Задание 1.

Для того, что бы выяснить насколько может быть опасен альпинизм собрали данные о 130 альпинистах в возрасте от 14 до 35 лет. На основе собранных данных были оценены две модели (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\textbf{Модель №1:}\ \hat{Y}_i = \underset{(1,9)}{4,3} + \underset{(0,05)}{0,21} * Age_i - \underset{(0,23)}{2,48} * D_i + \underset{(1,60)}{0,80} * V_i + \underset{(0,81)}{0,03} * S_i, \quad R^2 = 0,52$$

$$\textbf{Модель №2:}\ \hat{Y}_i = \underset{(1,2)}{5,4} + \underset{(0,02)}{0,18} * Age_i - \underset{(0,19)}{2,50} * D_i, \quad R^2 = 0,500$$

Y — состояние здоровья альпиниста, измеренное по специальной 10-балльной шкале (1 — очень плохо, 10 — очень хорошо); Age — возраст альпиниста; D — количество лет занятий альпинизмом; V – количество покоренных вершин; S – количество тренировок в неделю.

Какая из двух моделей является более удачной? Помогите исследователям в выборе наилучшей модели, используя тест «короткая»-«длинная» регрессия и пятипроцентный уровень значимости.

Выберите один ответ:

a. Расчетное значение составит 2,604 и следует выбрать модель №2

b. Расчетное значение составит 2,604 и следует выбрать модель №1

c. Расчетное значение составит 33,85 и следует выбрать модель №2

d. Расчетное значение составит 33,85 и следует выбрать модель №1

Задание 2. Менеджер чебуречной не уверен в правильности выбранной цены, поэтому в течение нескольких недель он меняет ее и записывает количество проданных чебуреков. Для изучения зависимости объема продаж от цены менеджер использует следующую модель: $q_t = \beta_1 + \beta_2 p_t + \varepsilon_t$. В таблице представлены данные о наблюдениях за продажами в течение 5 недель (q_t — количество проданных чебуреков, p_t — цена одного чебурека):

q	p
20	8
71	3
69	3
10	9
30	7

Вопрос 2.1. Найдите МНК-оценку коэффициента $\widehat{\beta_2}$.

Вопрос 2.2. Найдите МНК-оценку коэффициента $\widehat{\beta_1}$.

Вопрос 2.3. Менеджера интересует, сколько чебуреков будет продано по цене $p_6 = 6$. Используя оцененную вами модель, постройте прогноз объема продаж по этой цене.

Задание 3.

По 30 наблюдениям было оценено следующее уравнение регрессии (в скобках указаны стандартные отклонения оценок коэффициентов):

$$y_i = \underset{(1,0)}{1,5} - \underset{(0,4)}{0,9}\, x_i^{(1)} + \underset{(0,01)}{0,04}\, x_i^{(2)} + \underset{(0,02)}{0,09}\, x_i^{(3)} + \underset{(0,6)}{2,0}\, x_i^{(4)}. \qquad R^2 = 0.59$$

Проверьте (при уровне значимости 5%) гипотезу о том, что все коэффициенты при переменных уравнения одновременно равны нулю.

Выберите один ответ:

a. Расчетная статистика составит 8,99 и следует принять тестируемую гипотезу

b. Расчетная статистика составит 8,99 и следует отклонить тестируемую гипотезу

c. Расчетная статистика составит 0,23 и следует принять тестируемую гипотезу

d. Расчетная статистика составит 0,23 и следует отклонить тестируемую гипотезу

Задание 4.

На основе 10 наблюдений была оценена следующая модель:

$$\hat{y}_i = \underset{(1,911)}{5,6} + \underset{(0,018)}{0,56}\, x_i^{(2)} - \underset{(21,756)}{23}\, x_i^{(3)}$$

На однопроцентном уровне значимости проверьте гипотезу о том, что коэффициент при переменной x(2) равен 0,5.

Выберите один ответ:

a. Расчетное значение составит 3,33 и следует отклонить тестируемую гипотезу

b. Расчетное значение составит 31,11 и следует принять тестируемую гипотезу

c. Расчетное значение составит 31,11 и следует отклонить тестируемую гипотезу

d. Расчетное значение составит 3,33 и следует принять тестируемую гипотезу

Задание 5.

В Тридевятом королевстве местный ученый обеспокоился популяцией зайцев, которая неуклонно снижалась ввиду того, что король очень любил охотится на зайцев. Чтобы продемонстрировать Его Величеству пагубность такого пристрастия, на основании 45 наблюдений ученый оценил следующую зависимость (в скобках указаны стандартные ошибки оценок коэффициентов):

$$\hat{y}_i = \underset{(9,01)}{5,9} + \underset{(0,87)}{6,7}\, x_i^{(1)} + \underset{(3,23)}{8,7}\, x_i^{(2)} + \underset{(5,34)}{3,2}\, x_i^{(3)}$$

Где y_i – количество убитых на королевской охоте зайцев, $x_i^{(1)}$ - число королевских охотников, $x_i^{(2)}$ - количество борзых, участвующих в травле, $x_i^{(3)}$ - среднее число лет участвующих в охоте собак.

