

## **ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №6**

### **Гипотезы и процедуры проверки**

Используя следующую последовательность шагов, при выполнении проверки гипотезы, решите задачу индивидуального варианта.

1. Опишите характеристику, гипотеза о которой должна быть проверена.
2. Сформулируйте нулевую гипотезу  $H_0$ .
3. Сформулируйте альтернативную гипотезу  $H_1$ .
4. Выберите уровень значимости.
5. Укажите статистику теста, которая будет использоваться, с заменой гипотетического значения, указанного на шаге 2, но без каких-либо вычислений на этом этапе.
6. Убедитесь, что все предположения, необходимые для теста, являются разумными.
7. Вычислите все величины, фигурирующие в статистике теста, а затем значение самой статистики теста.
8. Определите Р-значение, связанное с наблюдаемым значением тестовой статистики. Р-значение (также иногда называемое наблюдаемым уровнем значимости) является мерой несоответствия между предполагаемым значением для характеристики генеральной совокупности и наблюдаемой выборкой. Эта вероятность, при условии, что  $H_0$  истинно, получить значение тестовой статистики, по крайней мере, столь же несовместимое с  $H_0$ , как то, что наблюдалось.
9. Сформулируйте вывод (который состоит в том, чтобы отклонить  $H_0$ , если Р-значение  $\leq \alpha$ , и не отклонять  $H_0$  в противном случае). Затем следует сформулировать вывод в контексте проблемы и указать уровень значимости.

Шаги 1–4 составляют формулировку проблемы, шаги 5–8 дают анализ, который приводит к решению, а шаг 9 дает заключение.

1. В статье «7 миллионов подростков США провалят тест на беговой дорожке» (Associated Press, декабрь 2005 г.) обобщены результаты исследования, в котором 2205 подростков в возрасте от 12 до 19 лет прошли тест на беговой дорожке проверки сердечно-сосудистой системы. Исследователи, проводившие тест, указали, что выборка была отобрана таким образом, чтобы ее можно было рассматривать как репрезентативную для подростков по всей стране. Из 2205 протестированных подростков 750 показали плохой уровень сердечно-сосудистой системы. Подтверждает ли этот тест на уровне значимости 0,05 утверждение о том, что более 30% подростков имеют низкий уровень сердечно-сосудистой системы? Используйте девятиэтапную процедуру проверки гипотез.
2. Отчет «Разрыв в образовательных навыках в Калифорнии: скромные улучшения могут принести большие выгоды» (Институт государственной политики Калифорнии, апрель 2008 г.) утверждает, что по всей стране 61 % выпускников средней школы продолжают учиться два или четыре года в колледже через год после окончания.. Процент поступающих в колледж для выпускников средних школ Калифорнии оценивается в 55%. Предположим, что оценка в 55% была основана на случайной выборке 1500 выпускников средних школ Калифорнии в 2009 году. Можем ли мы сделать статистический вывод о том, что доля выпускников средней школы Калифорнии в 2009 году, которые поступили в колледж через год после выпуска, отличается от общенационального показателя? Используйте девятиэтапную процедуру проверки гипотез и уровень значимости 0,01, чтобы ответить на этот вопрос.

3. Эксперимент, проведенный исследователями из Университета штата Пенсильвания, изучал вопрос, нарушается ли восприятие времени, как показатель способности человека концентрироваться, во время отмены никотина. Результаты исследования были представлены в статье “Воздержание от курения снижает точность оценки времени у курильщиков сигарет” (Бюллетень психофармакологии, 2003). После 24-часового воздержания от курения 20 курильщиков попросили оценить, сколько времени прошло за 45-секундный период. Были полученные данные о воспринимаемом прошедшем времени (в секундах):

69 65 72 73 59 55 39 52 67 57 56 50 70 47 56 45 70 64 67 53

Исследователи хотели определить, оказывает ли воздержание от курения негативное влияние на восприятие времени, заставляя переоценивать прошедшее время. Чтобы ответить на этот вопрос, проведите проверку гипотезы с уровнем значимости 0,05, используя девятиэтапную процедуру.

