Подготовка к экзамену.

1. В задаче используются данные (Mroz, 1975). Пуассоновская регрессия для моделирования количества детей в семье:

$$P(Nkids = k_i) = e^{\lambda_i} \frac{\lambda_i^{k_i}}{k_i!},$$

где $\lambda_i = exp(\beta_1 + \beta_2 AGE_i + \beta_3 AGE_i^2 + \beta_4 WE_i + \beta_5 INCOME_i)$,
k – количество детей в семье,
AGE – возраст женщины (в годах),
$AGE2$ – квадрат возраста женщины,
WE – образование женщины (в годах),
$INCOME$ – доход семьи в \$10000.

Ниже в таблице приведены результаты оценивания меотодом максимального правдоподобия.

Poisson Regression	
Dependent variable	NKIDS
Number of observations	753
Iterations completed	7
Log likelihood function	-1083.397
Number of parameters	5
Restricted log likelihood	-1279.522

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
Constant	-7.64180956	1.14268278	-6.688	.0000	
AGE	.49624655	.05663388	8.762	.0000	42.5378486
AGE2	-.00686403	.00069963	-9.811	.0000	1874.54847
WE	-.03430021	.01448182	-2.369	.0179	12.2868526
INCOME	.01193400	.02569902	.464	.6424	2.30805950

Matrix Cov.Mat. has 5 rows and 5 columns.

	1	2	3	4	5
1	1.30572	-.06373	.00078	-.00319	.00284
2	-.06373	.00321	-.3948059D-04	.3794861D-04	-.00012
3	.00078	-.3948059D-04	.4894781D-06	-.3460068D-06	.1216040D-05
4	-.00319	.3794861D-04	-.3460068D-06	.00021	-.00014
5	.00284	-.00012	.1216040D-05	-.00014	.00066

(a) Оцените эффект увеличения возраста на 1 год на среднее (expected) количество детей.

(b) Покажите, что выборочное среднее оценок $\hat{\lambda}_i$ равно выборочному среднему k_i.

(c) Протестируйте на 5% уровне значимости гипотезу о совместной незначимости всех регрессоров $AGE, AGE^2, WE, INCOME$ при помощи теста отношения правдоподобия (LR-тест).

(d) Укажите ограничения Пуассоновской регрессии. Какие модели Вы можете предложить для преодоления этих ограничений.

2. Рассмотрим следующую модель для панельных данных

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + u_{it}.$$

Обозначим $\bar{y}_i = \frac{1}{T}\sum\limits_{t=1}^{T} y_{it}$, $\varepsilon_{it} = \alpha_i + u_{it}$.

Рассмотрим следующее преобразование модели

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{it} - \lambda\bar{x}_i) + (\varepsilon_{it} - \lambda\bar{\varepsilon}_i).$$

(a) Какие модели получатся при $\lambda = 0$ и $\lambda = 1$?

(b) Пусть $\alpha_i \sim i.i.d.(0, \sigma_\alpha^2)$, $u_{it} \sim i.i.d.(0, \sigma_u^2)$, $Cov(\alpha_i, u_{jt}) = 0$ для всех i и j. Определим $\lambda = 1 - [\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}]^{1/2}$. Покажите, что $\varepsilon_{it} - \lambda\bar{\varepsilon}_i$ имеют нулевое математическое ожидание, постоянную дисперсию и серийно некоррелированы.

3. Рассмотрим модель с фиксированными эффектами:

$$y_{it} = \alpha_i + u_{it}.$$

Случайные ошибки u_{it} предполагаются независимыми по i и по t, но гетероскедастичными, то есть $V(u_{it}) = \sigma_i^2$. Панельные данные не являются сбалансированными, для каждой индустрии наблюдается разное число наблюдений T_i

(a) Покажите, что OLS– и GLS–оценки для α_i совпадают.

(b) Пусть $\sigma^2 = \sum_{i=1}^{N} T_i \frac{\sigma_i^2}{n}$, где $n = \sum_{i=1}^{N} T_i$. Покажите, что OLS-оценка для σ^2 является смещенной. Покажите так же, что смещение пропадет, если панельные данные будут сбалансированны, а ошибки u_{it} – гомоскедастичны, т.е. $V(u_{it}) = \sigma^2$.