1. [10] In a white room with soft walls Michael has found four vectors from $\mathbb{R}^n$: a, b, c and d. He discovered that $a + 2b = 3c + 6d$ and $\|a\| = \|c\| = \|d\| = 3$ with scalar product $\langle a, c\rangle = 1$. The vector d is orthogonal to a and c.

   Michael's identities would like to estimate two different regressions using OLS: $\hat{b} = \hat{\beta}_1 a + \hat{\beta}_2 c$ and $\hat{a} = \hat{\gamma} c$.

   (a) [5] Provide estimates of coefficients where possible.

   (b) [5] Calculate sum of squared residuals $\mathrm{SS}^{\mathrm{res}}$ and total sum of squares SST where possible.

2. [10] Consider the model $y = X\beta + u$ where β is non-random, $\mathbb{E}(u \mid X) = m(X) \neq 0$. The matrix X of size $n \times k$ has rank $X = k$ and $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$. Let $\hat{\beta}$ be the standard OLS estimator of β.

   (a) [3] Find $\mathbb{E}(\hat{\beta} \mid X)$. Is it possible that $\hat{\beta}$ is unconditionally unbiased?

   (b) [3] Find $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X)$.

   (c) [1] Will the default confidence interval for β be valid in this case? Explain shortly why.

   (d) [3] Find $\mathbb{C}\mathrm{ov}(\hat{u}, \hat{\beta} \mid X)$.

3. [10] The whole dataset of $n = 600$ observations is split into three parts. Donald Trump estimated the regression $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 w_i$ on these three parts separately and on the whole dataset.

   He has obtained $\mathrm{SS}_1^{\mathrm{res}} = 100$, $\mathrm{SS}_2^{\mathrm{res}} = 200$, $\mathrm{SS}_3^{\mathrm{res}} = 300$ correspondingly and for the whole dataset $\mathrm{SS}^{\mathrm{res}} = 650$ and $\mathrm{SST} = 800$.

   (a) [4] Test H_0: $\beta_1 = 0$ and $\beta_2 = 0$ on the whole dataset against H_1: $\beta_1 \neq 0$ or $\beta_2 \neq 0$.

   (b) [6] Test H_0 that the linear model is the same on the whole dataset against three different linear models.

   You are free to use these 5% critical values: $F_{1,597} = 3.9$, $F_{2,597} = 3.0$, $F_{3,597} = 2.6$, $F_{4,597} = 2.4$, $F_{5,597} = 2.2$, $F_{6,597} = 2.1$, $F_{1,591} = 3.9$, $F_{2,591} = 3.0$, $F_{3,591} = 2.6$, $F_{4,591} = 2.4$, $F_{5,591} = 2.2$, $F_{6,591} = 2.1$.

4. [10] The true model is $y_i = \beta_0 + \beta_1 x_i + u_i$ with $\mathbb{E}(u \mid x) = m(x)$, $\mathbb{V}\text{ar}(u \mid x) = \sigma^2 I$. Observations are independent. Winnie-the-Pooh observes y, x and a strange variable z such that $\mathbb{C}\text{ov}(x_i, z_i) \neq 0$, but $\mathbb{C}\text{ov}(z_i, u_i) = 0$.

   Consider regression A: $\hat{x}_i = \hat{\gamma}_0 + \hat{\gamma}_1 z_i$ and regression B: $\hat{y}_i = \hat{\delta}_0 + \hat{\delta}_1 z_i$.

   (a) [3] Find plim $\hat{\gamma}_1$ in terms of $\mathbb{C}\text{ov}(x_i, z_i)$ and $\mathbb{V}\text{ar}(z_i)$.

   (b) [5] Find plim $\hat{\delta}_1$ in terms of β_1, $\mathbb{C}\text{ov}(x_i, z_i)$ and $\mathbb{V}\text{ar}(z_i)$.

   (c) [2] Construct a consistent estimator of β_1 using $\hat{\gamma}_1$ and $\hat{\delta}_1$.

5. [10] Consider the following joint distribution of the regressor x_i and random error u_i:

	$x_i = -1$	$x_i = 0$	$x_i = 1$
$u_i = -1$	0.2	0.2	0.1
$u_i = 1$	0.1	0.2	0.2

   (a) [4] Find $\mathbb{E}(u_i \mid x_i)$, $\mathbb{V}\text{ar}(u_i \mid x_i)$.

   (b) [3] Find $\mathbb{E}(u_i)$, $\mathbb{C}\text{ov}(x_i, u_i)$.

   (c) [1] Which Gauss — Markov assumptions are violated?

   (d) [2] Is the OLS estimator $\hat{\beta}_1$ in the model $y_i = \beta_0 + \beta_1 x_i + u_i$ conditionally unbiased?

6. [10] (from LSE past exams) SAT-test (Scholastic Assessment Test) is used for colledge admissions in the US. Consider the following regression of `sat` (SAT-test score):

$$\widehat{\text{sat}}_i = \underset{(6.3)}{1028} + \underset{(3.8)}{19.3}\text{hsize}_i - \underset{(0.5)}{2.2}\,\text{hsize}_i^2 - \underset{(4.3)}{45}\,\text{fem}_i - \underset{(13)}{170}\text{black}_i + \underset{(13)}{62}\,\text{fem}_i\text{black}_i$$

   Here `fem` is a dummy variable equal to 1 for females and 0 for males and `black` is a race dummy variable equal to 1 for black and 0 otherwise and `hsize` is the class size. Standard errors are supplied in brackets.

   (a) [2] Why is it reasonable to include hsize^2?

   (b) [2] For which class size the SAT-score is maximal ceteris paribus?

   (c) [2] Let's fix `hsize`. Estimate the SAT-score difference between non-black females and non-black males. Is this difference statistically significant?

   (d) [2] Let's fix `hsize` once again. Estimate the SAT-score difference between black females and non-black females. Which information would you need to test statistical significance of this difference?

   (e) [2] Describe the problem that you will encounter during estimation if all females in your sample are black.