МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «Тувинский государственный университет»

ФИЗИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА МАТЕМАТИКИ И МЕТОДИКИ ПРЕПОДАВАНИЯ МАТЕМАТИКИ

Выпускная квалификационная работа
(бакалаврская работа)

Проверка статистических гипотез в динамике успеваемости студентов ФМФ 2015-2019 годов поступления

Работа допущена к защите
Зав. кафедрой _____
Танзы М.В., к.п.н., доцент
(фамилия, и.о., уч.степень, звание)

Работа защищена «__»_____20__г.
С оценкой _____
Председатель ГЭК _____
(подпись)
Сенашов В.И. д.ф-м.н., профессор
ведущий научный сотрудник Института
вычислительного моделирования
СО РАН,Красноярск

Члены комиссии _____

_____

_____

_____
(подписи)

Студенты 4 курса 5 группы
направления подготовки 01.03.01
профиль «Математика»
очной формы обучения
Сундуй Чимис Юрьевна
Тюлюш Ай-Суу Артемовна

_____
(Ф. И. О.)

_____
(подпись)

«__»_____20___г.

Научный руководитель: _____
(подпись)
Жданок А.И., д.ф.-м.н., профессор

(фамилия, и. о., должность, уч.степень, звание)

Кызыл – 2020г.

Содержание

Введение…… 3
ГЛАВА 1. ИСПОЛЬЗУЕМЫЕ СВЕДЕНИЯ ИЗ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ…… 5
§1. Первичная обработка статистических данных…….. 5
§2. Статистическое оценивание…… 7
2.1. Понятие оценки параметров…….. 7
2.2. Свойства статистических оценок…….. 8
2.3. Точечные оценки математического ожидания и дисперсии……... 9
§3. Корреляционный анализ…….. 11
3.1. Коэффициент корреляции…….. 11
3.2. Парная линейная регрессия…… 13
3.3. Нелинейная (квадратичная) регрессия……. 17
ГЛАВА 2. СПЕЦИАЛИЗИРОВАННЫЙ ПАКЕТ STATISTICA…….. 19
§1. Сведения о пакете STATISTICA……. 19
§2. Общая характеристика непараметрических методов…… 22
§3. Статистическая проверка гипотез…….. 25
3.1. Постановка задачи…….. 25
3.2. Проверка гипотез об однородности выборки……... 28
3.2.1. Критерий серий Вальда – Вольфовица (Wald-Wolfowitz)….. 29
3.2.2. Двухвыборочный тест Колмогорова – Смирнова (Kolmogorov – Smirnov two – sampletest)…… 30
3.2.3. Критерий Манна – Уитни (Mann-Whitney test)…….. 31
3.2.4. Однофакторный дисперсионный анализ Краскела – Уоллиса (Kruskal – Wallis ANOVA)…….. 32
ГЛАВА 3. СБОР И СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ……. 34
§1. Оценки параметров распределения случайных величин…….. 34
§2. Корреляционный анализ……... 38
§3. Статистическая проверка гипотез…… 42
ЗАКЛЮЧЕНИЕ…… 49
ЛИТЕРАТУРА……. 51
ПРИЛОЖЕНИЕ 1…… 54
ПРИЛОЖЕНИЕ 2…… 67
ПРИЛОЖЕНИЕ 3…… 75
ПРИЛОЖЕНИЕ 4…… 88

Введение

В настоящей бакалаврской работе исследуются характеристики успеваемости студентов ФМФ за 5 лет обучения. Данная работа является частью комплексной темы научно–исследовательской работы студентов, осуществляемой с 2001 года на кафедре математического анализа и МПМ ТувГУ, руководимой д.ф.-м.н., профессором А.И. Жданком. В рамках данной тематики в 2001/02 уч. г. первой была выполнена курсовая работа Никольской О.И., Кондрашовой Н.С. «Формирование банка данных по успеваемости студентов ФМФ и его первичный статистический анализ».

История исследований по данной теме описана в статье [9].

В нашей работе мы использовали некоторые данные, полученные ранее в дипломной работе студентов Даваа Б.Н. и Монгуш А.Д. (2019 г.) [7], а также из [13] и [16].

Была проведена большая предварительная работа по уточнению и отбору оценок студентов по ведомостям в деканате ФМФ за 2015-2020 учебные годы, всего в исследование включены результаты за двенадцать семестров.

Цель бакалаврской работы заключается в том, чтобы сформулировать и проверить статистические гипотезы о характеристиках процесса успеваемости студентов ФМФ 2015-2019 годов поступления методами математической статистики.

Объект исследования – процесс обучения студентов на ФМФ.

Предмет исследования – оценки аттестатов, вступительных и сессионных экзаменов студентов ФМФ.

Для достижения поставленной цели нами выполнены следующие задачи:

– создание соответствующего банка данных на компьютере, первичная обработка данных методами математической статистики;

– установление корреляционных зависимостей между результатами ЕГЭ и результатами сессионных экзаменов, между оценками аттестата и результатами сессионных экзаменов студентов;

– построение соответствующих уравнений регрессии;

– проведение проверки гипотезы об однородности выборок (различных учебных групп студентов) с помощью методов пакета STATISTICA.

Работа состоит из введения, трех глав, заключения, списка литературы и приложений.

В первых двух Главах дипломной работы изложена теоретическая часть. В нем представлен необходимые теоретические сведения, используемые для статистической обработки собранной информации. В данном разделе мы рассматриваем такие вопросы, как статистическое оценивание и свойства статистических оценок, точечные оценки математического ожидания и дисперсии, корреляционный анализ, а также проверка статистических гипотез о неизвестных параметрах распределений, ориентируясь на источники [1], [3], [5], [6], [8], [10], [11], [12], [14], [15], [17], [18].

Последняя третья Глава является главной, она посвящена практическим результатам наших исследований. В этой главе представлены данные, по которым проводились исследования, дополнительные формулы, используемые при их обработке, а также результаты численных расчетов на ЭВМ с использованием пакета STATISTICA, представленные в виде соответствующих таблиц, графиков, диаграмм и формул на примере нескольких выбранных нами студенческих групп.

В Приложениях представлены динамика экзаменационных оценок и корреляционная зависимость, построение уравнений регрессий и их графиков, проверка гипотез об однородности выборок методами пакета STATISTICA, а также следующие из этого выводы по всем группам студентов ФМФ.

Список Литературы содержит 19 наименований использованных источников, расположенных в алфавитном порядке.

ГЛАВА 1.

ИСПОЛЬЗУЕМЫЕ СВЕДЕНИЯ ИЗ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ.

§1. Первичная обработка статистических данных.

Изучается некоторая случайная величина X. С этой целью производится ряд независимых опытов, или, как принято говорить, наблюдений, в каждом из которых величина X принимает то или иное значение. Совокупность полученных значений

$x_1, x_2, \dots, x_n$ (1)

величины X (где n — число опытов) и есть произведенная нами *выборка*. Эту совокупность часто называют *статистическим рядом*; последний играет роль исходного числового материала, подлежащего дальнейшей обработке и анализу.

Первый этап обработки ряда (1) - составление так называемого вариационного ряда. Его получают следующим образом: среди чисел (1) отбирают все различные и располагают их в порядке возрастания:

$\alpha_1, \alpha_2, \dots, \alpha_m,$ (2)

где $\alpha_1 < \alpha_2 < \cdots < \alpha_m$.

Следующий этап обработки ряда (1) – составление эмпирического закона распределения. Форма его записи зависит от характера изучаемой случайной величины X.

Пусть X – дискретная случайная величина. Тогда эмпирический закон распределения составляется в виде таблицы частот, показывающей, с какой частотой наблюдалось то или иное значение.

Таблица частот имеет вид:

В первой строке

α_1	α_2	...	α_m
μ_1	μ_2	...	μ_m

(3) записаны числа вариационного ряда (2), а во второй – их частоты, т.е. числа вида:

$$\mu_i = \frac{k_i}{n} (i = 1,2, \dots, m),$$

где n- число всех опытов, а k_i- число опытов, в котором наступало событие $X = \alpha_i$. Можно ожидать, что с увеличением числа n опытов таблица частот будет все более приближаться к истинному закону распределения величины X.

Когда величина X является непрерывной случайной величиной, то таблица частот примет вид:

x_1	x_2	...	x_m
1/n	*1/n*	...	*1/n*

Ясно, что такая таблица дает весьма малую информацию о распределении значений величины x .

Принимая во внимание сказанное, эмпирический закон распределения задают в этом случае по-другому. Его записывают с помощью, так называемой интервальной таблицы частот, имеющей вид:

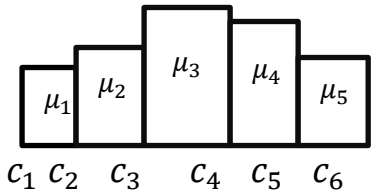
$c_1; c_2$	$c_2; c_3$	...	$c_1; c_{i+1}$
$\tilde{\mu}_1$	$\tilde{\mu}_2$	...	$\tilde{\mu}_1$

(4)

Эта запись означает, что весь диапазон изменения величины x разбит на интервалы (границами -го интервала являются c_i и c_{i+1});

число $\mu_i (i = 1,2, \dots ,1)$ есть частота попадания в -й интервал, т. е. $\tilde{\mu}_i = \frac{\tilde{k}_i}{n}$, где k_i - количество чисел в исходном ряде (1) приходящихся на i-й интервал. На практике число интервалов выбирается обычно в пределах одного-двух десятков; длины интервалов не обязаны быть одинаковыми.

Интервальная таблица частот часто изображается графически в виде так называемой гистограммы. Гистограмма представляет собой ступенчатую линию; основанием i-й ступеньки является интервал (c_i, c_{i+1}), а площадь этой ступеньки равна μ_i (рис.1).



(Рис.1)

§2. **Статистическое оценивание.**

2.1. Понятие оценки параметров.

Пусть изучается случайная величина X с законом распределения, зависящим от одного или нескольких параметров. Например, параметры α и σ для нормального закона распределения.

Требуется по выборке $x_1, x_2, \dots, x_n$, полученной в результате n наблюдений (опытов), оценить неизвестный параметр θ.

Напомним, что $x_1, x_2, \dots, x_n$ – случайные величины: x_1 – результат первого наблюдения, x_2 – результат второго и т.д., причем случайные величины $x_i, i = 1,2, \dots, n,$ имею такое же распределение, что и случайная величина x.

Статистической оценкой $\tilde{\theta}_n$ (далее просто – оценкой $\tilde{\theta}$) параметра θ теоретического распределения называют его приближенное значение, зависящее от данных выбора.

Очевидно, что оценка $\tilde{\theta}$ есть значение некоторой функции результатов наблюдения над случайной величиной, т.е. $\tilde{\theta} = \tilde{\theta}(x_1, x_2, \dots, x_n)$.

Функцию результатов наблюдений (т.е. функцию выборки) называют *статистикой*.

Можно сказать, что оценка $\tilde{\theta}$ параметра θ есть статистика, которая в определенном смысле близка к истинному значению θ.

Оценка $\tilde{\theta}$ является случайной величиной, так как является функцией независимых случайных величин $x_1, x_2, \dots, x_n$; если произвести другую выборку, то функция примет другое значение.

К оценке любого параметра предъявляется ряд требований, которым она должна удовлетворять, чтобы быть «близкой» к истинному значению параметра.

2.2. Свойства статистических оценок

Качество оценки определяют, проверяя, обладает ли она свойствами несмещенности, состоятельности, эффективности.

Оценка $\tilde{\theta}$ параметра θ называется *несмещенной*, если $M[\tilde{\theta}] = \theta$.

Если $M[\tilde{\theta}] \neq \theta$, то оценка $\tilde{\theta}$ называется *смещенной*.

Желательно, пользуясь величиной $\tilde{\theta}$ вместо θ не делать систематических ошибок ни в сторону занижения, ни в сторону завышения.

Оценка $\tilde{\theta}$ параметра θ называется *состоятельной*, если она сходится по вероятности к оцениваемому параметру: $\tilde{\theta}_n \to \theta$ при $n \to \infty$.

Это означает, что с увеличением объема выборки мы все ближе приближаемся к истинному значению параметра θ, т.е. практически достоверно $\tilde{\theta}_n \approx \theta$. Т.е. если оценка $\tilde{\theta}_n$ параметра θ является несмещенной и $D[\tilde{\theta}] \to 0$ при $n \to \infty$, то $\tilde{\theta}_n$- состоятельная оценка.

Свойство состоятельности обязательно для любого правила оценивания.

Несмещенная оценка $\tilde{\theta}_n$ параметра θ называется *эффективной*, если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра θ, т.е. оценка $\tilde{\theta}_n$ эффективна, если ее дисперсия минимальна.

На практике не всегда удается удовлетворить перечисленным требованиям, так как при прочих условиях желательно, чтобы выражение для

функции $\tilde{\theta}(x_1, x_2, \dots, x_{n)}$ было не слишком сложно. В некоторых случаях, например, применяют незначительно смещенные оценки.

2.3. Точечные оценки математического ожидания и дисперсии.

Пусть изучается случайная величина x с математическим ожиданием $M[x]$ и дисперсией $D[x]$; оба параметра неизвестны.

Точечной называют оценку, которая определяется одним числом. При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, т. е. приводить к грубым ошибкам. По этой причине при небольшом объеме выборки следует пользоваться интервальными оценками. То есть точечная оценка характеристики генеральной совокупности – это число, определяемое по выборке.

Пусть $x_1, x_2, \dots x_n$ – это выборка, полученная в результате проведения n независимых наблюдений за случайной величиной x. Чтобы подчеркнуть случайный характер величин $x_1, x_2, \dots, x_n$, перепишем их в виде $X_1, X_2, \dots, X_n$, т.е. под x_iбудем понимать значение случайной величины X в i–м опыте.

Пусть $x_1, x_2, \dots, x_n$ – выборка из генеральной совокупности и $M[x_i] = M[X] = a,\ D[x_i] = D[X](i = 1,2,..,n)$. Тогда выборочное среднее $\widetilde{m} = \frac{1}{n}\sum_{i=1}^{n} x_i$ - это несмещенная и состоятельная оценка математического ожидания $M[X]$.

Следуя из того, что случайные величины $x_1, x_2, \dots, x_n$ имеют один и тот же закон распределения (он совпадает с законом распределения величины X), поэтому $M[\widetilde{m}] = \frac{1}{n}\sum_{i=1}^{n} M[x_i] = \frac{1}{n} nM[X] = m$ (1)

Дисперсия этой оценки $D[\widetilde{m}] = \frac{1}{n^2}\sum_{i=1}^{n} D[x_i] = \frac{1}{n^2} nD[X] = \frac{D}{n}$ (2)

где D – дисперсия случайной величины X. Отсюда вытекает состоятельность

оценки $\widetilde{m}$.

Так как по самому определению D дисперсия есть математическое ожидание случайной величины $(X - \widetilde{m})^2$, то естественной оценкой для D представляется выражение $\widetilde{D} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widetilde{m})^2$ – среднее арифметическое квадратов отклонений от эмпирического среднего $\widetilde{m}$; эту оценку называют *выборочной дисперсией*.

Можно показать, что $\widetilde{D}$ является состоятельной оценкой. Однако, $\widetilde{D}$ смещена относительно D. Подсчет, который мы проведем в заключительной части параграфа, показывает, что математическое ожидание величины $\widetilde{D}$ не совпадает с числом D, а несколько меньше последнего: $M\left[\widetilde{D}\right] = \frac{n-1}{n}D,$ (3)

Отсюда видно, что несмещенной оценкой дисперсии является величина $\widetilde{\widetilde{D}} = \frac{n-1}{n}\widetilde{D} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widetilde{m})^2$.

Действительно, $M[\widetilde{\widetilde{D}}] = \frac{n}{n-1}M\left[\widetilde{D}\right] = D$.

Оценку $\widetilde{\widetilde{D}}$ условимся называть *несмещенной выборочной дисперсией*. Так как $\widetilde{\widetilde{D}}$ отличается от $\widetilde{D}$ множителем $\frac{n}{n-1}$, стремящимся к единице при $n \to \infty$, то при больших значениях n практически безразлично, какой из двух оценок мы пользуемся.

С чисто технической точки зрения вычисление оценки $\widetilde{D}$ удобно производить с помощью формулы $\widetilde{D} = \widetilde{m}_x^2 - (\widetilde{m}_x)^2,$ (4)

являющейся аналогом формулы $D[X] = M[X^2] - M^2[X]$ (5)

Вывод формулы (4) аналогичен выводу формулы (5) имеем:

$\left[\widetilde{D}\right] = \frac{1}{n}\sum(x_i - \widetilde{m})^2 = \frac{1}{n}\sum(x_i^2 - 2\widetilde{m}x_i + \widetilde{m}^2) = \frac{1}{n}(\sum x_i^2 - 2\widetilde{m}\sum x_i + n\widetilde{m})$.

§3. Корреляционный анализ.

3.1 Коэффициент корреляции.

Пусть в некотором опыте наблюдается две случайные величины ***x*** и ***y***. Тот факт, что ***x*** и ***y*** обусловлены одним и тем же опытом, вообще говоря, создает между этими величинами некоторого рода связь; как принято говорить **x и y с коррелированы** (согласованы)друг с другом. Такого рода зависимость между переменными величинами называется **корреляционной**, или **корреляцией**.

Первой характеристикой корреляции служит корреляционный момент (момент связи).

Корреляционным моментом двух случайных величин ***x***и ***y*** называется математическое ожидание произведения отклонений этих случайных величин от их математического ожидания и обозначается через K_{xy}.

$$K_{xy} = M[(x - m_x)(y - m_y)]$$

Где m_x и m_y – математические ожидания величин *x* и *y* соответственно.

Корреляционный момент (от латинского слова correlation – соответствие, взаимосвязь), играет определенную роль при оценке зависимости x и *y*. Если величины *x* и *y* независимы, то их корреляционный момент равен нулю. Поэтому неравенство нулю величины K_{xy} свидетельствует о наличии связи между *x* и *y*.

Две случайные величины x и y называют ***коррелированными,*** если их корреляционный момент (коэффициент корреляции r) отличен от нуля; x и указывают ***некоррелированными,*** если их корреляционный момент равен нулю.

Так как K_{xy} является размерной величиной, то вводится вторая безразмерная характеристика связи между x и y – *коэффициент корреляции*.

Коэффициентом корреляции случайных величин x и y называется число

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}$$

— отношение корреляционного момента к произведению средних квадратичных отклонений величин x и y, где σ_x и σ_y – среднеквадратическое отклонение величин *x* и *y,* вычисляемое по формуле:

$$\sigma_x = \sqrt{D[x]}$$

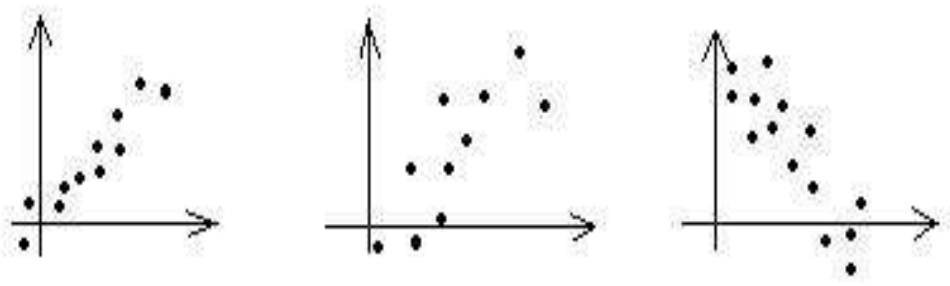
Коэффициент корреляции всегда заключен между -1 и 1:

В случае, когда r=1, величины x и y связаны линейной зависимостью

$$y = ax + b\ (a, b = const), \text{причем } a > 0$$

При r=-1 между величинами x и y имеет место линейная зависимость с a<0.

Близость величины r 2 к единице есть признак того, что зависимость между x и y *близка к линейной.* Если при этом r>0, то с возрастанием x возрастает *в среднем* и y, тогда говорят о *положительной корреляции* между величинами x и y; если же r<0, то при возрастании x величина y в среднем убывает, в этом же случае говорят об *отрицательном корреляции.* Иллюстрацией служит рисунок.



Точки на этом рисунке изображают результаты наблюдений над системой (x,y); каждое наблюдение дает пару значений (x,y), т.е. точку на плоскости**.**

Максимальное возможное абсолютное значение коэффициента корреляции |***r***| = ***1,00***; минимальное |***r***| = ***0,00*** (это следует из того, что в этом случае $K_{xy} = 0$).

Общая классификация корреляционных связей : сильная, или **тесная** при коэффициенте корреляции $|r| > 0{,}70$;

Сильная	$0{,}70 < \|r\| < 1{,}00$
Средняя	$0{,}50 < \|r\| < 0{,}69$
умеренная	$0{,}30 < \|r\| < 0{,}49$
Слабая	$0{,}20 < \|r\| < 0{,}29$
очень слабая	$0{,}00 < \|r\| < 0{,}19$

Переменные x и y могут быть измерены в разных шкалах, именно это определяет выбор соответствующего коэффициента корреляции.

3.2 Парная линейная регрессия

Под регрессионным анализом понимают исследование закономерностей связи между явлениями (процессами), которые зависят от многих, иногда неизвестных, факторов. Часто между переменными x и y существует связь, но не вполне определенная, при которой одному значению x соответствует несколько значений (совокупность) y. Таким образом, функция y=f(x) является регрессионной, если каждому значению аргумента соответствует статистический ряд распределения y.

Задача регрессионного анализа состоит в определении случайной величины результата, если случайные величины факторов, от которых статистически зависит этот результат, приняли конкретные значения. Для

этого корреляционную зависимость результата от факторов требуется отобразить с помощью функциональной зависимости, которая представляется уравнением регрессии. Необходимо определить такой вид и параметры уравнения регрессии, при которых наиболее точно осуществляется такое представление с помощью коэффициента частной корреляции или с помощью t-критерия (критерия Стьюдента). Качество уравнения регрессии характеризуется коэффициентом корреляции или с помощью критерия Фишера. Экономический анализ на основе уравнений регрессии проводится посредством интерпретации полученных коэффициентов этих уравнений, определения вклада факторов в изменение результирующего показателя, а также прогнозирования результатов экономической деятельности.

Для расчета коэффициентов линейного уравнения регрессии используется метод наименьших квадратов.

Уравнение парной регрессии. Связь зависимой переменной с одной или несколькими независимыми переменными представляют в виде уравнения

регрессии $\hat{y}=f\ (x_1\ ...x_n\ )$. Уравнение регрессии - наиболее часто встречающийся в практике вид статистической модели. Подобные модели применяются:

1) непосредственно для экономического и технико- экономического анализа, где с помощью уравнений регрессии измеряют влияние отдельных факторов на зависимую переменную. Тем самым анализ становится более конкретным, познавательная его ценность существенно увеличивается.

2) уравнения регрессии интенсивно применяются в прогностических работах.

Уравнения регрессии: решение двух основных задач:

Первая заключается в выборе независимых переменных, существенно влияющих на зависимую величину, и в определении вида уравнения

регрессии. Этот этап в разработке регрессии называется спецификацией. Данная задача решается с помощью качественного анализа изучаемой взаимосвязи.

Большую пользу выборе вида уравнения могут оказать графики. Рассмотрим простой случай. Пусть необходимо описать в виде некоторой функции зависимость двух переменных величин x и y.

Если зависимость y от x существует, то эмпирические данные покажут, что Y_x (этим символом мы обозначаем условную среднюю переменной y для фиксированных значений x) изменяется с изменением значения x. Соединим на корреляционном поле значения Y_x , получим ломаную линию. Поскольку Y_x являются выборочными оценками, то с увеличением объема наблюдений ошибки выборки этих средних будут уменьшаться, а ломаная линия приближаться к некоторой сглаженной кривой - линии регрессии.

При неограниченно большом числе наблюдений, все значения Y_x должны лежать на кривой, если, повторяем, между переменными действительно существует зависимость соответствующего вида. Практически, однако, число наблюдений всегда ограничено и эмпирическая линия, характеризующая зависимость Y от x будет в большей или меньшей мере отличаться от плавной кривой.

Вторая задача-оценивание параметров (коэффициентов) уравнения-решается с помощью того или иного математико-статистического метода обработки данных.

Линейное уравнение парной регрессии.

Пусть между переменными x и y теоретически существует некоторая линейная зависимость. Однако в действительности между x и y наблюдается не столь жесткая связь. Отдельные наблюдения y будут отклоняться от линейной зависимости в силу воздействия различных причин. Обычно зависимая переменная находится под влиянием целого ряда факторов, в том числе и не известных исследователю, а также

случайных причин (возмущении и помехи); существенным источником отклонений в ряде случаев являются ошибки измерения. Отклонения от предполагаемой формы связи, естественно, могут возникнуть и в силу неправильной спецификации уравнения, т.е. неправильного выбора вида самого уравнения, описывающего эту зависимость.

Полагаем, что спецификация выполнена правильно. Учитывая возможные отклонения, линейное уравнение связи двух переменных (парную регрессию) представим в виде:

$$y = \alpha x + \beta + \varepsilon \quad (1)$$

где y характеризует некоторое среднее значение y для данного значения x. Соответственно уравнение (1) показывает значения y с учетом возможных отклонений от средних значений.

