

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
АМУРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(ФГБОУ ВО «АмГУ»)

**МЕТОДЫ МАТЕМАТИЧЕСКОЙ ОБРАБОТКИ В ПЕДАГОГИЧЕСКИХ
ИССЛЕДОВАНИЯХ**
сборник учебно-методических материалов
для направления подготовки

44.06.01 – Образование и педагогические науки

2017 г.

*Печатается по решению
редакционно-издательского совета
факультета математики и информатики
Амурского государственного
Университета*

Составители: Юрьева Т.А.

Методы математической обработки в педагогических исследованиях:
сборник учебно-методических материалов для направления подготовки 44.06.01
– Образование и педагогические науки. – Благовещенск: Амурский гос. ун-т,
2017.

Рассмотрен на заседании кафедры общей математики и информатики
03.11.2017, протокол № 3.

© Амурский государственный университет, 2017

© Кафедра общей математики информатики, 2017

© Юрьева Т.А., составление

ВВЕДЕНИЕ

Цель дисциплины: формирование готовности использовать математические и статистические методы обработки и анализа экспериментальных данных в практических и научных педагогических исследованиях.

Задачи дисциплины:

- на примерах математических понятий и методов продемонстрировать сущность научного подхода, специфику математики, ее роль в развитии педагогических исследований;
- научить аспирантов приемам исследования и решения, математически формализованных задач педагогики.

В результате освоения обучающийся должен демонстрировать следующие результаты образования:

- 1) Знать: основные математические и статистические методы обработки данных, полученных при решении основных профессиональных задач;
- 2) Уметь: получать, обрабатывать и интерпретировать данные исследований с помощью математико-статистического аппарата;
- 3) Владеть: навыками использования в профессиональной деятельности базовых знаний в области математической статистики.

1 КРАТКОЕ ИЗЛОЖЕНИЕ ЛЕКЦИОННОГО МАТЕРИАЛА

Тема: Структура курса. Анализ данных на компьютере

Математическая статистика занимается математическим описанием случайных явлений, т.е. построением вероятностных моделей, а также проверкой их пригодности. Поэтому выделяют два раздела: описательную статистику и статистику «проверяющую» (статистическую проверку гипотез); соответственно разделяется и методический аппарат. Понятия и методы описательной статистики создаются в теории вероятностей, а понятия и методы статистической проверки гипотез создаются в специальных теориях (например, в теории статистических решений) либо в приложениях теории вероятностей к конкретным наукам.

В словесном описании результатов наблюдения и эксперимента трудно избежать элементов субъективизма, которые проявляются чаще всего в преждевременных выводах и необоснованных обобщениях. Известно, что качественное описание является недостаточно точным, поскольку с помощью языковых средств сложно передать дифференцированность изучаемых явлений и особенно своеобразие их динамики. Одно только качественное описание не позволяет определить также и степень ошибки наблюдения или эксперимента. Но это вовсе не означает, что в психологии необходимо отказаться от качественного анализа в пользу оперирования исключительно количественными показателями. Мы хотим подчеркнуть лишь то обстоятельство, что количественный анализ результатов исследования должен не только предшествовать, но и обязательно следовать за качественным анализом.

Это особенно важно для интерпретации результатов исследования. Именно таким образом может быть преодолен субъективизм, так как формулируемые суждения и выводы становятся более независимыми от личности исследователя и обеспечивается возможность их проверки. Знание различных приемов обработки и анализа результатов наблюдений и эксперимента с помощью статистических показателей является обязательным для педагога. Поэтому первый раздел нашего практикума посвящен анализу видов психоло-педагогических измерений и способов статистической обработки получаемых при этом результатов. Однако следует иметь в виду, что знакомство с материалом данного раздела не может заменить студенту систематического изучения математической статистики. Нами будут рассмотрены лишь элементарные статистические методики, без которых нельзя обойтись на практических занятиях по общей и экспериментальной психологии.

Процесс измерения лежит в основе любой эмпирической науки. Беглый взгляд на историю науки показывает, что совершенствование принципов и техники измерения было основным фактором, обеспечивающим ее движение вперед. Самого высокого уровня развития на сегодняшний день достигли те ее области, которым быстрее удалось преодолеть трудности, связанные с разработкой методологических и методических проблем измерения. Прежде чем рассмотреть приемы измерения, используемые в психолого-педагогическом эксперименте, и способы статистической обработки его результатов, познакомимся с основными методами психолого-педагогического исследования.

Исходя из порядка операций с объектами в научном исследовании, Б. Г. Ананьев разработал классификацию методов современной психологии. В основу ее он положил целостный цикл психологического исследования и все методы распределил по четырем группам. В первую группу, которую можно назвать группой организационных методов, Ананьев относит сравнительный, лонгитюдный (т. е. исследование одних и тех же лиц в течение длительного времени) и комплексный методы: «Они действуют на протяжении всего исследования, и их эффективность определяется по конечным результатам исследования...». Вторая группа методов включает известные также по традиционным классификациям эмпирические способы добывания научных данных. В эту группу входят: наблюдательные методы (наблюдение и самонаблюдение), экспериментальные методы (лабораторные, полевые, психолого-педагогические), психодиагностические методы (тесты, анкеты, опросники, интервью, беседы), праксиметрические методы (приемы анализа процессов и продуктов деятельности: хронометрии, профессиографическое описание, оценка выполненных работ), моделирование

(математическое, кибернетическое) и биографические методы (приемы исследования жизненного пути, изучение документации). Третью группу методов составляют приемы обработки результатов эксперимента и наблюдений. Ананьев в эту группу относит как стандартные приемы статистической обработки данных (количественная обработка), так и приемы качественного анализа, включая дифференциацию материала по классам, разработку типологии, составление психологической казуистики (описание случаев). Четвертая группа методов — интерпретационные методы — представлена в классификации Ананьева вариантами генетического и структурного методов. Генетический метод интерпретирует весь обработанный материал исследования в характеристиках развития, а структурный метод — в характеристиках типов связей между отдельными компонентами структуры изучаемой личности или структуры социальной группы.

Приступая к выбору методики, экспериментатор должен иметь четкое представление о том, что именно он хочет измерить, и удовлетворят ли результаты измерения требованиям адекватного решения исследовательской или практической задачи. В первую очередь ему надлежит доказать валидность, надежность и объективность избранной методики. Под валидностью методики понимается адекватность ее предмету исследования. Количественно валидность определяется путем установления взаимосвязи между результатами, полученными с помощью данной методики, и каким-либо из внешних критериев. Поясним сказанное примером. Очевидно, что успешность обучения в какой-то степени обусловлена уровнем интеллектуального развития обучаемого, и поэтому в качестве внешнего критерия правомерно рассматривать оценку его успеваемости. Допустим, что было проведено тестовое исследование умственного развития группы лиц, например студентов, с помощью избранной методики. Так вот, применяемая методика может считаться валидной лишь в том случае, если между результатами тестирования и оценкой успеваемости в обучении будет обнаружена положительная взаимосвязь. Не менее важным аспектом оценки качества методики является ее надежность. Под надежностью психологической методики понимается точность производимых с ее помощью измерений. Иначе говоря, через надежность определяется пригодность данной методики в качестве измерительного инструмента. Наконец, объективность методики характеризует степень независимости результатов измерения от пользователя данной методики. Объективными результаты будут лишь в том случае, если, во-первых, они независимы от личностных особенностей пользователя и, во-вторых, исключен произвол в их обработке и интерпретации.

