

Задача 1. Файл «Данные к задаче 1.ods» содержит сведения о продолжительности грудного вскармливания в неделях в выборке рожениц. Число после названия переменной соответствует номеру варианта (так, duration_6 — данные для шестого варианта). Ваша цель — оценить среднюю продолжительность вскармливания.

а) Рассчитайте 90% доверительный интервал для средней продолжительности, считая распределение признака нормальным.

б) Постройте график «квантиль-квантиль» и попробуйте понять, соответствует ли распределение времени вскармливания нормальному закону.

в) Прочитайте ниже описание бутстрапа и рассчитайте этим методом 90% доверительный интервал для средней продолжительности вскармливания, сгенерировав 1000 перевыборок.

г) При выполнении предыдущего пункта вы получите 1000 средних значений продолжительности вскармливания в перевыборках. Постройте гистограмму для этих значений. Похоже ли распределение среднего в перевыборках на нормальное?

Требования к оформлению.

Отчёт должен быть представлен в текстовом документе (.doc, .docx, .pdf, .odt) и представлять собой связный текст: краткое описание решения всех частей задания с выводами по каждому пункту.

Отчёт отсылается проверяющему. Распределение проверяющих:

Варианты 1 и 2 — Коломенцов А.М., amkolomentsov@edu.hse.ru.

Вариант 3 проверяет Исхакова Л.Р., lriskhakova_1@edu.hse.ru.

Вариант 4 проверяет Румянцева Е.В., evrumyantseva.2006@yandex.ru.

Вариант 5 проверяет Чернышёва И.К., riina_mustika@mail.ru.

Вариант 6 проверяет Фурманов К.К., furmach@inbox.ru.

Вариант 7 проверяет Дубина Н.А., nadubina@edu.hse.ru.

Вариант 8 проверяет Дерябин Ф.В., fvderyabin@edu.hse.ru.

Вариант 9 проверяет Смирнов Н.С., nssmirnov_1@edu.hse.ru.

Вариант 10 проверяет Гуломкодиров Н.Ф., nfgulomkodiroy@edu.hse.ru.

Срок сдачи: до 27 февраля включительно.

Штрафы за опоздание:

не больше чем на сутки — 1 балл,

от суток до недели — 3 балла,

больше недели — 5 баллов.

Бутстрап

В случаях, когда предпосылки стандартных статистических процедур доверительного оценивания и проверки гипотез нарушаются, полезной альтернативой может оказаться бутстрап. Термин «бутстрап» (bootstrap) объединяет совокупность процедур статистического вывода, в которых распределение тех или иных статистик оценивается с помощью статистического эксперимента — расчёта статистик по выборкам, полученным из имеющегося выборочного распределения анализируемого показателя («перевыборкам»).

Допустим, нужно рассчитать 90% доверительный интервал для среднего по следующим десяти наблюдениям:

9 5 10 13 10 13 8 10 9 14

Взяв десять раз случайно отобранное число из имеющихся данных (одно и то же число может быть выбрано сколько угодно раз), получаем перевыборку — вот несколько примеров перевыборок:

9	14	9	13	10	10	13	8	10	9
10	10	13	13	5	10	9	9	9	13
5	10	10	13	10	13	8	10	8	14

Число 10 встречается в исходной выборке трижды, поэтому в каждом наблюдении перевыборки оно появляется с вероятностью 0.3. Числа 13 и 9 появляются с вероятностью 0.2, числа 5, 8 и 14 — с вероятностью 0.1.

Рассчитав большое количество — например, тысячу — перевыборок, мы можем оценить распределение выборочного среднего (точнее, перевыборочного среднего). Пример гистограммы, полученной по тысяче перевыборок, представлен на рис. 4.11.

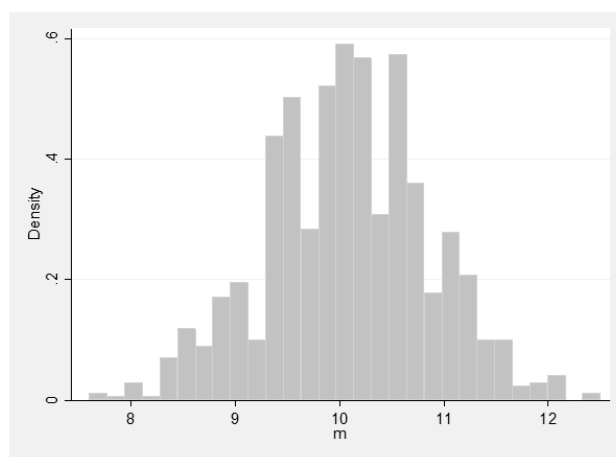


Рис. 4.11. Гистограмма для среднего в перевыборках

Допустим, средние в перевыборках оказались распределены так, что квантили порядка 5% и 95% для средних оказались равны 8.7 и 11.4 соответственно. Тогда интервал (8.7; 11.4) и будет 90% бутстраповским доверительным интервалом. В общем случае, выборочные квантили $\hat{Q}(\alpha/2)$ и $\hat{Q}(1 - \alpha/2)$ для средних в перевыборках образуют бутстраповский доверительный интервал для среднего в генеральной совокупности с уровнем доверия $1 - \alpha$.

Бутстрап — асимптотический метод, не предназначенный для применения к выборкам малого размера. Однако он даёт хорошие результаты при весьма общих предпосылках, не требуя нормальности распределения, и часто показывает себя лучше стандартного доверительного интервала. Здесь описана лишь одна из множества процедур бутстрапа.