По поводу возмущения ε сделаем следующие предположения:

1) возмущение является нормально распределенной случайной переменной;

2) математическое ожидание равно нулю: M(ε)=0

3) дисперсия возмущений постоянна: $\sigma_\varepsilon^2 = const$;

4) Последовательные значения ε не зависят друг от друга.

При построении регрессии (в данном случае линейной парной регрессии) применяется гипотеза о том, что для каждого наблюдения I справедлива следующая зависимость:

$$y = \alpha x_i + \beta + \varepsilon_i$$

В итоге статистического наблюдения исследователь имеет ряд характеристик независимой переменной x и соответствующие значения зависимой переменной y. Задача заключается в определении параметров α и β. Однако истинные значения этих параметров получить нельзя, так как мы опираемся на выборку ограниченного объема. Поэтому найденные значения

параметров являются статистическими оценками истинных (неизвестных нам) параметров. Обозначим соответствующие оценки как a и b. Таким образом уравнение парой регрессии $\hat{y} = ax + b$ есть оценка модели $y = \alpha x + \beta + \varepsilon$

При построении уравнения линейной регрессии используются

$$a = \frac{\frac{1}{n}\sum y \frac{1}{n}\sum x^2 - \frac{1}{n}\sum x \frac{1}{n}\sum yx}{\frac{1}{n}\sum x^2 - (\frac{1}{n}\sum x)^2}$$

$$b = \frac{\frac{1}{n}\sum xy - \frac{1}{n}\sum x \frac{1}{n}\sum yx}{\frac{1}{n}\sum x^2 - (\frac{1}{n}\sum x)^2}$$

3.3. Нелинейная (квадратичная) регрессия.

Рассмотрим нелинейную регрессию с двумя независимыми переменными *x* и *y*, т.е. парную регрессию, с уравнением вида:

$$y = a_0 + a_1 x + a_2 x^2 \quad (1)$$

Оценка параметров этого уравнения производится на основе выборочных данных; в каждом наблюдении (обозначим номер наблюдения через *i*).

Значения a_0 , a_1 и a_2 , очевидно, должны быть подобраны так, чтобы минимизировать сумму квадратов отклонений $\sum(y_i - y_i)^2$, где i =1, …, *n;n*— число наблюдений.

Взяв частные производные от $\sum(y_i - y_i)^2$ по a_0 , a_1 и a_2 , и приравняв их нулю, получим после несложных преобразований следующую систему нормальных уравнений.

$\sum y_i = n a_0 + a_1 \sum x_i + a_2 \sum x_i^2$;

$$\sum y_i x_i = a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i x_i^2; \quad (2)$$

$$\sum y_i x_i^2 = a_0 \sum x_i^2 + a_1 \sum x_i x_i^2 + a_2 \left(\sum (x_i^2)^2\right).$$

Оценку искомых параметров можно осуществлять, применяя систему (2). Удобнее, однако, найти готовые формулы для расчета параметров, а не решать систему каждый раз. Для этого решим первое уравнение относительно a_0:

$$a_0 = \frac{\sum x_i}{n} - a_1 \frac{\sum y_i}{n} - a_2 \frac{\sum x_i^2}{n} \qquad (3)$$

Во (2) и (3) уравнениях системы заменим исходные переменные, т. е. y_i,x_{i1} и $2x_{i2}$,на соответствующие отклонения от средних y_i',x'_iих$'_i{}^2$:

$$\sum y'_i x'_i = a_1 \sum x'_i + a_2 \sum x'_i x'^2_i \;;$$

$$\sum y'_i x'_{i\,2} = a_1 \sum x'_i x'^{\,2}_i + a_2 \left(\sum (x'^{\,2}_i)^2\right). \qquad (4)$$

Решим систему (4) относительно a_1и a_2:

$$a_1 = \frac{(\sum x'^2_i)(\sum y'_i x'_i) - (\sum x'_i x'^2_i)(\sum (y'_i x'^2_i))}{D} \;; \qquad (5)$$

$$a_2 = \frac{(\sum x'^2_i)(\sum y'_i x'^2_i) - (\sum x'_i x'^2_i)(\sum (y'_i x'_i))}{D},$$

$$\text{где, } D = (\sum x'^{\,2}_i)(\sum (x'^{\,2}_i)^2) - (\sum x'_i x'^{\,2}_i)^2. \qquad (6)$$

ГЛАВА 2.

СПЕЦИАЛИЗИРОВАННЫЙ ПАКЕТ STATISTIKA.

§1. Сведения о пакете STATISTIKA.

Изложенные материалы в данной главе ориентированы на источники [4],[19].

Универсальный статистический пакет STATISTICA (разработчик – компания StatSoft, с 2008 г. Золотой партнер MicrosoftCorp) предназначен для всеобъемлющего статистического и графического анализа, управления, прогнозирования и добычи данных, создания пользовательских приложений и их интеграции с другими программными системами.

Анализ данных пакета проводится в режиме последовательно открывающихся диалоговых окон, каждое из которых сконструировано так, что на первой вкладке содержатся только самые необходимые кнопки, а на последующих вкладках – углубленные методы и специальные опции.

Структура пакета – *модульная.* Модули содержат уникальные процедуры и методы анализа и отображения данных.

Достоинства пакета STATISTICA

- *Удобный интерфейс и уникальные возможности настройки* –интуитивно понятны, аналогичны интерфейсу Windows-приложений.

- *Непревзойденная графика* –поддержка более10 000графиков различных типов, удобных для редактирования и обработки с использованием операций вращения, масштабирования, изменения прозрачности и др.

- *Высокая скорость обработки данных при работе с файлами большого объема–*

используются собственные технологии оптимизации, а также ключевые

алгоритмы интеллектуального моделирования.

•*Повышенная точность вычислений*–обеспечивается проведение расчетов для переменных с небольшой относительной дисперсией.

•*Широкие возможности интеграции и совместимости, простой*

импорт/экспорт данных, легкий доступ к БД–поддержка технологий связи,

обработки и динамического обмена объектов и документов ActiveX, OLE, DDE.

•*Автоматизация любых процедур с помощью STATISTICA VisualBasic*–в

дополнение к стандартным доступны более 13 000 новых функций; создание макросов, интеграция с MS Excel, языками программирования C++, Java и др.

•*Поддержка Web-технологий* –используется многоуровневая архитектура клиент-сервер, обеспечивающая проведение анализа и обновление содержания HTML-страниц на Web-сервере.

Возможности пакета STATISTICA:

1.*Группировка*–категоризация данных по различным факторам

2.*Корреляция*–вычисление характеристик связи между переменными–коэффициентов корреляции Пирсона, ранговой корреляции Спирмена, сопряженности, неопределенности, автокорреляции, мер расхождения и др.

3.*Таблицы частот, сопряженности, флагов и заголовков, анализ многомерных откликов*

4.*Методы множественной регрессии*–простая, множественная, пошаговая, иерархическая, нелинейная, гребневая, взвешенных наименьших квадратов.

5.*Непараметрические статистики*–различные критерии

6.*Дисперсионный анализ*–одномерный, многомерный и ковариационный анализ (Ковариата– переменная, которая может влиять на взаимосвязь между изучаемыми переменными, однако исследуется сама по себе).

7.*Подгонка распределений*-для сравнения с теоретическими распределениями.

Перевод *STATISTICA* на русский язык выполнен группой технических специалистов StatSoftRussia под руководством д-ра Владимира Боровикова.

Пакет *STATISTICA* содержит большое число статистических процедур. Однако разобраться с особенностями процедуры, предположениями, при которых она может быть использована, и , наконец, с вопросами выбора той или иной процедуры при решении конкретной задачи можно только обратившись к специальной литературе.

Для студента или специалиста, желающего использовать статистические методы в своей области знаний, такую литературу либо трудно найти, либо весьма сложно читать, так как она написана на достаточно высоком математическом уровне. Описание статистических методов в руководствах, имеющихся в документации пакетов обычно не полно и мало понятно.

Наиболее полной для версии *STATISTICA*6.1.478 является книга Вуколова Э.А. [4], которую мы и использовали в своей работе.

§2. Общая характеристика непараметрических методов.

В своем исследовании по статистике успеваемости студентов ФМФ мы использовали, в основном, непараметрические методы, имеющиеся в пакете STATISTICA.

Основные методы математической статистики: оценка параметров распределения, проверка статистических гипотез, дисперсионный анализ применяются в предположении, что распределение генеральной совокупности известно. В частности t-критерий для сравнения средних двух генеральных совокупностей и однофакторный дисперсионный анализ, для сравнения средних нескольких совокупностей пригодны только в случае нормального распределения последних. Часто встречаются данные, для которых эти предположения не выполняются. Например, результаты социологических опросов имеют форму ответов вида «да» или «нет» и представляются в виде таблиц. Традиционные методы математической статистики не могут быть использованы для обработки таких данных. В этих случаях используются непараметрические методы, т.е. методы независящие от распределения генеральной совокупности.

Такие методы применяются для качественных данных, представленных в номинальной шкале и для данных, измеряемых в порядковой шкале (то есть представленных в виде рангов), а также для количественных данных в том случае, когда распределение генеральной совокупности неизвестно.

Приступая к решению конкретной задачи, необходимо выбрать тот или иной метод. Помощь в таком выборе может оказать следующая классификация непараметрических методов, используемых для проверки гипотезы о том, что анализируемые данные – это выборки из однородных генеральных совокупностей.

Первым критерием для выбора метода является вид шкалы,в которой представлены исходные данные.

Вторым критерием является вид выборок(независимые или связанные)*и их количество.*

Поясним понятие связанной выборки. Если над каждым из n *объектов проводятся два наблюдения: одно до, а другое после некоторого воздействия (приема лекарства, обучения, рекламной компании и т.д.), то результаты наблюдений представляют две связанные (зависимые) выборки объема* n*.*

Таким образом, рассматриваемые ниже методы можно классифицировать следующим образом.

1. *Исходные данные: две независимые выборки объемов* n_1 *и* n_2*.*

Проверяемая гипотеза H_0: выборки принадлежат однородным генеральным совокупностям.

Методы:

1.Критерий серий Вальда-Вольфовица

2. Двух выборочный критерий Колмогорова – Смирнова

3.Критерий Манна-Уитни

2. *Исходные данные*: *пары наблюдений* (x_i,y_i),i=1,2,…,n *двух признаков X и У, измеренных в порядковых или количественных шкалах.*

Проверяемая гипотеза H_0: признаки X и У некоррелированные.

Меры статистической зависимости: ранговый коэффициент корреляции Спирмена, коэффициент корреляции τ Кендалла.

3. *Исходные данные :*k *независимых выборок объемов* n_1, n_2,...,n_k.

Проверяемая гипотеза H_0*:* выборки принадлежат однородным генеральным совокупностям.

Методы:

1) Однофакторный дисперсионный анализ КраскелаУоллиса;

2) Медианный критерий.

4. *Исходные данные: две связанные выборки объемов* n.

Проверяемая гипотеза H_0: выборки принадлежат однородным генеральным совокупностям.

Методы:

1) Критерий знаков;

2) Критерий Вилкоксона.

5. *Исходные данные*: k *связанных выборок объемов* n.

Проверяемая гипотеза H_0: выборки принадлежат однородным генеральным совокупностям.

Методы:

1) Двухфакторный анализ Фридмана;

2) Меры связи – коэффициент конкордации Кендалла

(из социологического словаря: Конкордация- процесс формирования установок, интересов, умений и навыков личности, позволяющих ей осуществлять свою жизнедеятельность согласованно с деятельностью других членов общества.)

6. *Связанные выборки, измеряемые в номинальной шкале*

1) *Исходные данные*: две связанные выборки объемов n переменных X и У, каждая из которых принимает два значения (0,1; +,-; и т.д.).

Проверяемая гипотеза H_0*:* эффект воздействия отсутствует.

Метод: критерий Макнимара.

2)*Исходные данные*: k связанных выборок объемов n переменных каждая из которых принимает два значения.

Проверяемая гипотеза H_0*:* эффект воздействия отсутствует.

Метод: критерий Кокрена.

7. *Выборки, измеряемые в номинальной шкале*

1)Исходные данные: выборки двух случайных переменных X и У, каждая из которых принимает два значения.

Проверяемая гипотеза H_0:X и У независимы.

Метод: анализ таблицы сопряженности 22 (точный критерий Фишера, критерий χ^2).

Таблица сопряженности - средство представления совместного распределения двух переменных, предназначенное для исследования связи между ними. Таблица сопряженности является наиболее универсальным средством изучения статистических связей, так как в ней могут быть представлены переменные с любым уровнем измерения. Строки таблицы сопряженности соответствуют значениям одной переменной, столбцы - значениям другой переменной (количественные шкалы предварительно должны быть сгруппированы в интервалы). На пересечении строки и столбца указывается частота совместного появления f_{ig} соответствующих значений двух признаков x_i и x_j.

2) Исходные данные: выборки двух переменных X и У, представленных в номинальных шкалах. X принимает k значений, Y – r значений.

Проверяемая гипотеза H_0 : X и У независимы.

Метод: анализ таблицы сопряженности k×r (критерий χ^2).[17;18]

§3. Статистическая проверка гипотез.

3.1. Постановка задачи

Статистической гипотезой называется любое предположение о виде неизвестного закона распределения или о параметрах известных распределений. Предположим, что на основании имеющихся данных есть основания выдвинуть предположения о *законе распределения* или о параметре закона распределения случайной величины (или генеральной совокупности, на множестве объектов которой определена эта случайная величина).

Задача проверки статистической гипотезы заключается в подтверждении или опровержении этого предположения на основании выборочных (экспериментальных) данных.

Проверка статистической гипотезы означает проверку соответствия выборочных данных выдвинутой гипотезе. Статистическая гипотеза, которая проверяется, называется *основной* (нулевой) и обозначается H_0. Гипотеза, которая противопоставляется основной, называется *альтернативной* (конкурирующей) и обозначается H_1. Цель статистической проверки гипотез: на основании выборочных данных принять решение о справедливости основной гипотезы или отклонить в ее пользу альтернативной.

Проверка осуществляется на основании выборки, а не всей генеральной совокупности, то существует вероятность, возможно, очень малая, ошибочного заключения.

Нулевая гипотеза может быть отвергнута, в то время как в действительности в генеральной совокупности она является справедливой. Такую ошибку называют ***ошибкой первого рода,*** а ее вероятность— ***уровнем значимости*** и обозначают α.Возможно, что нулевая гипотеза принимается, в то время как в генеральной совокупности справедлива альтернативная гипотеза. Такую ошибку называют ***ошибкой второго рода,*** а ее вероятность обозначают β (табл.1).

Таблица1. Результаты проверки статистической гипотезы.

Принятое	В генеральной совокупности гипотеза Но	
	Верна	Неверна
$H_{0\ отвергнута}$	Ошибка 1 рода $P(H_1\|H_0)=\alpha$	Правильное решение $P(H_1\|H_1)$= 1-β
Н0 принята	Правильное решение $P(H_0\|H_0)$= 1- α	Ошибка 2 рода $P(H_0\|H_1)$= β

Статистическим критерием *К называется случайная величина,* которая *используется с целью проверки нулевой гипотезы. Вероятность 1* - β называют ***мощностью*** критерия. *Проверка статистических гипотез осуществляется с помощью статистического критерия.*

Проверяя статистические гипотезы, принято задавать заранее уровень значимости α (стандартные значения: 0.1, 0.05, 0.01, 0.001). Тогда из двух критериев, характеризующихся одной и той же вероятностью α, выбирают тот, которому соответствует меньшая ошибка 2-го рода, т.е. большая мощность. Уменьшить вероятности обеих ошибок α и β одновременно можно, увеличив объем выборки.

Очевидно, значения критерия К разделяются на две части; область ***допустимых значений*** *(область принятия гипотезы* H_0*)и* ***критическую область****(область принятия гипотезы* H_1*).Критическая область состоит из тех же значений критерия К, которые маловероятны при справедливости гипотезы* H_0*. Если значение К набл. Критерия* К*, рассчитанное по выборочным данным, попадает в критическую область, то гипотеза* H_0 *отвергается в пользу альтернативной* H_1*; в противном случае мы утверждаем, что нет оснований отклонять гипотезу* H_0**.**

При построении критической области, нужно знать закон распределения статистики К при условии, что гипотеза H_0 *справедлива. Уровень значимости* α *(вероятность наблюдаемому значению К* набл. *попасть в критическую область) определяет «размер» критической области, а конкурирующая гипотеза* H_1*— «форму» критической области. Например, если проверяется гипотеза* H_0*:*$\theta=\theta_0$*,а в качестве альтернативы —* H_1*:*$\theta>\theta_0$*, то критическая область будет правосторонней (рис.1, а). При альтернативе* H_1*:*$\theta< \theta_0$ *критическая область — левосторонняя (рис.1, б). При альтернативе критическая область* H_1*:*$\theta=\theta_0$*— двусторонняя (рис.1,в). Во всех этих случаях при заданном уровне значимости* α *заштрихованная*

*площадь составляет 100*α% от всей площади под кривой плотности распределения статистики К.*

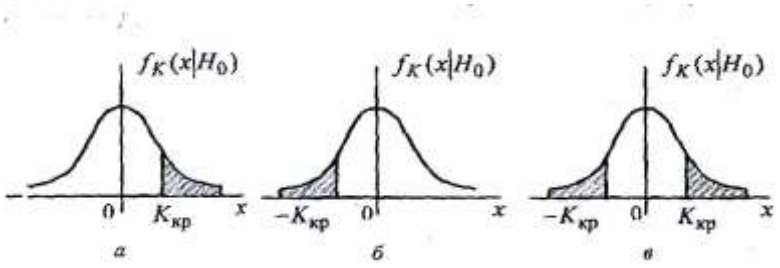


Рис.1 Правосторонняя(а), левосторонняя (б) и двусторонняя (в) критические области.

Алгоритм проверки статистических гипотез сводится к следующему:

1) сформулировать основную Н0 и альтернативную Н1гипотезы
2) выбрать уровень значимости α
3) в соответствии с видом гипотезы Н0 выбрать статистический критерий для ее проверки, т.е. случайную величину К, распределение которой известно
4) по таблицам распределения случайной величины К найти границу критической области $K_{крит}$ (вид критической области определить по виду альтернативной гипотезы H_1);
5) по выборочным данным вычислить наблюдаемое значение критерия *К набл.*
6) принять статистическое решение: если К набл. Попадает в критическую область – отклонить гипотезу H_0 в пользу альтернативной H_1; если $K_{набл}$ Попадает в область допустимых значений, то нет оснований отклонять основную гипотезу.

3.2. Проверка гипотез об однородности выборки.

Допустим, что выборки извлечены из одной генеральной совокупности требует проверки. Мы имеем гипотезу об однородности данных. Для проверки гипотез об однородности выборок служат специальные критерии — ***критерии согласия.*** Они отвечают на вопрос: представляют ли случайные независимые выборки группы данных из одной генеральной совокупности.

3.2.1 Критерий серий Вальда-Вольфовица (Wald – Wolfowitz).

Критерий серий применяется для проверки гипотезы H_0,утверждающей, что две группы данных представляют случайные независимые выборки с объемами n_1 и n_2 из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку.

Результаты наблюдений записываются в виде вариационного ряда объединенной выборкой, а принадлежность данных к той или иной группе определяется с помощью кодирующей переменной, принимающей два значения (0 и 1,+ и - , 1 и 2 и т.д.). Полученную таким образом последовательность назовем *последовательностью кодов*.

Серией в последовательности кодов называется всякая под последовательность, состоящая из одинаковых кодов и ограниченная противоположными кодами, либо находящаяся в начале или конце исходной последовательности. Например, в последовательности кодов: 00111011111000111 имеется шесть серий: (00), (111), (0), (11111),(000),(111).

Статистикой критерия является число серий N в последовательности кодов. Если гипотеза H_0 верна, то обе выборки должны быть хорошо перемешаны в общем вариационном ряду и число серий N должно быть велико. Если же выборки получены из генеральной совокупности с разными распределениями, различающимися средними значениями или разбросом, то число серий N , по-видимому, будет мало.

Критическая область определяется неравенствами: $N \leq N_1$ и $N \geq N_2$, где значения N_1 и N_2 определяются по объему выборок n_1 и n_2 и уровню значимости α .

При больших объемах выборок ($n_1>20$ и /или $n_2>20$) для проверки гипотезы H_0 можно использовать статистику Z:

$$Z = \frac{\left|N - (\frac{2n_1n_2}{n_1+n_2} + 1)\right| - \frac{1}{2}}{\sqrt{\frac{2n_1n_1(2n_1n_2 - n_1 - n_2)}{(n_1n_2)^2(n_1+n_2-1)}}}$$

Если гипотеза H_0верна, то Z имеет (приближенно) стандартное нормальноераспределение N(0,1). Гипотеза H0 принимается на уровне значимости α, если выборочное значение статистики Z, Zв удовлетворяет неравенству: $|z_B| \leq u_{1-\frac{\alpha}{2}}$ где $u_{1-\frac{\alpha}{2}}$ квантиль нормального распределения N(0,1) порядка $1-\frac{\alpha}{2}$ если $|z_B| \leq u_{1-\frac{\alpha}{2}}$ то гипотеза H_0 отклоняется.

3.2.2. Двухвыборочный тест Колмогорова - Смирнова

(Kolmogorov – Smirnov two-sampletest)

Применяется для проверки гипотезы о том, что две независимые выборки x_1, x_2, x_n и y_1, y_2, y_n получены из одной генеральной совокупности, то есть функции распределения $F_1(x)$ и $F_2(x)$ двух генеральных совокупностей равны (в этом случае говорят, что генеральные совокупности однородны):$H_0: F_1(x) \equiv F_2(y)|_{y=x}$ при альтернативной гипотезе H_1: $F_1(x) \neq F_2(y)|_{y=x}$.

Статистикой критерия является максимальная разница между эмпирическими функциями распределения, построенными по выборкам

$$D = \max|F_1 * (x) - F_2 * (y)|$$

При больших значениях n_1и n_2 (>40) используется следующие критические значения:

При $\alpha = 0.05$ $D_{\text{крит}} = 1.36\sqrt{\frac{n_1n_2}{n_1+n_2}}$

При $\alpha = 0.1$ $D_{\text{крит}} = 1.22\sqrt{\frac{n_1 n_2}{n_1+n_2}}$

3.2.3.Критерий Манна-Уитни (Mann-Whitney test).

Критерий применяется для сравнения двух независимых выборок объемов n_1 и n_2 и проверяют гипотезу H_0, утверждающую, что выборки получены из однородных генеральных совокупностей и, в частности, имеют равные средние и медианы, т.е. применяется в тех же условиях, что и критерий серий.

Статистика W критерия определяется следующим образом. Расположим n_1 n_2*значений объединенной выборки в порядке возрастания, т.е. в виде вариационного ряда. Каждому элементу ряда поставим в соответствие номер в ряду -* ранг**.**

Если несколько элементов ряда совпадают по величине, то каждому из них присваивается ранг, равный среднему арифметическому их номеров. Последний элемент в ранжированной объединенной выборке должен иметь ранг n_1+n_2. Этот факт можно использовать при проверке правильности ранжирования.

Пусть R_1-сумма рангов первой выборки, R_2 - сумма рангов второй выборки. Вычислим значения w_1 и w_2, которые определяются формулами:

$$w_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$w_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

Правильность вычислений проверяется по формуле: $w_1 + w_2 = n_1 n_2$

Выборочное значение w_B статистики критерия W есть наименьшее из чисел w_1 и w_2.

Если объем каждой из выборок больше 8, то проверку гипотезы Но можно проводить, используя статистику

$$Z = \frac{W - \frac{1}{2}n_1 n_2}{\sqrt{\frac{1}{2}n_1 n_2 (n_1 + n_{2+1})}}$$

имеющую (при условии, что верна гипотезаH_0) приблизительно стандартное нормальное распределение N(0,1). В этом случае гипотезаH_0отклоняется науровне значимости α, если выборочное значение z_B статистики Z удовлетворяет неравенству

$z_B<u_a(z_B>u_{1-a})$

При левосторонней (правосторонней) альтернативной гипотезе H_1 и если

$$|z_B| > z_{1-\frac{a}{2}}$$

При двусторонней альтернативной гипотезе H_1.

3.2.4. Однофакторный дисперсионный анализ Краскела-Уоллиса *(Kruskal-Wallis)*

Критерий Краскела-Уоллиса служит для проверки гипотезы H_0: k выборок объемов n_1, n_2,…, n_k получены из одной генеральной совокупности, т.е. является обобщением U-критерия Манна-Уитни на случай, когда число выборок k>2.