4. Растущую озабоченность работодателей вызывает время, затрачиваемое на такие виды деятельности, как серфинг в Интернете и переписка в мессенджерах с друзьями в рабочее время. Газета San Luis Obispo Tribune обобщила результаты опроса большой выборки работников в статье под заголовком “Кто бездельничает 2 часа в день? Большинство работников, говорится в опросе” ( Август 2016 г.). Предположим, что генеральный директор крупной компании хочет определить, является ли среднее количество потерянного времени в течение 8-часового рабочего дня для сотрудников его компании меньше заявленных 120 минут. Были опрошены 10 случайных сотрудников о ежедневном потерянном времени на работе. В результате получают следующие данные:

108 112 117 130 111 131 113 113 105 128

Свидетельствуют ли эти данные о том, что среднее потерянное время для этой компании составляет менее 120 минут? Чтобы ответить на этот вопрос, проведите проверку гипотезы используя девятиэтапную процедуру, на уровне значимости 0,05.

5. В статье «Сытые сверчки привлекают девиц» (Nature Science, февраль 1999 г.) сообщается, что самок полевых сверчков привлекают самцы, у которых наблюдается высокий уровень щебетания, и высказывается предположение, что частота щебетания связана со статусом питания. Сообщалось, что обычная частота щебетания самцов полевых сверчков колеблется в среднем около 60 щебетаний в секунду. Чтобы выяснить, связана ли частота щебетания со статусом питания, исследователи держали сверчков-самцов на высокобелковой диете в течение 8 дней, после чего измеряли частоту щебетания. Средняя частота щебетания сверчков на диете с высоким содержанием белка составляла 109 щебетаний в секунду. Является ли это убедительным доказательством того, что средняя частота щебетания сверчков на диете с высоким содержанием белка превышает 60 (что тогда означает преимущество в привлечении самок)? Предположим, что размер выборки  $n = 32$  и стандартное отклонение выборки  $s = 40$ . Проверьте соответствующие гипотезы с уровнем значимости 0,01, используя девятиэтапную процедуру.
6. По-прежнему ли женщинам платят меньше, чем мужчинам, за сопоставимую работу? Авторы статьи “Пол и зарплата: Опрос специалистов по закупкам и снабжению” (Journal of Purchasing and Supply Management, 2008) провели исследование, в котором данные о заработной плате были собраны из случайной выборки мужчин и случайной выборки женщин, которые работали менеджерами по закупкам и были подписчиками Journal of Purchasing and Supply Management. Данные о заработной плате, соответствующие сводным количествам, приведенным в документе, приведены ниже (фактические размеры выборки для исследования были намного больше):

Зарплата (в тысячах долларов)

Мужчины 81 69 81 76 76 74 69 76 79 65

Несмотря на то, что выборки были отобраны из подписчиков конкретного журнала, авторы статьи полагали, что было разумно рассматривать выборки в исследовании как репрезентативные для двух групп — мужчин - менеджеров по закупкам и женщин - менеджеров по закупкам. Для целей этого примера предположим, что также разумно рассматривать две приведенные здесь выборки как репрезентативные для популяций. Используя приведенные данные и уровень значимости 0,05 определите, имеются ли убедительные доказательства того, что средняя годовая зарплата менеджеров по закупкам мужчин больше, чем средняя годовая зарплата менеджеров по закупкам женщин.

7. Статья “Насколько я был счастлив, в любом случае? Смещение ретроспективного воздействия” (Social Cognition, 2003) сообщили об эксперименте, предназначенном для оценки степени, в которой люди оправдывают плохую производительность. В этом исследовании 246 студентов колледжа были случайным образом распределены в одну из двух групп - группу с отрицательной обратной связью или группу с положительной обратной связью. Каждый участник прошел тест, в ходе которого ему было предложено угадать эмоции, отображаемые на фотографиях лиц. В конце теста участникам группы с отрицательной обратной связью было сказано, что они правильно ответили на 21 из 40 вопросов и получили “плохую” оценку. Тем, кто был в группе с положительной обратной связью, сказали, что они правильно ответили на 35 из 40, и им была присвоена "отличная" оценка. Спустя короткое время участникам было предложено ответить на два набора вопросов. Один набор вопросов касался достоверности теста, а другой набор вопросов касался важности умения читать по лицам. Исследователи предположили, что те, кто находится в группе с отрицательной обратной связью, будут склонны объяснять свою плохую работу оценивая как валидность теста, так и важность умения читать с лица ниже, чем в группе с положительной обратной связью. Подтверждают ли данные этого эксперимента гипотезы исследователей?