Для проверки валидности и надежности методик чаще всего привлекаются количественные (статистические) критерии оценки. Объективность методики можно обосновать исходя из положений общей теории измерений и специфики их в отношении психолого-педагогического исследования. Конечно, это вовсе не означает, что психологическое исследование исчерпывается измерением. Однако знание разнообразных измерительных процедур вооружает педагога исследовательским инструментом, с помощью которого он способен решать психолого-педагогические задачи.

Тема: Начала теории измерений

С точки зрения теории измерения, все множество различных измерительных процедур, применяемых в педагогике и психологии, является процедурами построения шкалы эмпирической переменной, иначе говоря, процедурами психологического шкалирования. В понимании большинства психологов шкалирование — это совокупность экспериментальных и математических приемов для измерения особенностей психических процессов и состояний. Вслед за С. С. Стивенсоном в настоящее время понятие «шкалирование» рассматривают в качестве синонима понятия «измерение». Под шкалированием психологических процессов, свойств, объектов или событий понимается процесс приравнивания к этим процессам, свойствам, объектам или событиям чисел по определенным правилам, а именно таким образом, чтобы в отношениях чисел отображались отношения явлений, подлежащих измерению.

Итак, измерение состоит в отображении эмпирических систем с помощью математических систем, а целью такого рода отображения является частичная замена действий, произ-

водимых с реальными предметами, формальными действиями с числами. Область чисел выполняет функцию модели определенных свойств предметов и в качестве средства познания дает возможность более глубоко проникать в объективно существующие свойства и взаимосвязи. В этом смысле шкалирование (измерение) служит главной силой, преобразующей психологию из науки описательной, следующей за фактами, в науку, умеющую предсказывать новые факты.

Понятно, что относительно разных эмпирических систем мы должны использовать разные методики измерения, т. е. применять измерительные шкалы разных типов. Понимание исследователем формальных аспектов измерения является необходимым условием для адекватного выбора им измерительных инструментов и процедур, а также для применения адекватных методов анализа полученных в наблюдении и эксперименте данных. Основываясь на правилах измерения, принято различать несколько типов шкал, с каждым из которых могут быть соотнесены конкретные процедуры шкалирования. При этом каждый тип шкалы может быть охарактеризован соответствующими числовыми свойствами. Рассмотрим более подробно основные свойства разных типов шкал, эмпирические операции, допустимые на уровне этих шкал, а также статистические приемы обработки и анализа исходных или, как их чаще называют, первичных результатов исследования.

Шкалы наименований, или номинативные шкалы. Шкала наименований представляет собой взаимнооднозначное отображение некоторой эмпирической системы в числовой системе. Таким образом, шкала наименований отображает взаимнооднозначное соответствие между классами эквивалентности, т. е. классами эмпирических объектов — обозначений. Само название «шкала наименований» указывает на то, что в этом случае шкальные значения играют роль лишь названий классов эквивалентности.

Шкалы порядка, или ординальные шкалы. В порядковых измерениях символы, в частности числа, присваивают классам объектов так, чтобы первые отображали не только равенство или неравенство, эквивалентность или неэквивалентность, но и упорядоченность объектов в отношении измеряемого свойства. В шкалах порядка массы объектов, как и в случае шкал наименований, являются дискретными. И хотя числа можно сравнивать, всегда надо помнить, что в шкалах порядка их величины имеют лишь относительное, а не абсолютное значение. Например, если какой-то один класс объектов обозначен большим числом, чем другой, то мы понимаем, что по измеряемой характеристике первый превосходит второй, но при этом нам неизвестно, насколько велико это различие. Дело в том, что в самих измерительных операциях, связанных с установлением порядка, не содержится никаких данных о величине различий. Шкала порядка отображает монотонное возрастание или убывание измеряемого признака с помощью монотонно возрастающих или монотонно уменьшающихся чисел. Оценить направление изменения признака можно только в том случае, если шкала порядка содержит не меньше трех классов, которые образуют последовательность. Из-за того, что в шкале порядка устанавливается последовательность классов, любые преобразования, связанные с перестановками элементов этой шкалы, недопустимы.

Шкалы интервалов. Когда шкала обладает всеми свойствами к порядковой шкалы и дополнительно к этому определены еще расстояния между ее единицами, то такую шкалу называют шкалой интервалов. Иначе говоря, классы объектов шкал интервалов всегда дискретны и упорядочены по степени возрастания (или убывания) измеряемого свойства. Кроме того, в эти шкалах одинаковым разностям степени выраженности измеряемого свойства соответствуют равные разности между приписываемыми им числам. Шкалы интервалов имеют равные единицы измерения, однако способ их определения является произвольным, следовательно, и сами единицы произвольны. При этом неизвестна абсолютная величина отдельных значений по шкале, поскольку шкала интервалов не имеет естественной нулевой точки отсчета. Последняя может быть произвольно смещена. Шкалам интервалов присущи все те отношения, которые характерны для номинативных и порядковых шкал. Кроме того, для них возможно использование арифметических действий.

Шкалы отношений. Конструирование шкал отношений предполагает наряду с наличием свойств предыдущих шкал существование постоянной естественной нулевой точки отсчета, в которой измеряемый признак полностью отсутствует. Следовательно, шкалы отношений характеризуются тем, что в них, во-первых, классы объектов разделены и упорядочены согласно измеряемому свойству, во-вторых, равным разностям между классами объектов соответствуют равные разности между приписываемыми им числами, в-третьих, числа, приравняемые классам объектов, пропорциональны степени выраженности измеряемого свойства. Последнее не было свойственно рассмотренным выше шкалам.

Тема: Критическая статистика

В процессе статистического анализа иногда бывает необходимо сформулировать и проверить предположения относительно величины независимых параметров или закона распределения изучаемой генеральной совокупности (совокупностей). Такие предположения называются статистическими гипотезами.

Статистические гипотезы подразделяются на нулевые и альтернативные.

Выдвинутая гипотеза называется нулевой (основной). Ее принято обозначать H_0 . Обычно нулевая гипотеза – это гипотеза об отсутствии различий.

По отношению к высказанной нулевой гипотезе всегда можно сформулировать альтернативную (конкурирующую), противоречащую ей. Альтернативную гипотезу принято обозначать H_1 .

Цель статистической проверки гипотез состоит в том, чтобы на основании выборочных данных принять решение о справедливости нулевой гипотезы H_0 .

Так как проверка статистических гипотез осуществляется на основании выборочных данных, то такое решение неизбежно сопровождается некоторой, хотя возможно и очень малой, ошибкой.

Ошибка, состоящая в том, что мы отклонили нулевую гипотезу, в то время как она верна, называется ошибкой I рода, а ее вероятность – уровнем значимости α .

Ошибка, состоящая в том, что мы приняли нулевую гипотезу, в то время как она неверна, называется ошибкой II рода, а ее вероятность обозначают β . Величину равную $1 - \beta$ называют мощностью критерия.

Мощность критерия определяется эмпирическим путем, а уровень значимости задается исследователем. В психологических и социологических исследованиях низшим уровнем значимости принято считать $\alpha = 0,05$ а достаточным $\alpha = 0,01$.

Статистический критерий — это правило (формула), по которому определяется мера расхождения результатов выборочного наблюдения с высказанной гипотезой H_0 .

Значение критерия, рассчитываемое по специальным правилам на основании выборочных данных, называется наблюдаемым значением критерия.

Значения критерия, определяемые на заданном уровне значимости α по таблицам распределения случайной величины, выбранной в качестве критерия, называются критическими точками.