Статистика критерия H определяется следующим образом. Все выборки записываются в одну последовательность. Эта последовательность записывается в порядке возрастания, т.е. в виде вариационного ряда. Для каждого элемента выборки определяется ранг (так же как в U-критерии). Пусть R_1- сумма рангов i-й выборки, i=1,2,…,k. Для контроля можно использовать тождество

$$\sum_{i=1}^{k} R_i = \frac{n(n+1)}{n}$$

где n- число элементов объединенной выборки: $n=\sum_{i=1}^{k} n_i$

Статистика критерия H определяется так:

$$H = \frac{12}{n(n+1)}\left(\sum_{i=1}^{k}\frac{R_i^2}{n_i}\right) - 3(n+1)$$

Если гипотеза H_0 верна, то при $n_i \geq 5$ и $k \geq 4$ статистика H имеет приблизительно распределение χ^2с(k-1) степенями свободы. Гипотеза H_0 отклоняется на уровне значимости α, если выборочное значение H_B статистики H удовлетворяет условию

$$H_B > \chi^2_{1-\alpha}(k-1)$$

где$\chi^{\,2}_{1-\alpha}(k-1)$квантильраспределения χ^2порядка*(1-a)* с (*k-1)* степенью свободы.

ГЛАВА 3.

СБОР И СТАТИСТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ.

§1. Оценки параметров распределения случайных величин.

Изучение статистических закономерностей в динамике процесса успеваемости студентов ФМФ мы произвели по следующей схеме:

- собрали информацию по студентам ТувГУ физико-математического факультета 1, 2, 3, 4 и 5 курса 2015-2019 годов поступления;
- провели первичный анализ собранной информации путем вычисления числовых характеристик генеральной совокупности данных по различным признакам;
- изучили полученные корреляционные зависимости между результатами школьного аттестата, ЕГЭ и результатами сессионных экзаменов студентов физико-математического факультета ТувГУ;
- была отслежена динамика в изменении по сессиям и по дисциплинам средних оценок по группам студентов ФМФ: «Математика-информатика», «Информатика-математика», «Физика и Астрономия», «ФИИТ», «Математика»;
- выявлены статистические закономерности, произведен корреляционный анализ по полученным данным, т.е. между результатами ЕГЭ (оценки аттестата и вступительные оценки) и результатами сессионных экзаменов студентов 2015-2019 годов поступления ФМФ ТувГУ;
- далее мы произвели проверку гипотезы об однородности выборок, с помощью четырех методов пакета STATISTICA: критерий серий Вальда-Вольфовица, критерий Манна-Уитни, двух выборочный тест Колмогорова-Смирнова, однофакторный дисперсионный анализ Краскела-Уоллиса.

Собранная нами информация была извлечена из личных дел студентов, экзаменационных ведомостей и занесена в электронный носитель на

компьютере с помощью программы MicrosoftExcel (Банк данных размещен в памяти Сервера локальной сети ФМФ для всеобщего доступа).

Обработку данных по указанной схеме мы покажем на примере данных 3 курса 2017 года поступления по специальности —Математика-информатика, т.е. группы МИ-307 (21 студентов).

Таблица 1

3 курс 2017 года поступления

Математика-информатика 1семестр

МИ - 307	Вст(х)	Атт(у)	Ист	матан	алг	Сумма
1	10	12	5	4	4	13
2	12	12	5	3	3	11
3	10	12	4	3	3	10
4	10	15	5	4	4	13
5	12	15	4	2	3	9
6	14	12	5	3	4	12
7	10	15	3	3	3	9
8	10	12	5	4	4	13
9	10	10	5	3	3	11
10	11	11	5	3	4	12
11	12	10	5	3	4	12
12	11	13	5	3	4	12
13	10	12	5	4	4	13
14	10	13	5	3	3	11
15	12	12	5	4	4	13
16	12	10	5	4	4	13
17	10	12	5	3	3	11
18	10	12	4	3	3	10
19	10	15	3	3	3	9
20	13	13	5	4	4	13
21	13	15	5	3	3	11
Сумма	232	263	98	69	74	241
M	11,04762	12,52381	4,666667	3,285714	3,52381	11,47619
D	1,569161	2,630385	0,412698	0,29932	0,249433	1,963719
sqrt(D)	1,252662	1,621846	0,642416	0,547101	0,499433	1,401328

Здесь фамилии студентов зашифрованы под номерами (1,2,3…21) и занесены в компьютер:

Аттестат. – сумма оценок по дисциплинам в аттестате (математика, информатика, русский язык).

Вступит. -сумма оценок оценки вступительных экзаменов по всем профильным дисциплинам (математика, информатика, русский язык (ЕГЭ).

Мат.анализ, Алгебра- оценка по этим дисциплинам на экзаменах в 1 семестре.

Сумма - сумма оценок по этим дисциплинам за 1 семестр.

M - математическое ожидание $M = \frac{1}{n}\sum x_i$

D – дисперсия $D = \frac{1}{n}\sum (x_i - M)^2$

Sqrt (D) – корень из дисперсии, т.е. среднее квадратическое отклонение.

По данным таблицы 1 рассчитываем среднее значение оценок вступительных экзаменов за 1-5 семестры по всем математическим дисциплинам [7].

Таблица 2

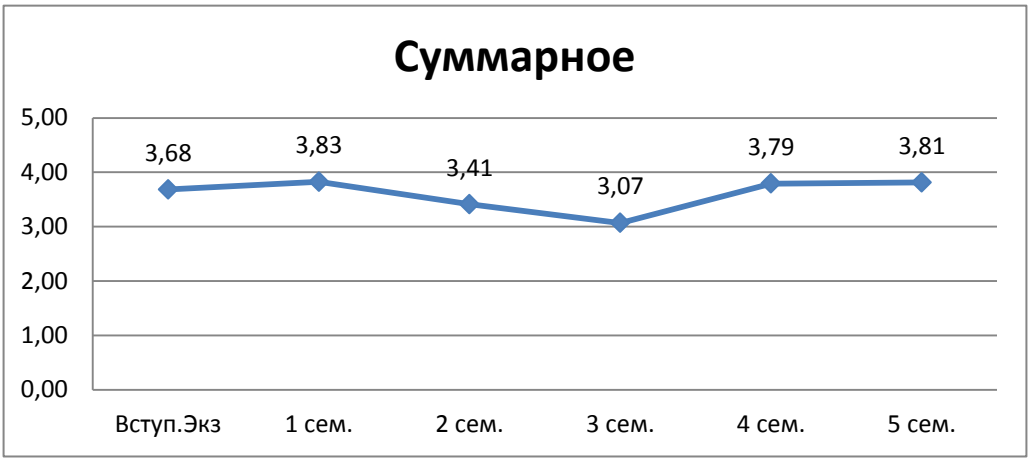
Средние оценки группы МИ-307 по всем дисциплинам

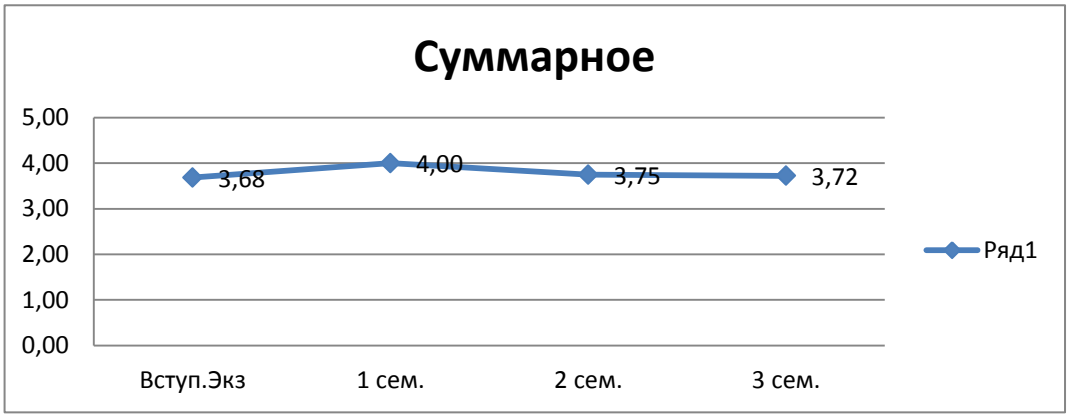
Вступ.Экз	3,68
1 сем.	3,83
2 сем.	3,41
3 сем.	3,07
4 сем.	3,79
5 сем.	3,81

Используя результаты данной таблицы, мы отследили динамику изменения по сессиям и дисциплинам средних оценок по всем группам студентов ФМФ всех специальностей.

Для сравнения посмотрим диаграмму для данной группы МИ-307 и для группы МИ-308 той же специальности 2016 года поступления (21 человек).

Средние оценки по всем семестрам





Из диаграммы видно, что уровень успеваемости студентов у обоих групп в первом семестре по сравнению вступительными баллами возрастает. У 2-го курса со 2 - го по 3-й семестры –уровень успеваемости падает. А у 3-го курса с 1- го по 3 - й семестры падает, но со 3 - го по падает, а с 4 - го по 5 - й растет.

Таким же способом нами построены диаграммы отдельно по основным математическим дисциплинам по всем группам ФМФ 2015-2019 гг. поступления.

§2. Корреляционный анализ.

Трудоемким и важным был проведенный нами расчет корреляционных зависимостей (Таблица 3).

При их вычислении мы использовали стандартную формулу коэффициента корреляции [9]:

$$r = \frac{\frac{1}{n}\sum xy - \frac{1}{n}\sum x \frac{1}{n}\sum y}{\sqrt{(\frac{1}{n}\sum x^2 - (\frac{1}{n}\sum x)^2) * (\frac{1}{n}\sum y^2 - (\frac{1}{n}\sum y)^2)}}$$

где $n -$ количество студентов;

x - оценки аттестат или (вступительных экзаменов);

y – оценки сессионных экзаменов или (аттестата).

Корреляционная зависимость между экзаменационными оценками по 3 – м экзаменам аттестата, вступительных экзаменов, сессионных экзаменов по группе МИ-307 3 курса студентов 2017 года поступления (21 студентов):

Таблица 3

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**-0,04**
Аттест.(x)	1 сем.(y)	**-0,44**
Аттест.(x)	2 сем.(y)	**0,47**
Аттест.(x)	3 сем.(y)	**0,14**
Аттест.(x)	4 сем.(y)	**0,39**
Аттест.(x)	5 сем.(y)	**0,29**
Вступит.(x)	1 сем.(y)	**0,20**
Вступит.(x)	2 сем.(y)	**0,22**
Вступит.(x)	3 сем.(y)	**0,04**
Вступит.(x)	4 сем.(y)	**0,35**
Вступит.(x)	5 сем.(y)	**0,33**

Из таблицы 3 можно сделать следующие выводы:

1. Между оценками аттестата и результатами вступительных экзаменов устудентов 3-го курса наблюдается слабая **обратная** зависимость;

2. Между оценками аттестата и результатами экзаменов сессий: в 1 семестре наблюдается очень слабая **обратная** зависимость, а во 2-ом – слабая, в 3-5– очень слабая, в 4-ом – слабая зависимость;

3. Между результатами вступительных экзаменов и результатами экзаменов сессий: в 1-2 семестрах – слабая зависимость, в 3-ем семестре – очень слабая в 4-5 семестрах - слабые зависимости.

4. Из предыдущих выводов следует, что результаты вступительных экзаменов в некоторой степени предопределяют уровень знаний и успеваемость студентов в будущих сессионных экзаменах.

5. Оценки аттестата не дают возможности прогнозировать будущую успеваемость студентов.

Для составления уравнения линейной регрессии использовались следующие известные формулы: $y = ax + b$

$$a = \frac{\frac{1}{n}\sum y \frac{1}{n}\sum x^2 - \frac{1}{n}\sum x \frac{1}{n}\sum yx}{\frac{1}{n}\sum x^2 - (\frac{1}{n}\sum x)^2}$$

$$b = \frac{\frac{1}{n}\sum xy - \frac{1}{n}\sum x \frac{1}{n}\sum y}{\frac{1}{n}\sum x^2 - (\frac{1}{n}\sum x)^2}$$

Для составления уравнения нелинейной (квадратичной) регрессии мы использовали формулы, приведенные в ГЛАВЕ 1: $y = a_0 + a_1 x + a_2 x^2$

$a_0 = \frac{\sum y_i}{n} - a_1 \frac{\sum x_{i1}}{n} - a_2 \frac{\sum x_{i2}}{n}$; $a_1 = \frac{(\sum x_{i2}'^2)(\sum y_i' x_{i1}') - (\sum x_{i1}' x_{i2}')(\sum (y_i' x_{i2}'))}{D}$

$a_2 = \frac{(\sum x_{i1}'^2)(\sum y_i' x_{i2}') - (\sum x_{i1}' x_{i2}')(\sum (y_i' x_{i1}^i)}{D}$,

где $D = \left(\sum x_{i1}'^2\right)(\sum x_{i2}'^2) - (\sum x_{i1}' x_{i2}')^2$.

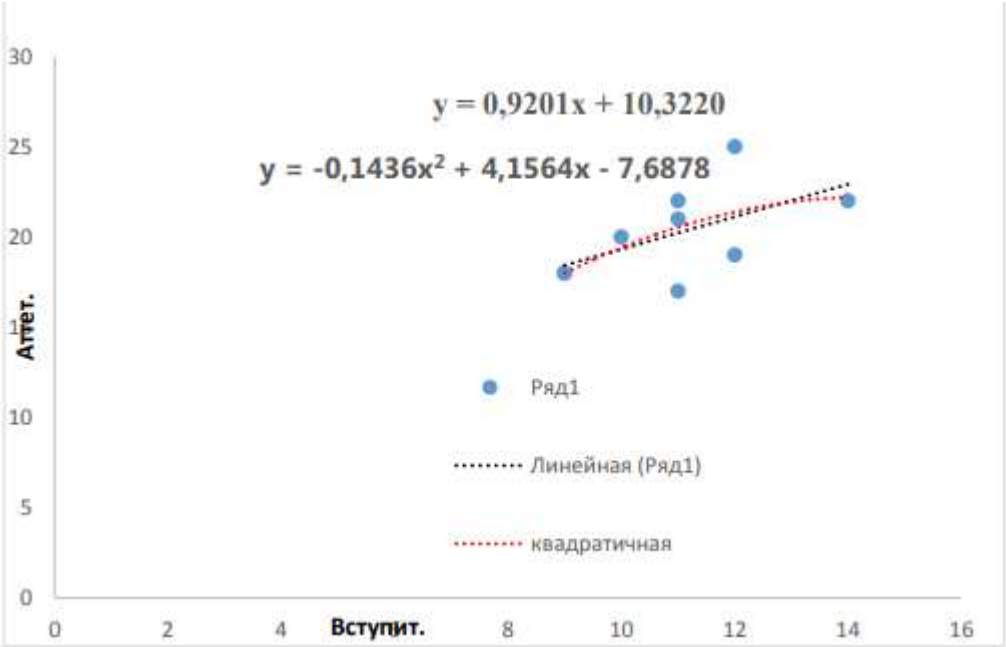
Далее приведем уравнения линейной и нелинейной (квадратичной) регрессии между экзаменационными оценками по математике аттестата, экзаменационными оценками ЕГЭ и сессионными экзаменами по 1 группе МИ-308 студентов 2-го курса (21 студентов) 2018 года поступления по системе ЕГЭ.

Таблица 4

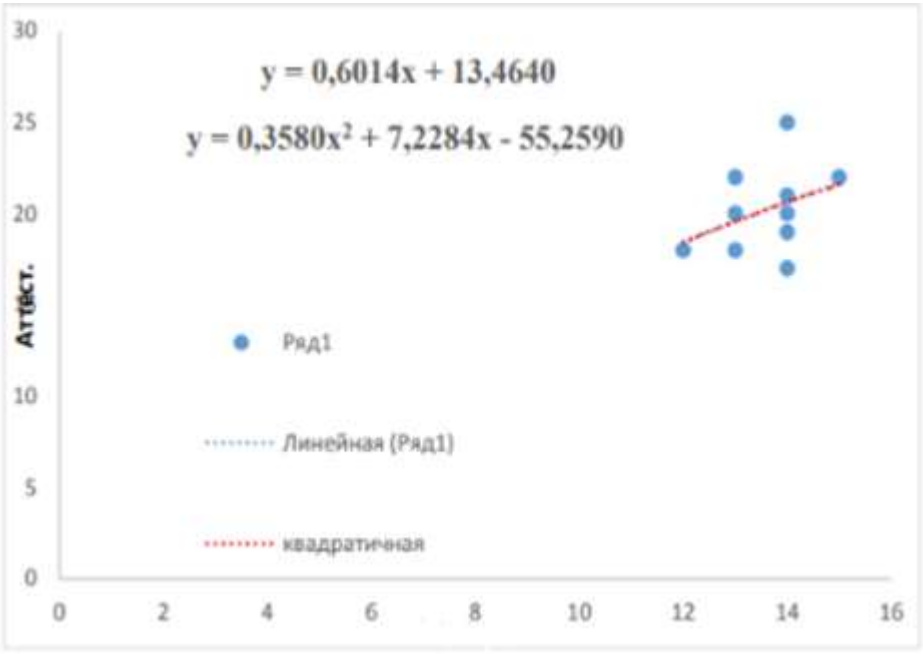
2 курс 1 группа	уравнение линейной регрессии	уравнение нелинейной регрессии
аттестат/1сем.	y = 0,1481x + 18,5560	$y = -0,3202x^2 + 7,5833x - 23,9211$
аттестат/2 сем.	y = 0,7458x + 8,8644	$y = 0,3060x^2 - 8,3203x + 75,2470$
аттестат/3сем.	y = 0,6014x + 13,4640	$y = 0,3580x^2 - 7,2284x + 55,2590$
вступит/1сем	y = 0,1429x + 9,7143	$y = -0,1184x^2 + 2,8929x - 5,9962$
вступит/2сем	y = -0,1525x + 13,6190	$y = 0,3439x^2 - 10,341x + 88,2210$
вступит/3сем	y = -0,2391x + 13,9780	$y = -0,1543x^2 + 3,1358x - 4,0370$
аттестат/вступит	y = 0,9201x + 10,3220	$y = -0,1436x^2 + 4,1464x - 7,6878$

Также нами были построены графики уравнений линейной и нелинейной (квадратичной) регрессии между оценками аттестата и оценками различных сессий, оценками вступительных экзаменов и сессионных. Они помещены ниже вместе с графиками линейной регрессии.

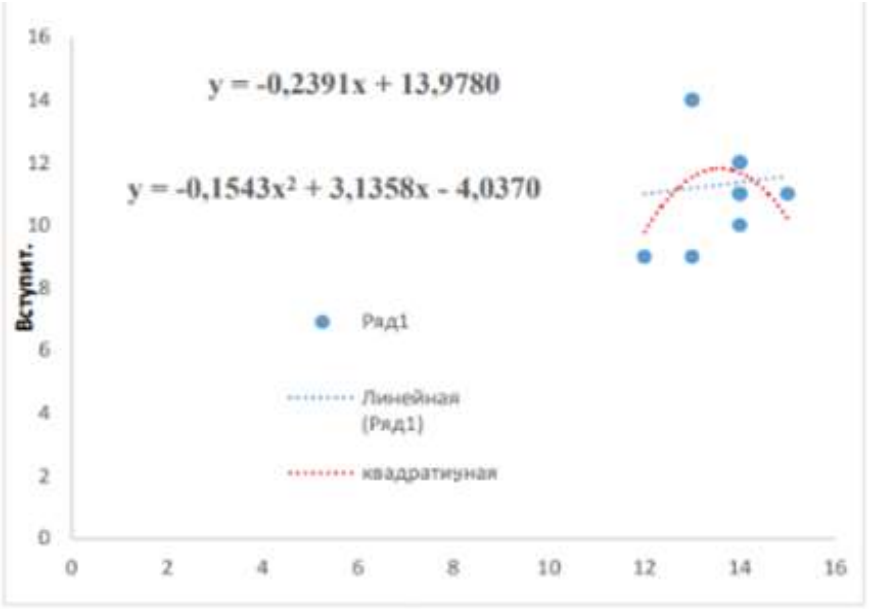
Зависимость вступительных оценок от оценок аттестата.



Зависимость между оценками 3-го семестра и оценками аттестата



Зависимость между оценками 3-го семестра и оценками вступительных экзаменов



- - экспериментальные данные;

_____- уравнение линейной регрессии;

_ _ _ _ -число уравнение нелинейной (квадратичной) регрессии.

Аналогичную работу провели в остальных группах студентов ФМФ 2015-

2019-ых поступления. Результаты представлены в Приложении 3.

§3. Статистическая проверка гипотез.

Проверка гипотез по статистике успеваемости в пакете STATISTICA.

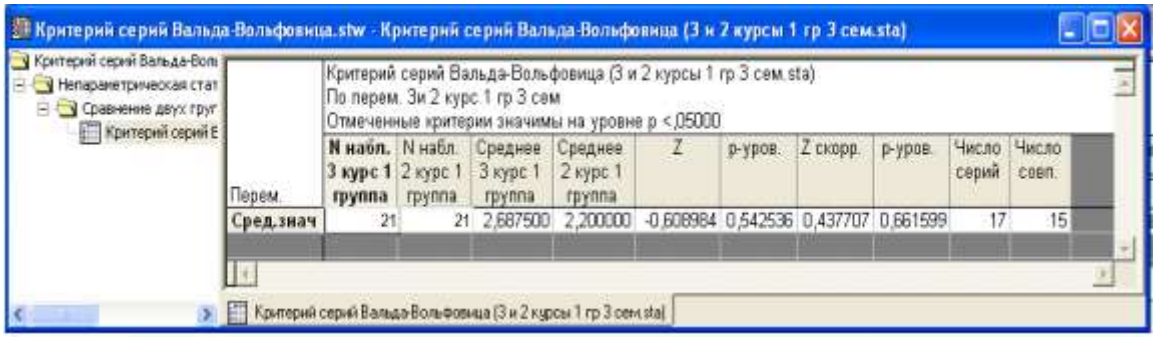
Задача 1: Даны две независимые выборки (средние оценки группы МИ-307 2017 года поступления (3 курс) и группы МИ-308 2018 года поступления (2 курс) по математическим дисциплинам за одинаковый 3 семестр (в обеих выборках оценки за 2 курс в 2017 и в 2018 году)) объемами $n_1 = 21$ и $n_2 = 21$. Проверяем гипотезу H_0, утверждающую, что две группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: две группы данных не принадлежат одной генеральной совокупности. Уровень значимости $\alpha = 0{,}10$.

Решение методом Критерий серий Вальда-Вольфовица:

Средние оценки по математическим дисциплинам записываются в виде вариационного ряда объединенной выборки. Чтобы решить задачу пакете STATISTICA надо записать данные в две переменные. В одну переменную (зависимую) надо занести объединенную выборку, а в другую (группирующую) –коды, определяющие принадлежность элементов к той или иной выборке. Задаем уровень значимости $\alpha = 0{,}10$.

Но программа не дает окончательного вывода о подтверждении или отклонении гипотезы, это нужно делать самим. В зависимости от уровня значимости α вычисляем по таблице [2] квантили U_p стандартного нормального распределения N(0,1). Сравниваем это значение с полученным в статистике значением скорректированного. Затем сравниваем α с полученным значением p-уровня.

Решение в STATISTICA выглядит на мониторе следующим образом:



$$1-\frac{\alpha}{2}=1-0{,}05=0{,}95 \qquad u_{1-\frac{\alpha}{2}}=u_{0{,}95}=1{,}645$$

В результате выполнения процедуры получим:

Число серий = 17;

Число совпадающих значений = 15;

Скорректированное на непрерывность выборочное значение Z=0,437707;

p-level(p-уровень)($p=P[|Z|>z_B]$) $=0{,}542536$.

- Вывод:

1. $z_B \approx -0{,}608984$; это значение меньше квантили распределения N(0,1) $u_{0{,}95}=1{,}645$, то есть $-0{,}608984<1{,}645$;
2. Выполняется условие $p>\alpha$, то есть, $0{,}542536>0{,}10$.

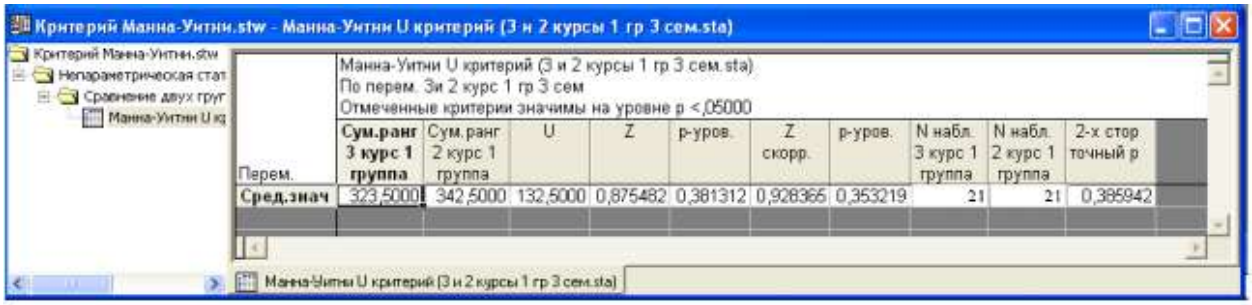
Исходя из первого и второго условий, получаем, что гипотеза H_0 не отклоняется, то есть две независимые выборки принадлежат одной генеральной совокупности.

