**Программа дисциплины**

**Методология исследования и базовая статистика**

для образовательной программы «Образование»

по направлению подготовки научно-педагогических кадров в аспирантуре

44.06.01 Образование и педагогические науки

Разработчик программы:

Осин Евгений Николаевич, канд. психол. н., [eosin@hse.ru](mailto:eosin@hse.ru)

Захаров Андрей Борисович, канд. пед. н., [abzakharov@hse.ru](mailto:abzakharov@hse.ru)

Утверждена Академическим советом Аспирантской школы по образованию

«11» октября 2018 г., протокол № 34

Академический директор

Е.А. Терентьев \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись)*

Москва – 2018

*Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.*

1. **Область применения и нормативные ссылки**

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения аспиранта и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и аспирантов, обучающихся по образовательной программе «Образование» и изучающих дисциплину «Методология исследования и базовая статистика».

Программа учебной дисциплины разработана в соответствии с:

* образовательным стандартом НИУ ВШЭ по направлению подготовки научно-педагогических кадров высшей квалификации 44.06.01 Образование и педагогические науки»;
* образовательной программой «Образование»;
* учебным планом образовательной программы «Образование».

1. **Цели и задачи освоения дисциплины**

Целями освоения дисциплины «Методология исследования и базовая статистика» являются:

* дать аспирантам четкое понимание принципов исследовательского дизайна и базовых статистических понятий;
* ознакомить с основными качественными и количественными методологическими парадигмами;
* развить умения формировать выборку, а также применять методы описательной статистики и статистического вывода, включая корреляционный и регрессионный анализ.

1. **Компетенции обучающегося, формируемые в результате освоения дисциплины**

В результате освоения дисциплины аспирант должен:

* знать различия между качественными, количественными и смешанными методами;
* уметь формулировать эмпирически проверяемые вопросы для количественных исследований и гипотезы для различных исследовательских задач;
* уметь выдвигать статистические гипотезы, соответствующие различным исследовательским вопросам;
* уметь предлагать статистические процедуры, подходящие для проверки различных исследовательских гипотез;
* понимать, чем отличаются различные вероятностные и невероятностные способы формирования выборки и каковы их сильные стороны;
* понимать сильные и слабые стороны исследовательских дизайнов, наиболее часто используемых в исследованиях в образовании;
* уметь проводить и интерпретировать статистические процедуры, изучаемые в рамках данного блока (например, ANOVA, ANCOVA, регрессия МНК).

В результате освоения дисциплины аспирант осваивает следующие компетенции:

|  |  |  |  |
| --- | --- | --- | --- |
| Компетенция | Код по ФГОС/ НИУ | Дескрипторы – основные признаки освоения (показатели достижения результата) | Формы и методы обучения, способствующие формированию и развитию компетенции |
| способность выбирать и применять методы исследования, адекватные предмету и задачам исследования; | УК-3 | Может критически оценить методы исследования в соответствии с его целями и задачами, в том числе и своего собственного исследования. | Лекции, семинары, самостоятельные работы |
| способность собирать, анализировать, обрабатывать и хранить данные в соответствии с общепринятыми научными и этическими стандартами; | УК-4 | Может собирать, обрабатывать, анализировать и хранить данные в соответствии с принятыми научными и этическими стандартами. | Лекции, семинары, самостоятельные работы |
| способность самостоятельно осуществлять научно-исследовательскую деятельность в области образования с использованием современных методов исследования и информационно-коммуникационных технологий; | ОПК-1 | Может самостоятельно проводить исследования, в том числе выбрать адекватный метод исследования. | Лекции, семинары, самостоятельные работы |
| способность самостоятельно определять цели и ставить исследовательские задачи для решения фундаментальных и прикладных проблем в области образования; | ПК-1 | Может самостоятельно поставить исследовательскую задачу, выдвинуть гипотезу и/или вопросы исследования. | Лекции, семинары, самостоятельные работы |

# Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к циклу дисциплин по выбору и изучается на 1-м году обучения. Обучение в рамках данного блока не требует предварительной подготовки сверх тех навыков и способностей, которые необходимы для поступления в аспирантскую школу по образованию (знание базовой статистики).