В психолого-педагогических и социологических исследованиях принято определять значения критерия при $\alpha = 0,01$ и $\alpha = 0,05$. Полученные критические точки делят совокупность значений критерия на область допустимых значений или зону незначимости (область принятия нулевой гипотезы), критическую область или зону значимости (область принятия альтернативной гипотезы) и зону неопределенности.

Чаще всего критерии, используемые при психологических и социологических исследованиях, имеют положительные значения. Поэтому для простоты при решении прикладных задач изображают только неотрицательную часть оси значимости.

Основной принцип проверки статистических гипотез состоит в следующем:

- если наблюдаемое значение критерия принадлежит критической области, то нулевая гипотеза H_0 отклоняется и принимается конкурирующая H_1 ;

- если наблюдаемое значение критерия принадлежит области допустимых значений, то нулевую гипотезу H_0 нельзя отклонить;

-если наблюдаемое значение критерия принадлежит зоне неопределенности, то мы уже можем отклонить нулевую гипотезу H_0 , но еще не можем принять конкурирующую H_1 .

Критерии делятся на параметрические – включающие в формулу расчета параметры распределения (средние и дисперсии) и непараметрические – основанные на оперировании частотами или рангами.

Тема: Параметрические критерии

1. t-критерий Стьюдента используется

а) для сравнения выборочной средней \bar{x} с некоторым известным числовым значением a_0 .

Возможны гипотезы:

H_0 : $\bar{x} = a_0$ выборочная средняя генеральной совокупности равна заданному числу a_0 .

H_1 : $\bar{x} \neq a_0$ ($\bar{x} < a_0$, $\bar{x} > a_0$) выборочная средняя генеральной совокупности не равна (меньше, больше) заданному числу a_0 .

Наблюдаемое значение t-критерия рассчитывается по формуле:

– если дисперсия генеральной совокупности неизвестна

$$t_{\text{набл}} = \frac{\bar{x} - a_0}{S} \sqrt{n} \quad (1)$$

– если дисперсия генеральной совокупности известна

$$t_{\text{набл}} = \frac{\bar{x} - a_0}{\sigma_{\text{ген}}} \sqrt{n} \quad (2)$$

где \bar{x} - выборочная средняя;

a_0 — числовое значение генеральной средней;

S — исправленное среднее квадратическое отклонение;

$\sigma_{\text{ген}}^2$ - известная дисперсия генеральной совокупности;

n — объем выборки.

Критическое значение $t_{\text{кр}}$ следует находить с помощью таблиц распределения Стьюдента по уровню значимости α и числу степеней свободы $k = n - 1$.

б) для обнаружения различия между средними значениями \bar{x} , \bar{y} двух выборок.

Возможны гипотезы:

H_0 : $\bar{x} = \bar{y}$ средние значения двух выборок равны,

H_1 : $\bar{x} \neq \bar{y}$ средние значения двух выборок не равны.

Наблюдаемое значение t-критерия рассчитывается по формуле:

– для независимых выборок

$$t_{\text{набл}} = \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \quad (3)$$

– для зависимых выборок

$$t_{\text{эм}} = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n - 1}}} \quad (4)$$

где $S_x^2 = \frac{1}{n_1 - 1} \sum (x - \bar{x})^2$ - выборочная дисперсия 1 выборки;

$S_y^2 = \frac{1}{n_2 - 1} \sum (y - \bar{y})^2$ - выборочная дисперсия 2 выборки;

\bar{x} - среднее значение признака для 1 выборки;

\bar{y} - среднее значение признака для 2 выборки;

n_1 - объем 1 выборки;

- n_2 — объем 2 выборки;
- d — разность между результатами в каждой паре («после»минус «до»);
- n — число пар данных в зависимых выборках

Критическое значение $t_{кр}$ следует находить с помощью таблиц распределения Стьюдента по уровню значимости α и числу степеней свободы $k = n_1 + n_2 - 2$.

t -критерий для независимых выборок можно использовать для сравнения средних показателей экспериментальной группы с контрольной группой.

t -критерий для зависимых выборок очень полезен в тех ситуациях, когда две сравниваемые группы основываются на одной и той же совокупности наблюдений (субъектов), которые тестировались дважды (например, до и после эксперимента).

2. F — критерий Фишера-Снедекора используют

а) для сравнения разброса значений двух выборок, т.е. для проверки гипотезы о равенстве дисперсий.

Возможны гипотезы:

$H_0 : S_x^2 = S_y^2$ - разброс значений признака относительно среднего одинаковый в обеих выборках.

$H_1 : S_x^2 \neq S_y^2$ - разброс значений признака не совпадает.

Наблюдаемое значение F — критерия рассчитывается по формуле:

$$F_{набл} = \frac{S_x^2}{S_y^2}, \quad (5)$$

где S_x^2 — большая (по величине) выборочная дисперсия;

S_y^2 — меньшая (по величине) выборочная дисперсия.

Критическое значение $F_{крит}$ следует находить с помощью таблицы распределения Фишера-Снедекора по уровню значимости α и числу степеней свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$,

где k_1 — число степеней свободы большей (по величине) дисперсии;

k_2 — число степеней свободы меньшей (по величине) дисперсии;

n_1 — объем выборки большей (по величине) дисперсии;

n_2 — объем выборки меньшей (по величине) дисперсии.

Тема: Основы корреляционного анализа

Задача корреляционного анализа сводится к установлению направления и формы между варьирующими признаками, измерению тесноты, и, наконец, к проверке значимости коэффициентов корреляции.

Переменные x и y могут быть измерены в разных шкалах. Это обстоятельство определяет выбор соответствующего коэффициента линейной корреляции.

Тип шкалы		Мера связи
Переменная x	Переменная y	
Интервальная или отношений	Интервальная или отношений	Коэффициент Пирсона r_{xy}
Ранговая, интервальная или отношений	Ранговая, интервальная или отношений	Коэффициент Спирмена ρ_{xy}
Ранговая	Ранговая	Коэффициент τ Кендалла
Дихотомическая	Дихотомическая	Коэффициент ассоциации ϕ
Дихотомическая	Ранговая	Рангово-бисериальный $R_{ГВ}$
Дихотомическая	Интервальная или отношений	Бисериальный $R_{бис}$

Все коэффициенты по абсолютной величине не могут превосходить 1.

а) Коэффициент корреляции Пирсона вычисляется по формуле:

$$r_{xy} = \frac{\sum_i (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 * \sum_i (y_i - \bar{y})^2}} = \frac{n * \sum_i x_i y_i - \sum_i x_i * \sum_i y_i}{\sqrt{(n * \sum_i x_i^2 - (\sum_i x_i)^2) * (n * \sum_i y_i^2 - (\sum_i y_i)^2)}} \quad (9)$$

где x_i – значения, переменных принимаемые переменной x ,

y_i – значения, переменных принимаемые переменной y ,

\bar{x} – средняя по x ,

\bar{y} – средняя по y .

Оценка значимости осуществляется при числе степеней свободы $k=n-2$.

б) Коэффициент корреляции рангов Спирмена вычисляется по формуле:

$$\rho = 1 - \frac{6 \cdot \sum_i (d_i)^2}{n \cdot (n^2 - 1)}, \quad (6)$$

где n – количество ранжируемых признаков

d_i – разность между рангами по двум переменным для каждого испытуемого.

При наличии одинаковых рангов в числитель добавляются поправки на одинаковые ранги:

$$D_1 = \frac{n^3 - n}{12}, \quad (7)$$

$$D_2 = \frac{k^3 - k}{12}, \quad (8)$$

где n – число одинаковых рангов в первом столбце,

k – число одинаковых рангов во втором столбце.