Группа	Размер выборки	Оценка валидности теста		Оценка важности чтения по лицу	
		Выборочное среднее	Стандартное отклонение	Выборочное среднее	Стандартное отклонение
с отрицательной обратной связью	123	5,51	0,79	5,36	1,00
с положительной обратной связью	123	6,95	1,09	6,62	1,19

Проверьте гипотезу об оценке валидности теста, используя девятиэтапную процедуру на уровне значимости 0,01.

8. Повышает ли диалог кровяное давление, способствуя тенденции к повышению кровяного давления при измерении в кабинете врача, чем при измерении в менее напряженной обстановке? (Этот хорошо документированный эффект называется “эффектом белого халата”.) В статье “Эффект диалога и эффект ”Белого халата" у пациентов с гипертонией: физические усилия или эмоциональное содержание" (Поведенческая медицина 2001) описано исследование, в котором пациенты с высоким кровяным давлением были случайным образом распределены в одну из двух групп. Тем, кто входил в первую группу (группа диалога), были заданы вопросы об их истории болезни и об источниках стресса в их жизни за несколько минут до того, как было измерено их кровяное давление. Тех, кто был во второй группе (группа подсчета), попросили посчитать вслух от 1 до 100 четыре раза, прежде чем измерять их кровяное давление. Следующие значения данных для диастолического артериального давления (в миллиметрах ртутного столба) соответствуют суммарным величинам, приведенным в документе:

Диалог 104 110 107 112 108 103 108 118

Подсчёт 110 9 6 103 9 8 100 109 9 7 105

Проверьте гипотезу что среднее диастолическое артериальное давление при разговоре выше среднего при подсчете, используя девятиэтапную процедуру на уровне значимости 0,05.

9. В статье “Счастье на продажу: делают ли "эмпирические покупки" потребителей счастливее, чем материальные покупки?” (Journal of Consumer Research, 2009), авторы проводят различие между расходом денег на впечатления (такие как путешествия) и расходом денег на материальные ценности (такие как автомобиль). В ходе эксперимента по определению того, влияет ли тип покупки на то, насколько люди довольны покупкой после ее совершения, 185 студентов колледжа были случайным образом распределены в одну из двух групп. Студентов из “эмпирической” группы попросили вспомнить время, когда они потратили около 300 долларов на впечатления (опыт), а затем оценили эту покупку по трем различным шкалам счастья, которые затем были объединены в общую меру счастья. Студентов, отнесенных к “материальной” группе, попросили вспомнить время, когда они потратили около 300 долларов на вещь. Средний балл счастья составил 5,75 для “эмпирической” группы и 5,27 для “материальной” группы.

"Эмпирическая"	"Материальная"
$n_1 = 92$	$n_2 = 93$
$s_1 = 1,2$	$s_2 = 1,5$

Используйте уровень значимости 0,05 для проверки гипотезы: подтверждают ли данные вывод авторов о том, что “покупки для приобретения чувственного опыта вызвали больше сообщений о счастье”.