Основные положения дисциплины должны быть использованы в дальнейшем при изучении дисциплин:

* Разработка инструментов измерения,
* Углубленная психометрика,
* Углубленная статистика.

# Формы контроля знаний аспирантов:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Тип контроля | Форма контроля | 1 год | | | | Параметры \*\* |
| 1 | 2 | 3 | 4 |
| Текущий  (неделя) | Контрольная работа | \* | \* |  |  | письменная работа на занятии |
| \* | \* |  |  | письменная работа на занятии |
|  |  |  |  |  |  |
| Итоговый | Экзамен |  | \* |  |  | Тест |

## Критерии оценки знаний, навыков

Контрольные работы проводятся в форме тестов с закрытыми вопросами по типу множественного выбора или кейсовых заданий с открытыми вопросами. Работы проводятся в формате open-book (допускается использование материалов курса). Оценка по каждой контрольной работе рассчитывается путем деления количества баллов аспиранта на максимальный возможный балл с последующим умножением на 10 (без округления). Общая оценка по текущему контролю рассчитывается как среднее арифметическое по всем работам.

Примеры заданий:

Закрытый вопрос:

Что угрожает валидности данных кросс-секционных исследований, но не является проблемой в лонгитюдных исследованиях?

A. Когортные различия

B. Эффект повторного тестирования

C. Эффект созревания

D. Взаимодействие тестирования и воздействия

Открытый вопрос (5 баллов):

Для исследования предикторов учебной успешности в городе были случайно отобраны 30 групп школьников по 25 человек в группе. Чему равен эффективный объем выборки, если интраклассовая корреляция составляет 0.2?

# Содержание дисциплины

**Тема 1. Общее представление о научном исследовании**

Философия о науке и научном методе. Позитивизм и постпозитивизм. Номотетическая и идиографическая стратегия исследований. Задачи исследований в социальных науках. Этические стандарты исследований с людьми, этические комиссии. Защита прав участников исследований: информированное согласие, конфиденциальность.

**Тема 2. Количественная и качественная парадигмы**

Общее представление о количественной и качественной парадигмах: их методологические различия, исследовательские задачи и процедуры. Исследование в количественной парадигме: обзор основных дизайнов. Смешанные исследования.

**Тема 3. Проблемы исследования, вопросы и гипотезы**

Теоретические обзоры: задачи, процедуры и применение в науке. Обнаружение проблем исследования. Формулировка исследовательских вопросов и гипотез: хорошие и плохие гипотезы. Теория как мост от исследовательского вопроса к проверяемой гипотезе.

**Тема 4. Переменные и измерительные шкалы**

Виды переменных и их место в научном исследовании: а) наблюдаемые и латентные, б) независимые, зависимые и побочные, г) предикторы, критерии и ковариаты, медиаторы и модераторы. Типичные переменные в социальных исследования. Виды шкал: номинативные, порядковые, интервальные, отношений. Параметрические и непараметрические статистики.

**Тема 5. Характеристики количественных данных**

Результат измерения как случайная величина: непрерывные и дискретные величины. Генеральная совокупность и выборка. Теоретические и эмпирические распределения. Распределение одной, двух и множества переменных. Описательные статистики: меры центральной тенденции и меры разброса, асимметрия и эксцесс. Нормальное распределение: свойства и критерии диагностики. Стандартное нормальное распределение и его вероятностная интерпретация.

**Тема 6. Проверка статистических гипотез**

Статистическая гипотеза: определение. Подход с проверкой нулевой гипотезы: нулевая и альтернативная гипотеза, процедура проверки и критерии принятия решений. Односторонние и двусторонние гипотезы. Квантили, критические значения и наблюдаемые значения статистик, уровень значимости. Новая статистика: размеры эффекта, доверительные интервалы и их связь с проверкой гипотез. Ошибки первого и второго рода и связанные с ними риски.

**Тема 7. Размер эффекта и статистическая мощность**

Понятие размера эффекта. Основные семейства размеров эффекта: корреляция, различие средних, соотношение шансов, доля объясненной дисперсии. Способы перевода размера эффекта в другие единицы. Оценка величины размера эффекта. Понятие и задачи мета-анализа. Понятие статистической мощности: основные задачи анализа мощности. Анализ мощности в GPower.

