гипотезы H_1 .

Этап 2. Задается вероятность α , которую называют *уровнем значимости*. Эта вероятность имеет следующий смысл.

Решение о том, можно ли считать высказывание H_0 справедливым для генеральной совокупности, принимается по выборочным данным, т. е. по ограниченному объему информации. Следовательно, это решение может быть ошибочным. При этом может иметь место ошибка двух родов:

- ошибка первого рода совершается при отклонении гипотезы H_0 (т. е. принимается альтернативная H_1), тогда как на самом деле гипотеза H_0 верна; вероятность такой ошибки обозначим α ;
- ошибка второго рода совершается при принятии гипотезы H_0 , тогда как на самом деле высказывание H_0 неверно и следовало бы принять гипотезу H_1 ; вероятность ошибки второго рода обозначим как

$$\beta = p(H_0 / H_1) . \quad (19)$$

Тогда уровень значимости α определяет ошибку первого рода, т. е.

$$\alpha = p(H_1 / H_0) . \quad (20)$$

Поэтому вероятность α задается малым числом, поскольку это вероятность ошибочного высказывания. При этом обычно используются стандартные значе-



ния: 0,05; 0,01; 0,005. Например, $\alpha = 0,05$ означает следующее: если гипотезу H_0 проверять по каждой из 100 выборок одинакового объема, то в среднем в пяти случаях из 100 совершим ошибку первого рода.

Обратим внимание на то, что в результате проверки гипотезы H_0 могут быть приняты *правильные решения* двух следующих видов:

- принимается гипотеза H_0 тогда, когда она верна (т. е. H_0 имеет место в генеральной совокупности); вероятность этого решения равна $p(H_0 / H_0) = 1 - \alpha$;
- не принимается гипотеза H_0 (т. е. принимается гипотеза H_1) тогда, когда и на самом деле она неверна (т. е. справедлива гипотеза H_1), вероятность этого решения равна $p(H_1 / H_1) = 1 - \beta$.

Этап 3. Определяют величину K такую, что:

- а) ее значения зависят от выборочных данных x_1, x_2, \dots, x_n , т. е. $K = K(x_1, x_2, \dots, x_n)$;
- б) будучи величиной случайной (в силу случайности выборки x_1, x_2, \dots, x_n), величина K подчиняется при выполнении гипотезы H_0 некоторому известному закону распределения;
- в) ее значения позволяют судить о расхождении гипотезы H_0 с выборочными данными. Величину K называют *критерием*.

Этап 4. В области всевозможных значений критерия K выделяют подобласть ω , называемую *критической областью*. Значения критерия, попавшие в критическую область, свидетельствуют о существенном расхождении выборки с гипотезой H_0 . Поэтому руководствуются следующим правилом: если вычисленное по выборке значение критерия попадает в критическую область ω , то гипотеза H_0 отвергается и принимается альтернативная H_1 . При этом следует помнить, что такое решение может быть ошибочным – на самом деле гипотеза H_0 может быть справедливой. Таким образом, ориентируясь на критическую область, можно совершить ошибку первого рода, вероятность которой задана заранее и равна α . Отсюда вытекает следующее требование к критической области ω .

Вероятность принятия критерием K значения из критической области ω при справедливости гипотезы H_0 должна быть равна α , т. е.

$$p(K \in \omega) = \alpha. \quad (21)$$

Контрольные вопросы

- 1 Что называется критерием, уровнем значимости, критической областью и областью допустимых значений критерия?
- 2 Что такое уровень значимости?
- 3 Что такое ошибки первого и второго рода?
- 4 Что называется мощностью критерия?
- 5 Сформулируйте этапы проверки статистических гипотез.



Лабораторная работа № 14. Проверка гипотез в EXCEL

Цель работы: получить навыки проверки статистических гипотез с использованием табличного процессора Excel.

В табличном процессоре Excel определены несколько функций и режимов работы *Пакета анализа*, которые можно использовать для проверки различных статистических гипотез.