Задача 2. Даны две независимые выборки (средние оценки группы МИ-307 2017 года поступления (3 курс) и группы МИ-308 2018 года поступления (2 курс) по математическим дисциплинам за одинаковый 3 семестр (в обеих выборках оценки за 2 курс в 2017 и в 2018 году)) объемами $n_1=21, n_2=21$. Проверяем гипотезу H_0, утверждающую, что две группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по

наблюдаемому признаку. Альтернативная гипотеза H_1: две группы данных не принадлежат одной генеральной совокупности. Уровень значимости $\alpha = 0{,}10$.

Решение методом Манна-Уитни U критерий:

Как и в предыдущем методе, пакет STATISTICA не дает окончательного вывода, которые мы делаем сами. Приводим «картинку» итога работы программы ниже (табл. 2), и уже без подробностей делаем вывод.



- Вывод:

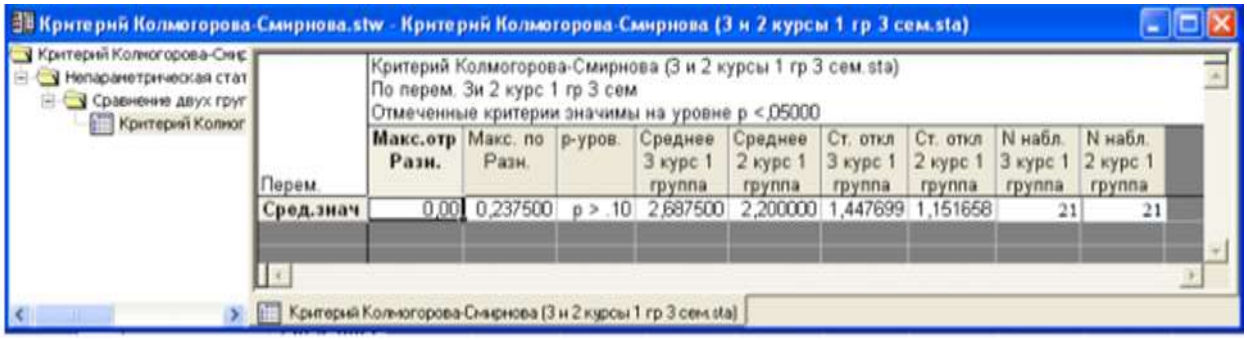
1. $z_B \approx 0{,}875482$; это значение меньше квантили распределения $u_{0,95} = 1{,}645$;
2. Выполняется условие $p > \alpha$, то есть, $0{,}381312 > 0{,}10$.

Из первого и второго условий следует, что гипотеза H_0 не отклоняется, то есть две независимые выборки принадлежат одной генеральной совокупности.

Задача 3. Даны две независимые выборки (средние оценки группы МИ-307 2017 года поступления (3 курс) и группы МИ-308 2018 года поступления (2 курс) по математическим дисциплинам за одинаковый 3 семестр (в обеих выборках оценки за 2 курс в 2017 и в 2018 году)) объемами $n_1 = 21, n_2 = 21$. Проверяем гипотезу H_0, утверждающую, что две группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: две группы данных не

принадлежат одной генеральной совокупности. Уровень значимости $\alpha = 0{,}10$.

Решение методом Критерий Колмогорова – Смирнова:



- Вывод:

1. $D_{\text{отр}} = 0{,}00$;
2. $p > 0{,}10$.

Из первого и второго условий следует, что гипотеза H_0 не отклоняется, то есть две независимые выборки принадлежат одной генеральной совокупности.

Задача 4. Даны две попарно связанные выборки (средние оценки группы МИ-307 2017 года поступления (3 курс) и группы МИ-308 2018 года поступления (2 курс) по математическим дисциплинам за одинаковый 3 семестр (в обеих выборках оценки за 2 курс в 2017 и в 2018 году)) с одинаковым объемом $n_1 = 21$. Проверяем гипотезу H_0, утверждающую, что две группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: две группы данных не принадлежат одной генеральной совокупности. Уровень значимости $\alpha = 0{,}10$.

Решение методом Критерия знаков:

Чтобы решить задачу в пакете STATISTICA надо записать данные в две переменные. В первую переменную надо занести средние оценки

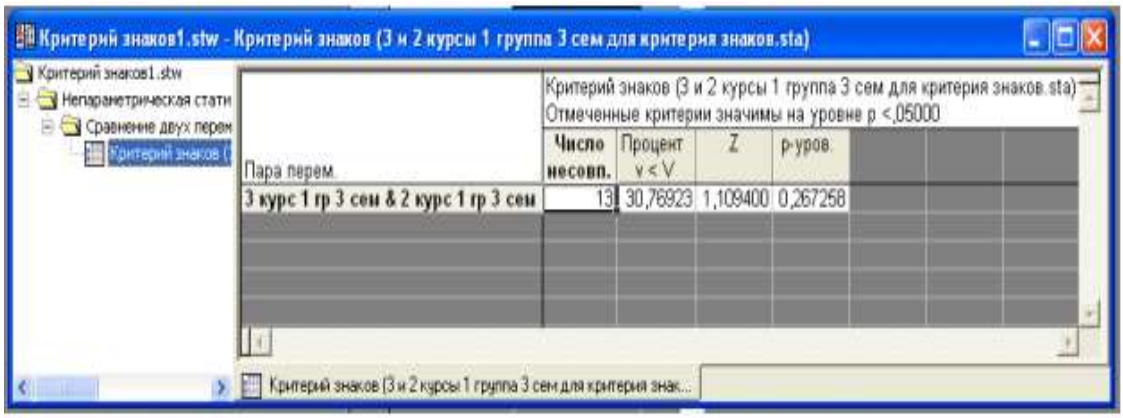
группы МИ-307 3 курса, а во вторую – средние оценки группы МИ-308 2 курса. Задачу решаем, применяя критерий знаков.

Последовательность знаков разностей $\vartheta_1 - \vartheta_2$: $+, -, 0, 0, -, +, +, +, +, -, +, +, 0, +, -, +$. Число ненулевых разностей $l = 13$, а число положительных разностей $r = 9$. Выборочное значение статистики критерия z_B равно:

$$z_B = \frac{\left|3 - \frac{9}{2}\right| + 0{,}5}{\sqrt{\frac{9}{4}}} = 1{,}109400$$

Вычисленный уровень значимости $p = 0{,}267258$ больше заданного уровня значимости $\alpha = 0{,}10$, следовательно, гипотеза не отклоняется. Следует считать, что группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку.

В пакете STATISTICA *решение выглядит следующим образом:*



- Вывод:

1. $z_B = 1{,}109400$;
2. $p > 0{,}10$.

Из первого и второго условий следует, что гипотеза H_0 не отклоняется, то есть две независимые выборки принадлежат одной генеральной совокупности.

Решив задачи разными четырьмя методами, мы пришли к общему

выводу, что во всех случаях гипотеза H0 не отклоняется. Можно считать, что две независимые выборки принадлежат одной генеральной совокупности. Наши выборки – это студенты ФМФ 3-его курса группы МИ-308 и 2-ого курса группы МИ-308 одного направления: «Математика - Информатика».

Задача 5. Даны три независимые выборки (средние оценки за 7 семестр у 4 курса по специальностям «Математика – Информатика» (1 группы МИ-306), «Информатика – Математика» (2 группы ИМ-306), «Физика»(3 группы) по математическим дисциплинам объемами $n_1 = 14, n_2 = 10, n_3 = 8$. Проверяем гипотезу H_0, утверждающую, что группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: выборки получены из разных ГС. Уровень значимости α=0,10.

Решение методом «Однофакторный дисперсионный анализ Краскела – Уоллиса»:

Запишем все выборки в одну последовательность.

Эта последовательность записывается в порядке возрастания, т.е. в виде вариационного ряда. Для каждого элемента выборки определяется ранг.



$H_B = 0{,}223$.

Квантиль распределения $x^2_{0,9}(3-1) = 4{,}6$

$p = P[x^2(3) > H_B] = 0{,}8946$

- Вывод:

1. $H_B = 0{,}223$ меньше чем квантиль распределения $x^2_{0,9}(3-1) = 4{,}6$;
2. Выполняется условие $p > \alpha$, то есть, $0{,}8945 < 0{,}10$.

Из первого и второго следует, что гипотеза H_0 не отклоняется, то есть три независимые выборки принадлежат одной генеральной совокупности.

Решив задачу данным методом, мы пришли к выводу, что гипотеза H_0 не отклоняется. Можно считать, что три независимые выборки принадлежат одной генеральной совокупности.

Всего мы применили 5 методов проверки гипотез и все они подтвердили гипотезу Но о том, что все исследованные группы принадлежат одной генеральной совокупности.

Заключение

В настоящей работе приведены результаты полного первичного статистического анализа динамики успеваемости студентов ФМФ всех специальностей 2015-2019 годов поступления, т.е. за пять лет.

Представлены также расчеты по корреляционному и регрессионному анализу взаимосвязей между итогами экзаменов на разных этапах.

При сравнении полученных характеристик с результатами подобных исследований в этой части результатов, полученными нашими студентами – предшественниками по НИРС, в прошлые годы, каких – либо значительных отклонений не обнаружено.

Для выборочных пар групп студентов одной специальности, но разных годов поступления было проведено статистическое исследование более высокого и сложного уровня по проверке статистических гипотез об их однородности по фактору сессионных оценок. Были использованы 5 методов проверки гипотез:

1.Критерий серий Вальда - Вольфовица;

2.Критерий Манна-Уитни;

3.Двухвыборочный тест Колмогорова-Смирнова;

4.Однофакторный дисперсионный анализ Краскела – Уоллиса;

5. Критерий знаков (этот критерий не использовали в предыдущей ВКР по этой теме).

Эти методы вложены в программу STATISTICA. Все пять методов подтвердили гипотезу об однородности групп по успеваемости. Этот важный результат подтвердил результаты аналогичных исследований студентов – дипломников в прошлые 3 года, то есть у студентов одной специальности, но разных годов поступления отмечается схожий уровень успеваемости.

В предыдущей ВКР [7] три статистических метода из четырёх подтвердили гипотезу об однородности групп по успеваемости, но четвёртый

метод отклонил данную гипотезу. В нашей работе все пять методов подтвердили гипотезу по однородности, т.е. мы пришли к более убедительному выводу о том, что студенты одной и той же специальности, но разных годов поступления на учебу в ФМФ ТувГУ, принадлежат одной генеральной совокупности.

Некоторые полученные в настоящей дипломной работе результаты были нами доложены на ежегодной научной конференции студентов ТувГУ, в апреле 2020 года и представлены в поданной печать работе. Наш доклад занял призовое место в рейтинге докладов на этой конференции.

Литература

1. Андронов А.М., Копытов Е.А., Гринглаз Л.Я. Теория вероятностей и математическая статистика: учебник для вузов .- Спб.: Питер,2004. (серия «Учебник для вузов»).
2. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.– М.: Наука, 1983.
3. Вентцель Е.С. Теория вероятностей. – М.: Издательский центр «Академия», 2005-576 с.
4. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: М.: ФОРУМ, 2012. – 464 с.
5. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для вузов. – 8-ое изд., стер. – М.: Высш.шк.; 2002.
6. Гнеденко Б.В.Курс теории вероятностей. – М.:Наука, 1988.
7. Даваа Б. Н. Монгуш А.Д. Статистический анализ и проверка гипотез в успеваемости студентов ФМФ 2014-2018 годов поступления методами пакета STATISTICA. Выпускная квалификационная работа студенток 5 курса ФМФ (научный руководитель А.И.Жданок), Кызыл, ТывГУ, каф.мат. анализа и МПМ, 2019.
8. Жданок А.И. Дистанционный курс «Математическая статистика 2» на платформе Moodle в системе дистанционного обучения ТувГУ (ЦДО ТувГУ), 2020.
9. Жданок А.И. Некоторые итоги многолетнего коллективного исследования по статистике успеваемости студентов ТывГУ.// Научные труды ТывГУ, Выпуск VIII, Том II, Кызыл, изд-во ТывГУ, 2010 г.

10. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. 2- ое изд. – М.: Наука. Главная редакция физико-математической литературы, 1985.
11. Румшинский Л.З.. Математическая обработка результатов эксперимента.- М.:Наука, 1971.
12. Пугачев В.С. Теория вероятностей и математическая статистика. М.,1979.
13. Салгын-оол Ч. Р. Сарыглар Ш. А. Статистические зависимости в успеваемости студентов ФМФ, поступивших в 2013 -2017г.г. по системе ЕГЭ.Курсовая работа студенток 3 курса ФМФ (научный руководитель А.И. Жданок), Кызыл, ТывГУ, кафедра математического анализа и МПМ, 2017.
14. Свешников А.А. Сборник задач по теории вероятностей, математической статистике и теории случайных функций. 4-ое изд. –Спб. 2007.
15. Солодовников А.С. Теория вероятностей. – М.: Просвещение, 1983.- 207с.
16. Тумат А. Наскыл Ч. Статистические закономерности в успеваемости студентов ФМФ поступивших 2015-2018 годов. Курсовая работа студенток 3 курса ФМФ (научный руководитель А.И.Жданок), Кызыл, ТывГУ, каф.мат. анализа и МПМ, 2019.
17. Четыркин Е.М., Калихман И.Л. Вероятность и статистика. – М.:Финансы и статистика, 1982
18. Шеффе Г. Дисперсионный анализ. М.: Физматгиз, 1980.
19. Электронное руководство в пакете STATISTICA 6 (StatSoft STATISTICA 6.1.478 Russian, Enterprise Single User)

ПРИЛОЖЕНИЕ

Приложение 1

Корреляционная зависимость между оценками аттестата (по математике), вступительных экзаменов, сессионных экзаменов студентов ФМФ специальности «Математика и Информатика» 2015-2019 г.г. поступления.

Корреляционная зависимость между оценками аттестата (алгебра, обществознание, русский язык), вступительных экзаменов (математика, русский язык) по результатам ЕГЭ, сессионных экзаменов по группам 5 курса студентов 2015 г. п.

Математика-информатика группа МИ-305 (14 студентов)

Таблица 1

Зависимость между		**r =**
Вступит.(x)	Аттест.(y)	-0,42
Аттест.(x)	1 сем.(y)	-0,06
Аттест.(x)	2 сем (y)	-0,05
Аттест.(x)	3 сем (y)	0,18
Аттест.(x)	4 сем (y)	-0,01
Аттест.(x)	5 сем (y)	-0,16
Аттест.(x)	6 сем (y)	0,25
Аттест.(x)	7 сем (y)	0,25
Аттест.(x)	8 сем (y)	0,12
Аттест.(x)	9 сем (y)	-0,27
Вступит.(x)	1 сем.(y)	0,39
Вступит.(x)	2 сем (y)	0,08
Вступит.(x)	3 сем (y)	-0,14
Вступит.(x)	4 сем (y)	-0,02
Вступит.(x)	5 сем (y)	0,18
Вступит.(x)	6 сем (y)	-0,02
Вступит.(x)	7 сем (y)	-0,02
Вступит.(x)	8 сем (y)	0,05
Вступит.(x)	9 сем (y)	0,07

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1, 2, 4, 5, 9 семестрах отслеживается слабая **обратная** зависимость, в 3, 6, 7, 8 семестрах очень слабая прямая зависимость.

3. Между результатами вступительных экзаменов и результатами экзаменов в 1 семестре отслеживается слабая прямая зависимость, во 2, 5, 8, 9 – семестрах очень слабая прямая зависимость, в 3, 4, 6, 7 – семестрах слабая **обратная** зависимость.
4. Из предыдущих выводов следует, что оценки аттестата в средней степени прогнозируют уровень знаний и успеваемость студентов. Результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Информатика-математика группа ИМ-305 (14 студентов)

Таблица 1

Зависимость между		**r =**
Вступит.(x)	Аттест.(y)	-0,10
Аттест.(x)	1 сем.(y)	0,58
Аттест.(x)	2 сем (y)	0,23
Аттест.(x)	3 сем.(y)	0,28
Аттест.(x)	4 сем (y)	0,38
Аттест.(x)	5 сем.(y)	0,05
Аттест.(x)	6 сем (y)	0,26
Аттест.(x)	7 сем.(y)	0,26
Аттест.(x)	8 сем.(y)	0,19
Аттест.(x)	9 сем.(y)	0,12
Вступит.(x)	1 сем.(y)	-0,53
Вступит.(x)	2 сем (y)	-0,37
Вступит.(x)	3 сем (y)	-0,36
Вступит.(x)	4 сем (y)	-0,14
Вступит.(x)	5 сем (y)	-0,19
Вступит.(x)	6 сем (y)	0,02
Вступит.(x)	7 сем (y)	-0,18
Вступит.(x)	8 сем (y)	-0,23
Вступит.(x)	9 сем (y)	-0,20

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается средняя зависимость, в 2, 3, 4, 6, 7 - семестрах слабая прямая зависимость, а в 5, 8, 9 - семестрах очень слабая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2, 3 - семестрах отслеживается **обратная** зависимость, в

4, 5, 7, 9 – семестрах очень слабая **обратная** зависимость, в 6 – семестре слабая прямая зависимость.

4. Из предыдущих выводов следует, что оценки аттестата в средней степени прогнозируют уровень знаний и успеваемость студентов. Результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Корреляционная зависимость между оценками аттестата (алгебра, обществознание, русский язык), вступительных экзаменов (математика, русский язык) по результатам ЕГЭ, сессионных экзаменов по группам 4 курса студентов 2016 г. п.

Математика и Информатика группа МИ-306 (14 студентов)

Таблица 3

Зависимость между		r =
Вступит.(x)	Аттест.(y)	-0,199996
Аттест.(x)	1 сем.(y)	0,658334
Аттест.(x)	2 сем.(y)	0,559184
Аттест.(x)	3 сем.(y)	0,441537
Аттест.(x)	4 сем.(y)	0,337616
Аттест.(x)	5 сем.(y)	0,260189
Аттест.(x)	6 сем.(y)	0,305773
Аттест.(x)	7 сем.(y)	0,281050
Вступит.(x)	1 сем.(y)	0,220209
Вступит.(x)	2 сем.(y)	0,288572
Вступит.(x)	3 сем.(y)	0,240985
Вступит.(x)	4 сем.(y)	0,475247
Вступит.(x)	5 сем.(y)	0,177816
Вступит.(x)	6 сем.(y)	0,235480
Вступит.(x)	7 сем.(y)	0,265632

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1, 2 - семестрах отслеживается средняя прямая зависимость, в 3 семестре слабая прямая зависимость, в 4, 5, 6, 7 – семестрах очень слабая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1,2,3 - семестрах наблюдается очень слабая прямая зависимость, в 4-ом семестре слабая прямая зависимость, в 5, 6, 7 - семестрах наблюдается очень слабая прямая зависимость.

4. Из предыдущих выводов следует, что оценки аттестата не дают возможности прогнозировать будущую успеваемость студентов. Результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Информатика-математика группа ИФ-306 (9 студентов)

Таблица 4

Зависимость между		r =
Вступит.(x)	Аттест.(y)	0,286299
Аттест.(x)	1 сем.(y)	0,289901
Аттест.(x)	2 сем.(y)	0,175776
Аттест.(x)	3 сем.(y)	0,364779
Аттест.(x)	4 сем.(y)	0,483168
Аттест.(x)	5 сем.(y)	0,799155
Аттест.(x)	6 сем.(y)	0,519248
Аттест.(x)	7 сем.(y)	0,519248
Вступит.(x)	1 сем.(y)	0,712556
Вступит.(x)	2 сем.(y)	0,102327
Вступит.(x)	3 сем.(y)	0,021235
Вступит.(x)	4 сем.(y)	0,571920
Вступит.(x)	5 сем.(y)	0,683110
Вступит.(x)	6 сем.(y)	0,163733
Вступит.(x)	7 сем.(y)	0,163733

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая прямая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1, 2, 3 - семестрах отслеживается очень слабая прямая зависимость, в 4 семестре слабая прямая зависимость, в 5-ом семестре высокая прямая зависимость, в 6, 7- семестрах средняя прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 5 - семестрах наблюдается высокая прямая зависимость, во 2, в 3, 6, 7 - семестрах очень слабая прямая зависимость, в 4-ом семестре средняя прямая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Физика группа 4. 3 (8 студентов)

Таблица 5

Зависимость между		r =
Вступит.(x)	Аттест.(y)	-0,028116
Аттест.(x)	1 сем.(y)	0,098342
Аттест.(x)	2 сем.(y)	0,545443
Аттест.(x)	3 сем.(y)	-0,023616
Аттест.(x)	4 сем.(y)	-0,496390
Аттест.(x)	5 сем.(y)	0,365377
Аттест.(x)	6 сем.(y)	0,474201
Аттест.(x)	7 сем.(y)	0,482524
Вступит.(x)	1 сем.(y)	-0,091245
Вступит.(x)	2 сем.(y)	-0,305083
Вступит.(x)	3 сем.(y)	-0,503972
Вступит.(x)	4 сем.(y)	-0,506295
Вступит.(x)	5 сем.(y)	-0,449704
Вступит.(x)	6 сем.(y)	-0,531008
Вступит.(x)	7 сем.(y)	-0,416691

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 - семестре отслеживается очень слабая прямая зависимость, во 2-ом семестре средняя зависимость, в 3 – семестре отслеживается очень слабая **обратная** зависимость, в 4 – семестре слабая **обратная** зависимость, в 5, 6, 7 - семестрах слабая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1 семестре наблюдается очень слабая **обратная** зависимость, в 2, 7 - семестрах слабая зависимость, в 3, 4, 6 – семестрах средняя **обратная** зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

ФИИТ группа 4.4 (6 студентов)

Таблица 6

Зависимость между		r =
Вступит.(x)	Аттест.(y)	0,176777
Аттест.(x)	1 сем.(y)	0,481125
Аттест.(x)	2 сем.(y)	0,516398
Аттест.(x)	3 сем.(y)	0,878310
Аттест.(x)	4 сем.(y)	0,876141
Аттест.(x)	5 сем.(y)	0,819946

Аттест.(x)	6 сем.(y)	0,549927
Аттест.(x)	7 сем.(y)	0,610341
Вступит.(x)	1 сем.(y)	0,000000
Вступит.(x)	2 сем.(y)	-0,547723
Вступит.(x)	3 сем.(y)	0,000000
Вступит.(x)	4 сем.(y)	0,075094
Вступит.(x)	5 сем.(y)	-0,193263
Вступит.(x)	6 сем.(y)	-0,544400
Вступит.(x)	7 сем.(y)	-0,460857

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая прямая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1-ом семестре отслеживается слабая прямая зависимость, в 2,6,7 - семестрах средняя прямая зависимость, в 3, 4, 5 – семестрах высокая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 3, 4 - семестрах наблюдается очень слабая прямая зависимость, в 2, 5, 6, 7 - семестрах очень слабая **обратная** зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Математика группа 3.5 (8 студентов)

Таблица 7

Зависимость между		r =
Вступит.(x)	Аттест.(y)	-0,251124
Аттест.(x)	1 сем.(y)	0,000000
Аттест.(x)	2 сем.(y)	0,284747
Аттест.(x)	3 сем.(y)	0,551819
Аттест.(x)	4 сем.(y)	0,436251
Аттест.(x)	5 сем.(y)	0,466372
Аттест.(x)	6 сем.(y)	0,562564
Аттест.(x)	7 сем.(y)	-0,179374
Вступит.(x)	1 сем.(y)	-0,042524
Вступит.(x)	2 сем.(y)	0,377964
Вступит.(x)	3 сем.(y)	-0,169031
Вступит.(x)	4 сем.(y)	0,267261
Вступит.(x)	5 сем.(y)	-0,142857
Вступит.(x)	6 сем.(y)	0,203653
Вступит.(x)	7 сем.(y)	0,142857

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1, 2 - семестрах отслеживается очень слабая прямая зависимость, в 3, 6 – семестрах средняя прямая зависимость, в 4, 5 – слабая зависимость, в 7 – семестре отслеживается очень слабая **обратная** зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 3, 5 - семестрах наблюдается очень слабая **обратная** зависимость, в 2, 4, 6, 7 – семестрах отслеживается очень слабая прямая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Корреляционная зависимость между оценками аттестата (алгебра, обществознание, русский язык), вступительных экзаменов (математика, русский язык) по результатам ЕГЭ, сессионных экзаменов по группам 3 курса студентов 2017 г. п.

Математика и Информатика группа МИ-307 (21 студентов)

Таблица 8

Зависимость между		r =
Вступит.(х)	Аттест.(у)	**-0,04**
Аттест.(х)	1 сем.(у)	**-0,44**
Аттест.(х)	2 сем.(у)	**0,47**
Аттест.(х)	3 сем.(у)	**0,14**
Аттест.(х)	4 сем.(у)	**0,39**
Аттест.(х)	5 сем.(у)	**0,29**
Вступит.(х)	1 сем.(у)	**0,20**
Вступит.(х)	2 сем.(у)	**0,22**
Вступит.(х)	3 сем.(у)	**0,04**
Вступит.(х)	4 сем.(у)	**0,35**
Вступит.(х)	5 сем.(у)	**0,33**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая **обратная** зависимость.