По каждой группе одинаковых рангов вводится своя поправка.

Критически значения определяются при уровне значимости равном числу значений признака, по таблице критических значений ρ Спирмена.

в) Коэффициент ассоциации φ вычисляется по формуле:

$$\varphi = \frac{p_{xy} - p_x \cdot p_y}{\sqrt{p_x \cdot (1 - p_x) \cdot p_y \cdot (1 - p_y)}}, \quad (9)$$

где p_x – частота или доля признака, имеющего 1 по x , $(1-p_x)$ – частота или доля признака, имеющего 0 по x , p_y – частота или доля признака, имеющего 1 по y , $(1-p_y)$ – частота или доля признака, имеющего 0 по y , p_{xy} – частота или доля признака, имеющих 1 и по x и по y .

г) Коэффициент корреляции τ Кендалла вычисляется по формуле:

$$\tau = 1 - \frac{4 \cdot Q}{N \cdot (N - 1)}, \quad (10)$$

где Q – число инверсий (подсчет инверсий осуществляется суммированием числа рангов второго признака меньше каждого из рангов второго признака, при условии, что ранги первого признака упорядочены по возрастанию);

N – число ранжируемых признаков.

д) Бисериальный коэффициент корреляции вычисляется по формуле:

$$R_{\text{бисэм}} = \frac{\bar{x}_1 - \bar{x}_0}{\sigma_y} * \sqrt{\frac{n_1 * n_0}{N * (N - 1)}}, \quad (11)$$

где \bar{x}_1 - среднее по тем элементам переменной y , которым соответствует признак 1 в переменной x ,

\bar{x}_0 - среднее по тем элементам переменной y , которым соответствует признак 0 в переменной x ,

n_1 - число единиц в переменной x ,

n_0 - число нулей в переменной x ,

$N=n_1+n_0$,

σ_y - среднее квадратическое отклонение переменной y .

е) Рангово-бисернальный коэффициент корреляции вычисляется по формуле:

$$R_{rb\text{эм}} = \frac{(\bar{x}_1 - \bar{x}_0)^2}{N}, \quad (12)$$

где \bar{x}_1 - средний ранг по тем элементам переменной y , которым соответствует признак 1 в переменной x ;

\bar{x}_0 - средний ранг по тем элементам переменной y , которым соответствует признак 0 в переменной x ;

N - количество элементов в переменной x .

Тема: Основы регрессионного анализа

Взаимосвязь между переменными величинами может быть описана разными способами. Например, эту связь можно описать с помощью различных коэффициентов корреляции (линейных, частных, корреляционного отношения и т. п.). В то же время эту связь можно выразить по-другому: как зависимость между аргументом (величиной X) и функцией Y . В этом случае задача будет состоять в нахождении зависимости вида $X=F(Y)$ или, напротив, в нахождении зависимости вида $X=F(Y)$. При этом изменение функции в зависимости от изменений одного или нескольких аргументов называется регрессией.

Графическое выражение регрессионного уравнения называют линией регрессии. Линия регрессии выражает наилучшее предсказание зависимой переменной (Y) по независимым переменным (X). Эти независимые переменные, а их может быть много, носят название *предикторов*.

Регрессию выражают с помощью двух уравнений регрессии, которые в самом простом случае выглядят, как уравнения прямой, а именно так:

$$Y = a_0 + a_1 \cdot X \quad (13)$$

$$X = b_0 + b_1 \cdot Y \quad (14)$$

В уравнении 1 Y – зависимая переменная, а X – независимая переменная, a_0 свободный член, а a_1 – коэффициент регрессии, или угловой коэффициент, определяющий наклон линии регрессии по отношению к осям координат.

В уравнении 2 X – зависимая переменная, а Y – независимая переменная, b_0 свободный член, а b_1 – коэффициент регрессии, или угловой коэффициент, определяющий наклон линии регрессии по отношению к осям координат.

Линии регрессии пересекаются в точке $O(\bar{x}, \bar{y})$, с координатами, соответствующими средним арифметическим значениям корреляционно связанных между собой переменных X и Y . Линия AB , проходящая через точку O , соответствует линейной функциональной зависимости между переменными величинами X и Y равен $r_{xy} = 1$. При этом наблюдается такая закономерность: чем сильнее связь между X и Y , тем ближе обе линии регрессии к прямой AB . При отсутствии связи между X и Y линии регрессии оказываются под прямым углом по отношению друг к другу и в этом случае $r_{xy} = 0$.

Количественное представление связи (зависимости) между X и Y (между Y и X) называется регрессионным анализом. Главная задача регрессионного анализа заключается, собственно говоря, в нахождении коэффициентов a_0 , b_0 , a_1 и b_1 и определении уровня значимо-

сти полученных аналитических выражений (13) и (14), связывающих между собой переменные X и Y .

При этом коэффициенты регрессии a_1 и b_1 показывают, насколько в среднем величина одной переменной изменяется при изменении на единицу меры другой. Коэффициент регрессии a_1 в уравнении (1) можно подсчитать по формуле:

$$a_1 = r_{xy} \cdot \frac{S_y}{S_x} \quad (15)$$

а коэффициент b_1 в уравнении (2) по формуле (4)

$$b_1 = r_{xy} \cdot \frac{S_x}{S_y} \quad (16)$$

где r_{yx} – коэффициент корреляции между переменными X и Y ;

S_x – среднее квадратическое отклонение, подсчитанное для переменной X ;

S_y – среднее квадратическое отклонение, подсчитанное для переменной Y .

Тема Непараметрические критерии

а) Q-критерий Розенбаума

Критерий используется для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного.

Возможны гипотезы:

H_0 : Уровень признака в выборке 1 не превышает уровня признака в выборке 2.

H_1 : Уровень признака в выборке 1 превышает уровень признака в выборке 2.

Ограничения критерия Q

1) В каждой из сопоставляемых выборок должно быть не менее 11 наблюдений.

При этом объемы выборок должны примерно совпадать. Если в обеих выборках меньше 50 наблюдений, то абсолютная величина разности между объемами выборок n_1 и n_2 , соответственно, не должна быть больше 10 наблюдений. Если в каждой из выборок больше 51 наблюдения, но меньше 100, то абсолютная величина разности между объемами выборок n_1 и n_2 , соответственно, не должна быть больше 20 наблюдений. Если в каждой из выборок больше 100 наблюдений, то допускается, чтобы одна из выборок была больше другой не более чем в 1,5-2 раза.

2) Диапазоны разброса значений в двух выборках должны не совпадать между собой, в противном случае применение критерия бессмысленно.

3) Измерение может быть проведено по шкале порядка, интервалов или отношений.

4) Выборки должны быть независимыми.

Эмпирическое значение критерия подсчитывается по формуле:

$$Q_{эм} = S_1 + S_2, \quad (17)$$

где S_1 – количество наблюдений в выборке 1, которые выше максимального значения в выборке 2;

S_2 – количество наблюдений в выборке 2, которые ниже минимального значения выборки 1.

Значения в выборках должны быть упорядочены по возрастанию признака.

Критические значения Q-критерия определяются по таблице для данных n_1 и n_2 и для выбранного уровня значимости. Если $Q_{эм}$ не меньше $Q_{кр}$, то H_0 отвергается.

б) U- критерий Манна-Уитни

Критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного.

Возможны гипотезы:

H_0 : Уровень признака в группе 2 не ниже уровня признака в группе 1.