Проверка гипотезы о числовом значении математического ожидания нормального распределения при известной дисперсии.

В качестве нулевой гипотезы H_0 принимается $H_0: a = a_0$, в качестве альтернативной $H_1: a \neq a_0$. Уровень значимости α принимается равным 0,05.

Используется функция ZТЕСТ, обращение к которой имеет вид:

$$= \text{ZТЕСТ}(\text{массив}; a_0; \sigma),$$

где *массив* – адреса ячеек, содержащих выборочные данные случайной величины, математическое ожидание которой сравнивается с заданной величиной a_0 ;

a_0 – задаваемое значение математического ожидания;

σ – задаваемое среднее квадратичное отклонение случайной величины (если этот параметр опущен, то используется выборочная дисперсия, вычисленная по той же выборке).

Результатом работы функции является корень $x_{np,0,05/2}$, т. е.

$$x_{np,0,05/2} = \text{ZТЕСТ}(\text{массив}; a_0; \sigma).$$

Величины $x_{np,0,05/2}$, $x_{лев,0,05/2} = -x_{np,0,05/2}$ определяют критические области $(-\infty, x_{лев,0,05/2}]$, $[x_{np,0,05/2}, \infty)$.

Проверка гипотезы о равенстве математических ожиданий двух нормальных распределений с известными дисперсиями.

Изучаются две нормально распределенные случайные величины $X \sim N(a_X, \sigma_X)$, $Y \sim N(a_Y, \sigma_Y)$. Числовые значения дисперсий σ_X^2 , σ_Y^2 известны. Проверяется основная гипотеза $H_0: M(X) = M(Y)$.

Для проверки этой гипотезы используется режим работы *Двухвыборочный z-тест для средних*. Для вызова этого режима необходимо обратиться к пункту **Сервис** строки меню Excel, команде *Пакет анализа*. Затем в появившемся списке режимов выбрать данный режим и щелкнуть ОК. В диалоговом окне (рисунок 19) задаются следующие параметры:

Интервал переменной 1: – адреса ячеек, содержащих выборочные значения случайной величины X .

Интервал переменной 2: – адреса ячеек, содержащих выборочные значения случайной величины Y .

Гипотетическая средняя разность: – задает число, равное предполагаемой разности математических ожиданий $a_X - a_Y$ (при проверке гипотезы о равенстве



математических ожиданий задается 0).

Рисунок 19 – Задание параметров режима *Двухвыборочный z-тест для средних*

Дисперсия переменной 1 (известная): – вводится известное значение σ_X^2 .

Дисперсия переменной 2 (известная): – вводится известное значение σ_Y^2 .

Метки – включается, если первая строка содержит заголовки столбцов.

Альфа: – задается уровень значимости.

Выходной интервал: / *Новый рабочий лист:* / *Новая рабочая книга* – указывается, куда выводятся результаты вычислений. При включении *Выходной интервал:* вводится адрес ячейки, начиная с которой выводятся результаты, которые оформлены в виде таблицы (пример такой таблицы приведен на рисунке 20).

	А	В
1	Автомат 1	Автомат 2
2	182,3	185,3
3	183,0	185,6
4	181,8	184,8
5	181,4	186,2
6	181,8	185,8
7	181,6	184,0
8	183,2	185,2
9	182,4	184,2
10	182,5	184,2
11	179,7	
12	179,9	
13	181,9	
14	182,8	
15	183,4	
16	Среднее	Среднее
17	182,0	185,0

Рисунок 20 – Исходные данные к задаче 1

Задача 1. Выборочные данные о диаметре валиков, изготовленных автоматом 1 и автоматом 2, приведены в столбцах А, В документа Excel (см. рисунок 20). Предварительным анализом установлено, что размер валиков, изготовленных каждым автоматом, имеет нормальное распределение с дисперсиями $\sigma_X^2 = 5 \text{ мм}^2$ (автомат 1) и $\sigma_Y^2 = 7 \text{ мм}^2$ (автомат 2).

