Домашнее задание по курсу “Теория вероятностей и математическая статистика”
ОП “Бизнес-информатика”
Выполнил
Кузнецов Вадим Олегович ББИ212
Вариант 10

Задание 1.

Рассчитайте выборочное среднее, смещенную оценку выборочной дисперсии, несмещенную оценку выборочной дисперсии, минимальное и максимальное значения в выборке.

Решение:

n = 1000

Формула выборочного среднего:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Значение каждого x_i дано в таблице, рассчитаем в “HomeWork_Kuznetsov.pynb”.

Получим $\bar{x}$ = 7.590100102044744

Формула смещенной оценки выборочной дисперсии:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Получили, что сумма равна 170558.6411818631, умножим на 1 / 1000.

Тогда s^2 = 170.5586411818631.

Формула несмещенной оценки выборочной дисперсии:

$$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Получили, что сумма равна 170558.6411818631, умножим на 1 / 999.

Тогда $\widehat{\sigma^2}$ = 170.7293705524155.

Минимальное значение нашли по формуле .min(), аналогично с максимальным (.max())

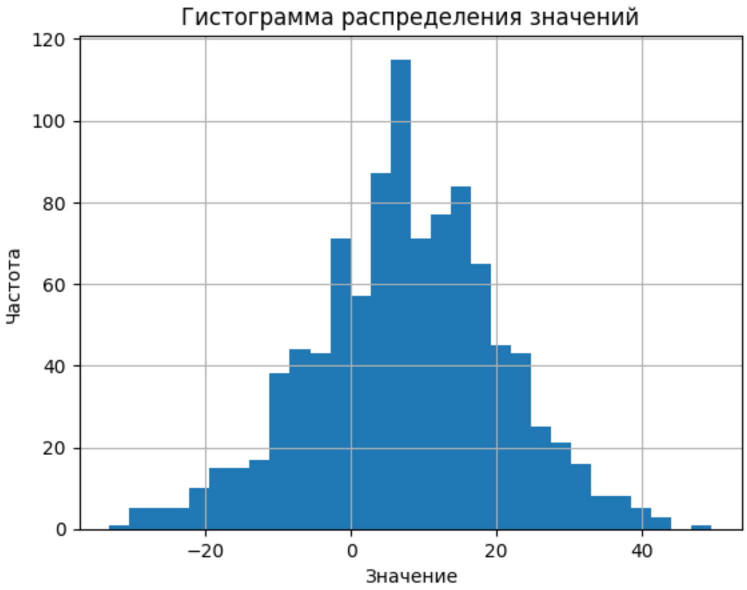
min = -33.26398528367281

max = 49.64002200216055

Задание 2.

Постройте гистограмму. Предположите закон распределения генеральной совокупности, из которой была получена выборка. Обоснуйте ответ

Решение:



Предположим, что это в генеральной совокупности нормальный закон распределения, так как видно распределение с пиком в центре и симметричными боковыми сторонами, что со ответствует описанию нормального распределения.

Далее предположим, что выборка получена из нормальной генеральной совокупности N(μ,σ2).

Задание 3.

Постройте 95% доверительный интервал для математического ожидания.
а) выпишите формулу доверительного интервала, которую будете использовать. Поясните почему именно такую формулу выбрали;
б) выпишите критическое значение статистики.

Решение:

Пусть наша величина генеральной совокупности распределена нормально, учитывая, что дисперсия неизвестна.

Тогда по данным выборки можно строить случайную величину, которая имеет распределение Стьюдента с n – 1 степенями свободы.

$$\frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \sim t_{n-1}$$

Формула:

$$\left(t_{n-1,\frac{\alpha}{2}} \leq \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t_{n-1,1-\frac{\alpha}{2}} \right) = 0.95$$

Само распределение Стьюдента с k степенями свободы выглядит следующим образом:

Пусть Z – случайная величина, имеющая стандартное нормальное распределение, Y – с.в., распределённая по закону хи-квадрат с k степенями свободы, Z и Y независимы. Пусть $U = \frac{Z}{\sqrt{\frac{1}{k}Y}}$.

Рассмотрим величину $U = \frac{\overline{X} - \mu}{\left(\hat{\sigma} / \sqrt{n}\right)}$. Иначе её можно представить так:

$$U = \frac{\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}}\right)}{\sqrt{\frac{1}{n-1}\left(\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right)}}$$

Теперь в выражении есть истинная дисперсия наряду с выборочной. Выше уже говорилось, что $\hat{\sigma}^2$ и $\overline{X}$ независимы, а значит, независимы числитель и знаменатель этой дроби. В числителе стоит случайная величина, имеющая стандартное нормальное распределение, а в знаменателе – число 1/(n-1), умноженное на случайную величину, имеющую хи-квадрат распределение с (n-1) степенью свободы. Таким образом, величина U имеет распределение Стьюдента.