Но у него возникли подозрения, что в такой модели может быть мультиколлинеарность. Для того, чтобы это выяснить, он решил оценить вспомогательные регрессии одного ресрессора на другие, в результате чего получил следующие результаты:

$$x_i^{(1)} = 5{,}0 + 4{,}5x_i^{(2)} + 3{,}2x_i^{(3)}, \quad R^2 = 0{,}8$$

$$x_i^{(2)} = 1{,}1 + 2{,}7x_i^{(1)} + 0{,}6x_i^{(3)}, \quad R^2 = 0{,}75$$

$$x_i^{(3)} = 3{,}2 + 0{,}9x_i^{(1)} + 5{,}6x_i^{(2)}, \quad R^2 = 0{,}85$$

Можно ли на основании этих данных утверждать, что в модели действительно присутствует мультиколлинеарность?

Выберите один ответ:

a. Да, можно, так как хотя бы в одной из вспомогательных регрессий VIF больше 10.

b. Нет, нельзя, так как VIF для всех вспомогательных регрессий меньше 10.

c. Да, можно, так как высокие значения коэффициентов детерминации в вспомогательных регрессиях это подтверждают

d. В условии отсутствуют коэффициенты корреляции между переменными, следовательно мы не можем сделать вывод о наличии/отсутствии мультиколлинеарности

Задание 6.

Фармацевтическая компания изучает эффективность нового лекарства от горной болезни, с которой сталкиваются люди, оказавшись на большой высоте. Компания собрала данные о 130 альпинистах в возрасте от 18 до 35 лет, часть из которых, находясь на высоте, получали новое лекарство, а часть — нет. Альпинисты, получающие лекарство, выбирались случайным образом. На основе собранных данных были оценены две модели (в скобках указаны робастные стандартные ошибки оценок коэффициентов):

Модель №1: $$\hat{Y}_i = \underset{(1{,}9)}{2{,}3} + \underset{(0{,}05)}{0{,}18} * Age_i + \underset{(1{,}33)}{2{,}48} * D_i + \underset{(1{,}60)}{0{,}13} * D_i * Age_i, \quad R^2 = 0{,}504$$

Модель №2: $$\hat{Y}_i = \underset{(1{,}2)}{3{,}4} + \underset{(0{,}02)}{0{,}20} * Age_i, \quad R^2 = 0{,}485$$

Y — состояние здоровья альпиниста, измеренное по специальной 10-балльной шкале (1 — очень плохо, 10 — очень хорошо); Age — возраст альпиниста; D — фиктивная переменная, равная единице, если альпинист получил лекарство, и равная нулю в противном случае.

Действительно ли лекарство оказывает влияние на здоровье альпиниста (на пятипроцентном уровне значимости)?

Выберите один ответ:

a. Да, оказывает. Хотя бы одна из добавленных переменных оказывает значимое влияние

b. Да, оказывает. При учете в модели фиктивной переменной R^2 вырос.

c. Нет, не оказывает. Все добавленные переменные влияют незначимо

Задание 7.

Имеются данные о 150 абитуриентах, сдававших вступительный экзамен в магистратуру некоторого экономического факультета:

Y — количество баллов за вступительный экзамен по экономической теории.

D — фиктивная переменная равная единице, если соответствующий абитуриент посещал подготовительные курсы для поступающих, и равная нулю в противном случае.

EF — фиктивная переменная равная единице, если соответствующий абитуриент является выпускником бакалавриата данного экономического факультета, и равная нулю в противном случае.

Используя эти данные, исследователь оценил параметры линейной регрессионной модели:

$$\hat{Y}_i = \underset{(0,1)}{20} + \underset{(4,5)}{30}\,EF_i - \underset{(1,3)}{10}\,D_i + \underset{(1,4)}{15}\,D_i * EF_i$$

Вопрос 1 В соответствии с полученными результатами, определите, какое количество баллов в среднем получает абитуриент, который не заканчивал бакалавриат данного экономического факультета и не посещал курсы?

Вопрос 2. В соответствии с полученными результатами, определите, какое количество баллов в среднем получает абитуриент, который заканчивал бакалавриат данного экономического факультета и не посещал курсы?

Вопрос 3. В соответствии с полученными результатами, определите, какое количество баллов в среднем получает абитуриент, который заканчивал бакалавриат данного экономического факультета и посещал курсы?

Задание 8.

. Оценивание модели $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \varepsilon_i$ методом наименьших квадратов по 30 наблюдениям дало следующие результаты (в скобках указаны стандартные отклонения оценок коэффициентов):

Модель №1: $\hat{y}_i = \underset{(3,1)}{9,1} - \underset{(2,1)}{10,2}\,x_i^{(2)} + \underset{(3,0)}{4,5}\,x_i^{(3)}$,
сумма квадратов остатков 790, общая сумма квадратов TSS=1000

Оценивание модели $y_i = \beta_1 + \beta_2 x_i^{(2)} + \beta_3 x_i^{(3)} + \beta_4 x_i^{(4)} + \beta_5 x_i^{(5)} + \varepsilon_i$ методом наименьших квадратов по тем же самым 30 наблюдениям дало следующие результаты:

Модель №2: $\hat{y}_i = 8,2 - 8,5x_i^{(2)} + 9,0x_i^{(3)} + 5,0x_i^{(4)} + 6,1x_i^{(5)}$, $R^2 = 0,213$

Вопрос 3.1. Для модели №1 проверьте значимость уравнения в целом при уровне значимости 5%.

Вопрос 2. Сравните модели №1 и №2, используя в качестве критерия подходящий статистический тест. Какую из моделей следует предпочесть в соответствии с результатами теста?

Вопрос .3. Сравните модели №1 и №2, используя в качестве критерия скорректированный коэффициент детерминации R^2_{ADJ}. Какую из моделей следует предпочесть в соответствии с этим критерием?