Необходимо проверить нулевую гипотезу $H_0: a_x = a_y$: при альтернативной гипотезе $H_1: a_x \neq a_y$.

Решение

Обратимся к режиму *Двухвыборочный z-тест для средних* и в появившемся диалоговом окне зададим необходимые параметры, а затем щелкнем на ОК. Результаты работы режима показаны на рисунке 21.

Двухвыборочный z-тест для средних		
	Автомат 1	Автомат 2
Среднее	181,979	185,03
Известная дисперсия	5,000	7,00
Наблюдения	14,000	9,00
Гипотетическая разность средних	0,000	
z	-2,867	
P(Z<=z) одностороннее	0,002	
z критическое одностороннее	1,645	
P(Z<=z) двухстороннее	0,004	
z критическое двухстороннее	1,960	

Рисунок 21 – Результаты работы режима *Двухвыборочный z-тест для средних*

Величина z является расчетным значением $K_{наб} = z = -2,867$. Это значение попадает в критическую область $|K_{наб}| > |z_{кр}| = 1,96$. Поэтому нулевая гипотеза с уровнем значимости $\alpha = 0,05$ отвергается и принимается альтернативная.

Проверка гипотезы о равенстве математических ожиданий двух нормальных распределений с неизвестными, но равными дисперсиями.

Изучаются две нормально распределенные случайные величины $X \sim N(a_X, \sigma_X)$ и $Y \sim N(a_Y, \sigma_Y)$. Дисперсии равны, но не известны, т. е. $\sigma_X^2 = \sigma_Y^2 = \sigma^2$. Необходимо проверить статистическую гипотезу $H_0: a_X = a_Y$ при альтернативной гипотезе $H_1: a_X \neq a_Y$.

Для проверки этой гипотезы используется режим *Двухвыборочный t-тест с одинаковыми дисперсиями*. Для вызова режима необходимо обратиться к пункту **Сервис** строки меню Excel, команде *Пакет анализа*. Затем в появившемся списке режимов выбрать данный режим и щелкнуть ОК. В появившемся диалоговом окне этого режима задаются следующие параметры (рисунок 22).

Рисунок 22 – Задание параметров режима *Двухвыборочный t-тест с одинаковыми дисперсиями*

Интервал переменной 1: – адреса ячеек, содержащих выборочные значения случайной величины X .

Интервал переменной 2: – адреса ячеек, содержащих выборочные значения случайной величины Y .

Гипотетическая средняя разность: – задает число, равное предполагаемой разности математических ожиданий $\mu_X - \mu_Y$ (при проверке гипотезы $\mu_X = \mu_Y$ задается 0).

Метки – включается, если первая строка содержит заголовки столбцов.

Альфа: – задает уровень значимости α .

Выходной интервал: / *Новый рабочий лист:* / *Новая рабочая книга* – указывается, куда выводятся результаты вычислений. При включении *Выходной интервал:* вводится адрес ячейки, начиная с которой выводятся результаты, представленные в виде таблицы.

Проверка гипотезы о равенстве дисперсий двух нормальных распределений.

В качестве границ критической области выступают квантили $f_\gamma(l, k)$ распределения Фишера. Для вычисления этих квантилей используется функция ФРАСПОБР, обращение к которой имеет вид:

$$= \text{ФРАСПОБР}(\text{вероятность}; \text{степень1}; \text{степень2}),$$

где *вероятность* – уровень значимости α при построении правосторонней критической области; *степень1* – число степеней свободы l ; *степень2* – число степеней свободы k .