Б) Найдем же теперь критическое значение статистики

$$P\left(t_{n-1,1-\frac{\alpha}{2}} \leq \frac{\overline{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n}}} \leq t_{n-1,\frac{\alpha}{2}}\right) = 0.95$$

1 - α = 0.95

α = 0.05

Выясним критические значения распределения Стьюдента с 999 степенями свободы

$t_{999,0.025}$ = 1.962341461133449

$t_{999,0.975}$ = -1.962341461133449

$$P\left(\overline{x} - t_{n-1,\frac{\alpha}{2}} * \frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{x} - t_{n-1,1-\frac{\alpha}{2}} * \frac{\hat{\sigma}}{\sqrt{n}}\right) = 0.95$$

$$P\left(\overline{x} - t_{999,0.025} * \frac{\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \overline{x} - t_{999,0.975} * \frac{\hat{\sigma}}{\sqrt{n}}\right) = 0.95$$

$$P(6.779272173212221 \leq \mu \leq 8.400928030877267) = 0.95$$

Ответ: $6.779272173212221 \leq \mu \leq 8.400928030877267$

Задание 4.

Постройте 95% доверительный интервал для дисперсии.

а) выпишите формулу доверительного интервала, который будете использовать. Поясните почему именно такую формулу выбрали;

б) выпишите критическое значение статистики.

$$P(\chi 2_{(n-1),1-\frac{\alpha}{2}} \leq \frac{(n-1) * \hat{\sigma}^2}{\sigma^2} \leq \chi 2_{(n-1),\frac{\alpha}{2}}) = 0.95$$

$$P(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}) = 0.95$$

Пусть наша величина генеральной совокупности распределена нормально, учитывая, что дисперсия неизвестна. Тогда по данным выборки можно строить случайную величину, имеющую хи-квадрат распределение с k степенями свободы. Очевидно, что

$$P(\chi^2_{k,1-\alpha/2} < \chi^2_k < \chi^2_{k,\alpha/2}) = 1-\alpha\ .$$

Известно, что в случае нормальной генеральной совокупности случайная величина $\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$ имеет хи-квадрат распределение со степенями свободы n-1. Таким образом,

$$1-\alpha = P(\chi^2_{n-1,1-\alpha/2} < \chi^2_{n-1} < \chi^2_{n-1,\alpha/2}) = P(\chi^2_{n-1,1-\alpha/2} < \frac{(n-1)\hat{\sigma}^2}{\sigma^2} < \chi^2_{n-1,\alpha/2}) =$$

$$= P(\frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,\alpha/2}} < \sigma^2 < \frac{(n-1)\hat{\sigma}^2}{\chi^2_{n-1,1-\alpha/2}}).$$

Б) Расчитаем

$\chi^2_{n-1,\frac{\alpha}{2}} = 1088.4870677259353$
$\chi^2_{n-1,1-\frac{\alpha}{2}} = 913.3009983021134$

$$P(156.69330967633252 \leq \sigma^2 \leq 186.74964934774277) = 0.95$$

Ответ: $156.69330967633252 \leq \sigma^2 \leq 186.74964934774277$

Задание 5.

Проверьте гипотезу о равенстве математического ожидания нулю. Используйте двустороннюю альтернативу и уровень значимости 10%.
а) сформулируйте основную и альтернативную гипотезы;
б) выпишите расчетную статистику. Почему была выбрана именно такая статистика?

А) Сформулируем гипотезы
H0: $\mu_0 = 0$
H1: μ_0 != 0 (так как используем двустороннюю альтернативу, то $\mu_0 > 0$ или $\mu_0 < 0$)
Б) Распределение расчетной статистики:

$$T = \frac{\bar{X} - \mu 0}{\sqrt{\hat{\sigma}^2/n}} \sim t(n-1)$$

Эта статистика была выбрана, так как при больших n распределение Стьюдента почти не отличается от нормального.
Найдем критическое значение функции:
$T = \frac{7.590100102044744}{\sqrt{170.7293705524155/1000}} = 18.369332869233308$
Уровень значимости 10%
$\alpha = 0.1$
Тогда найдем z_{95} и z_5
$t_{999,0.05} = -1.6463803454275356$
$t_{999,0.95} = 1.646380345427535$

Критическое значение 18.369332869233308 выходит за пределы интервала (-1.6463803454275356, 1.646380345427535)

H_0 тогда отвергается в пользу H1.

Задание 6.

