1. James Bond estimates the model $y_i = \beta x_i + u_i$ with OLS. We know that u_i are normal, $\mathcal{N}(0;\sigma^2)$ and independent. We also know that $x_1 = 1$ and $x_i = 2$ for $i \geq 2$.

   (a) Find expected value and variance of $\hat{\beta}$.

   (b) Is $\hat{\beta}$ efficient amond all linear unbiased estimators? Why or why not?

   (c) Calculate probability limit of $\hat{\beta}$. Is $\hat{\beta}$ consistent?

2. James Bond uses robust to heteroskedasticity White standard errors in the case of homoskedastic model $y_i = \beta x_i + u_i$, where $u_i \sim \mathcal{N}(0;\sigma^2)$ and u_i are independent. We also know that $x_1 = 1$ and $x_i = 2$ for $i \geq 2$.

   (a) Calculate $se_{HC}(\hat{\beta})$;

   (b) Will $\hat{\beta}$ remain unbiased, efficient among all linear unbiased estimators? Why or why not?

   (c) Is it correct to test significance of coefficients using t-distribution? How many degrees of freedom for t distribution should James use? If it is not correct then what should James do?

3. James Bond assumes that observations y_i are independent and probability distribution for y_i is given by the table

y_i	$y_i = 1$	$y_i = 2$	$y_i = 3$
probability	$2a$	$3a$	$1 - 5a$

   James Bond has collected 400 observations: 100 observations are equal to 1, 100 observations equal to 2, and 200 observations equal to 3.

   (a) Estimate unknown a using maximum likelihood;

   (b) Construct 95% confidence interval for a.

4. Using 400 observation James Bond estimated the logistic regression for probability of passing econometrics exam. Predictors are: x_i — number of study hours and d_i — dummy variable for having at least one dinner in nearby Khatchapuri restaurant:

$$\hat{\mathbb{P}}(\text{pass}) = \Lambda(0.15 + 0.03x_i - 0.2d_i),$$

   The estimate of coefficient covariance matrix is:

$$\begin{pmatrix} 0.0004 & -0.0001 & 0 \\ -0.0001 & 0.0001 & 0 \\ 0 & 0 & 0.0009 \end{pmatrix}$$

   (a) Test whether the number of study hours influences the probability of passing the exam.

   (b) Construct 95% confidence interval for the difference of probabilities of passing the exam for two students. We know that both students have studied 20 hours, but only one of the students had a dinner in Khatchapuri restaurant.

   (c) For which value of d_i the marginal effect of increasing x_i is bigger for $x_i = 20$?

5. James Bond would like to check whether the econometrics final exam score, y_i, varies for students that had at least one dinner at Khatchapuri and those who had not. Once again x_i is the number of study hours.

   He estimates the same model three datasets:

dataset	equation	RSS	observations
all students	$\hat{y}_i = 40 + 2x_i$	7000	400
at least one Khatchapuri dinner	$\hat{y}_i = 50 + 3x_i$	3000	200
no Khatchapuri dinner	$\hat{y}_i = 37 + 1.8x_i$	2000	200

   (a) Assume d_i is those who have visited Khatchapuri at least once. What estimates will John obtain in the regression
   $$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 d_i + \hat{\beta}_4 d_i x_i$$

   (b) Using appropriate test check the econometrics final exam score varies for two type of students.

   $F_{a,b}$ distribution 5% critical values are given below:

	$b = 100$	$b = 200$	$b = 500$
$a = 1$	3.94	3.89	3.86
$a = 2$	3.09	3.04	3.01
$a = 3$	2.70	2.65	2.62
$a = 200$	1.34	1.26	1.21

6. The last problem :)

   James Bond would like to estimate the model $y_i = \beta_1 + \beta_2 x_i + u_i$, where y_i — is the econometrics final exam result, x_i is the number of study hours.

   The problem is that James has only data on self-reported $x_i^* = x_i \cdot f_i$ where f_i is the famous Russian exageration factor. We know that $x_i \sim \mathcal{N}(20; 9)$, $u_i \sim \mathcal{N}(0; 1)$, $f_i \sim \mathcal{N}(2; 4)$. Variables (x_i), (u_i) and (f_i) are independent.

   James Bond uses the OLS regression $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i^*$.

   (a) Find probability limit of $\hat{\beta}_2$. Is $\hat{\beta}_2$ consistent?

   (b) Briefly describe the possible way to obtain consistent estimator.