Граница x_{np} правосторонней критической области вычисляется с помощью выражения

$$x_{np, \alpha} = \text{ФРАСПОБР}(\alpha; l; k).$$

Граница $x_{np, \alpha/2}$ при построении двухсторонней критической области вычисляется с помощью выражения

$$x_{np, \alpha/2} = \text{FPАСПОБР}(\alpha/2; l; k).$$

Проверить гипотезу о равенстве дисперсий двух случайных величин $X \sim N(a_X, \sigma_X)$, $Y \sim N(a_Y, \sigma_Y)$ можно с использованием режима **Двухвыборочный F-тест для дисперсии**. Для вызова режима необходимо обратиться к пункту **Сервис** строки меню Excel, команде **Пакет анализа**. Затем в появившемся списке режимов выбрать данный режим и щелкнуть ОК. В появившемся диалоговом окне этого режима задаются следующие параметры (рисунок 23).

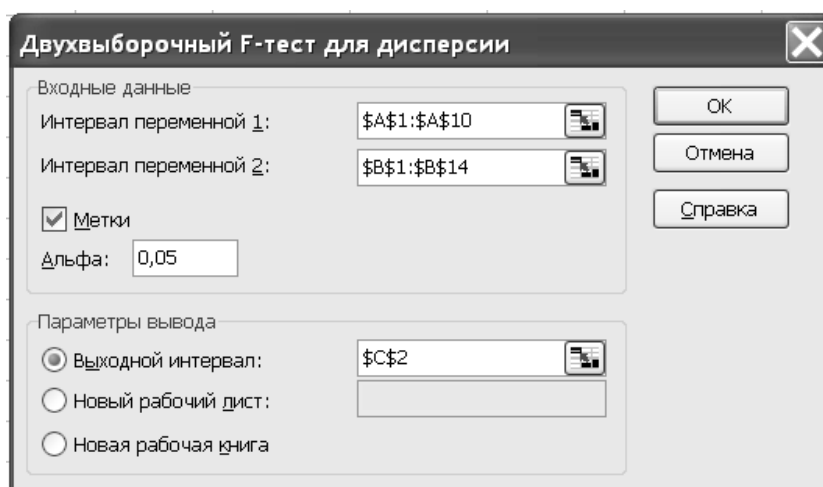


Рисунок 23 – Задание параметров режима **Двухвыборочный F-тест для дисперсии**

Интервал переменной 1: – адреса ячеек, содержащих выборочные значения случайной величины X .

Интервал переменной 2: – адреса ячеек, содержащих выборочные значения случайной величины Y .

Метки – включается, если первая строка содержит заголовки столбцов.

Альфа: – задает уровень значимости α .

Выходной интервал: / **Новый рабочий лист:** / **Новая рабочая книга** – указывается, куда выводятся результаты вычислений. При включении **Выходной интервал:** вводится адрес ячейки, начиная с которой выводятся результаты, представленные в виде таблицы.

Задача для самостоятельного решения

1 Выборочные данные о расходе сырья при производстве продукции по старой и новой технологии приведены в столбцах А, В документа Excel (рисунок 24). Предполагая, что расход сырья по старой и новой технологии распределен по нормальному закону и имеет одинаковую дисперсию, проверить статистическую гипотезу $a_X = a_Y$ при уровне значимости $\alpha = 0,05$.

	А	В
1	Старая технология	Новая технология
2	308	308
3	308	304
4	307	306
5	308	306
6	304	306
7	307	304
8	307	304
9	308	306
10	307	306
11		304
12		303
13		304
14		303

Рисунок 24 – Исходные данные

2 Выборочные данные о расходе сырья при производстве продукции по старой и новой технологии приведены в столбцах А, В документа Excel (см. рисунок 24). Предполагая, что расход сырья по старой и новой технологии распределен по нормальному закону, нужно проверить статистическую гипотезу $\sigma_X^2 = \sigma_Y^2$ при уровне значимости $\alpha = 0,05$.

Контрольные вопросы

- 1 Проверку какой гипотезы осуществляет функция Excel ZTEST?
- 2 Как выполнить проверку в табличном процессоре Excel гипотезы о равенстве математических ожиданий при известных дисперсиях?
- 3 Как выполнить проверку в табличном процессоре Excel гипотезы о равенстве математических ожиданий при неизвестных, но равных дисперсиях?
- 4 Как выполнить проверку в табличном процессоре Excel гипотезы о равенстве дисперсий двух нормальных распределений?