**Тема 8. Выборки**

Свойства и задачи выборки. Критерии качества выборки: репрезентативность и точность выборочных оценок. Выборочная ошибка. Единицы выборки: наблюдения и кластеры. Невероятностные и вероятностные стратегии построения выборок. Невероятностные стратегии: удобная выборка, целевая выборка, метод снежного кома, самоотбор. Систематические стратегии.

**Тема 9. Вероятностные стратегии и комплексные выборки**

Целевые и доступные генеральные совокупности. Простая случайная выборка с возвращением и без. Стратифицированная случайная выборка. Кластерная случайная выборка. Двухэтапная процедура построения выборки. Комплексные выборки. Оценка влияния кластеров на выборочную ошибку и эффективный объём выборки, расчёт эффекта дизайна. Выборочные веса: теория и процедуры.

**Тема 10. Валидность научного исследования**

Понятие внутренней и внешней валидности. Основные угрозы внутренней и внешней валидности. Роль валидности статистического вывода. Понятие валидности измерительного инструмента: валидность инструмента и валидность исследования.

**Тема 11. Коэффициенты корреляции и корреляционные исследования**

Описательные исследования. Корреляция и каузальность. Параметрические и непараметрические коэффициенты корреляции: допущения и источники ошибки. Проблемы корреляционных исследований (выбросы, надежность измерения, аггрегация, проблема третьей переменной). Коэффициент детерминации. Общее представление о регрессионных моделях.

**Тема 12. Дизайны корреляционных исследований**

Дизайны без воздействия. Срезовые дизайны: основные исследовательские вопросы и угрозы валидности. Лонгитюдные дизайны и временные серии: основные исследовательские вопросы и угрозы валидности. Иерархические структуры данных и общее представление о задачах многоуровневых методов. Исследования без воздействия на одном наблюдении: исследование значимого случая. Исторические исследования.

**Тема 13. Теория причинно-следственного вывода**

Необходимые условия причинности. Каузальная модель Неймана-Рубина: единицы, воздействия, потенциальные следствия. Что такое эффект: ITE, ATE, ATT, ATU. SUTVA. Способы назначения: unconfounded, ignorable и non-ignorable. Проблема эндогенности и ошибки: почему рандомизованные эксперимент является «золотым стандартом» в каузальном анализе.

**Тема 14. Угрозы внутренней валидности и доэкспериментальные дизайны**

Угрозы внутренней валидности. Доэкспериментальные (слабые) дизайны и как они обходятся с этими угрозами: дизайн с пост-тестом на одной группе, дизайн на одной группе с претестом и пост-тестом, дизайны с неэквивалентными группами, отсутствующим воздействием, повторным воздействием, case-control.

**Тема 15. Рандомизованные контролируемые дизайны**

Рандомизованные контролируемые дизайны и угрозы внутренней валидности: дизайн с посттестом и контрольной группой, дизайн с претестом и посттестом и контрольной группой, множественные воздействия и контрольные группы, факторные дизайны, дизайны с повторными измерениями, лонгитюдные дизайны. Ситуации, когда рандомизация желательна и затруднена.

**Тема 16. Шаги в осуществлении рандомизованного контролируемого исследования**

Разработка исследовательского вопроса. Выбор генеральной совокупности и стратегии построения выборки. Варианты рандомизации: простая, кластерная, стратифицированная. Оценка минимального выявляемого размера эффекта и статистической мощности для разных вариантов дизайна. Как организовать воздействие: наличие воздействия, разные виды воздействий, дозировка (уровни) воздействия. Достоверность (fidelity). Оценка эффекта воздействия.

**Тема 17. Дизайн прерванных временных серий**

Структура дизайна прерванных временных серий. Изменения наклона (slope) и остаточного среднего (intercept). Отложенные эффекты. Осложнение дизайна дополнительными условиями. Частые проблемы дизайна прерванных временных серий.