2. Между оценками аттестата и результатами экзаменов в 1-ом семестре отслеживается очень слабая **обратная** зависимость, в 2-5 – семестрах очень слабая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2, 3 - семестрах наблюдается очень слабая прямая зависимость, в 4, 5-семестрах слабая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Физика группа 3.3 (16 студентов)

Таблица 9

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**0,60**
Аттест.(x)	1 сем.(y)	**0,53**
Аттест.(x)	2 сем.(y)	**0,71**
Аттест.(x)	3 сем.(y)	**0,44**
Аттест.(x)	4 сем.(y)	**0,42**
Аттест.(x)	5 сем.(y)	**0,03**
Вступит.(x)	1 сем.(y)	**0,44**
Вступит.(x)	2 сем.(y)	**0,58**
Вступит.(x)	3 сем.(y)	**0,33**
Вступит.(x)	4 сем.(y)	**0,37**
Вступит.(x)	5 сем.(y)	**0,01**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается средняя прямая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1, 3, 4 - семестрах отслеживается средняя прямая зависимость, а во 2-ом высокая прямая зависимость, в 5 семестре очень слабая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 3, 4 - семестрах наблюдается слабая прямая зависимость, а во 2 средняя прямая зависимость, а в 5-ом семестре очень слабая прямая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах. Результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

ФИИТ группа 3.4 (6 студентов)

Таблица 10

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**0,48**
Аттест.(x)	1 сем.(y)	**0,27**
Аттест.(x)	2 сем.(y)	**0,63**
Аттест.(x)	3 сем.(y)	**0,14**
Аттест.(x)	4 сем.(y)	**-0,12**
Аттест.(x)	5 сем.(y)	**-0,32**
Вступит.(x)	1 сем.(y)	**-0,57**
Вступит.(x)	2 сем.(y)	**-0,14**
Вступит.(x)	3 сем.(y)	**-0,66**
Вступит.(x)	4 сем.(y)	**-0,90**
Вступит.(x)	5 сем.(y)	**-0,94**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается слабая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1,3- семестрах отслеживается очень слабая прямая зависимость, во 2-ом средняя прямая зависимость, а в 4, 5 – семестрах очень слабая **обратная** зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1,3 - семестрах наблюдается средняя обратная зависимость, во 2 очень слабая обратная зависимость, а в 4, 5 – семестрах очень высокая обратная зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Математика группа 3.5 (9 студентов)

Таблица 11

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**0,73**
Аттест.(x)	1 сем.(y)	**0,49**
Аттест.(x)	2 сем.(y)	**0,53**
Аттест.(x)	3 сем.(y)	**0,57**
Аттест.(x)	4 сем.(y)	**0,49**
Аттест.(x)	5 сем.(y)	**0,49**
Вступит.(x)	1 сем.(y)	**0,56**
Вступит.(x)	2 сем.(y)	**0,59**

Вступит.(х)	3 сем.(у)	**0,43**
Вступит.(х)	4 сем.(у)	**0,36**
Вступит.(х)	5 сем.(у)	**0,36**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается высокая прямая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1-5 - семестрах отслеживается средняя прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2 - семестрах наблюдается средняя прямая зависимость, а в 3, 4, 5 – семестрах слабая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов не дают возможности прогнозировать будущую успеваемость студентов.

Корреляционная зависимость между оценками аттестата (алгебра, обществознание, русский язык), вступительных экзаменов (математика, русский язык) по результатам ЕГЭ, сессионных экзаменов по группам 2 курса студентов 2018 г. п.

Математика и Информатика группа МИ-308 (20 студентов)

Таблица 12

Зависимость между		r =
Вступит.(х)	Аттест.(у)	**-0,05**
Аттест.(х)	1 сем.(у)	**0,46**
Аттест.(х)	2 сем.(у)	**0,20**
Аттест.(х)	3 сем.(у)	**0,10**
Вступит.(х)	1 сем.(у)	**0,03**
Вступит.(х)	2 сем.(у)	**0,16**
Вступит.(х)	3 сем.(у)	**-0,30**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается средняя зависимость, а во 2-ом и 3-ем – семестрах слабая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2 - семестре наблюдается очень слабая прямая зависимость, а в 3-ем семестре слабая **обратная** зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют

уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Физика группа 2.2 (17 студентов)

Таблица 13

Зависимость между		r =
Вступит.(x)	Аттест.(y)	-0,15
Аттест.(x)	1 сем.(y)	-0,06
Аттест.(x)	2 сем.(y)	0,62
Аттест.(x)	3 сем.(y)	0,03
Вступит.(x)	1 сем.(y)	0,60
Вступит.(x)	2 сем.(y)	0,62
Вступит.(x)	3 сем.(y)	0,61

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая **обратная** зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается очень слабая **обратная** зависимость, во 2-ом семестре средняя прямая зависимость, а в 3-ем семестре очень слабая прямая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2, 3 - семестрах наблюдается высокая прямая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

ФИИТ группа 2.3 (15 студентов)

Таблица 14

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**-0,07**
Аттест.(x)	1 сем.(y)	**0,43**
Аттест.(x)	2 сем.(y)	**-0,09**
Аттест.(x)	3 сем.(y)	**-0,02**
Вступит.(x)	1 сем.(y)	**0,09**
Вступит.(x)	2 сем.(y)	**0,27**
Вступит.(x)	3 сем.(y)	**0,46**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается слабая обратная зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается слабая зависимость, а в 2, 3 – семестрах наблюдается слабая обратная зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1, 2 - семестрах наблюдается очень слабая прямая зависимость, а 3 семестре слабая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Корреляционная зависимость между оценками аттестата (алгебра, обществознание, русский язык), вступительных экзаменов (математика, русский язык) по результатам ЕГЭ, сессионных экзаменов по группам 1 курса студентов 2019 г. п.

Математика-информатика группа МИ – 309 (23 студентов)

Таблица 15

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**0,09**
Аттест.(x)	1 сем.(y)	**0,63**
Вступит.(x)	1 сем.(y)	**0,01**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается очень слабая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается средняя зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1 семестре наблюдается очень слабая прямая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Информатика-математика группа ИМ- 309 (19 студентов)

Таблица 16

Зависимость между		r =
Вступит.(x)	Аттест.(y)	**0,65**
Аттест.(x)	1 сем.(y)	**0,76**
Вступит.(x)	1 сем.(y)	**0,71**

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается средняя зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается высокая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1 семестре наблюдается высокая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

Физика и Астрономия группа ФиА – 309 (23 студентов)

Таблица 17

Зависимость между		r =
Вступит.(x)	Аттест.(y)	0,83
Аттест.(x)	1 сем.(y)	0,81
Вступит.(x)	1 сем.(y)	0,88

Из таблицы можно сделать следующие выводы:

1. Между оценками вступительных экзаменов и аттестата отслеживается высокая зависимость.
2. Между оценками аттестата и результатами экзаменов в 1 семестре отслеживается высокая зависимость.
3. Между результатами вступительных экзаменов и результатами экзаменов в 1 семестре наблюдается высокая зависимость.
4. Из предыдущих выводов следует, что оценки аттестата и результаты вступительных экзаменов в наибольшей степени предопределяют уровень знаний и работоспособность студентов в будущих сессионных экзаменах.

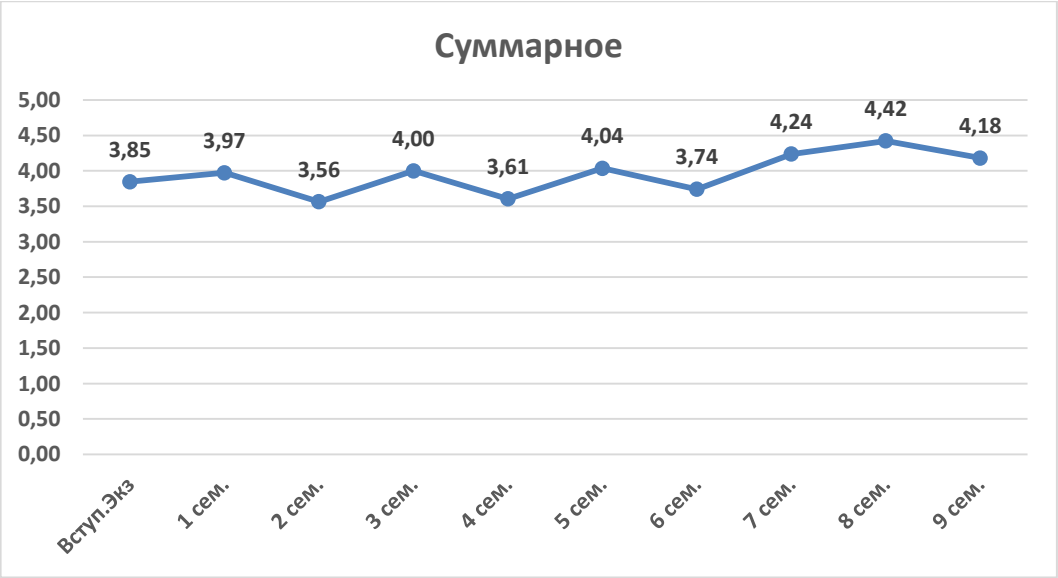
Приложение 2

Динамика экзаменационных оценок по математическим дисциплинам студентов ФМФ ТувГУ с 2015 по 2019 уч. г.г. по всем учебным группам.

Динамика экзаменационных оценок за 1-9 семестры. Средние оценки 5 курса МИ-305 группы (14 студентов) специальности «Математика и Информатика» 2015 года поступления. (отчисленные и восстановленные студенты не учитываются)

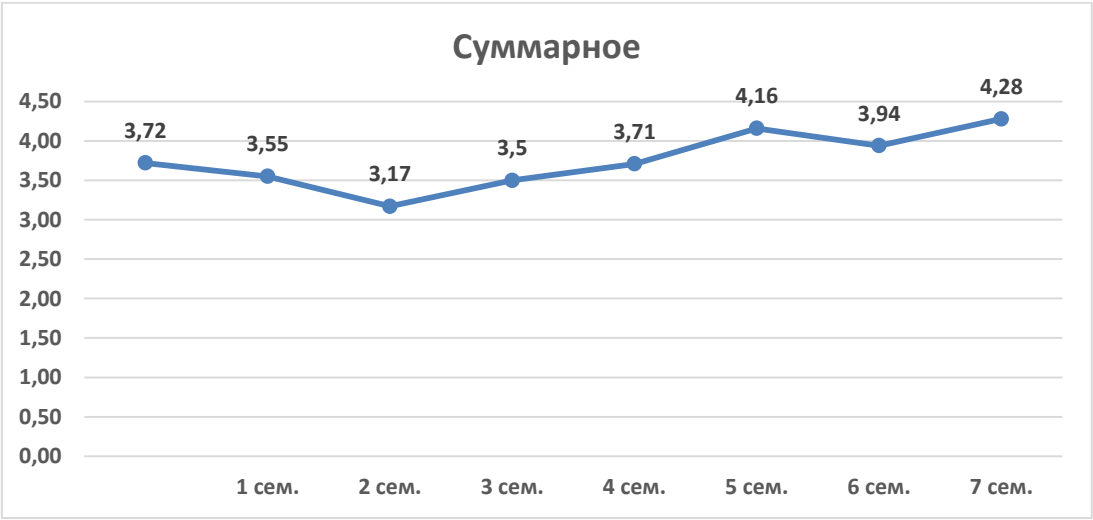


Динамика экзаменационных оценок за 1-9 семестры. Средние оценки 5 курса ИМ-305 группы (13 студентов) специальности «Математика и Информатика» 2015 года поступления. (отчисленные и восстановленные студенты не учитываются)

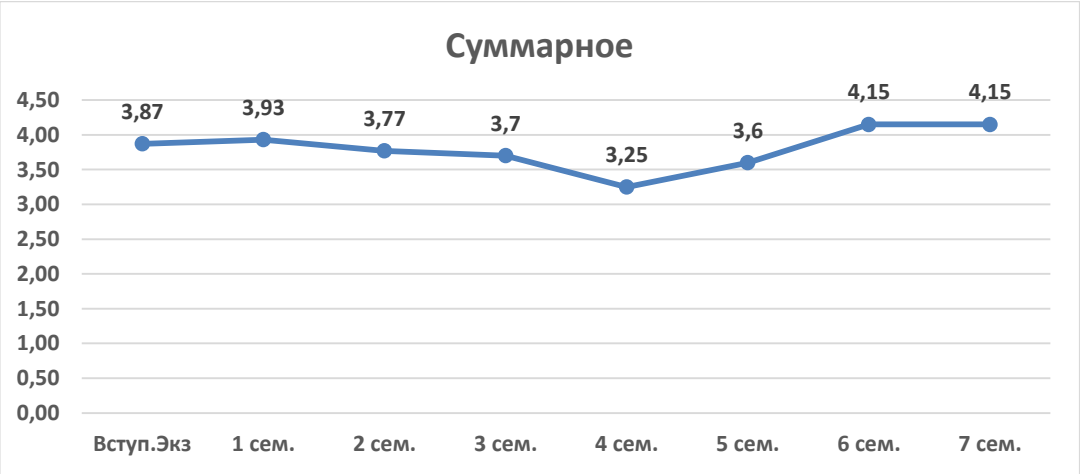


Динамика экзаменационных оценок за 1-7 семестры.

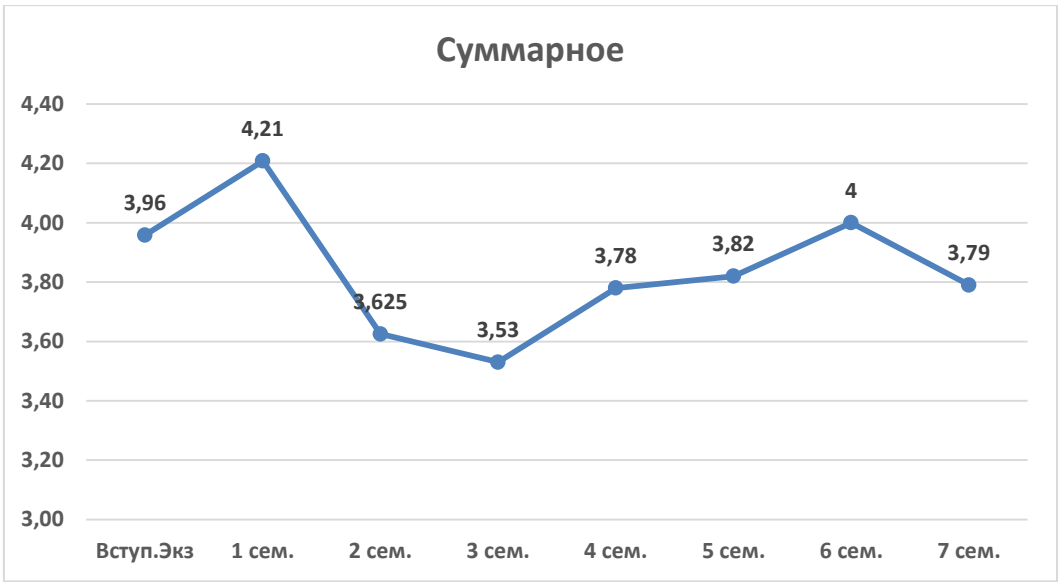
Средние оценки 4 курса МИ-306 группы (14 студентов) специальности «Математика и Информатика» 2016 года поступления. (отчисленные и восстановленные студенты не учитываются)



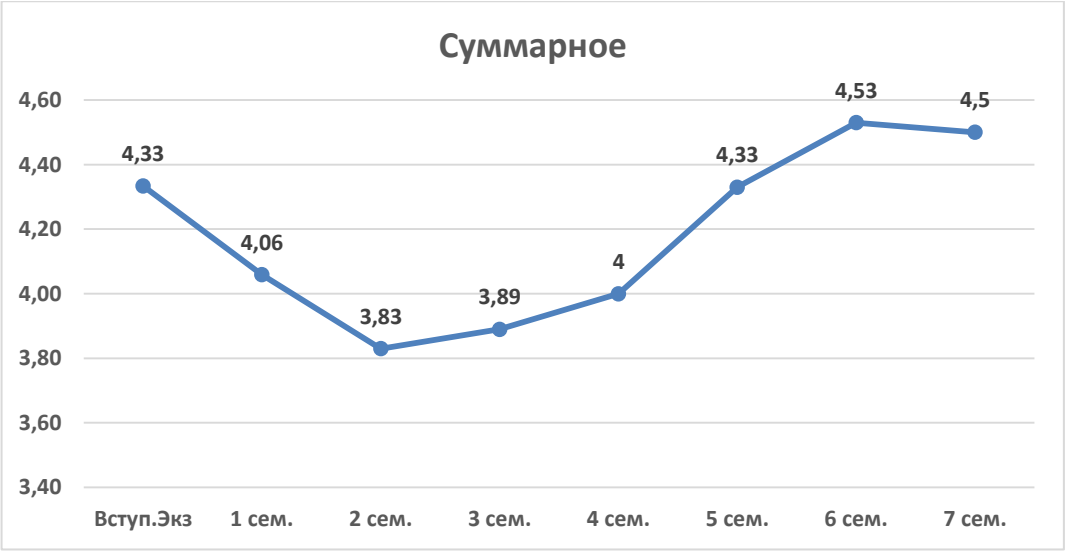
Средние оценки 4 курса ИМ-306 группы (10 студентов) специальности «Информатика и Математика» 2016 года поступления. (отчисленные и восстановленные студенты не учитываются)



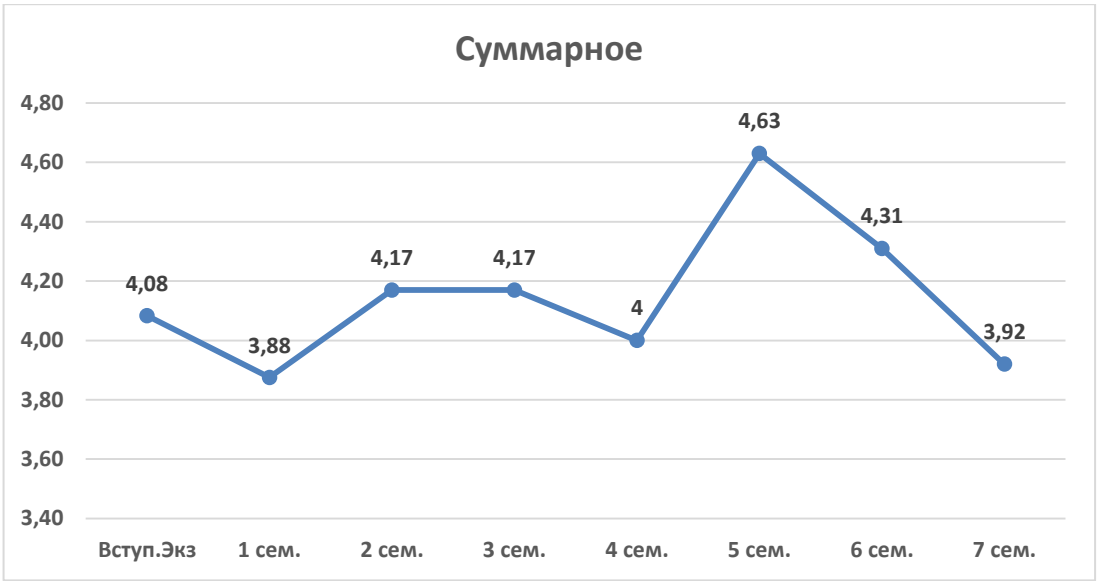
Средние оценки 4 курса 3 группы (8 студентов) специальности «Физика» 2016 года поступления. (отчисленные и восстановленные студенты не учитываются)



Средние оценки 4 курса 4 группы (6 студентов) специальности «ФИИТ» 2016 года поступления. (отчисленные и восстановленные студенты не учитываются)

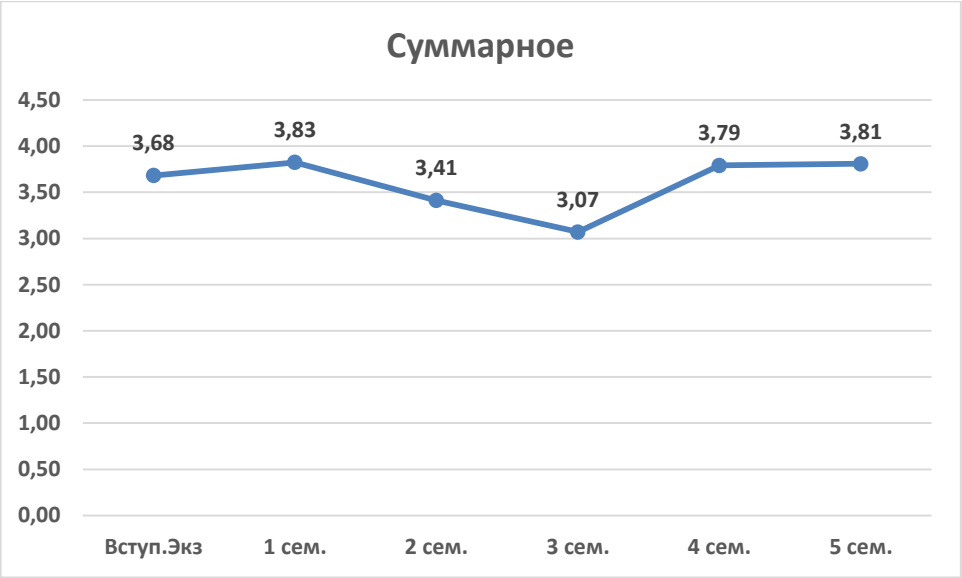


Средние оценки 4 курса 5 группы (8 студентов) специальности «Математика» 2016 года поступления. (отчисленные и восстановленные студенты не учитываются)

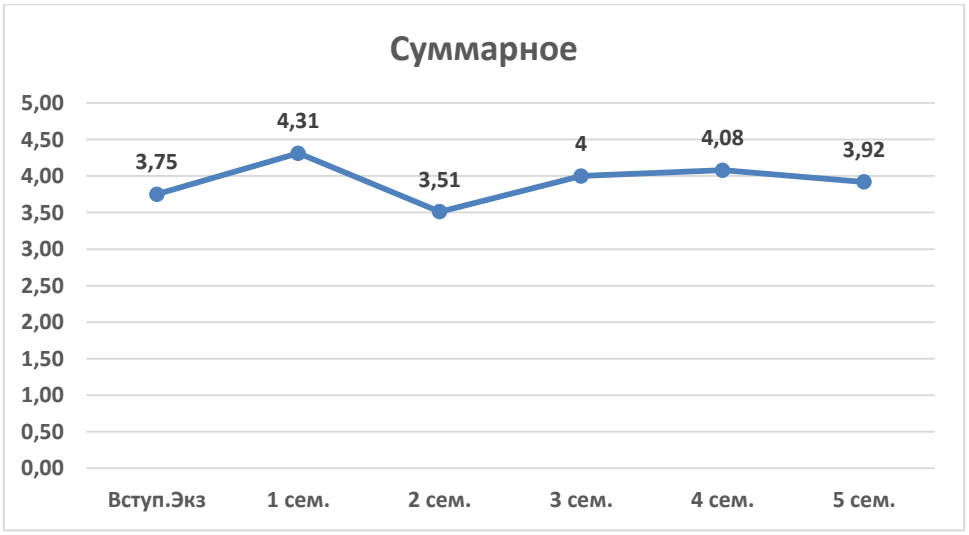


Динамика экзаменационных оценок за 1-5 семестры.