H_1 : Уровень признака в группе 2 ниже уровня признака в группе 1.

Ограничения критерия U

- 1) В каждой выборке должно быть не менее 3 наблюдений. Допускается, чтобы в одной выборке было 2 наблюдения, но тогда во второй их должно быть не менее 5.
- 2) В каждой выборке должно быть не более 60 наблюдений.
- 3) Измерение может быть проведено по шкале интервалов или отношений.
- 4) Выборки должны быть несвязными.

Наблюдения обеих выборок необходимо объединить и проранжировать по степени нарастания признака.

Эмпирическое значение критерия рассчитывается по формуле:

$$U_{\text{эмп}} = (n_1 \cdot n_2) + \frac{n_x \cdot (n_x + 1)}{2} - T_x, \quad (18)$$

где n_1, n_2 - количество испытуемых в выборках 1 и 2 соответственно,

T_x - большая из ранговых сумм,

n_x - количество испытуемых в группе с большей суммой рангов.

Критическое значение критерия определяется по таблице для данных n_1 и n_2 и выбранного уровня значимости. Если $U_{\text{эмп}}$ больше $U_{\text{кр}}$, то принимается H_0 .

в) Н - критерий Крускала – Уоллиса

Критерий предназначен для оценки различий между тремя и более выборками по уровню какого-либо признака. Он позволяет установить, что уровень признака изменяется при переходе от группы к группе, но не указывает направление этих изменений.

Возможны гипотезы:

H_0 : Между выборками 1, 2, 3 и т. д. существуют лишь случайные различия по уровню исследуемого признака.

H_1 : Между выборками 1, 2, 3 и т. д. существуют неслучайные различия по уровню исследуемого признака.

Ограничения критерия Н

- 1) Измерение может быть проведено по шкале интервалов или отношений.
- 2) Выборки должны быть независимыми.
- 3) При сопоставлении 3-х выборок допускается, чтобы в одной из них $n=3$, а в двух других $n=2$. Но при таких численных составах установить различия можно лишь на низшем уровне значимости.

4) При большем 5 количестве выборок и испытуемых в каждой выборке необходимо пользоваться таблицей критических значений критерия χ^2 . Число степеней свободы при этом определяется как $v=c-1$, где c – количество сопоставляемых выборок.

Наблюдения всех выборок необходимо объединить и проранжировать по степени нарастания признака.

Эмпирическое значение критерия Н подсчитывается по формуле:

$$H_{\text{эмп}} = \frac{12}{N(N+1)} \cdot \sum_{i=1}^c \frac{T_i^2}{n_i} - 3 \cdot (N+1), \quad (19)$$

где N – общее количество испытуемых в объединенной выборке,

n_i – количество испытуемых в каждой выборке,

T_i^2 – квадраты сумм рангов по каждой i -ой выборке.

Если эмпирическое значение критерия меньше критического значения, то H_0 принимается.

г) S – критерий тенденций Джонкира

Критерий S предназначен для выявления тенденций изменения признака при переходе от выборки к выборке при сопоставлении трех и более выборок.

Гипотезы:

H_0 : Тенденция возрастания значений признака при переходе от выборки к выборке является случайной.

H_1 : Тенденция возрастания значений признака при переходе от выборки к выборке не является случайной.

Ограничения критерия

1) Измерение может быть проведено по шкале интервалов или отношений.

2) Выборки должны быть независимыми.

3) Количество наблюдений в каждой выборке должно быть одинаковым.

4) Нижняя граница применимости критерия: не менее трех выборок и не менее двух элементов в каждом наблюдении. Верхняя граница определяется таблицей приложения: не более 6 выборок и не более 10 наблюдений в каждой выборке.

Выборки необходимо располагать по возрастанию суммы значений признака слева на право.

Эмпирическое значение критерия рассчитывается по формуле:

$$S_{эмп} = 2A - B, \quad (20)$$

где A – общая сумма инверсий,

$$B = \frac{c \cdot (c - 1)}{2} \cdot n^2 - \text{максимально возможное значение величины } A.$$

Под числом инверсий понимается число значений признака, больших каждого конкретного значения рассматриваемой выборки и расположенных правее от нее.

Критические значения критерия определяются по таблице, в соответствии с выбранным уровнем значимости, количеством выборок (c) и числом наблюдений (n) в каждой выборке.

Если эмпирическое значение критерия меньше критического значения, то принимается гипотеза H_0 .

д) G-критерий знаков

Критерий знаков предназначен для установления общего направления сдвига исследуемого признака. Он позволяет установить, в какую сторону в выборке в целом изменяются значения признака при переходе от первого измерения ко второму: изменяются ли показатели в сторону улучшения, повышения или усиления или, наоборот, в сторону ухудшения, понижения или ослабления.

Возможны гипотезы:

H_0 : Преобладание типичного направления сдвига является случайным.

H_1 : Преобладание типичного направления сдвига не является случайным.

Ограничения критерия G

1) Измерение может быть проведено по шкале порядка, интервалов или отношений.

2) Выборка должна быть однородной и связной.

3) Объем выборки должен быть равным от 5 до 300.

4) При равенстве типичных и нетипичных сдвигов критерий знаков неприменим.

Эмпирическое значение критерия $G_{эмп}$ принимают равным числу нетипичных сдвигов, т. е. не преобладающих сдвигов в сторону увеличения или уменьшения показателя.

Критическое значение критерия $G_{кр}$ определяют по таблице в соответствии с выбранным уровнем значимости и объемом выборки без учета нулевых сдвигов. Если $G_{эмп}$ не превосходит $G_{кр}$, то гипотеза H_0 отвергается.

ж) Парный критерий T – Вилкоксона

Критерий применяется для сопоставления показателей, измеренных в двух разных условиях на одной и той же выборке испытуемых. Он позволяет установить не только направленность изменений, но и их выраженность. С его помощью мы определяем, является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Возможны гипотезы:

H_0 : Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

H_1 : Интенсивность сдвигов в типичном направлении превышает интенсивности сдвигов в нетипичном направлении.

Ограничения критерия T

- 1) Измерение может быть проведено по любой шкале, кроме номинальной.
- 2) Выборка должна быть связной.
- 3) Объем выборки должен быть равным от 5 до 50.

Эмпирическое значение критерия подсчитывают по формуле:

$$T_{эмп} = \sum R_r, \quad (21)$$

где R_r – ранговые значения сдвигов с более редким знаком.

Критическое значение критерия $T_{кр}$ определяется для данного объема выборки и выбранного уровня значимости по таблице. Если $T_{эмп}$ не превосходит $T_{кр}$, то гипотеза H_0 отвергается.

з) Критерий χ^2 Фридмана

Критерий применяется для сопоставления показателей, измеренных в трех или более условиях на одной и той же выборке испытуемых. Он позволяет установить, что величины показателей от условия к условию изменяются, но при этом не указывает на направление изменений.

Возможны гипотезы:

H_0 : Между показателями, полученными (измеренными) в разных условиях, существуют лишь случайные различия.

H_1 : Между показателями, полученными (измеренными) в разных условиях, существуют неслучайные различия.

Ограничения критерия χ^2

- 1) Измерение может быть проведено по шкале интервалов или отношений.
- 2) Выборка должна быть связной.
- 3) В выборке должно быть не менее двух испытуемых, каждый из которых имеет не менее трех показателей. Количество измерений не может превышать 100.