**Тема 18. Дизайн разорванной регрессии**

Общая структура дизайна. Разорванная регрессия как эффект воздействия в рандомизованном контролируемом эксперименте. Моделирование разорванной регрессии. Соответствие точки отсечения. Неточная разорванная регрессия. Дизайн разорванной регрессии и угрозы внутренней валидности.

**Тема 19. Балансировка данных при помощи propensity scores, взвешивание**

Способы балансировки данных в неэкспериментальном дизайне. Conditional ignorability. Проблема размерности и свойства propensity scores. Способы анализа с использованием propensity scores. Оценка propensity scores: определение ковариатов и выбор корректной модели. Многомерный анализ с использованием propensity scores в качестве весов.

**Тема 20. Стратификация и мэтчинг**

Техники мэтчинга: разные виды «жадного мэтчинга», оптимальный мэтчинг. Баланс и common support. Диагностика баланса. Анализ на парных (matched) выборках: техники и выбор ковариатов. Стратификация при помощи propensity scores и анализ внутри страт. Угрозы внутренней валидности при использовании propensity scores.

# Образовательные технологии

Лекции, семинары, самостоятельные работы, письменные самостоятельные работы, работа по группам, общегрупповая дискуссия, мастер-классы.

# Оценочные средства для текущего контроля и аттестации аспиранта

Для текущего контроля используются короткие самостоятельные работы.

## 10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

***Примеры заданий контрольных работ***:

* Прочитайте краткое описание исследования и определите угрозы внутренней валидности. Как можно было бы спланировать исследование, чтобы избежать их?
* По заданным параметрам оцените статистическую мощность анализа и потребный объём выборки.
* По таблице ANOVA рассчитайте эффект дизайна для кластерной выборки. Во сколько раз необходимо увеличить выборку?

## *Вопросы для оценки качества освоения дисциплины*

Задания для итоговой оценки качества освоения дисциплины представляют собой вопросы-кейсы и закрытые вопросы с множественным выбором. Все тесты и экзамены в формате open-book (можно пользоваться материалами курса).

## 10.2 Примеры заданий промежуточной аттестации

1. Стандартное нормальное распределение – это теоретическое нормальное распределение со следующими дополнительными свойствами:

A. Среднее = 0, стандартное отклонение = 1, площадь под кривой = 1

B. Асимметрия = 0, эксцесс = 0

C. Симметрия относительно среднего

D. Кривая никогда не соприкасается с осью X

12. Какая стратегия контроля релевантна для межгрупповых планов?

A. Балансировка

B. Контрбалансировка

C. Позиционное уравнивание

D. Мэтчинг

Кейс. Руководство ВУЗа создало центр психологической помощи студентам и провело исследование эффективности программы психологической помощи. Трое психологов использовали батарею психологических тестов и показатель тревожности, чтобы измерить изменения в личности, связанные с прохождением 3-месячной программы консультирования. Анализ различий между показателями до и после консультирования у 85 студентов, которые полностью прошли программу, продемонстрировал статистически значимое повышение средних показателей доверия, эмоциональной стабильности, экстраверсии при значимом снижении показателей истерии и депрессии. По оценкам кураторов курсов, помощь оказалась успешной в более чем 80% случаев. Есть ли основания усомниться в том, что центр психологической помощи эффективен? Какие альтернативные объяснения можно дать полученным данным?

# Порядок формирования оценок по дисциплине

В накопленную оценку входят оценка активности на занятиях (вес 0,3) и оценка по текущим контрольным / самостоятельным работам / тестам (вес 0,7).

Оценка активности на занятиях ставится по 5-балльной шкале (максимум 2 балла за присутствие: 0 – аспирант отсутствовал, 1 – аспирант присутствовал частично, 2 – аспирант присутствовал полностью и максимум 3 балла за активность: 0 – аспирант не принимал участия в занятии, 1 – аспирант участвовал пассивно, 2 – участвовал активно или выполнил часть заданий, 3 – участвовал очень активно или выполнил все задания). Оценку за каждое занятие аспиранты ставят себе в ведомость самостоятельно. В конце оценки за курс суммируются и нормируются (максимум 10 баллов).