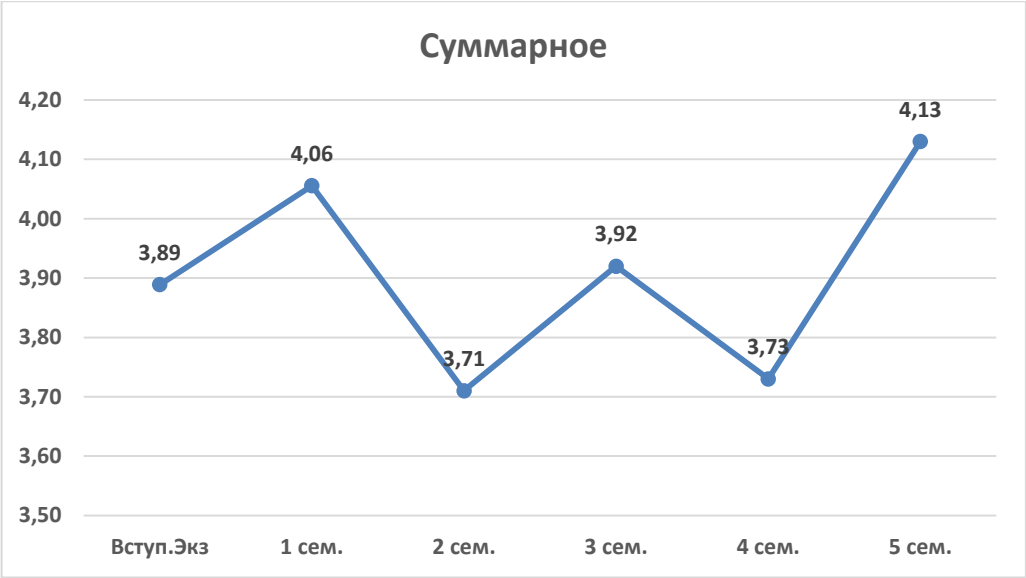
Средние оценки 3 курса МИ-307 группы (21 студентов) специальности «Математика и Информатика» 2017 года поступления. (отчисленные и восстановленные студенты не учитываются)



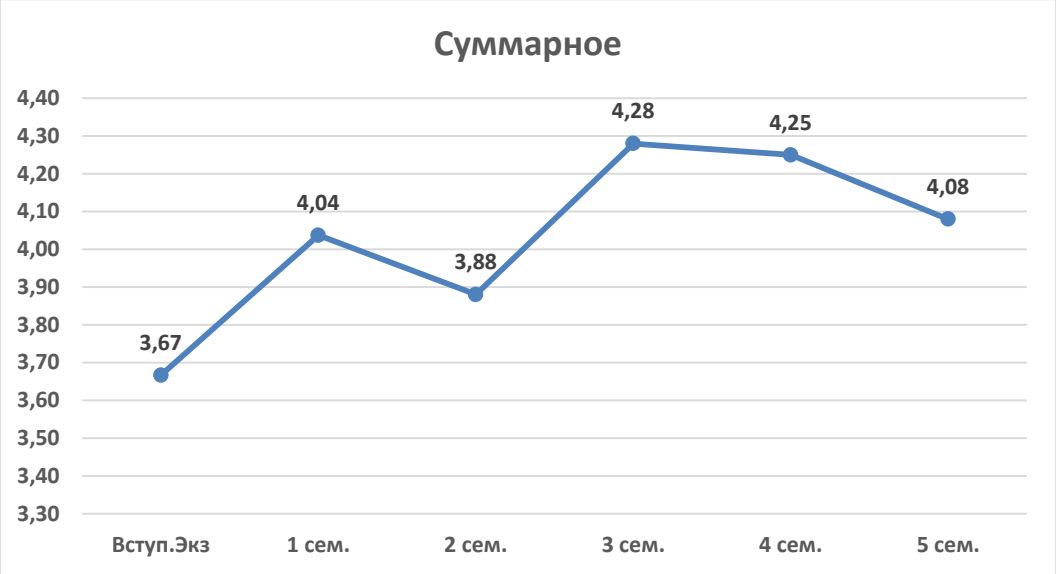
Средние оценки 3 курса 3 группы (16 студентов) специальности «Физика» 2017 года поступления. (отчисленные и восстановленные студенты не учитываются)



Средние оценки 3 курса 4 группы (6 студентов) специальности «ФИИТ» 2017 года поступления. (отчисленные и восстановленные студенты не учитываются)

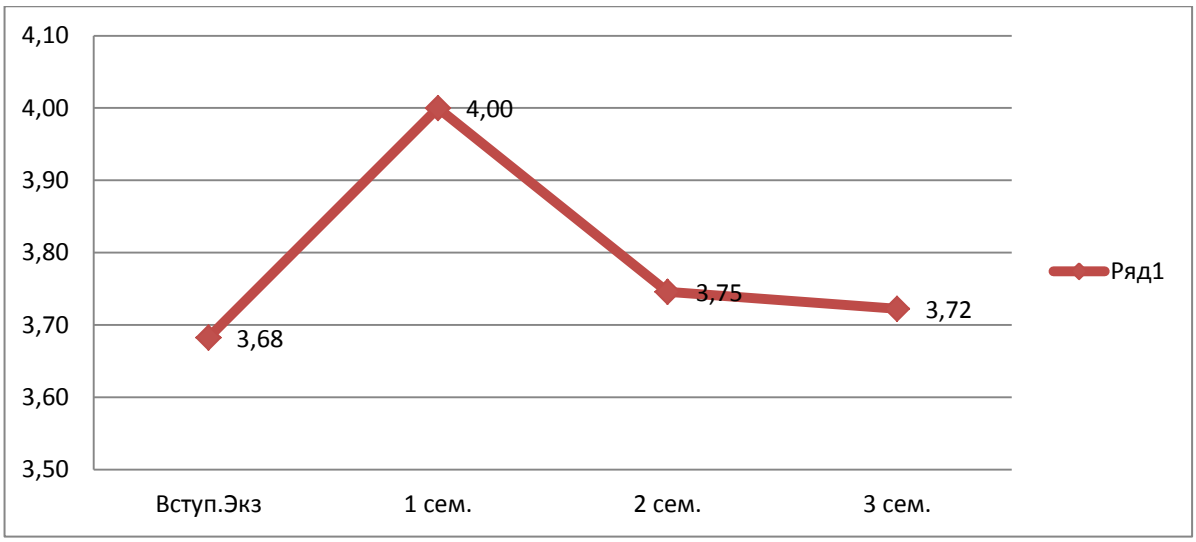


Средние оценки 3 курса 5 группы (9 студентов) специальности «Математика» 2017 года поступления. (отчисленные и восстановленные студенты не учитываются)

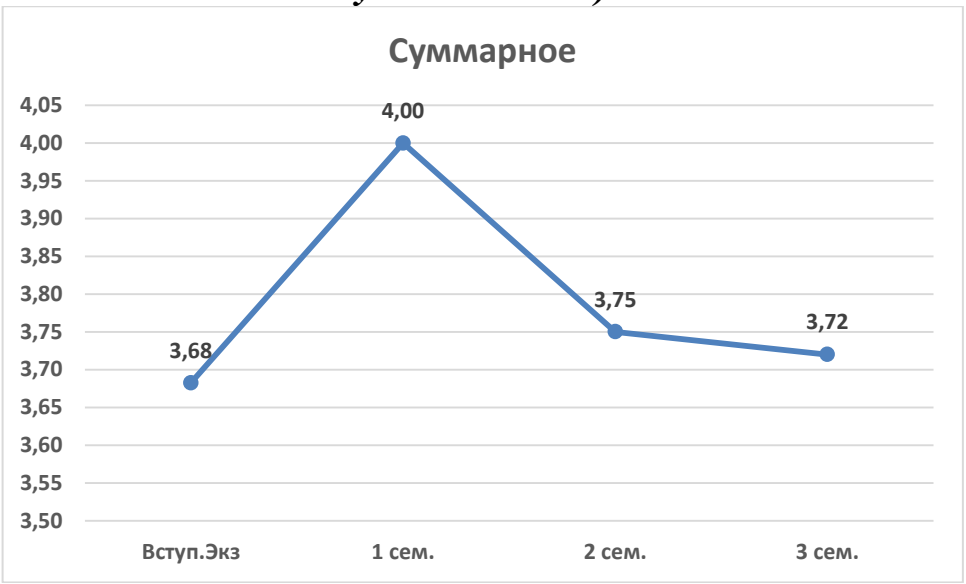


Динамика экзаменационных оценок за 1-3 семестры.

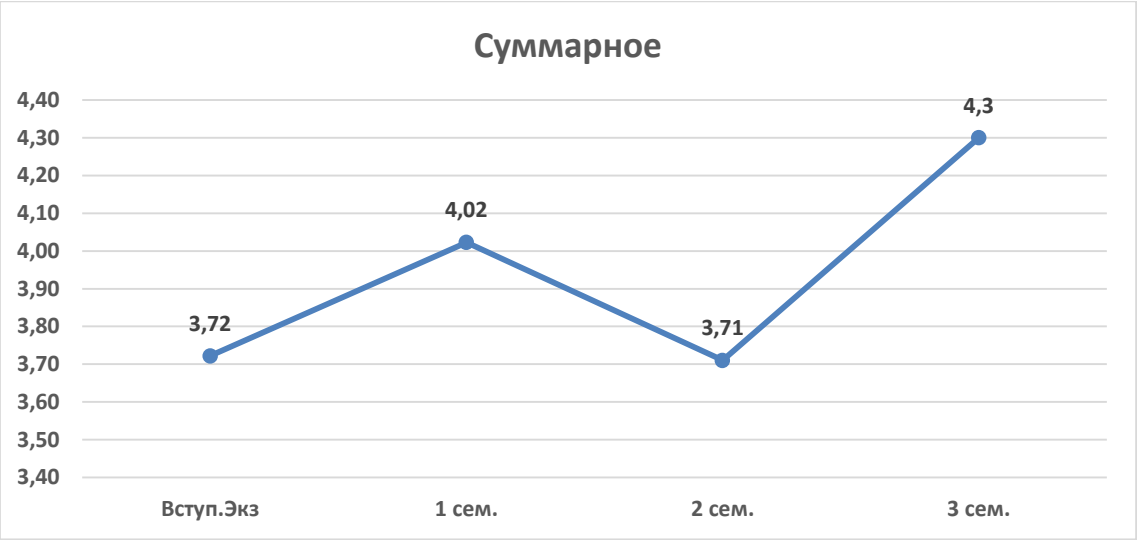
Средние оценки 2 курса МИ-308 группы (21 студентов) специальности «Математика и Информатика» 2018 года поступления. (отчисленные и восстановленные студенты не учитываются)



Средние оценки 2 курса 3 группы (17 студентов) специальности «Физика» 2018 года поступления. (отчисленные и восстановленные студенты не учитываются)

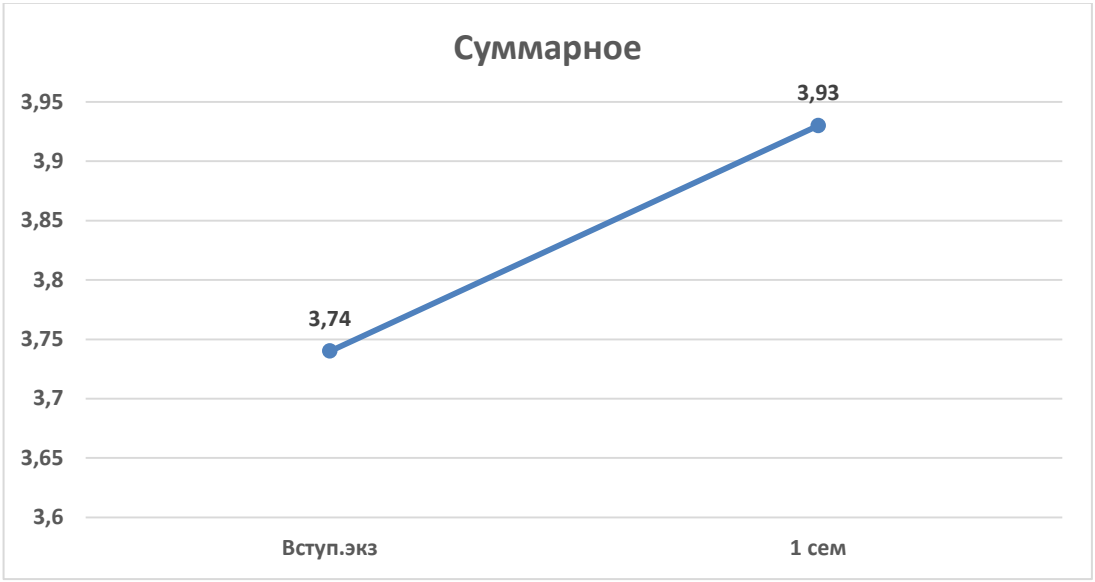


Средние оценки 2 курса 4 группы (15 студентов) специальности «ФИИТ» 2018 года поступления. (отчисленные и восстановленные студенты не учитываются)

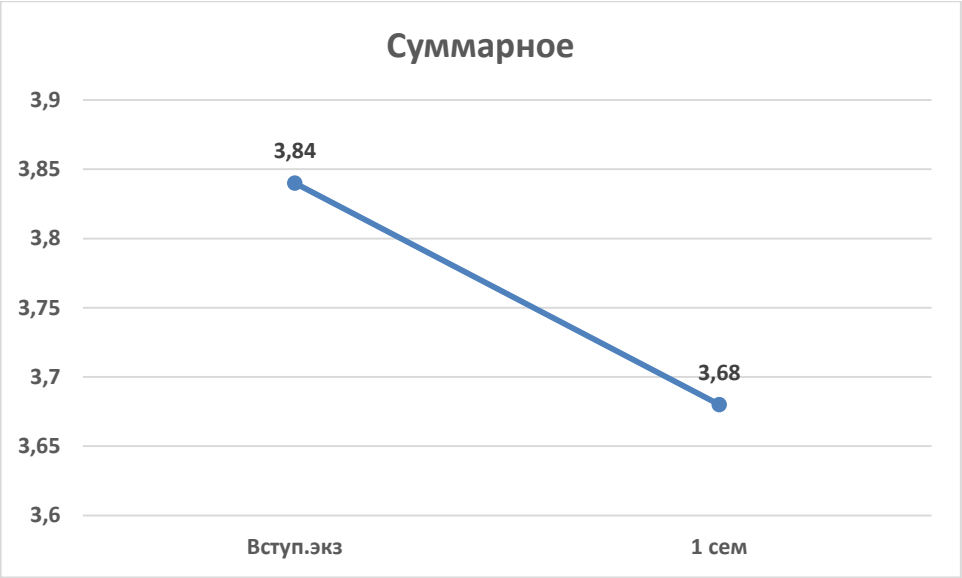


Динамика экзаменационных оценок за 1 семестр.

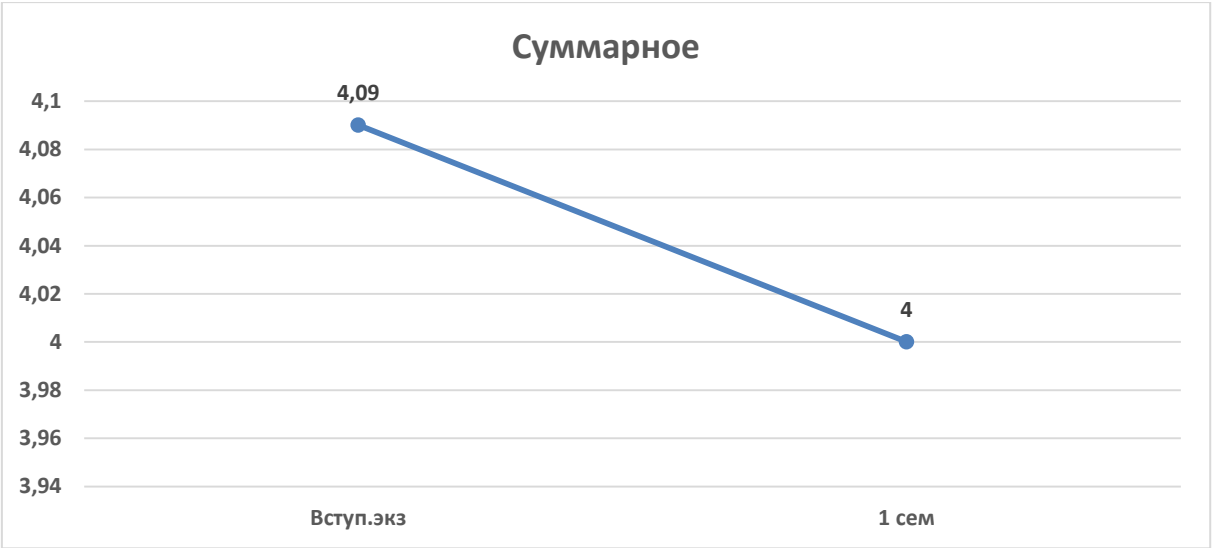
Средние оценки 1 курса МИ-309 группы (23 студентов) специальности «Математика и Информатика» 2019 года поступления. (отчисленные и восстановленные студенты не учитываются)



Средние оценки 1 курса ИМ-309 группы (19 студентов) специальности «Информатика и Математика» 2019 года поступления. (отчисленные и восстановленные студенты не учитываются)



Средние оценки 1 курса ФиА-309 группы (23 студентов) специальности «Физика и Астрономия» 2019 года поступления. (отчисленные и восстановленные студенты не учитываются)



Приложение 3

Уравнения между линейной и нелинейной (квадратичной) регрессии между экзаменационными оценками по математике аттестата, вступительных экзаменов, сессионных экзаменов у студентов 5 курса ФМФ 2015 года поступления.

Уравнения между линейной и нелинейной (квадратичной) регрессии между экзаменационными оценками по математике аттестата, вступительных экзаменов, сессионных экзаменов у студентов 5 курса 1-ой группы ФМФ 2015 года поступления.

Таблица 1

5 курс 1 группа	Ур-е лин. Регрессии	Ур-е нелин. Регрессии
Аттест/ 1 семестр	$y = 0,1481x + 18,556$	$y = 0,3202x^2 – 7,5833x - 23,921$
Аттест/ 2 семестр	$y = 0,7458x + 8,8644$	$y = 0,306x^2 – 8,3203x +75,247$
Аттест/ 3 семестр	$y = 0,6014x + 13,464$	$y = 0,358x^2 - 7,2284x+55,259$
Аттест/ 4 семестр	$y = 0,7869x + 9,6557$	$y = 0,1553x^2 - 3,5525x+39,548$
Аттест/ 5 семестр	$y = 1,083x + 7,0964$	$y = 0,271x^2 - 5,6905x+48,572$
Аттест/ 6 семестр	$y = 0,4498x + 15,207$	$y = 0,9747x^2 - 20,925x+129,17$
Аттест/ 7 семестр	$y = 0,75x + 11,5$	$y = 0,6333x^2 - 14,383x+99,9$
Аттест/ 8 семестр	$y = 0,4383x + 15,642$	$y = -0,274x^2+6,9165x - 21,143$
Аттест/ 9 семестр	$y = 1,0625x+5,75$	$y = -0,0588x^2+2,6471x - 4,8824$
Вступит/ 1 семестр	$y = 0,1429x + 9,7143$	$y = -0,1184x^2+2,8929x - 5,9962$
Вступит/ 2 семестр	$y = -0,1525x + 13,619$	$y = 0,3439x^2 - 10,341x+8821$
Вступит/ 3 семестр	$y = -0,2391x + 13,978$	$y = -0,1543x^2+3,1358x - 4,037$
Вступит/ 4 семестр	$y = -0,2131x + 14,156$	$y = 0,2945x^2 - 8,4452x+70,863$
Вступит/ 5 семестр	$y = -0,0568x + 11,987$	$y = 0,0319x^2 - 0,8528x+16,882$
Вступит/ 6 семестр	$y = -0,313x + 14,775$	$y = 0,4704x^2 -$

		10,63x+69,779
Вступит/ 7 семестр	y = -0,092x + 12,368	$y = 0,2417x^2$-5,8667x+46,1
Вступит/ 8 семестр	y = -0,1296x + 12,648	$y = 0,0525x^2$-1,3677x+19,678
Вступит/ 9 семестр	y = 0,1875x+8,75	$y = -0,8039x^2$+21,843x-136,56
Аттестат/Вступит	y = 0,9x + 10,332	$y = -0,1436x^2$+4,1464x-7,6878

Уравнения линейной и нелинейной (квадратичной) регрессии между экзаменационными оценками по математике аттестата, вступительных экзаменов, сессионных экзаменов у студентов 4 курса 1-ой группы ФМФ 2016 года поступления.

Таблица 2

4 курс 1 группа	Ур-е лин. Регрессии	Ур-е нелин. Регрессии
Аттест/ 1 семестр	y = 1,5676x + 10,108	$y = -0,6288x^2$ +10,795x - 23,167
Аттест/ 2 семестр	y = 0,1316x + 19,804	$y = -0,0603x^2$ +1,8539x +8,2923
Аттест/ 3 семестр	y = 0,2517x + 18,869	$y = -0,2404x^2$ +5,6109x-10,103
Аттест/ 4 семестр	y = 0,2249x + 18,354	$y = -0,0579x^2$+1,9678x+5,7343
Аттест/ 5 семестр	y = 0,623x + 14,164	$y = -0,0122x^2$+0,9146x+12,439
Аттест/ 6 семестр	y = 0,2391x + 18,935	$y = 0,1618x^2$-3,682x+41,947
Аттест/ 7 семестр	y = 0,2154x + 19,385	$y = -0,0476x^2$+1,3582x+12,718
Вступит/ 1 семестр	y = 0,6757x + 8,0811	$y = -0,1212x^2$+2,4545x+1,6667
Вступит/ 2 семестр	y = 0,1754x + 10,48	$y = -0,077x^2$+2,3767x-4,2334
Вступит/ 3 семестр	y = 0,1374x + 11,517	$y = -0,155x^2$+3,593x-7,1612
Вступит/ 4 семестр	y = 0,0487x + 12,37	$y = 0,0712x^2$-2,096x+27,899
Вступит/ 5 семестр	y = 0,3525x + 8,8033	$y = -0,2581x^2$+6,5264x-27,707
Вступит/ 6 семестр	y = 0,125x + 11,625	$y = 0,1082x^2$-

		2,4967x+27,011
Вступит/ 7 семестр	y = 0,1x + 12	$y = 0,0714x^2$-1,6143x+22
Аттестат/Вступит	y = 1,1224x + 7,0612	$y = 0,9048x^2$-21,884x+152,08

Уравнения линейной и нелинейной (квадратичной) регрессии между экзаменационными оценками по математике аттестата, вступительных экзаменов, сессионных экзаменов у студентов 4 курса 2-ой группы ФМФ 2016 года поступления.

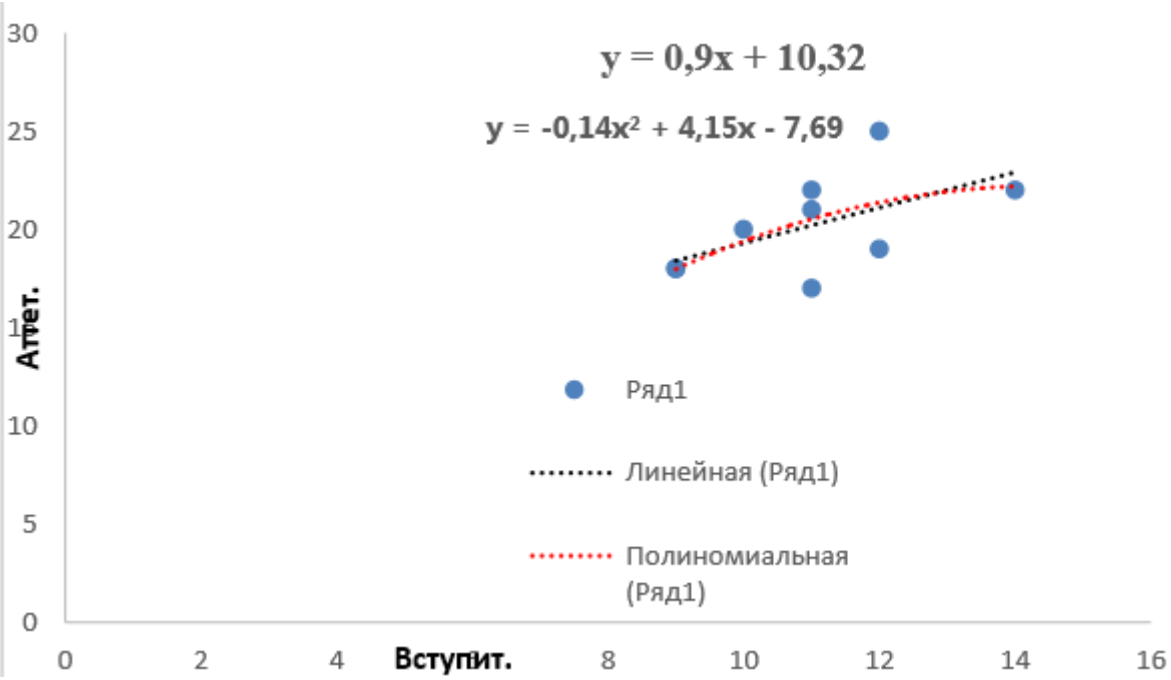
Таблица 3

4 курс 2 группа	Ур-е лин. Регрессии	Ур-е нелин. Регрессии
Аттест/ 1 семестр	y = 0,3x + 17,3	$y = -0,2857x^2$ +4,5x +2,0714
Аттест/ 2 семестр	y = 0,4504x + 14,376	$y = 0,0311x^2$ -0,2777x +18,551
Аттест/ 3 семестр	y = 0,3494x + 15,937	$y = -0,2941x^2$ +6,7941x-18,412
Аттест/ 4 семестр	y = 0,3494x + 15,937	$y = -0,2941x^2$ +6,7941x-18,412
Аттест/ 5 семестр	y = 0,1719x + 17,516	$y = -0,1283x^2$+3,2722x-0,7656
Аттест/ 6 семестр	y = 0,1358x + 16,864	$y = -0,0497x^2$-2,2004x-4,2344
Аттест/ 7 семестр	y = 0,3571x + 15,253	$y = -0,0743x^2$+2,1081x+5,1778
Вступит/ 1 семестр	y = -0,2333x + 12,433	$y = -0,5714x^2$+8,1667x-18,024
Вступит/ 2 семестр	y = 0,0496x + 10,124	$y = -0,0311x^2$+0,7777x+5,9486
Вступит/ 3 семестр	y = 0,0063x + 10,628	$y = 0,0123x^2$-0,2623x+12,059
Вступит/ 4 семестр	y = 0,0063x + 10,628	$y = 0,0123x^2$-0,2623x+12,059
Вступит/ 5 семестр	y = 0,0469x + 10,141	$y = -0,1678x^2$+4,1012x-13,766
Вступит/ 6 семестр	y = -0,0298x + 11,28	$y = -0,0006x^2$-0,0054x+11,03
Вступит/ 7 семестр	y = 10,692	$y = -0,0337x^2$+0,7947x+6,12
Аттестат/Вступит	y = 0,247x + 16,898	$y = -0,2886x^2$+6,6041x-17,792

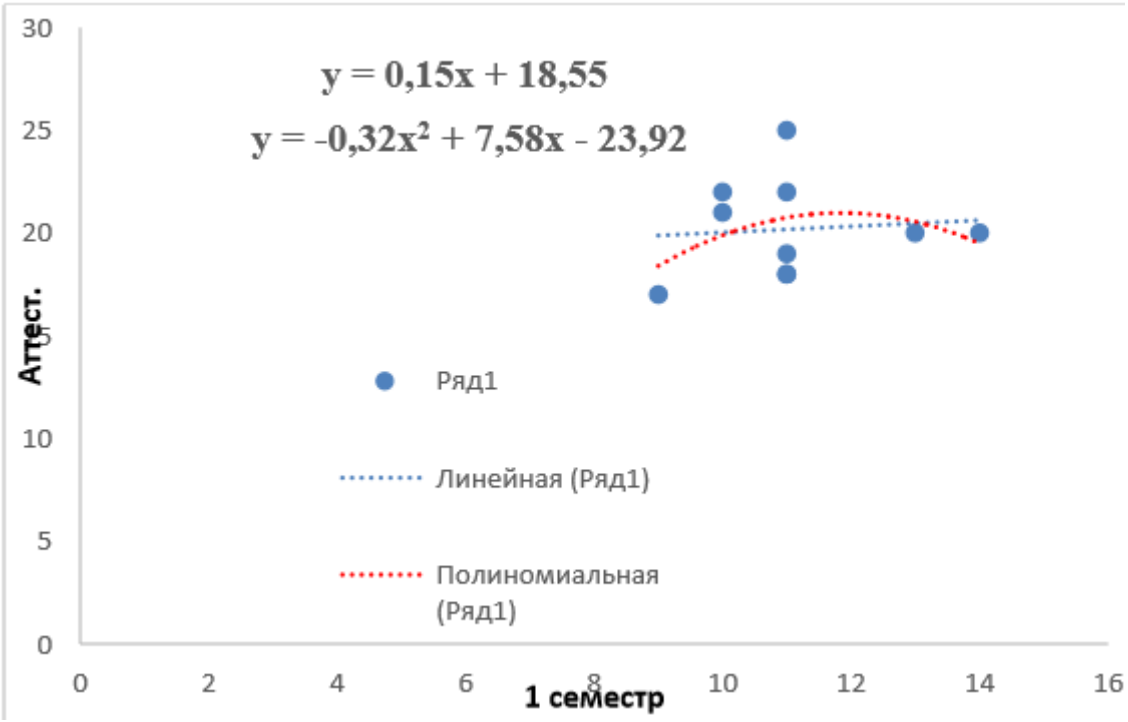
Графики корреляционной зависимости (линейной регрессии) между оценками аттестата и оценками различных сессий, оценками вступительных экзаменов и сессионных экзаменов у студентов ФМФ.

