Контрольная работа №1.

7 ноября 2023 г.

Задача 1. (50 баллов) В вашем распоряжении имеются следующие данные о 229 работниках (111 мужчин и 118 женщин): $lnEARNINGS$ — логарифм текущего часового заработка в долларах США, S — продолжительность обучения (число полных лет обучения), EXP — общий стаж работы после окончания учебы, $FEMALE$ — пол респондента (0 для мужчин, 1 для женщин). Ваша цель состоит в том, чтобы выявить влияние опыта работы и образования на доход индивида. Для этого вы оцениваете модель регрессии:

$$lnEARNINGS_i = \beta_1 + \beta_2 EXP_i + \beta_3 S_i + \beta_4 FEMALE_i + \beta_5 EXP_i * FEMALE_i + \varepsilon_i, i = 1,...,229.$$

Ниже представлены результаты оценивания модели:

Source	Sum of Squares	Degrees of Freedom
Explained	13.0563901	???A
Residual	???B	224
Total	57.617198	???C

Number of obs	= 229
F(4, ???D)	= ???E
Prob > F	= 0.0000
R-squared	= ???F
Adj R-squared	= ???G
Standard error	= ???H

lnearnings	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
exp	???I	.0166437	3.55	0.000	.0262485	.0918449
s	.0962928	.0132887	7.25	???J	.0701059	.1224797
female	-.2069607	.1627401	???K	NA	???L	???M
exp*female	-.0086271	???N	???O	0.687	NA	NA
const	???P	NA	NA	NA	.6054517	1.67635

NA — скрытые значения, которые **не нужно восстанавливать**.

Матрица $(X'X)^{-1}$ имеет следующий вид:

	const	exp	s	female	expfemale
const	.37113295				
exp	-.01718743	.00139249			
s	-.01636512	.00048446	.00088769		
female	-.05522604	.00783554	-.00077057	.13313258	
expfemale	.00762025	-.00110927	.00003449	-.01629641	.00230557

(а) (19) Заполните отмеченные знаками «???» пропуски в таблице (A–P). Обоснуйте ответ. При расчетах используйте значения из таблицы, округленные до тысячных.

(б) (3) Дайте содержательную интерпретацию коэффициента β_3 при переменной S.

(в) (3) Дайте содержательную интерпретацию коэффициента β_2 при переменной EXP.

(г) (5) Проверьте гипотезу о значимости коэффициента β_4 для переменной $FEMALE$.

(д) (10) Проверьте гипотезу о том, что отдача от образования у мужчин в два раза выше отдачи от опыта работы, который не различается в зависимости от пола.

(е) (10) Постройте 95% доверительный интервал для $2\beta_2 - \beta_4$.

Задача 2. (20 баллов) В программе исследований k разных удобрений, предназначенных для повышения урожайности огурцов, использованы в опытах на $n = n_1 + ... + n_k$ опытных участках. Удобрение номер s ($s = 1,...,k$) использовалось на n_s опытных участках. Для изучения влияния удобрений использовалась регрессионная модель:

$$Y_i = \beta_1 D_{1i} + ... + \beta_k D_{ik} + \varepsilon_i, i = 1,...,n.$$

Здесь Y — урожайность, D_k — фиктивная переменная, равная 1 для участка номер k и 0 в других случаях. Известны выборочные средние $\bar{Y}_s$ и стандартные отклонения $\hat{\sigma}_s$:

$$\bar{Y}_s = \frac{1}{n_s} \sum_{D_{si}=1} Y_i, \hat{\sigma}_s = \frac{1}{n_s - 1} \sum_{D_{si}=1} (Y_i - \widehat{Y}_s)^2.$$

Покажите, что F-статистика для тестирования гипотезы о равном влиянии всех удобрений ($\beta_1 = \beta_2 = ... = \beta_k$) выражается через известные величины следующим образом:

$$F = \frac{\sum_{s=1}^{k} n_s (\bar{Y}_s - \bar{Y})^2}{\sum_{s=1}^{k} (n_s - 1)\hat{\sigma}_s^2} \cdot \frac{n-k}{k},$$

где $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

Задача 3. (30 баллов) Рассмотрим модель регрессии, которая включает в себя два набора объясняющих переменных X_1 и X_2. Тогда

$$y = X\beta = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Здесь y — вектор размерности $n \times 1$, X_1 — матрица размерности $n \times k_1$, X_2 — матрица размерности $n \times k_2$, β_1 — вектор размерности $k_1 \times 1$, β_2 — вектор размерности $k_2 \times 1$, ε — вектор размерности $n \times 1$.

(а) (15 баллов) Покажите, что справедлива следующая формула:

$$\hat{\beta}_2 = (X_2' M_1 X_2)^{-1}(X_2' M_1 y),$$

где $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$.

(б) (15 баллов) Пусть теперь мы хотим построить регрессию y на набор k переменных, записанных в матрицу X, и на константу. Для того чтобы вычислить МНК-оценку вектора коэффициентов при X, мы можем реализовать следующий алгоритм оценивания:

(1) преобразовать вектор y, перейдя к отклонениям от среднего, то есть к $y - \bar{y}$;

(2) преобразовать каждый столбец матрицы X, перейдя к отклонениям от среднего значения соответствующего столбца;

(3) оценить регрессию преобразованного вектора y в соответствии с (1) на преобразованную матрицу X из (2) без константы.

Даст ли выше описанный алгоритм оценивания тот же результат для вектора коэффициентов, если теперь мы преобразуем только матрицу X в исходной модели с константой?

Подсказка: для решения пункта (б) используйте результат, описанный в пункте (а).