Лабораторная работа № 15. Вычисление коэффициента корреляции Пирсона в EXCEL

Цель работы: научиться определять силу линейной связи с помощью коэффициента линейной корреляции Пирсона.

Методические указания

При проведении статистического анализа часто необходимо отразить степень линейной зависимости между двумя массивами данных. Таким показателем является коэффициент корреляции Пирсона r – безразмерный индекс в интервале от $-1,0$ до $1,0$ включительно, который отражает степень линейной зависимости между двумя множествами данных.

В Excel имеется несколько возможностей вычисления такого результата. Одним из таких вариантов является встроенная функция Пирсона.

Синтаксис функции: ПИРСОН(массив1;массив2).

Массив1 — множество независимых значений.

Массив2 — множество зависимых значений.

Замечания:

- аргументы должны быть либо числами, либо содержащими числа именами, массивами или ссылками;
- если аргумент, который является массивом или ссылкой, содержит тексты, логические значения или пустые ячейки, то такие значения игнорируются; однако ячейки, которые содержат нулевые значения, учитываются;
- если массив1 или массив2 пуст, либо число точек данных в этих массивах не совпадает, функция ПИРСОН возвращает значение ошибки #Н/Д.
- коэффициента корреляции Пирсона r вычисляется по следующей формуле:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{(\sum (x - \bar{x})^2)(\sum (y - \bar{y})^2)}}, \quad (22)$$

где x и y – выборочные средние значения СРЗНАЧ(массив1) и СРЗНАЧ(массив2).

Для того, чтобы рассчитать коэффициент корреляции Пирсона в Excel, необходимо сделать следующие шаги:

- 1) внести значения для двух переменных в таблицу (Например, *Переменная 1* и *Переменная 2*);
- 2) установить курсор в пустую ячейку;
- 3) на панели инструментов нажать кнопку **fx** (*вставить формулу*);
- 4) в открывшемся окне «Мастер функций» в поле «Категории» выбрать *Полный алфавитный перечень*;



5) затем в поле «*Выберите функцию*» найти функцию **PEARSON**. Нажать Ок;

б) в открывшемся окне «*Аргументы функции*» в поле Массив1 внести **номера ячеек**, содержащие значения Переменной 1, в поле Массив2 внести **номера ячеек**, содержащие значения Переменной 2. Нажать Ок.

На рисунке 25 представлен результат расчета коэффициента корреляции Пирсона в EXCEL.

	A	B	C	D	E
	Переменная1,	Переменная2,			
1	x	y			
2	13	15			
3	11	14			
4	9	10			
5	7	6			
6	5	1			
7	3	5			
8	1	3			
9					
10	PEARSON	0,90			
11					

Рисунок 25 – Результат расчета коэффициента корреляции Пирсона

Контрольные вопросы

- 1 Какие виды связей между признаками в статистике Вы знаете?
- 2 Что называется корреляционной связью?
- 3 По какой формуле вычисляется коэффициент линейной корреляции?
- 4 Запишите формулу коэффициента корреляции Спирмена.

Лабораторная работа № 16. Вычисление уравнения линейной регрессии в EXCEL

Цель работы: научиться определять коэффициенты парной линейной регрессионной модели с помощью формул и с помощью встроенных функций EXCEL.

Методические указания

Для вычисления уравнения линейной регрессии в Excel можно воспользоваться несколькими способами: инструмент анализа данных *Регрессия*; добавление линии тренда на точечном графике с отображением уравнения на диаграмме;

применение встроенных функций Excel.

Рассмотрим особенности использования инструмента анализа данных *Регрессия* на примере.

Задача 1. В ходе проведения наблюдения изучена среднедневная заработная плата и среднедушевой прожиточный минимум в день одного трудоспособного жителя 12 регионов.

Вычислить уравнение линейной регрессии, отражающей зависимость среднедневной заработной платы от среднедушевого прожиточного минимума. Исходные данные представлены в таблице 8.

