

Задача 0.

А) Напишите фамилию своего семинариста.

Б) Напишите свое ФИО.

Выполнение данной задачи обязательно и является необходимым условием проверки.

Задача 1.

(4 балла) Исследователь оценил зависимость продолжительности жизни Y от концентрации промышленных выбросов в атмосфере X и ежегодных частных расходов на медицинскую помощь Z .

Для 300 жителей индустриальных центров, $\hat{Y}_i = 65.91 - \frac{0.03}{(0.0001)} X_i - \frac{0.036}{(0.019)} Z_i$, $RSS = 300$.

Для 200 сельских жителей, $\hat{Y}_i = 58.4 - \frac{0.017}{(0.006)} X_i - \frac{0.024}{(0.007)} Z_i$, $RSS = 200$.

А также по общей выборке, $\hat{Y}_i = 63.2 - \frac{0.02}{(0.005)} X_i - \frac{0.031}{(0.001)} Z_i$, $RSS = 900$.

В скобках приведены стандартные ошибки.

Можно ли считать, что зависимость едина для городских и сельских жителей?

Ответ обоснуйте подходящим тестом, аккуратно выписав тестируемую гипотезу.

Задача 2.

(5 баллов)

По 29 наблюдениям оценили функцию спроса на яблоки

$$\ln Q = 14 - 6 \ln P_{apple} + 3 \ln P_{orange} + 2 \ln P_{banana},$$

где Q — спрос на яблоки, а P_{apple} — цена яблок, P_{orange} — цена апельсинов и P_{banana} — цена бананов.

Известна оценка ковариационной матрицы коэффициентов регрессии:

$$\widehat{\text{Var}}(\hat{\beta}) = \begin{pmatrix} 1 & 0.1 & -0.2 & 0.3 \\ 0.1 & 2 & 0.5 & 0.7 \\ -0.2 & 0.5 & 3 & 0.6 \\ 0.3 & 0.7 & 0.6 & 4 \end{pmatrix}$$

На уровне значимости 5% проверьте гипотезу о том, что $\beta_{orange} = \beta_{banana}$.

Задача 3.

(6 баллов)

Из предыдущих исследований и откровений внеземного разума известно, что время, проведённое студентом за игрой в Cyberpunk 2077 $Time_i$, зависит от числа экзаменов в ближайшую неделю $Exams_i$ и объёма выпитого кофе $Coffee_i$. Известно также, что $D(Exams_i) = 8, D(Coffee_i) = 20, corr(Exams_i, Coffee_i) = 0.7$.

Совместное распределение переменных хорошо аппроксимируется многомерным нормальным.

Исследователь Стёпчик оценил соответствующую регрессию и получил результат:

$$\widehat{Time}_i = \underset{(0.01)}{20} - \underset{(0.00075)}{3} Exams_i + \underset{(0.57)}{1.2} Coffee_i$$

В скобках указаны Р-значения.

Увидев, что Р-значение коэффициента при $Coffee$ больше уровня значимости в 5%, исследователь Стёпчик решил выкинуть из модели эту переменную и остановиться на простой парной регрессии, оценив которую он получил результат:

$$\widehat{Time}_i = \underset{(0.005)}{24} - \underset{(0.0081)}{2.2} Exams_i$$

а) Известно, что обычно невключение существенных переменных приводит к несостоятельности оценок коэффициентов при оставшихся в модели переменных. Будет ли наблюдаться смещение в нашем случае? Если нет, докажите, если да, найдите величину смещения, зная, что истинная зависимость имеет вид:

$$Time_i = 15 - 2.5Exams_i + 1.5Coffee_i + \epsilon_i$$

б) Стёпчик зачем-то ещё оценил регрессию времени только на объём выпитого кофе. Найдите математическое ожидание коэффициента при переменной $Coffee$.

Задача 4.

(5 баллов) Исследователь Д'Артаньян стандартизировал (центрировал и нормировал) все имеющиеся регрессоры и поместил их в столбцы матрицы \tilde{X} . Выборочная корреляционная матрица регрессоров равна:

$$\begin{pmatrix} 1 & 0.85 & 0 \\ 0.85 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

- Найдите параметр обусловленности (condition number) матрицы $\tilde{X}^T \tilde{X}$.
- Вычислите одну или две главные компоненты, объясняющие не менее 70% суммарной дисперсии стандартизированных регрессоров. Выпишите найденные компоненты как линейные комбинации столбцов матрицы \tilde{X} .