Графики корреляционной зависимости линейной и нелинейной (квадратичной) регрессии между оценками аттестата и оценками различных сессий, оценками вступительных экзаменов и сессионных экзаменов с 1 по 9 семестры по студентам 5 курса 1 группы, ФМФ 2015 года поступления.

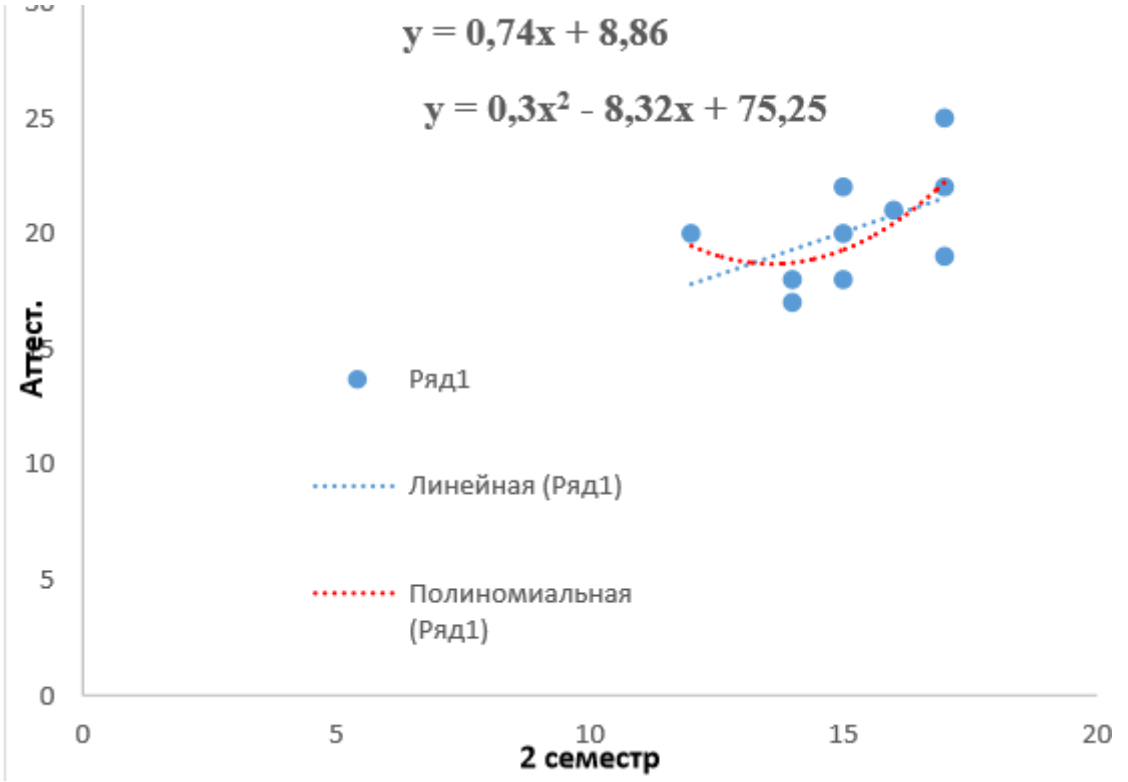
Зависимость вступительных оценок от оценок аттестата



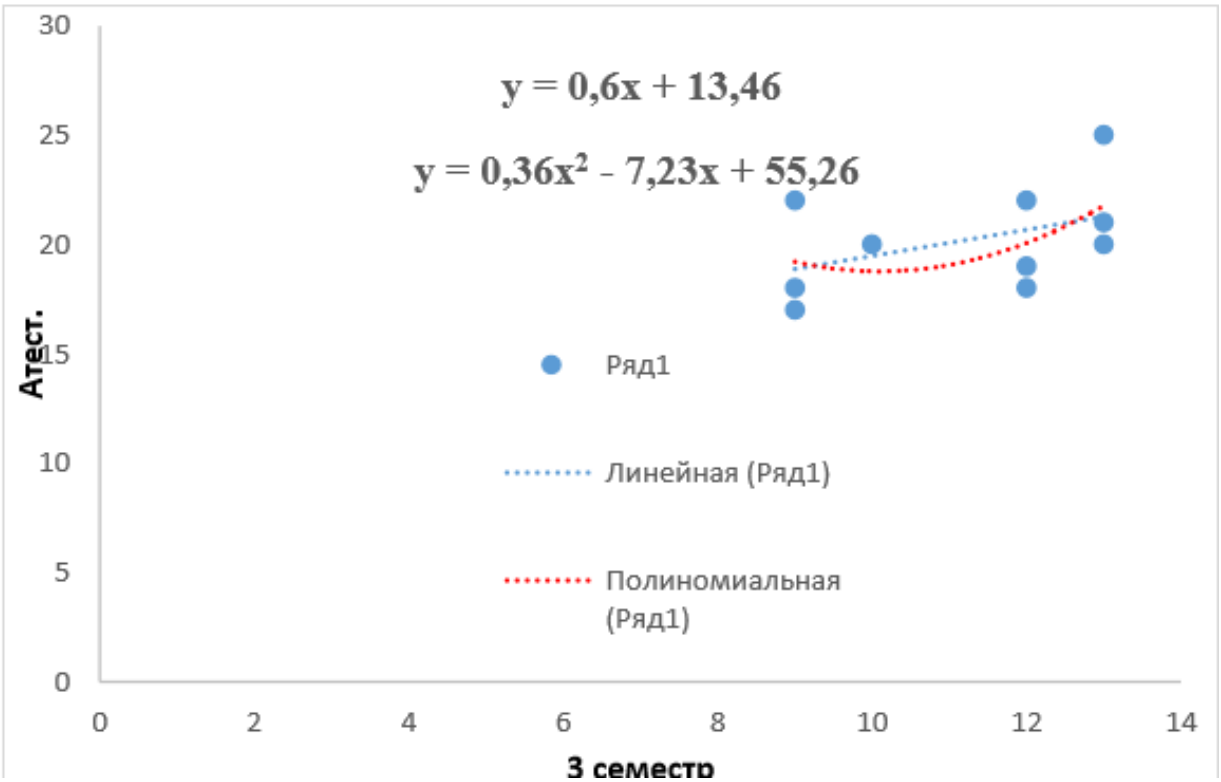
Зависимость между оценками аттестата и результатами 1-го семестра



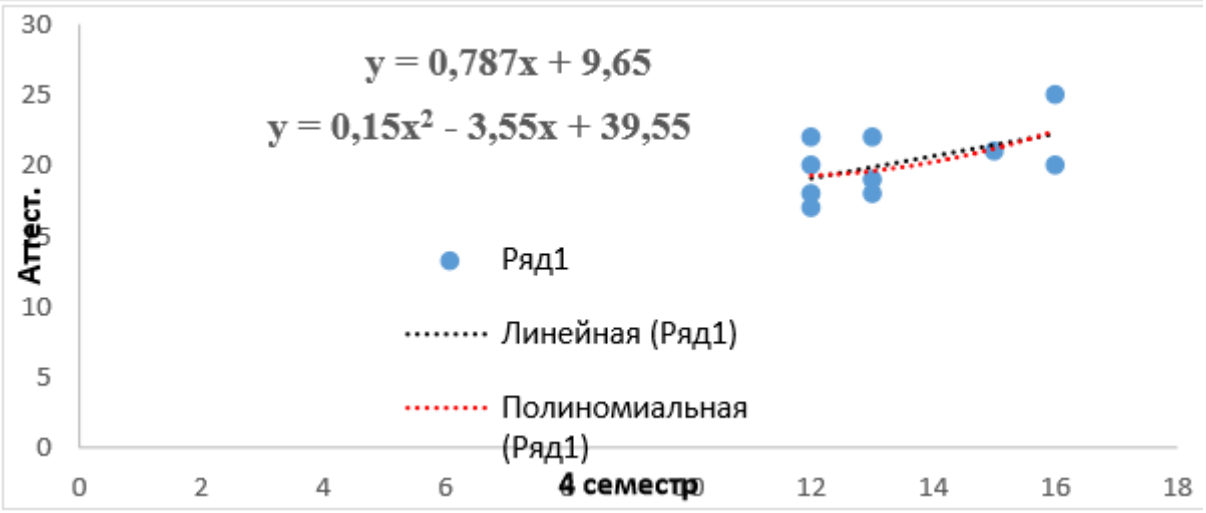
Зависимость между оценками аттестата и результатами 2-ого семестра



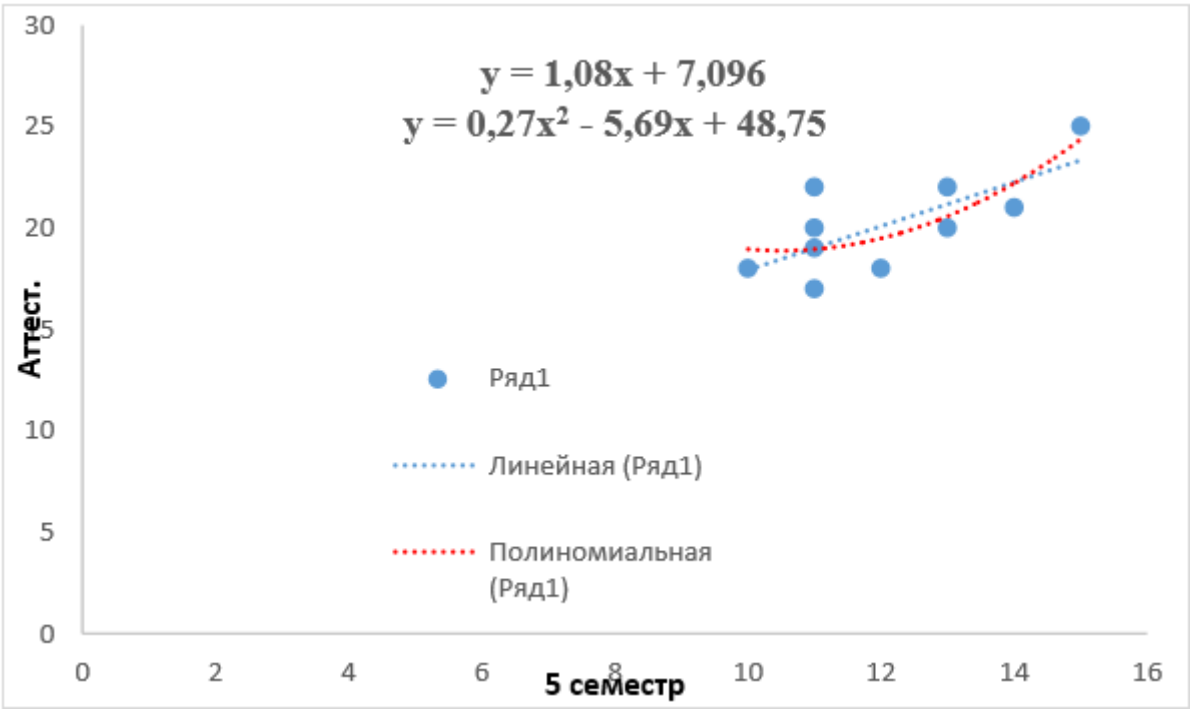
Зависимость между оценками аттестата и результатами 3-его семестра



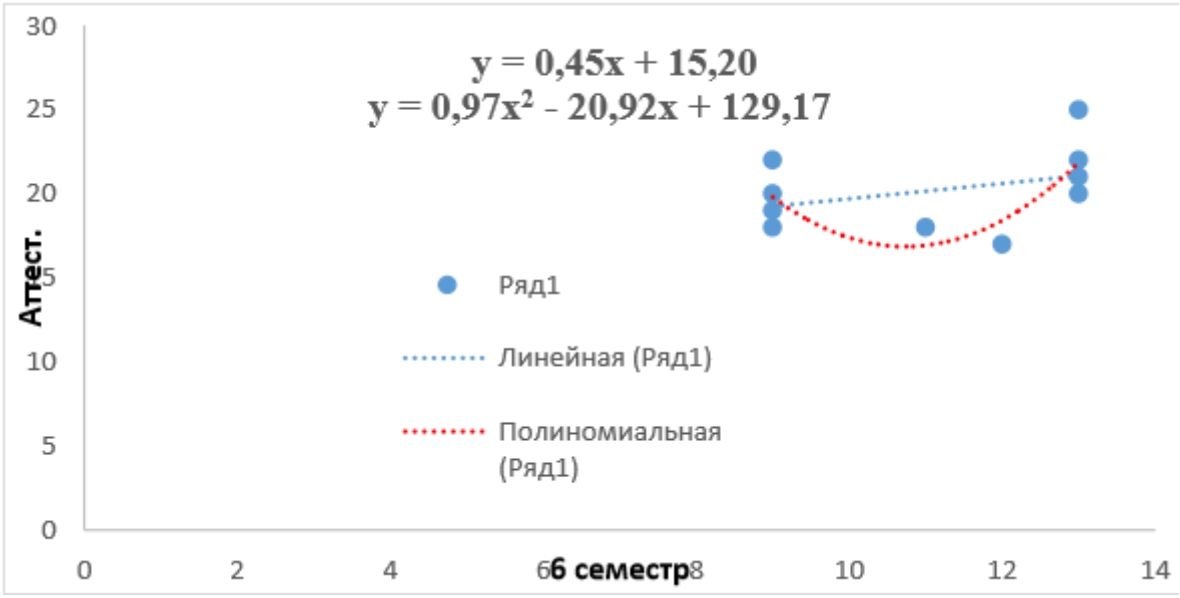
Зависимость между оценками аттестата и результатами 4-го семестра



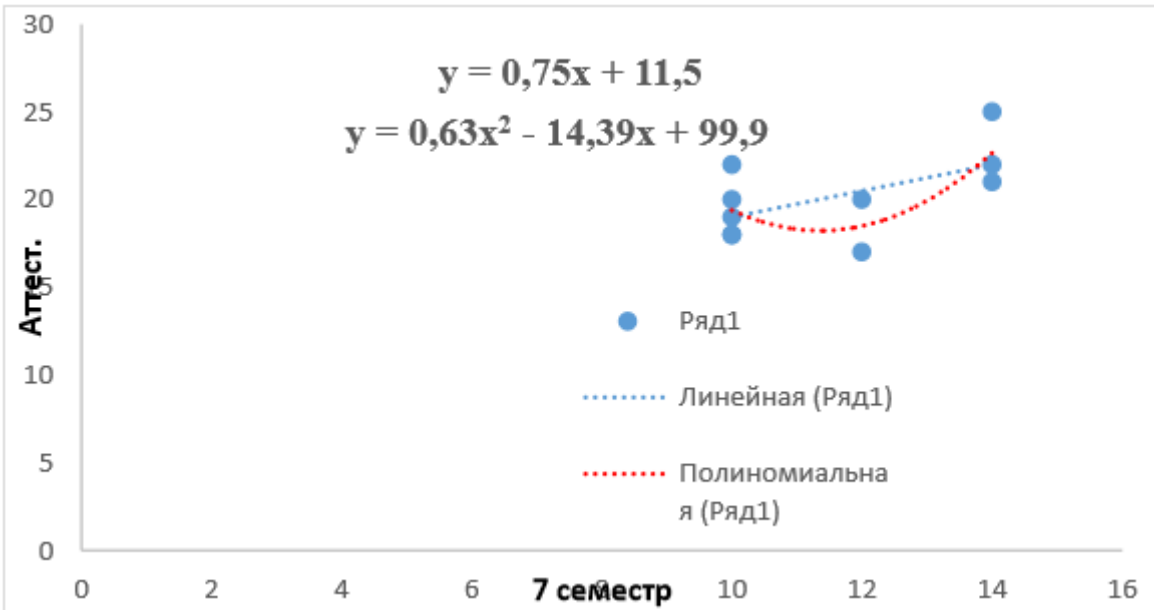
Зависимость между оценками аттестата и результатами 5-го семестра



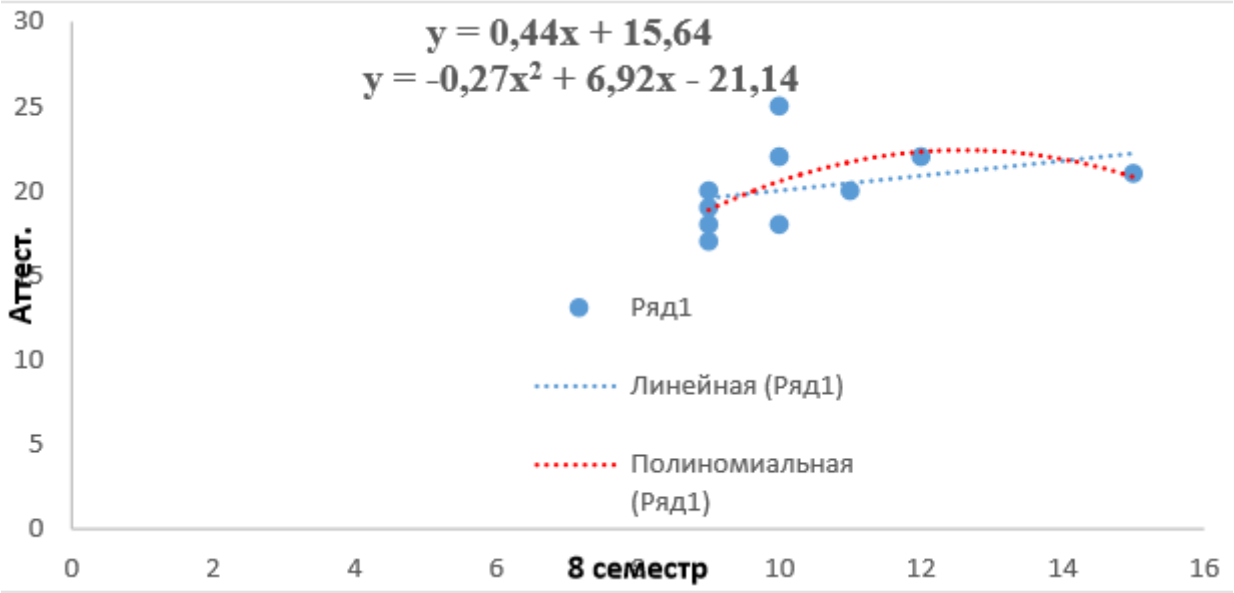
Зависимость между оценками аттестата и результатами 6-го семестра



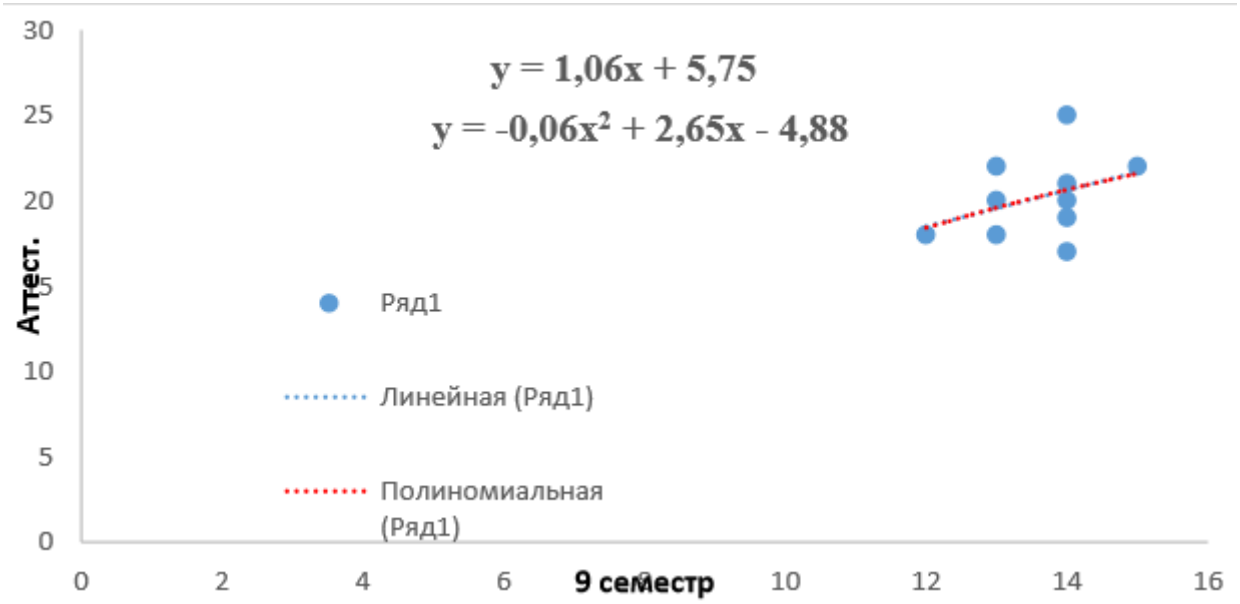
Зависимость между оценками аттестата и результатами 7-го семестра



