Фамилия, имя, номер группы:

. . . .

1. Исследователь рассматривает уравнение зависимости расходов на питание (W) от доходов (Income), с учетом сезона. Переменная сезон (S) принимает следующие значения: 1 – зима, 2 – весна, 3 – лето и 4 – осень. Исследователь предполагает, что в каждый сезон может выполняться своя линейная зависимость.

   а) (2 балла) Выпишите уравнение оцениваемой модели. Укажите смысл всех включенных в модель переменных.

   б) (2 балла) Как проверить гипотезу о единой линейной зависимости расходов на питание для всех сезонов? Выпишите аккуратно основную и альтернативную гипотезы, формулу расчета статистики и способ проверки.

2. Начинающий исследователь Елисей исследует зависимость успехов в учёбе своих однокурсников, G_i, от времени, которое они тратят на учёбу, T_i. По выборке из 100 человек он смог оценить следующую регрессию:

$$\hat{G}_i = 30 + 6T_i$$

   Елисей был бы рад полученному результату, но тут на лекции по эконометрике ему рассказали про эндогенность и пропущенные переменные, и он решил, что в его модели эти проблемы точно есть. Изучив литературу, он узнал, что на успехи в учёбе кроме времени влияют ещё и способности студента, A_i, при этом способности коррелированы со временем, которое студент тратит на учёбу.

   а) Проверьте, является ли найденная Елисеем оценка коэффициента при времени состоятельной;

   б) Если оценка не состоятельна, то предложите способ получения состоятельной оценки;

   в) Найдите асимптотическую величину смещения оценки, если $\mathbb{C}\mathrm{ov}(G_i, A_i) = 6$, $\mathbb{C}\mathrm{ov}(T_i, A_i) = 4$, $\mathbb{V}\mathrm{ar}(G_i) = 16$, $\mathbb{V}\mathrm{ar}(A_i) = 100$, $\mathbb{V}\mathrm{ar}(T_i) = 49$.

3. По 24 наблюдениям была оценена модель:

$$\widehat{Y}_i = 15 - 4Z_i + 3W_i$$

   Известно, что случайные ошибки нормально распределены, $RSS = 180$, и

$$(X'X)^{-1} = \begin{pmatrix} 0.216 & -0.112 & -0.075 \\ -0.112 & 0.119 & 0.021 \\ -0.075 & 0.021 & 0.047 \end{pmatrix}$$

   а) (1 балл) Проверьте гипотезу $H_0 : \beta_Z = 0$ против $H_a : \beta_Z \neq 0$ на уровне значимости 5%.

   б) (3 балла) Проверьте гипотезу $H_0 : \beta_Z + \beta_W = 0$ против $H_a : \beta_Z + \beta_W \neq 0$ на уровне значимости 5%.

   в) (2 балла) Выпишите использованные при проверке гипотез предпосылки о случайных ошибках модели.

4. Исследовательница Глафира изучает зависимость спроса на молоко от цены молока и дохода семьи. В её распоряжении есть следующие переменные:

- $price$ — цена молока в рублях за литр
- $income$ — ежемесячный доход семьи в тысячах рублей
- $milk$ — расходы семьи на молоко за последние семь дней в рублях

В данных указано, проживает ли семья в сельской или городской местности. Поэтому Глафира оценила три регрессии: (All) — по всем данным, (Urban) — по городским семьям, (Rural) — по сельским семьям.

	(All)	(Urban)	(Rural)
(Intercept)	-1.765	-4.059	-0.155
	(4.943)	(6.601)	(7.812)
income	0.308^{***}	0.341^{***}	0.281^{***}
	(0.052)	(0.072)	(0.079)
price	-0.383^{*}	-0.352	-0.391
	(0.161)	(0.253)	(0.221)
R-squared	0.304	0.356	0.273
adj. R-squared	0.290	0.325	0.245
sigma	4.912	4.857	5.036
F	21.216	11.593	9.744
P-value	0.000	0.000	0.000
RSS	2340.080	990.839	1318.741
n observations	100	45	55

а) (1 балл) Проверьте значимость в целом регрессии (All) на 5%-ом уровне значимости.

б) (2 балла) На 5%-ом уровне значимости проверьте гипотезу, что зависимость спроса на молоко является единой для городской и сельской местности.

5. Рассмотрим стационарный процесс, удовлетворяющий уравнению $Y_t = 1 + 0.6Y_{t-1} + u_t - 0.3u_{t-1}$, где u_t — белый шум с $u_t \sim \mathcal{N}(0; 4)$.

   а) Найдите $\mathbb{E}(Y_t)$, $\mathbb{V}\mathrm{ar}(Y_t)$.

   б) Найдите первые два значения автокорреляционной и частной автокорреляционной функций.

   в) Постройте 95% интервальный прогноз для Y_{101}, если известно, что $Y_{100} = 2$, $u_{100} = 1$.

6. По 200 наблюдениям исследователь Иннокентий оценил модель логистической регрессии для вероятности сдать экзамен по метрике:

$$\hat{\mathbb{P}}(Y_i = 1) = \Lambda(1.5 + 0.3X_i - 0.4D_i),$$

   где Y_i — бинарная переменная равная 1, если студент сдал экзамен; X_i — количество часов подготовки студента; D_i — бинарная переменная равная 1, если студент пробовал пиццу «четыре сыра» в новой столовой.

   Оценка ковариационной матрицы оценок коэффициентов имеет вид:

$$\begin{pmatrix} 0.04 & -0.01 & 0 \\ -0.01 & 0.01 & 0 \\ 0 & 0 & 0.09 \end{pmatrix}$$

   а) Проверьте гипотезу о том, что количество часов подготовки не влияет на вероятность сдать экзамен.

   б) Посчитайте предельный эффект увеличения каждого регрессора на вероятность сдать экзамен для студента не пробовавшего пиццу и готовившегося 24 часа. Кратко, одной-двумя фразами, прокомментируйте смысл полученных цифр.

   в) При каком значении D_i предельный эффект увеличения X_i на вероятность сдать экзамен максимален, если $X_i = 20$?