Эмпирическое значение критерия вычисляется по формуле:

$$\chi^2_{эмп} = \left[\frac{12}{n \cdot c \cdot (c + 1)} \cdot \sum_{i=1}^c T_i^2 \right] - 3 \cdot n \cdot (c + 1), \quad (22)$$

где c – количество условий,

n – количество испытуемых,

T_i – суммы рангов по каждому из условий.

Критическое значение критерия $\chi^2_{кр}$ определяем при выбранном уровне значимости и данном объеме выборки по правилам:

- 1) При $c=3$ и $n \leq 9$, критические значения определяются по таблице 8-А приложения.
- 2) При $c=4$ и $n \leq 4$, критические значения определяются по таблице 8-Б приложения.
- 3) При большем числе измерений и испытуемых критические определяются по таблице для критерия χ^2 . В этом случае число степеней свободы определяется по формуле $\nu = c - 1$.

Если $\chi^2_{эмп}$ не меньше $\chi^2_{кр}$ то гипотеза H_0 отклоняется.

и) L – критерий тенденций Пейджа.

Критерий L Пейджа применяется для сопоставления показателей, измеренных в трех и более условиях на одной и той же выборке испытуемых. Критерий позволяет выявить тенденции в изменении величин признака при переходе от условия к условию, а также указывает на направление этих изменений.

Возможны гипотезы:

H_0 : Увеличение индивидуальных показателей при переходе от первого условия ко второму, а затем к третьему и далее, случайно.

H_1 : Увеличение индивидуальных показателей при переходе от первого условия ко второму, а затем к третьему и далее, неслучайно.

Ограничения критерия L

- 1) Измерение может быть проведено по ранговой шкале, шкале интервалов или отношений.
- 2) Выборка должна быть связной.
- 3) В выборке должно быть не менее двух и не больше 12 испытуемых, каждый из которых имеет не менее трех показателей. Максимальное число условий – 6.

Эмпирическое значение критерия определяется по формуле:

$$L_{\text{эмп}} = \sum_{i=1}^c (T_i \cdot i), \quad (23)$$

где c – количество условий,

T_i – суммы рангов по каждому из условий,

i – порядковый номер, приписанный каждому условию, после упорядочения по возрастанию сумм рангов.

Критическое значение критерия $L_{кр}$ определяем при выбранном уровне значимости, данном объеме выборки и данном количестве условий по таблице. Если $L_{эмп}$ не меньше $L_{кр}$, то гипотеза H_0 отклоняется.

Тема Факторный анализ

Факторный анализ – статистический метод, который используется при обработке больших массивов экспериментальных данных. Задачами факторного анализа являются: сокращение числа переменных (редукция данных) и определение структуры взаимосвязей между переменными, т.е. классификация переменных, поэтому факторный анализ используется как метод сокращения данных или как метод структурной классификации.

Важное отличие факторного анализа от всех описанных выше методов заключается в том, что его нельзя применять для обработки первичных, или, как говорят, «сырых», экспериментальных данных, т.е. полученных непосредственно при обследовании испытуемых. Материалом для факторного анализа служат корреляционные связи, а точнее – коэффициенты корреляции Пирсона, которые вычисляются между переменными (т.е. психологическими признаками), включенными в обследование. Иными словами, факторному анализу подвергают корреляционные матрицы, или, как их иначе называют, матрицы интеркорреляций. Наименования столбцов и строк в этих матрицах одинаковы, так как они представляют собой перечень переменных, включенных в анализ. По этой причине матрицы интеркорреляций всегда квадратные, т.е. число строк в них равно числу столбцов, симметричные, т.е. на симметричных местах относительно главной диагонали стоят одни и те же коэффициенты корреляции.

Элементы факторной матрицы называются «*факторными нагрузками*, или *весами*»; и они представляют собой коэффициенты корреляции данного фактора со всеми показателями, использованными в исследовании. Факторная матрица очень важна, поскольку она показывает, как изучаемые показатели связаны с каждым выделенным фактором. При этом факторный вес демонстрирует меру, или тесноту, этой связи.

Тема Дисперсионный анализ.

Для выявления тенденций изменения признака в трех и более выборках при переходе от условия к условию применяется однофакторный дисперсионный анализ.

Возможны гипотезы:

для независимых выборок: (влияние разных условий на разных испытуемых)

H_0 : разные условия не влияют на изменение значений признака;

H_1 : условия влияют на изменение значений признака.

для зависимых выборок (одни и те же испытуемые, но в разных условиях) возможно два варианта гипотез:

а) H_0 : условия не влияют на изменение признака;

H_1 : условия влияют на изменение значений признака.

б) H_0 : индивидуальные различия испытуемых не влияют на изменение значений признака;

H_1 : индивидуальные различия между испытуемыми влияют на изменение значений признака.

Наблюдаемое значение критерия рассчитывается по формулам:

$$\text{для независимых выборок } F_{\text{набл}} = \frac{S_{\text{факт}}^2}{S_{\text{сл}}^2}, \quad (24)$$

$$\text{для зависимых выборок а) } F_{\text{набл}} = \frac{S_{\text{факт}}^2}{S_{\text{сл}}^2}, \quad (25)$$

$$\text{б) } F_{\text{набл}} = \frac{S_{\text{испыт}}^2}{S_{\text{сл}}^2}, \quad (26)$$

где $S_{\text{факт}}^2 = \frac{n \cdot \sum (\bar{x}_{\text{гр}} - \bar{x})^2}{K_{\text{факт}}}$ - характеризует изменение признака, обусловленное действием фактора (условия).

$S_{\text{общ}}^2 = \frac{\sum (x_{ij} - \bar{x})^2}{K_{\text{общ}}}$ - характеризует общее изменение признака по всем выборкам и наблюдениям.

$S_{\text{испыт}}^2 = \frac{m \cdot \sum (\bar{x}_{\text{испыт}} - \bar{x})^2}{K_{\text{испыт}}}$ - характеризует изменение признака, обусловленное индивидуальными особенностями испытуемых.

$$S_{\text{сл}}^2 = \frac{\sum (x_{ij} - \bar{x})^2 - n \cdot \sum (\bar{x}_{\text{гр}} - \bar{x})^2}{K_{\text{общ}} - K_{\text{факт}}} \text{ - для независимых выборок}$$

$$S_{\text{сл}}^2 = \frac{\sum (x_{ij} - \bar{x})^2 - n \cdot \sum (\bar{x}_{\text{гр}} - \bar{x})^2 - m \cdot \sum (\bar{x}_{\text{испыт}} - \bar{x})^2}{K_{\text{общ}} - K_{\text{факт}} - K_{\text{испыт}}} \text{ - для зависимых выборок.}$$

n – количество наблюдений,

m – количество выборок (групп),

$\bar{x}_{\text{гр}}$ - среднее значение признака в каждой группе,

\bar{x} - среднее значение признака по всей совокупности,

$\bar{x}_{\text{испыт}}$ - среднее значение признака каждого испытуемого,

k – число степеней свободы:

$$K_{\text{общ}} = n \cdot m - 1$$

$$K_{\text{факт}} = m - 1$$

$$K_{\text{испыт}} = n - 1$$

Критические значения $F_{\text{крит}}$ следует находить с помощью таблицы распределения Фишера-Снедекора по уровню значимости α и числу степеней свободы $k_1 = k_{\text{сл}}$, и в зависимости от проверяемых гипотез $k_1 = k_{\text{факт}}$ или $k_1 = k_{\text{испыт}}$

2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ

Практические занятия сопровождают лекционный курс дисциплины. Теоретические знания, представления, образы должны быть прожиты. Афоризм одного из известных физиков М. Лауэ: «знание есть то, что остается, когда все выученное уже забыто», характеризует важную роль практики.