10. Могут ли движения рук помочь детям выучить математику? Этот вопрос исследовался в статье «Жесты дают детям новые представления о математике» (Psychological Science, 2009). В эксперименте участвовали восемьдесят пять детей третьего и четвертого классов (американской начальной школы), которые не ответили правильно ни на какие вопросы в тесте с шестью задачами вида  $3 + 2 + 8 = \_\_ + 8$ . Детей случайным образом распределили либо в группу без жестов, либо в группу жестов. Всем детям был дан урок о том, как решать задачи такого типа, используя стратегию попытки уравнивать обе части уравнения. Детей в группе жестов также учили указывать на первые два числа в левой части уравнения указательным и средним пальцами одной руки, а затем указывать на пробел в правой части уравнения. Этот жест должен был подчеркнуть, что группировка участвует в решении задачи. Затем дети практиковали дополнительные задачи этого типа. После всем детям был предложен тест, в котором нужно было решить шесть задач, и для каждого ребенка было подсчитано количество правильных ответов. Сводная статистика представлена ниже.

Группа	n	$\bar{x}$	s
Без жестов	42	1,3	0,3
Жестов	43	2,2	0,4

Есть ли доказательства в поддержку теории о том, что изучение жестового подхода к решению задач такого типа приводит к более высокому среднему числу правильных ответов? Проверьте соответствующие гипотезы, используя девятиэтапную процедуру на уровне значимости 0,01.

11. Статья “Долговая грамотность, финансовый опыт и чрезмерная задолженность” (Social Science Research Network, 2008) включала анализ данных национальной выборки из 1000 американцев. Один из вопросов в ходе опроса был: “Вы должны 3000 долларов по своей кредитной карте. Вы платите минимальный платеж в размере 30 долларов США каждый месяц. При годовой процентной ставке 12% (или 1% в месяц), сколько лет потребуется для ликвидации долга по кредитной карте, если вы не будете делать никаких дополнительных расходов?”

Варианты ответов на этот вопрос были следующими: (а) менее 5 лет; (б) от 5 до 10 лет; (в) от 10 до 15 лет; (г) никогда — вы по-прежнему будете в долгах; (д) не знаете; и (е) предпочитаете не отвечать. Только 354 из 1000 респондентов выбрали правильный ответ "никогда". Для целей этого упражнения вы можете предположить, что выборка является репрезентативной для взрослых американцев. Есть ли убедительные доказательства того, что доля взрослых американцев, которые могут правильно ответить на этот вопрос, составляет менее 40 %? Используйте уровень значимости 0,05 для проверки соответствующих гипотез.

12. (Условие задачи 11). В статье также сообщается, что 37,8% тех, кто выбрали один из неправильных ответов (а, б и с) в качестве ответа на этот вопрос. Разумно ли заключить, что более трети взрослых американцев выбрали бы неправильный ответ на этот вопрос? Используйте уровень значимости 0,05 и девятиэтапную процедуру проверки гипотезы.
13. Исследование потребления фастфуда описано в статье "Что люди покупают в ресторанах быстрого питания" (Obesity, 2009). К взрослым посетителям трех сетей гамбургеров (McDonald's, Burger King и Wendy's) в обеденное время в Нью-Йорке подходили при входе в ресторан и просили предъявить чек при выходе. Затем по чекам определяли, что было куплено, и определяли количество потребленных калорий. Всего в исследовании приняли участие 3857 человек. Среднее выборочное количество потребленных калорий составило 857, а выборочное стандартное отклонение - 677.

Исходя из рекомендованного ежедневного потребления 2000 калорий, рекомендуется стремиться к потреблению 750 калорий в обед. Предполагая, что выборку из 3857 покупок в фастфуде разумно считать репрезентативной для всех покупок гамбургеров в обеденное время в Нью-Йорке, проведите проверку гипотезы, чтобы определить, дает ли выборка убедительные доказательства того, что среднее количество калорий в покупке гамбургеров в обеденное время в Нью-Йорке больше, чем рекомендованные 750 калорий. Используйте уровень значимости 0,01.

14. Медицинские исследования показали, что многократное разгибание запястья более чем на 20 градусов повышает риск травм запястья и кисти. Каждый из 24 студентов Корнельского университета использовал предложенную новую конструкцию компьютерной мыши, и во время использования мыши фиксировалось разгибание запястья каждого студента. Приведены данные, соответствующие суммарным значениям, приведенным в статье "Сравнительное исследование двух конструкций компьютерных мышей" (Cornell Human Factors Laboratory Technical Report).

