

## Задачи к семинару 1

1. Пусть  $Y_i, i = 1, \dots, n$  являются независимыми одинаково распределенными случайными величинами с  $E(Y_i) = \mu_Y$ . Покажите, что выборочное среднее  $\bar{Y}$  является несмещенной и состоятельной оценкой генерального среднего. Найдите дисперсию этой оценки.

### Задачи из учебника Стока и Уотсона:

- 3.5. Проводится опрос 1055 зарегистрированных избирателей, и их просят выбрать между кандидатом А и кандидатом В. Пусть  $p$  обозначает долю голосующих в генеральной совокупности, которые предпочитают кандидата А, и пусть  $\hat{p}$  обозначает долю голосующих в выборке, которые предпочитают кандидата А.

а. Вы заинтересованы в тестировании конкурирующих гипотез  $H_0: p = 0.5$  против  $H_1: p \neq 0.5$ . Предположим, что вы решаете отвергнуть  $H_0$ , если  $|\hat{p} - 0.5| > 0.02$ .

- Какой размер этого теста?
- Вычислите мощность этого теста, если  $p = 0.53$ .

б. В опросе  $\hat{p} = 0.54$ .

- Протестируйте гипотезу о том, что  $H_0: p = 0.5$  против  $H_1: p \neq 0.5$ , используя 5%-й уровень значимости.
  - Протестируйте гипотезу о том, что  $H_0: p = 0.5$  против  $H_1: p > 0.5$ , используя 5%-й уровень значимости.
  - Постройте 95%-й доверительный интервал для  $p$ .
  - Постройте 99%-й доверительный интервал для  $p$ .
  - Постройте 50%-й доверительный интервал для  $p$ .
- в. Предположим, что опрос проводился 20 раз, используя независимо выбранных голосующих в каждом опросе. Для каждого из этих 20 опросов построен 95%-й доверительный интервал для  $p$ .
- Какова вероятность того, что истинное значение  $p$  содержится во всех 20 доверительных интервалах?
  - Сколько этих доверительных интервалов, как вы ожидаете, содержат истинное значение  $p$ ?
- г. На жаргоне опроса, «погрешность» – это  $1.96 \times SE(\hat{p})$ ; т.е. это половина длины 95%-го доверительного интервала. Предположим, что вы хотите создать опрос, который имел погрешность не более 1%. Т.е. вы хотите, чтобы  $\Pr(|\hat{p} - p| > 0.01) \leq 0.05$ . Насколько большим должно быть  $n$ , если опрос использует простую случайную выборку?

- 3.8. Известны результаты новой версии теста SAT 1000 случайно выбранных старшеклассников. Выборочное среднее экзаменационной оценки составляет 1110, и выборочное стандартное отклонение – 123. Постройте 95%-й доверительный интервал для среднего экзаменационной оценки генеральной совокупности старшеклассников.

- 3.9. Предположим, что завод по производству лампочек производит лампы со средней продолжительностью работы, равной 2000 часов, и стандартным отклонением в 200 часов. Изобретатель утверждает, что разработал улучшенный процесс, который производит лампы с более длинной средней продолжительностью работы и тем же самым стандартным отклонением. Директор завода случайно выбирает 100 новых ламп. Она сказала, что она поверит в заявление изобретателя, если выборочное среднее продолжительности работы ламп будет больше, чем 2100 часов; в противном случае она заключит, что новый процесс не лучше,

чем старый. Пусть  $\mu$  обозначает среднее нового процесса. Рассмотрим нулевую и альтернативную гипотезы  $H_0 : \mu = 2000$  против  $H_1 : \mu > 2000$ .

- а. Какой размер тестовой процедуры директора завода?
- б. Предположим, что новый процесс фактически лучше и средняя продолжительность работы лампы равна 2150 часов. Какова мощность тестовой процедуры директора завода?
- в. Какая тестовая процедура должна использоваться директором завода, если она хочет получить размер теста, равный 5%?