Практические занятия должны проводиться в логичном единстве с теоретическим курсом, подкрепляя и уточняя понятийный аппарат.

Каждый практическое занятие начинается с теоретического опроса необходимого материала и проверки домашнего задания. Далее на конкретных примерах рассматриваются пути и способы применения тех математических методов, которые не требуют использования электронных вычислительных машин. При этом необходимо активизировать самостоятельную работу студентов. Задания и методические указания к ним выдаются студентам, каждый из которых выбирает оптимальный для себя темп работы. Преподавателю отводится роль консультанта и помощника. Задания, вызвавшие трудности у большинства студентов, разбираются на доске.

В конце занятия выдается домашнее задание, состоящее из теоретических вопросов, решение которых необходимо для следующего занятия и практических заданий по пройденному материалу.

При выполнении домашнего задания решать задачи удобнее поэтапно, в той последовательности, в какой эти задания сформулированы. В этом случае при возникновении трудностей будет легче обратиться к анализу тех тем, которые изложены в лекции и задач, разобранных на практическом занятии.

После выполнения практической части задания следует найти ответы на теоретические вопросы, заданные преподавателем и таким образом подготовиться к осознанному восприятию следующего материала.

Активная, регулярная самостоятельная работа над домашним заданием – путь к успешному усвоению дисциплины.

Тема: Структура курса. Анализ данных на компьютере

1. В таблице указано число студентов разного пола, обучающихся на факультетах института

Факультет	Мужчины	Женщины	Всего
Антропология	1	2	3
Искусствоведение	42	66	108
Биология	22	20	42
Деловое администрирование	100	22	122
Химия	20	8	28
Педагогика	27	42	69
Инженерное дело	75	3	78
Геология	8	0	8
История	25	25	50
Экономика	5	50	55
Математика	10	6	16
Музыка	3	10	13
Физика	38	18	56
Психология	24	28	52
<i>Всего</i>	400	300	700

Сколько процентов составляют мужчины среди студентов каждого факультета?

Постройте гистограмму распределения, используя столбцы белого цвета для мужчин и затемненные для женщин. Какая гистограмма лучше показывает половой состав института: выраженная в абсолютных числах или в процентах?

Каков модальный факультет для мужчин? Для женщин? Для всех студентов?

2. Постройте таблицу сопряженности, отражающую следующую информацию: В выпускных классах школы учатся 256 учеников, из них 125 – юноши. Высшее образование планируют получить 57 выпускников, 15 еще не определились, а остальные уверены, что не пойдут учиться дальше. Из тех, кто планирует получить высшее образование – 35 человек – мужчины, а 10 из 15 неопределившихся – женщины.

3. Для каждого вида измерений, перечисленных ниже, укажите соответствующую ему шкалу:

- Возраст человека в годах
- Число студентов в группе
- IQ
- Температура по Цельсию
- Температура по Кельвину
- Деление людей по полу.
- Оценка на экзамене

Тема: Параметрические критерии

Типовые задания

Построить частотное распределение, подсчитать стандартное отклонение для следующих рядов значений: (результаты тестирования интеллекта по тесту Векслера 2-х групп испытуемых, численностью 50 человек) :

Результаты тестирования 1-ой группы испытуемых:

85,93,93,99,101,105,109,110,111,115,115,116,116,117,117,117,118,119,121,121,122,124,124,124,124,125,125,125,127,127,127,127,127,128,130,131,132,132,133,134,134,135,138,138,140,143,144,146,150,158.

Результаты тестирования 2-й выборки:

70,75,76,78,78,81,82,82,83,84,84,84,85,86,86,86,89,89,90,91,91,91,91,92,92,93,93,95,95,95,96,96,98,98,100,101,103,103,103,105,108,110, 115,118,119,123,124,125,127,129.

Подсчитать t- критерий Стьюдента для двух выборок.

Проверить результаты на статистическую значимость.

Подсчитайте критерий Фишера для следующих результатов измерения тревожности 2-х различных групп

x	y
90	41
29	49
39	56
79	64
88	72
88	58
72	59
76	62
91	68
78	48

Тема: Основы корреляционного анализа

1. Был проведен опрос о количестве времени (в часах), которое тратит каждый студент на изучение учебной литературы и на просмотр художественных фильмов. Можно ли сделать вывод о существовании связи между 2-мя этими переменными?

№	Время на изучение учебной литературы	Время на просмотр фильмов
1	7	0

№	Время на изучение учебной литературы	Время на просмотр фильмов
2	8	4
3	5	10
4	1	7
5	15	0
6	5	1
7	2	10
8	4	6
9	1	9
10	1	8

2. Зависимость между величинами выражается в виде экспериментально полученной таблицы. Определить коэффициент корреляции Пирсона. Сделать выводы.

X	0,5	1	1,5	2	2,5	3
У	0,01	0,11	0,35	0,6	1,58	2,31

3. Подсчитайте корреляцию между физической привлекательностью студенток и их академическими достижениями.

№	Ранг по красоте	Ранг по академическим достижениям
1	3	7.5
2	2	2
3	6	9
4	8	4
5	4	4
6	10	10
7	7	6
8	1	1
9	9	7.5
10	5	4

4. Существует ли связь между лидерством и дружелюбностью? Исследователи отмечали наличие или отсутствие у человека лидерской позиции в группе, одновременно относя его либо к дружелюбным, либо к недружелюбным людям. Результаты показаны в таблице

Лидерская позиция	Дружелюбные	Недружелюбные	Всего
Лидер	2	4	6
Не явл. лидером	10	4	14
<i>Всего</i>	12	8	20

Что можно сказать о связи между этими переменными? Рассчитайте коэффициент ассоциации

Тема: Основы регрессионного анализа

Типовые задания

1. По заданной выборке найти линейную функцию методом наименьших квадратов. Построить чертеж. Обосновать правильность выбора степени многочлена, если относительная погрешность данных равна одному проценту среднего значения зависимой переменной.

x_i	68	62	65	70	63	63	67	68	55	66	58
y_i	-264	-240	-253	-275	-248	-243	-264	-267	-213	-218	-223

2. Врач исследователь выясняет зависимость площади пораженной части легкого у людей, заболевших эмфиземой легких, от числа лет курения. Статистические данные, со-

бренные им в некоторой области, имеют следующий вид:

Число лет курения	25	36	22	15	48	39	42	31	28	33
Площадь пораженной части легкого, %	55	60	50	30	75	70	70	55	30	35

Постройте график исходных данных и определите по нему характер зависимости. Постройте уравнение регрессии и дайте интерпретацию полученных результатов. Если человек курил 30 лет, то сделайте прогноз о степени поражения легких у случайно выбранного пациента.

Тема Непараметрические критерии

Типовые задания

1. Подсчитайте критерий Крускала -Уоллиса для следующих данных

23	45	34	21
20	12	24	22
34	34	25	26
35	11	40	27

2. Подсчитайте критерий Манна-Уитни для следующих данных:

x	y
6	
	8
25	
25	
30	
	31
	32
38	
39	
	41
	41
43	
44	
	45
	46
	50
	55

3. Две группы испытуемых решали техническую задачу. Показателем успешности было время решения задачи. Испытуемые 1 группы получали за выполнение денежное вознаграждение, испытуемые 2 группы – нет. Психолога интересует вопрос: влияет ли денежное вознаграждение на успешность решения задачи?

Результаты получены следующие (в секундах) :

1 группа: 39. 38..44. 6. 25.25.30.43.