Задания по самостоятельным / контрольным работам / тестам оцениваются по 10-балльной шкале и усредняются. В случае пропуска самостоятельной / контрольной работы аспирант имеет право дослать её не позднее чем в день итогового теста.

При накопленной оценке 7,5 и выше она может быть выставлена в качестве итоговой за курс по желанию аспиранта (автомат). В случае, если аспирант, имеющий право на автомат, сдает итоговый тест, результаты теста учитываются, если оценка оказывается более высокой.

Результирующая оценка по дисциплине: 0,5 \* накопленная + 0,5 \* итоговый тест.

# Учебно-методическое и информационное обеспечение дисциплины

**11.1 Основная литература**:

Cohen, L., Manion, L., & Morrison, K. (2000). Research methods in education. Routledge.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. Educational Researcher, 33(7), 14-26.

Gravetter, F.J., & Forzano, L.-A.B. (2010). Research Methods for the Behavioral Sciences. 4th Ed. Congage Learning.

Shadish, Cook, & Thomas (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton-Mifflin. Chapter 2. Statistical Conclusion Validity and Internal Validity.

Shadish, Cook, & Thomas (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton-Mifflin. Chapter 3. Construct Validity and External Validity.

Field, A. (2005). Discovering Statistics Using SPSS (and sex, drugs and rock’n’roll). London: Sage. Ch. 4 (Correlation)

**11.2 Дополнительная литература:**

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice, 40(5), 532-538.

Reshef, D. N., et al. (2011). Detecting novel associations in large data sets. Science, 334, 1518-1524.

Van de Vijver, F. J. R., van Hemert, D. A., & Poortinga, Y. H. (2008). Multilevel Analysis of Individuals and Cultures. Taylor & Francis. Chapter 1: Conceptual issues in multilevel models.

Van de Vijver, F. J. R., van Hemert, D. A., & Poortinga, Y. H. (2008). Multilevel Analysis of Individuals and Cultures. Taylor & Francis. Chapter 1: Conceptual issues in multilevel models.

Batchelder, J.S., & Rachal, J.R. (2000). Efficacy of a computer-assisted instruction program in a prison setting: An experimental study. Adult Education Quarterly, 50 (2), 120-133.

Greifeneder, R. Alt, A., Bottenberg, K., Seele, T., Zelt, S., & Wagener, D., (2010). On writing legibly: Processing fluency systematically biases evaluations of handwritten material. Social Psychological and Personality Science, 1 (3): 230-237.

Kiernan, N. E., Kiernan, M., Oyler, M. A., & Gilles, C. (2006). Is a web survey as effective as a mail survey? American Journal of Evaluation, 26 (2): 245-252.

Fraser, K., Wallis, M., & St. John, W. (2004). Improving children's problem eating and mealtime behaviours: An evaluative study of a single session parent education programme. Health Education Journal, 63 (3), 229-241.

Mathews, M. S., Peters, S. J., & Housand, A. M. (2012). Regression discontinuity design in gifted and talented education research. Gifted Child Quarterly, 56 (2): 105-112.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. Review of economics and statistics, 86(1), 226-244.

Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills?: A propensity score weighting analysis. Journal of School Psychology, 51(2), 243-260.

Dehejia, R.H. , & Wahba S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. The Review of Economics and Statistics. Vol. 84. No. 1. Pp. 151-161

Morgan, P. L., Frisco, M. L., Farkas, G., & Hibel, J. (2010). A propensity score matching analysis of the effects of special education services. The Journal of special education, 43(4), 236-254.

O'Muircheartaigh, C., & Hedges, L. V. (2014). Generalizing from unrepresentative experiments: a stratified propensity score approach. Journal of the Royal Statistical Society: Series C (Applied Statistics), 63(2), 195-210.

Clark, M. H., & Cundiff, N. L. (2011). Assessing the effectiveness of a college freshman seminar using propensity score adjustments. Research in Higher Education, 52(6), 616-639.

**11.3 Программные средства**

В рамках курса студенты могут выполнять задания в пакетах IBM SPSS или R (по желанию), знакомятся с пакетом GPower.

# Материально-техническое обеспечение дисциплины

Проектор (для лекций или семинаров) и компьютер (ноутбук) для преподавателя.