Таблица 8 – Результаты наблюдения

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного x , р.	Среднедневная заработная плата y , р.
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Вставим данные таблицы 8 на лист рабочей книги Excel.

В главном меню последовательно необходимо выбрать: Данные / Анализ данных / Инструменты анализа / Регрессия, а затем нажимаем кнопку ОК.

Далее заполняем диалоговое окно ввода данных и параметров вывода (рисунок 26):

– *Входной интервал Y* – диапазон, содержащий данные результативного признака;

– *Входной интервал X* – диапазон, содержащий данные факторного признака;

– *Метки* – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

– *Константа – ноль* – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

– *Выходной интервал* – достаточно указать левую верхнюю ячейку будущего диапазона;

– *Новый рабочий лист* – можно задать произвольное имя нового листа.



Рисунок 26 – Диалоговое окно ввода параметров инструмента Регрессия

Нажимаем кнопку ОК.

Результаты регрессионного анализа для данных задачи представлены на рисунке 27.

Регрессионная статистика								
Множественный R	0,721025214							
R-квадрат	0,519877359							
Нормированный R-квадр	0,471865095							
Стандартная ошибка	12,5495908							
Наблюдения	12							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	1	1705,327706	1705,327706	10,82801173	0,008141843			
Остаток	10	1574,922294	157,4922294					
Итого	11	3280,25						
Кoeffициенты		Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние
Y-пересечение	76,9764852	24,21156138	3,179327594	0,009830668	23,02976462	130,9232058	23,02976462	130,9
Переменная X 1	0,920430553	0,279715587	3,290594434	0,008141843	0,297185386	1,543675719	0,297185386	1,5436

Рисунок 27 – Результат применения инструмента Регрессия

Задание для самостоятельного решения

1 По представленным преподавателем данным определить параметры линейного уравнения регрессии между признаками x и y тремя способами:

- составив и решив систему нормальных уравнений по методу наименьших квадратов;
 - с помощью инструмента анализа данных *Регрессия*;
 - воспользовавшись встроенными функциями Excel ОТРЕЗОК(); НАКЛОН(); ТЕНДЕНЦИЯ ().
- 2 Сравнить результаты.

Контрольные вопросы

- 1 Какой признак считается факторным, а какой результативным?
- 2 Запишите систему уравнений для определения параметров уравнения парной линейной корреляции.
- 3 Какие формулы используются для расчёта коэффициентов в уравнении регрессии?
- 4 Можно ли строить уравнение парной корреляции для сгруппированных данных?
- 5 Что показывает коэффициент детерминации?

Список литературы

- 1 **Белько, И. В.** Теория вероятностей, математическая статистика, математическое программирование: учебное пособие / И. В. Белько, И. М. Морозова, Е. А. Криштапович. – Минск: Новое знание; Москва: ИНФРА-М, 2016. – 299 с.
- 2 **Вентцель, Е. С.** Задачи и упражнения по теории вероятностей: учебное пособие / Е. С. Вентцель, Л. А. Овчаров. – 8-е изд., стер. – Москва: КНОРУС, 2013. – 496 с.
- 3 **Вуколов, Э. А.** Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для вузов / Э. А. Вуколов. – 2-е изд., испр. и доп. – Москва: Форум, 2012. – 464 с.
- 4 **Гмурман, В. Е.** Теория вероятностей и математическая статистика: учебное пособие для бакалавров / В. Е. Гмурман. – 12-е изд. – Москва: Юрайт, 2012. – 479 с.
- 5 **Гусак, А. А.** Основы высшей математики: пособие для студентов вузов / А. А. Гусак, Е. А. Бричикова. – Минск: ТетраСистемс, 2012. – 208 с.
- 6 **Матальцкий, М. А.** Теория вероятностей и математическая статистика: учебник для вузов / М. А. Матальцкий, Г. А. Хацкевич. – Минск: Вышэйшая школа, 2017. – 591 с.