2 группа: 46.8.50.45.32.41.41.31.55.

Подсчитайте критерий Манна – Уитни.

Тема Факторный анализ

Типовые задания

Определите удовлетворенность, какой стороной жизни преобладает у разных респондентов. Пусть в эксперименте участвовало 10 человек, которые отвечали на вопросы, представленные в таблице

Возраст	Образование	Стаж непрерывной работы	Работа по специальности	Зарплата, тыс.руб	Кол-во человек в семье	Жилплощадь	Наличие хобби	Посещение учреждения вне работы

Образование: 2 – высшее, 1 – специальное, 0 – среднее.

Работа по специальности: 1 – да, 0 – нет.

Хобби: 1 – да, 0 – нет.

Посещение учреждения вне работы: 3 – часто, 2 – иногда, 1 – нет.

Проведите стандартизацию данных и определите оптимальное количество факторов, рассчитайте матрицы факторных нагрузок и факторных весов, сделайте вывод.

Тема Дисперсионный анализ.

Типовые задания

1. Провести дисперсионный анализ данных таблицы

Град. А	А1-короткие слова		А2- длинные слова	
	В1	В2	В1	В2
	9	4	5	7
	8	3	3	5
	6	3	3	6
	7	5	4	7
А-длина слов				
В-скорость предъявления слов				

2. Используйте двухфакторный анализ для следующей задачи:

4-м группам испытуемым предъявлялись списки из 10 слов:

1 группе – короткие слова с большой скоростью.

2-й - короткие с медленной скоростью.

3 –й - длинные слова с большой скоростью

стью

4-й - длинные слова с медленной скоростью.

В каждой группе было по 4 испытуемых, всего 16.

Гипотеза: При большой скорости лучше будут запоминаться короткие слова, при медленной скорости – длинные слова.

Результаты представлены в таблице:

фактор А	А1 – короткие слова		А2 – длинные слова	
	В1	В2	В1	В2
градации фактора В	9	4	5	7
	8	3	3	5
	6	3	3	6
	7	5	4	7
Сумма по ячейкам	30	15	15	25

3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Самостоятельная работа студентов предназначена для углубления сформированных знаний, умений, навыков. Самостоятельная работа развивает мышление, позволяет выявить причинно-следственные связи в изученном материале, решить теоретические и практические задачи. Самостоятельная работа студентов проводится с целью: систематизации и закрепления полученных теоретических знаний и практических умений студентов; углубления и расширения теоретических знаний; формирования умений использовать справочную документацию и специальную литературу; развития познавательных способностей и активности студентов: творческой инициативы, самостоятельности, ответственности и организованности; формированию самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации; развития исследовательских умений. Роль самостоятельной работы возрастает, т.к. перед учебным заведением стоит задача в т. ч. и по формированию у студента потребности к самообразованию и самостоятельной познавательной деятельности

Студентами практикуется два вида самостоятельной работы:

- аудиторная;
- внеаудиторная.

Аудиторная самостоятельная работа по дисциплине выполняется на учебных занятиях под непосредственным руководством преподавателя и по его заданию. В этом случае студенты обеспечиваются преподавателем необходимой учебной литературой, дидактическим материалом, в т. ч. методическими пособиями и методическими разработками.

Внеаудиторная самостоятельная работа выполняется студентом по заданию преподавателя, но без его непосредственного участия. Видами заданий для внеаудиторной самостоятельной работы могут быть:

- для овладения знаниями: чтение текста (учебника, методической литературы); составления плана текста; графическое изображение структуры текста, графическое изображение последовательности выполнения графической работы, выполнение графических работ; конспектирование текста; выписки из текста; работа со словарями и справочниками; ознакомление с нормативными документами; учебно-исследовательская работа; использование компьютерной техники, интернета и др.;

- для закрепления систематизации знаний: работа с конспектом лекции (обработки текста); повторная работа над учебным материалом (учебника, первоисточника, дополнительной литературы); составление плана выполнения работы в соответствии с планом, предложенным преподавателем; изучение ГОСТов; ответы на контрольные вопросы; тестирование, выполнение упражнений и графических работ;

- для формирования умений: решение задач и упражнений по образцу; решение вариативных задач и упражнений; выполнение чертежей, схем.

Основное содержание самостоятельной работы составляет выполнение индивидуальных заданий, подготовка к тестированию и зачету.

Прежде чем приступать к выполнению индивидуального задания, необходимо ознакомиться с содержанием теоретических вопросов по представленному списку литературы и по лекциям.

Работа пишется на стандартных листах писчей бумаги. Все листы заполняются только с одной стороны. Оформление индивидуального задания осуществляется в соответствии со стандартом. Каждая работа начинается с титульного листа, который служит обложкой работы. Сверху на нем указывается принадлежность студента к учебному заведению, факультету, специализации или кафедре. В середине листа указывается название изучаемой темы или раздела и название учебного задания, номер варианта. Ниже и справа указывается фамилия и инициалы студента, номер академической группы, фамилия и инициалы преподавателя. Внизу титульного листа отмечают год выполнения работы.

Эта страница служит также для отметок преподавателя о выполнении учебного задания и замечаний по поводу подготовленного студентом отчета.

При оформлении работы необходимо соблюдать нумерацию заданий. Задание переписывается полностью и ниже оформляется решение. Работа должна быть сдана на кафедру к назначенному преподавателем сроку.

Каждый учебный семестр заканчивается аттестационными испытаниями: зачетно - экзаменационной сессией.

Подготовка к экзаменационной сессии и сдача зачетов и экзаменов является ответственным периодом в работе студента. Seriously подготовиться к сессии и успешно сдать все экзамены - долг каждого студента. Рекомендуется так организовать свою учебу, чтобы перед первым днем начала сессии были сданы и защищены все лабораторные работы, сданы все зачеты, выполнены другие работы, предусмотренные графиком учебного процесса.

Основное в подготовке к сессии - это повторение всего материала, курса или предмета, по которому необходимо сдавать зачет или экзамен. Только тот успевает, кто хорошо усвоил учебный материал.

Если студент плохо работал в семестре, пропускал лекции, слушал их невнимательно, не конспектировал, не изучал рекомендованную литературу, то в процессе подготовки к сессии ему придется не повторять уже знакомое, а заново в короткий срок изучать весь материал. А это зачастую, оказывается, невозможно сделать из-за нехватки времени. Для такого студента подготовка к экзаменам будет трудным, а иногда и непосильным делом, а финиш - отчисление из учебного заведения.

В дни подготовки к экзаменам избегай чрезмерной перегрузки умственной работой, чередуй труд и отдых.

Примерные темы индивидуальных заданий:

1. Исследование факторов, определяющих специфику национального характера.
2. Исследование направленности и мотивов выбора профессии старшеклассников.
3. Исследование различий самооценки у девушек и юношей с помощью корреляционного анализа.
4. Исследование связи между вербальной агрессией и личностной тревожностью.
5. Влияния стиля руководства на климат в коллективе.
6. Мотивация «На оценку» и следования ее уровня на различных стадиях младшего школьного возраста.
7. Выявление факторов, определяющих положение человека в семье.
8. Исследование факторов влияющих на выбор страны для зарубежной поездки.

СОДЕРЖАНИЕ

<i>ВВЕДЕНИЕ</i>	3
1 КРАТКОЕ ИЗЛОЖЕНИЕ ЛЕКЦИОННОГО МАТЕРИАЛА	4
2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ	18
3 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ	23

Татьяна Александровна Юрьева,

доц. каф. общей математики и информатики АмГУ, канд. пед. наук