3.12. Для исследования гендерной дискриминации в фирмах случайно выбрана выборка из 100 мужчин и 64 женщин с похожими должностями. Информация об их месячных зарплатах говорит следующее:

	Средняя зарплата ( $\bar{Y}$ )	Стандартное отклонение ( $s_Y$ )	$n$
Мужчины	\$3100	\$200	100
Женщины	\$2900	\$320	64

- а. Что говорят эти данные по поводу гендерных различий в фирме? Представляют ли они статистически значимое свидетельство того, что средние зарплаты мужчин и женщин различны? (Чтобы ответить на этот вопрос, во-первых, сформулируйте нулевую и альтернативную гипотезы; во-вторых, вычислите соответствующую  $t$ -статистику; в-третьих, вычислите  $p$ -значение, связанное с  $t$ -статистикой; и наконец, используйте  $p$ -значение для ответа на вопрос.)
  - б. Свидетельствуют ли эти данные о том, что в фирме есть гендерная дискриминация в политике оплаты труда? Объясните.
- 3.16. Известно, что результаты стандартного школьного теста в США имеют среднее, равное 1000. Тест проходят 453 случайно выбранных школьников во Флориде; и в этой выборке среднее составляет 1013, а стандартное отклонение ( $s$ ) 108.
- а. Постройте 95%-й доверительный интервал для среднего результата теста школьников Флориды.
  - б. Можно ли говорить о наличии статистического свидетельства того, что школьники из Флориды выполняют тест иначе, чем другие школьники в Соединенных Штатах?
  - в. Другие 503 школьника были случайно выбраны во Флориде. Им был прочитан 3-часовой подготовительный курс до выполнения теста. Их средняя оценка составила 1019 со стандартным отклонением 95.
    - i. Постройте 95%-й доверительный интервал для изменения в средней оценке, связанной с подготовительным курсом.
    - ii. Есть ли статистическое свидетельство того, что подготовительный курс помог?
  - г. Первым 453 студентам также прочитали подготовительный курс, а затем попросили пройти тест во второй раз. Среднее изменение их оценки составило 9 пунктов, и стандартное отклонение изменилось на 60 пунктов.
    - i. Постройте 95%-й доверительный интервал для изменения в средней оценке.
    - ii. Есть ли статистически значимое свидетельство того, что студенты выполняют тест лучше со второй попытки после подготовительного курса?
    - iii. Студенты могут выполнить тест лучше со второй попытки из-за подготовительного курса или из-за своего полученного опыта при первой попытке. Опишите эксперимент, который будет количественной оценкой этих двух эффектов.

## Домашнее задание к семинару 1

- 3.6. Пусть  $Y_1, \dots, Y_n$  есть i.i.d., выбранные из распределения со средним значением  $\mu$ . Тест  $H_0: \mu = 5$  против  $H_1: \mu \neq 5$ , используя обычную  $t$ -статистику, приводит к  $p$ -значению, равному 0,03.
- Содержит ли 95%-й доверительный интервал  $\mu = 5$ ? Объясните.
  - Можете ли вы определить, содержится ли  $\mu = 6$  в доверительном интервале? Объясните.
- 3.11. Рассмотрим оценку  $\tilde{Y}$ , определенную в уравнении (3.1). Покажите, что (а)  $E(\tilde{Y}) = \mu_Y$  и (б)  $\text{var}(\tilde{Y}) = 1.25\sigma_Y^2 / n$ .
- 3.17. Прочитайте вставку «Гендерный разрыв в заработных платах выпускников колледжей в США» из раздела 3.5.
- Постройте 95%-й доверительный интервал для изменения в средних почасовых зарплатах мужчин между 1992 и 2008.
  - Постройте 95%-й доверительный интервал для изменения в средних почасовых зарплатах женщин между 1992 и 2008.
  - Постройте 95% доверительный интервал для изменения в гендерном разрыве средних почасовых зарплатах между 1992 и 2008. (Подсказка:  $\bar{Y}_{m,1992} - \bar{Y}_{w,1992}$  не зависит от  $\bar{Y}_{m,2008} - \bar{Y}_{w,2008}$ .)
- 3.18. Это упражнение показывает, что выборочная дисперсия является несмещенной оценкой дисперсии генеральной совокупности, когда  $Y_1, \dots, Y_n$  являются i.i.d. со средним  $\mu_Y$  и дисперсией  $\sigma_Y^2$ .
- Используйте уравнение (2.31), чтобы показать, что  $E[(Y_i - \bar{Y})^2] = \text{var}(Y_i) - 2\text{cov}(Y_i, \bar{Y}) + \text{var}(\bar{Y})$ .
  - Используйте уравнение (2.33), чтобы показать, что  $\text{cov}(Y_i, \bar{Y}) = \sigma_Y^2 / n$ .
  - Используйте результаты в (а) и (б), чтобы показать, что  $E(s_Y^2) = \sigma_Y^2$ .
- 3.19. а.  $\bar{Y}$  – несмещенная оценка  $\mu_Y$ . Является ли  $\bar{Y}^2$  несмещенной оценкой  $\mu_Y^2$ ?  
б.  $\tilde{Y}$  – состоятельная оценка  $\mu_Y$ . Является ли  $\tilde{Y}^2$  состоятельной оценкой  $\mu_Y^2$ ?