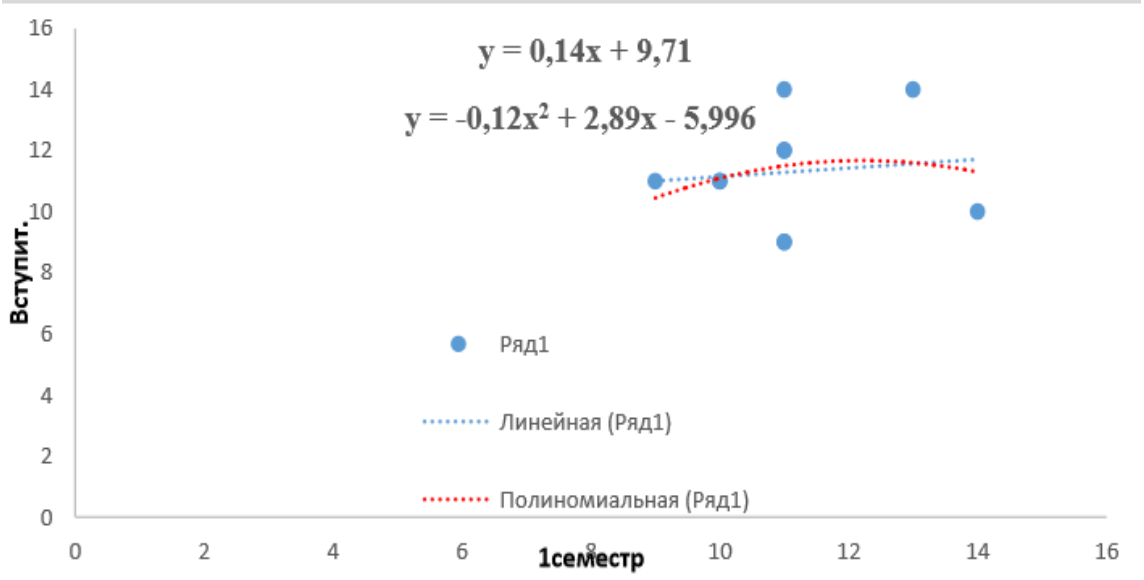
Зависимость между оценками аттестата и результатами 8-го семестра



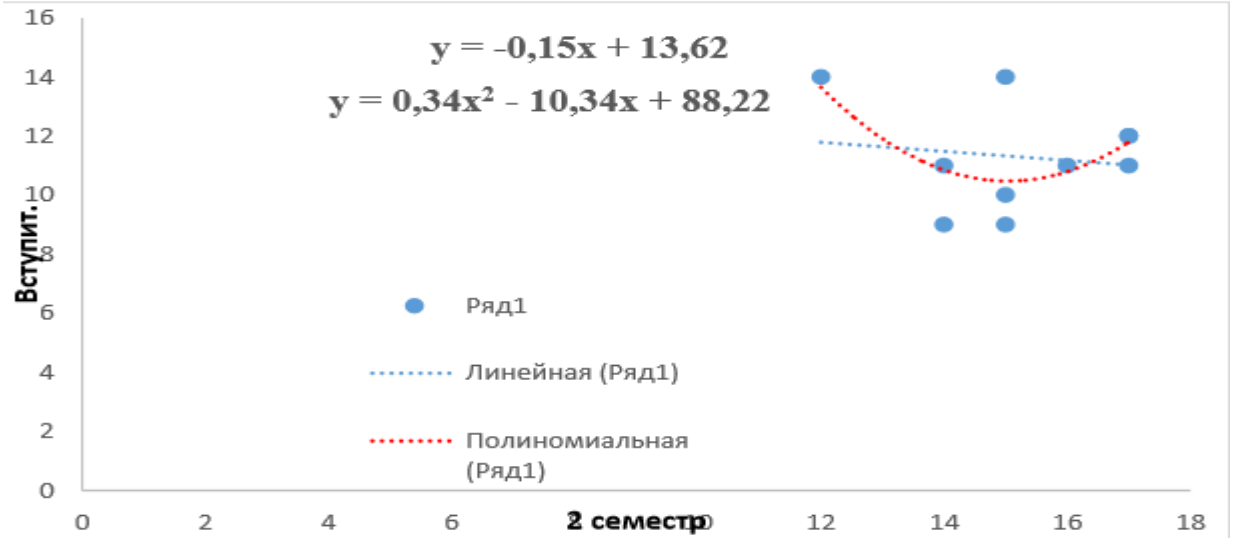
Зависимость между оценками аттестата и результатами 9-го семестра



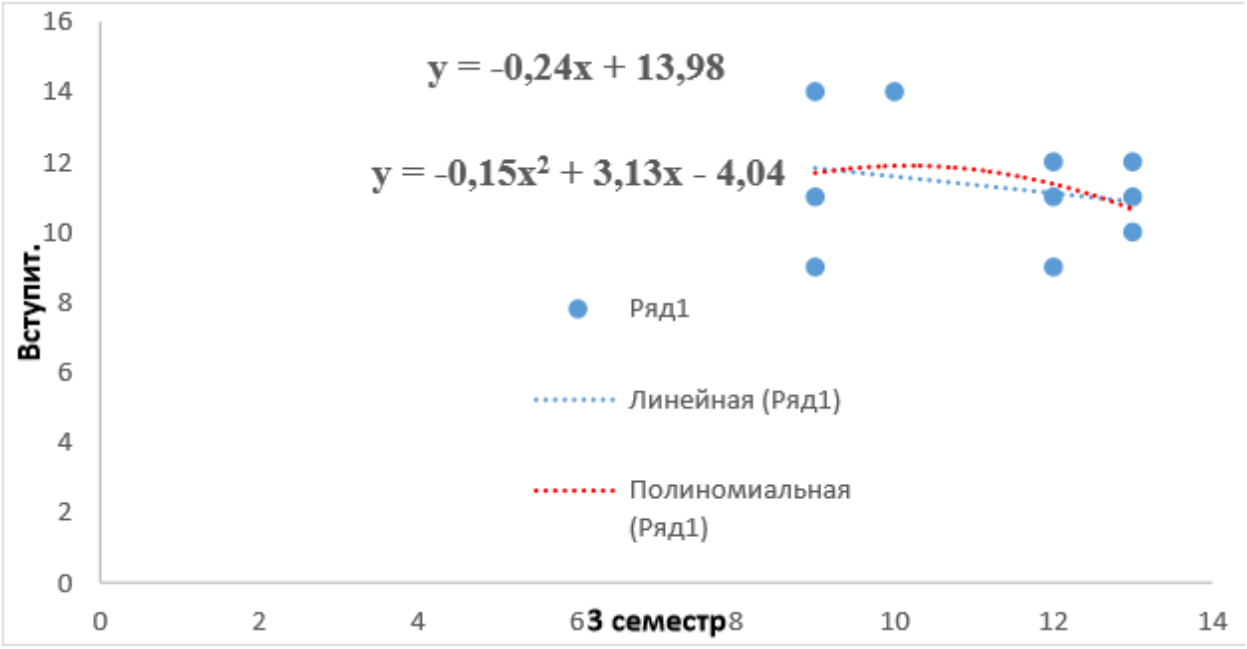
Зависимость между вступительных оценок и результатов 1-го семестра



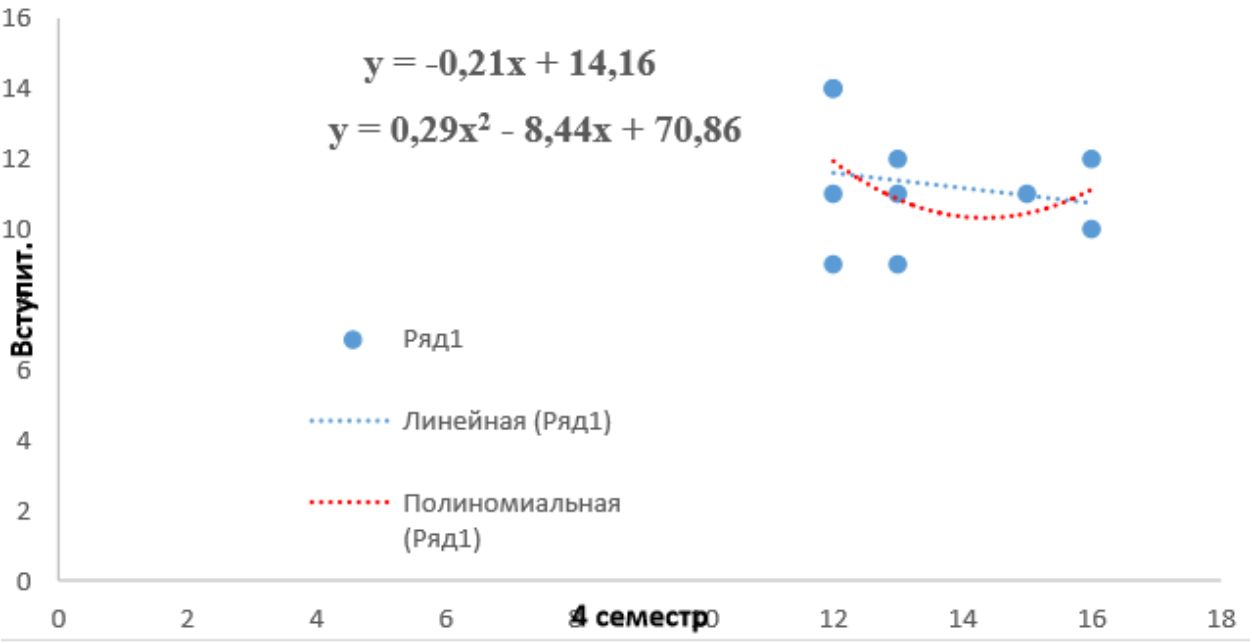
Зависимость между вступительных оценок и результатов 2-го семестра



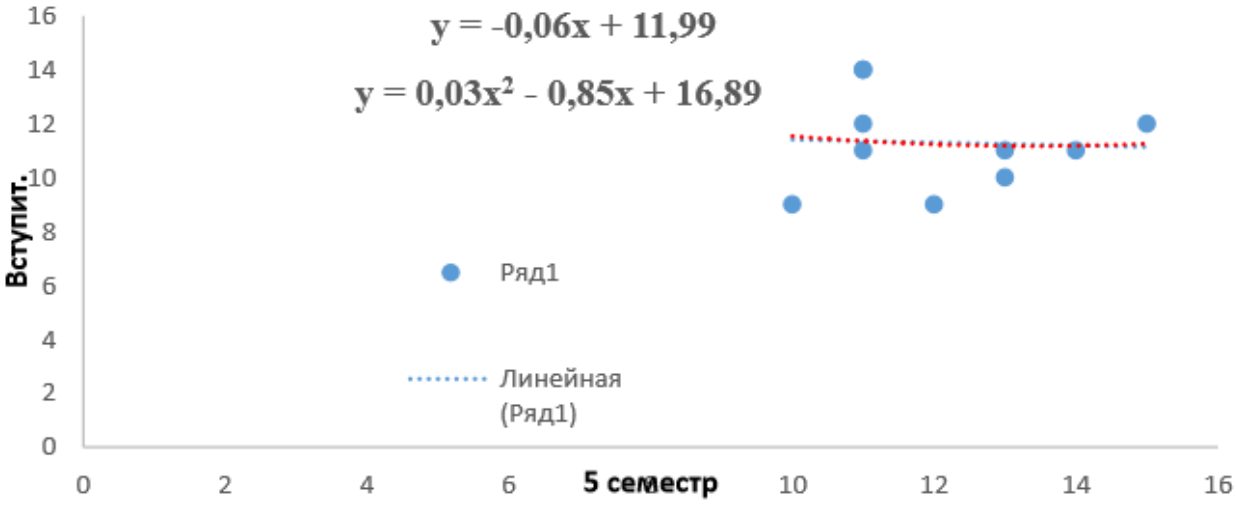
Зависимость между вступительных оценок и результатов 3-го семестра



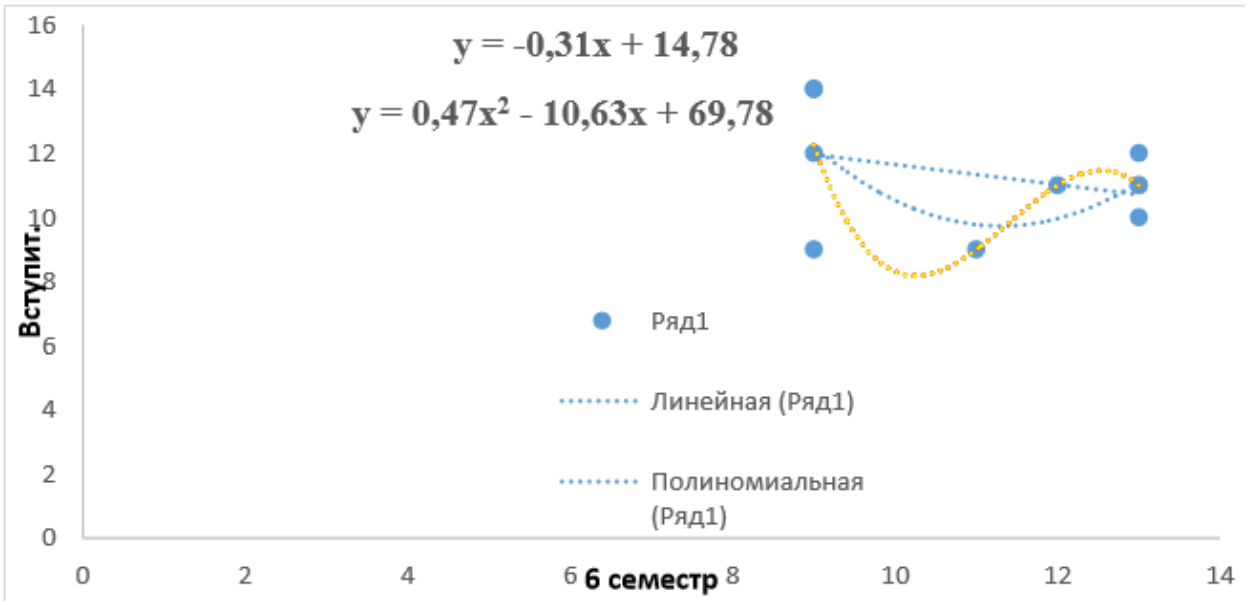
Зависимость между вступительных оценок и результатов 4-го семестра



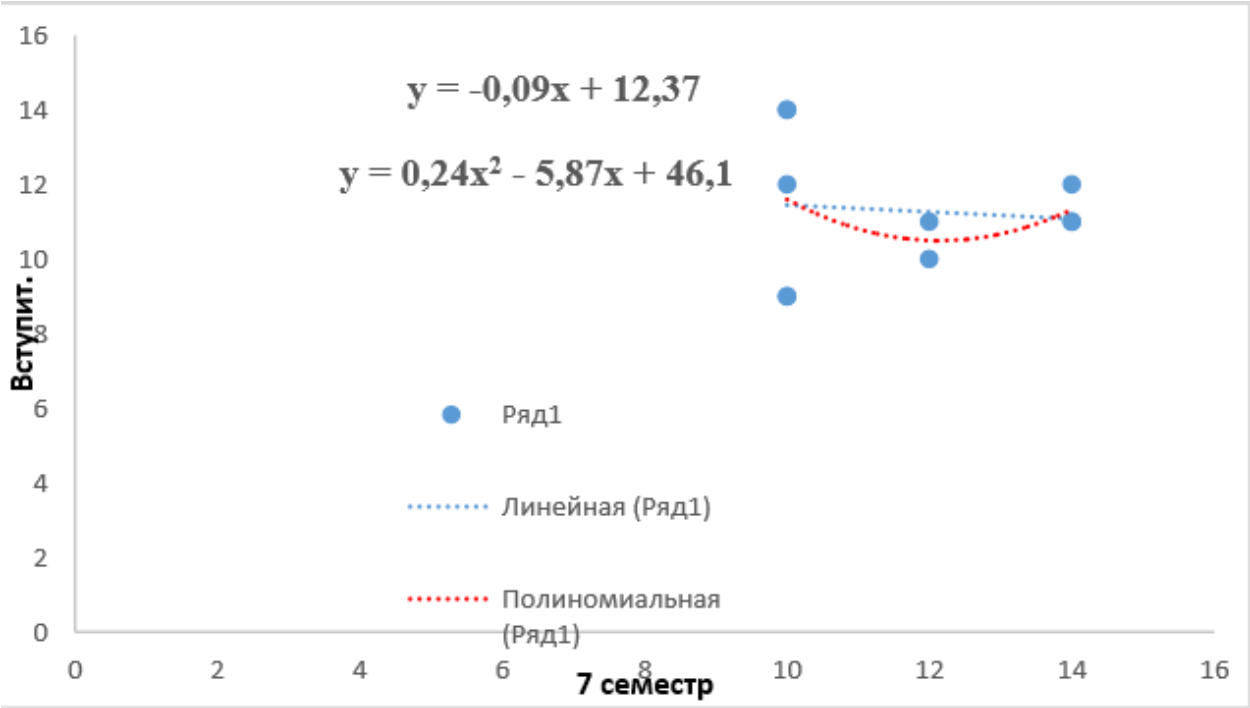
Зависимость между вступительных оценок и результатов 5-го семестра



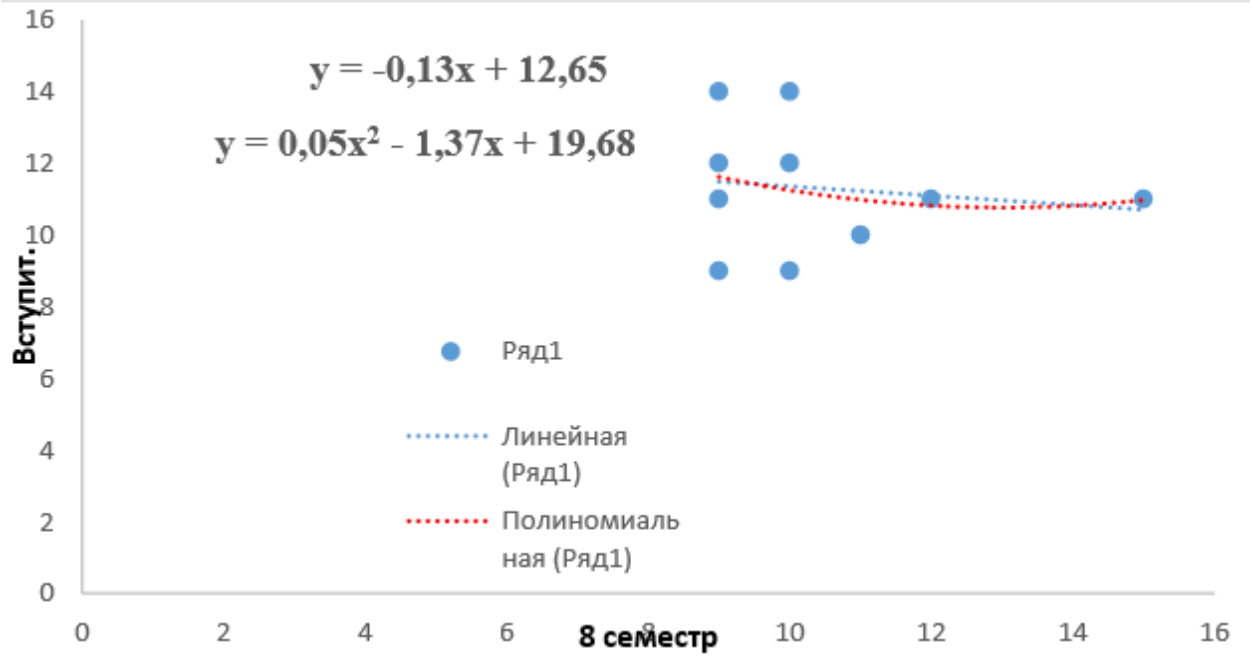
Зависимость между вступительных оценок и результатов 6-го семестра



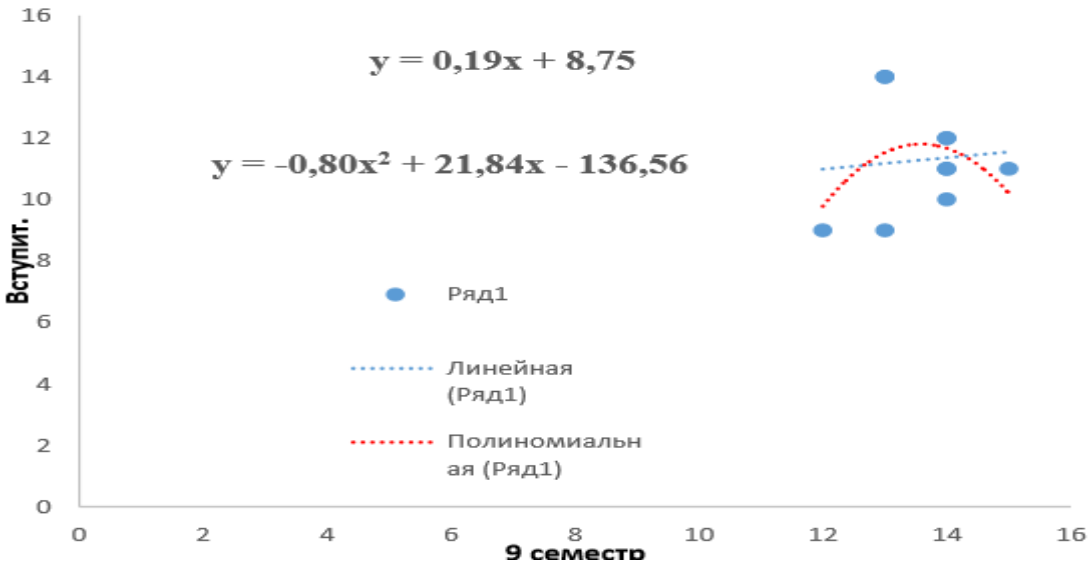
Зависимость между вступительных оценок и результатов 7-го семестра



Зависимость между вступительных оценок и результатов 8-го семестра



Зависимость между вступительных оценок и результатов 9-го семестра



Приложение 4

Проверка гипотез по статистике успеваемости в пакете STATISTICA. 4 курс 2016 года поступления.

Задача 1: Даны две независимые выборки(средние оценки за7семестр по специальности «математика и информатика», «информатика и математика» по математическим дисциплинам) объемами n_1=15, n_2=12.

Проверяем гипотезу Но, утверждающую, что группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: выборки получены из разных ГС.

Уровень значимости α=0,05.

Подробные решения аналогичных задач были рассмотрены выше. Здесь приведем краткие решения задач и сделаем выводы с помощью трех методов.

Решение 1: Метод Вальда-Вольфовица



➢ ***Вывод:***

1. $z_B \approx -2{,}915$;это значение меньше квантили распределенияN(0,1)

   $u_{0.95}$=1.960,то есть–2,915<1,960;

2. $p < \alpha$,то есть0,00<0,05.

Исходя из первого и второго условий, получаем, что гипотеза H_o отклоняется, то есть две независимые выборки не принадлежат одной генеральной совокупности.

Решение 2. Метод Колмогорова-Смирнова.



➢ ***Вывод:***

1. $D_{отр.}$=-0,10;
2. p <0,25,

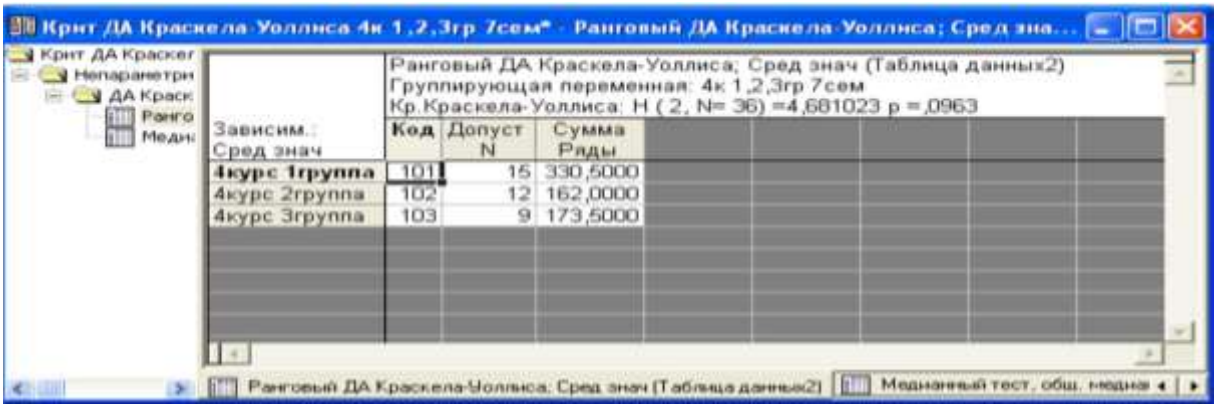
Из первого и второго условий следует, что гипотеза H_o отклоняется, то есть две независимые выборки не принадлежат одной генеральной совокупности.

<u>Задача 2.</u> Даны три независимые выборки(средние оценки за7семестров) у 4 курса по 3 группам по специальности «математика и информатика», «информатика-математика», «физика» по математическим дисциплинам объемами n_1=15,n_2=12, n_3=9. Проверяем гипотезу Но, утверждающую, что группы данных представляют случайные независимые выборки из одной генеральной совокупности, т.е. не отличаются друг от друга по наблюдаемому признаку. Альтернативная гипотеза H_1: выборки получены из разных ГС.

Уровень значимости α=0,10 .
Подробное решение аналогичной задачи была рассмотрена выше.

Здесь приведем краткое решение задачи и сделаем вывод с помощью четвертого метода.

Решение методом «Ранговый дисперсионный анализ Краскела-Уоллиса»:



➢ ***Вывод:***

1. *H_B=4,68 меньше чем квантиль распределения* $\chi_0^2{}_{,95}(4-1)=7{,}81$
2. *Условие не выполняется $p<\alpha$, то есть 0.09<0.10,*

   Из первого и второго условий следует, что гипотеза H_0 отклоняется, то есть три независимые выборки не принадлежат одной ГС.