27 28 24 26 27 25 25 24 24 24 25 28 22 25 24 28 27 26 31 25 28 27 27 25

Используйте эти данные для проверки гипотезы о том, что среднее значение разгибания запястья у людей, использующих новый дизайн мыши, больше 20 градусов. Уровень значимости 0,05.

15. Городской совет крупного города обеспокоен тенденцией к исключению арендаторов с детьми из квартир в пределах города. Координатор по жилищным вопросам решил отобрать случайную выборку из 125 квартир и определить для каждой из них, разрешено ли проживание детей. Пусть  $p$  - это доля всех квартир, в которых запрещено проживание детей. Если городской совет убедится, что  $p$  больше 0,75, он рассмотрит соответствующее законодательство.
- а) Если 102 из 125 отобранных квартир не допускают арендаторов с детьми, приведет ли вас тест уровня 0,05 к выводу, что более 75% всех квартир не допускают детей?
- б) Какова мощность теста, если  $p = 0,8$  и уровень значимости 0,05?

16. "Большинство любят погорячее" - так называется пресс-релиз, выпущенный Pew Research Center (2009 г.). В пресс-релизе говорится что "с подавляющим перевесом американцы хотят жить в солнечном месте". Это утверждение основано на данных опроса, проведенного по репрезентативной выборке 2260 взрослых американцев. Из опрошенных, 1288 человек указали,

что они предпочли бы жить в жарком климате, а не в холодном климате. Являются ли данные выборки убедительным доказательством того, что большинство всех взрослых американцев предпочитают жаркий климат чем холодный климат? Используйте девятиступенчатый процесс проверки гипотез с уровнем значимости 0,01 для ответа на этот вопрос.

17. Согласно опросу Washington Post-ABC News, 331 из 502 случайно выбранных опрошенных взрослых американцев, заявили, что их не беспокоило бы, если бы Национальное Агентство Безопасности собрало записи личных телефонных звонков, которые они сделали. Есть ли достаточные доказательства того, что большинство взрослых американцев чувствуют то же самое? Проверьте соответствующие гипотезы, используя уровень значимости 0,01. Используйте девятиступенчатый процесс.
18. Статья “Насколько я был счастлив, в любом случае? Смещение ретроспективного воздействия” (Social Cognition, 2003) сообщили об эксперименте, предназначенном для оценки степени, в которой люди оправдывают плохую производительность. В этом исследовании 246 студентов колледжа были случайным образом распределены в одну из двух групп - группу с отрицательной обратной связью или группу с положительной обратной связью. Каждый участник прошел тест, в ходе которого ему было предложено угадать эмоции, отображаемые на фотографиях лиц. В конце теста участникам группы с отрицательной обратной связью было сказано, что они правильно ответили на 21 из 40 вопросов и получили “плохую” оценку. Тем, кто был в группе с положительной обратной связью, сказали, что они правильно ответили на 35 из 40, и им была присвоена "отличная" оценка. Спустя короткое время участникам было предложено ответить на два набора вопросов. Один набор вопросов касался достоверности теста, а другой набор вопросов касался важности умения читать по лицам. Исследователи предположили, что те, кто находится в группе с отрицательной обратной связью, будут склонны объяснять свою плохую работу оценивая как валидность теста, так и важность умения читать с лица ниже, чем в группе с положительной обратной связью. Подтверждают ли данные этого эксперимента гипотезы исследователей?

Группа	Размер выборки	Оценка валидности теста		Оценка важности чтения по лицу	
		Выборочное среднее	Стандартное отклонение	Выборочное среднее	Стандартное отклонение
с отрицательной обратной связью	123	5,51	0,79	5,36	1,00
с положительной обратной связью	123	6,95	1,09	6,62	1,19

Проверьте гипотезу исследователей о том, что участники группы отрицательной обратной связи также не будут оценивать важность способности читать лица так же высоко, как и участники группы положительных отзывов, используя девятиэтапную процедуру на уровне значимости 0,01.