Проверьте гипотезу о равенстве математических ожиданий двух выборок, предполагая равенство дисперсий.
Используйте двустороннюю альтернативу и уровень значимости 10%.
а) сформулируйте основную и альтернативную гипотезы;
б) выпишите расчетную статистику. Почему была выбрана именно такая статистика?

n = 500, m = 500
$\bar{x} = 7.672579869842593$
$\bar{y} = 7.507620334246894$
$\widehat{\sigma x}^2 = 166.0071180387518$
$\widehat{\sigma y}^2 = 175.79376609323836$
А) Сформулируем гипотезы
H0: $\mu_0 = \mu_1$
H1: $\mu_0 \mathrel{!}= \mu_1$ (так как используем двустороннюю альтернативу, то $\mu_0 > 0$ или $\mu_0 < 0$)
Б) Расчетная статистика:

$$T = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\left(\frac{n-1}{n+m-2}\hat{\sigma}^2{}_X+\frac{m-1}{n+m-2}\hat{\sigma}^2{}_Y\right)}} \sim t(n+m-2)$$

Это две независимые нормальные одинаково распределённые случайные величины с математическими ожиданиями μ0 и μ1 и дисперсиями σ^2 0 и $\sigma^2$1, и на основании данных об их реализациях нужно построить доверительный интервал для величин d=Xi-Yi. Эти величины также образуют выборку из нормальной генеральной совокупности с математическим ожиданием E(d)=μ0-μ1 = 0 Таким образом, задача построения доверительного интервала для разности математических ожиданий сводится к задаче оценивания интервала для математического ожидания d, а это уже приводится к $t(n-1)$
Тогда расчетная статистика:

$$T = \frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\left(\frac{n-1}{n+m-2}\hat{\sigma}^2{}_X+\frac{m-1}{n+m-2}\hat{\sigma}^2{}_Y\right)}} \sim t(n+m-2)$$

Уровень значимости 10%, используем двустороннюю альтернативу
$\alpha = 0.1$
Тогда найдем z_{95} и z_5
$t_{998,0.05} = -1.6463803454275356$
$t_{998,0.95} = 1.646380345427535$

T = 0.04190924051968376

Критическое значение 0.04190924051968376 не выходит за пределы интервала (-1.6463818766348761, 1.6463818766348755)
Не отвергаем гипотезу H0

Задание 7.

Проверьте гипотезу о равенстве дисперсий двух выборок.

Используйте двустороннюю альтернативу и уровень значимости 10%.
а) сформулируйте основную и альтернативную гипотезы;
б) выпишите расчетную статистику. Почему была выбрана именно такая статистика?
А) Гипотезы:
H0: $\widehat{\sigma x^2} = \widehat{\sigma xy^2}$
H1: $\widehat{\sigma x^2}\ != \widehat{\sigma xy^2}$
Б) Расчетная статистика:

$$T = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}{\frac{1}{m-1}\sum_{i=1}^{m}(Y_i-\bar{Y})^2} \sim F(n-1, m-1)$$

Так как параметры матожидания нам не известны, а F-распределение основано на дополнительных предположениях о независимости и нормальности выборок данных, используем это распределение.

Область, в которой гипотеза H_0 не отвергается: $[F_{n-1,m-1;\alpha/2}; F_{n,m;1-\alpha/2}]$
$F_{n-1,m-1;\alpha/2}$ = 0.8629418796107595
$F_{n,m;1-\frac{\alpha}{2}}$ = 1.1586553765546113

T = 0.944328810560348

Критическое значение 0.944328810560348 не выходит за пределы интервала (0.862941879 6107595, 1.1586553765546113)
Не отвергаем гипотезу H0

Задание 8.

Проверьте гипотезу о равенстве долей единиц в двух выборках.
Используйте двустороннюю альтернативу и уровень значимости 10%.
а) Сформулируйте основную и альтернативную гипотезы.
б) Какие предположения должны быть сделаны о сформированных выборках для проверк и такой гипотезы?

$$\bar{x} = \frac{(\text{кол} - \text{во единиц})}{500}$$
$$\bar{y} = \frac{(\text{кол} - \text{во единиц})}{500}$$

$\bar{x} = 0.506$
$\bar{y} = 0.508$
А) Гипотеза:
H0: px = py
H1: px != py

Б) Предположим, что это распределение Бернулли. Мы должны проверить, что доли единиц равны. Доверительный интервал для p – частный случай доверительного интервала для μ. Выборочное среднее в этом случае есть выборочное отношение $\hat{p}$.

Расчетная статистика:

$$T = \frac{\bar{X} - \bar{Y} - (p_X - p_Y)}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}} \sim N(0; 1)$$

Уровень значимости 10%
Двусторнняя альтернатива

Тогда найдем z5 и z95
Z5 = -1.6448536269514729
Z95 = 1.6448536269514722
T = -0.06325187870752942
Критическое значение -0.06325187870752942 не выходит за пределы интервала (-1.644853 6269514729, 1.6448536269514722)
Не отвергаем гипотезу о равенстве долей единиц