Темы, выносимые на кр.3 и задачи к ним

1. Мультиколлинеарность данных и способы борьбы с ней

Мультиколлинеарность данных. Идеальная и практическая мультиколлинеарность (квазимультиколлинеарность). Теоретические последствия мультиколлинеарности для оценок параметров регрессионной модели. Нестабильность оценок параметров регрессии и их дисперсий при малых изменениях исходных данных в случае мультиколлинеарности. Признаки наличия мультиколлинеарности. Показатели степени мультиколлинеарности. Вспомогательные регрессии и показатель "вздутия" дисперсии (VIF). Индекс обусловленности информационной матрицы (CI) как показатель степени мультиколлинеарности. Полиномиальная регрессия. Методы борьбы с мультиколлинеарностью. Переспецификация модели (функциональные преобразования переменных). Исключение объясняющей переменной, линейно связанной с остальными. Понятие о методе главных компонент как средстве борьбы с мультиколлинеарностью данных. Понятие о методе LASSO.

Задачи

1) Признаком мультиколлинеарности служит:

1. маленькие t–статистики при R2, близком к 1
2. близкое к 0 значение коэффициента множественной детерминации
3. значительные изменения в оценках коэффициентов регрессии при небольших изменениях в данных
4. близкие к 0 значения коэффициентов корреляции регрессоров
5. все ответы верны

2) Оцененная с помощью МНК зависимость заработной платы индивида EARNINGS от его возраста AGE, опыта EXP, пола MALE, длительности обучения S, длительности обучения матери SM имеет вид (в скобках стандартные отклонения коэффициентов):

$$EA\hat{R}N = -\underset{(10.6)}{24} - \underset{(0.25)}{0.099}\,AGE + \underset{(0.249)}{2.49}\,S + \underset{(0.24)}{0.26}\,SM + \underset{(0.14)}{0.46}\,EXP + \underset{(1.11)}{6.23}\,MALE,\ R^2 = 0.247$$

Были оценены также вспомогательные регрессии:

$$A\hat{G}E = -.007 + 0.53AGE - 0.6S + 0.23SM + 1.23MALE,\ R^2 = 0.2,$$

$$\hat{S} = 8.47 + 0.095AGE + 0.4SM - 0.2EXP + 0.12MALE,\ R^2 = 0.25,$$

$$S\hat{M} = 6.16 - 0.045AGE + 0.42S + 0.08EXP + 0.42MALE,\ R^2 = 0.18,$$

$$E\hat{X}P = -.07 + 0.53AGE - 0.6S + 0.23SM + 1.23MALE,\ R^2 = 0.2,$$

VIF для переменной EXP равен ___.

Ответ. 1.25

3) При применении к модели, результаты оценки которой приведены ниже,

EARNINGS	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
S	2.578227	.2288185	11.27	0.000	2.128729	3.027726
AGE	-10.70493	9.211662	-1.16	0.246	-28.80062	7.390769
Agesq	.1300605	.1125515	1.16	0.248	-.0910395	.3511605
EXP	.4429137	.1442633	3.07	0.002	.159518	.7263094
ETHHISP	-1.078255	2.268688	-0.48	0.635	-5.534941	3.378432
ETHBLACK	-4.014172	2.152185	-1.87	0.063	-8.241996	.2136528
MALE	6.364055	1.111968	5.72	0.000	4.179668	8.548442
_cons	193.7202	187.6859	1.03	0.302	-174.9761	562.4165

. vif

Variable	VIF	1/VIF
AGE	1411.96	0.000708
Agesq	1411.13	0.000709
EXP	1.29	0.778114
S	1.14	0.875122
ETHBLACK	1.04	0.962602
MALE	1.03	0.966488
ETHHISP	1.02	0.983851
Mean VIF	404.09	

метода последовательного исключения, на ближайшем шаге из уравнения регрессии будет удалена переменная

1) S 2) AGE 3) EXPSQ 4) EXP 5) ETHWHITE 6)ETHHISP 7) FEMALE 8) ни одна из перечисленных

4) Первой главной компонентой системы показателей $X_1,...,X_k$ называется такая линейная комбинация этих показателей

1. в которой коэффициент при X_1 равен 1 2. которая обладает наименьшей дисперсией 3. которая обладает наибольшей дисперсией 4. которая ортогональна всем $X_j, j=1,...,k$

5) (Д.А.Борзых, Б.Б.Демешев, часть задачи 7.4)

Пионеры, Крокодил Гена и Чебурашка собирали металлолом несколько дней подряд. В распоряжение иностранной шпионки, гражданки Шапокляк, попали ежедневные данные по количеству собранного металлолома: вектор g – для Крокодила Гены, вектор h – для Чебурашки и вектор x – для пионеров.Гена и Чебурашка собирали вместе, поэтому выборочная корреляция $\hat{cor}(g,h)=-0.9$. Гена и Чебурашка собирали независимо от пионеров, поэтому $\hat{cor}(g,x)=0$, $\hat{cor}(h,x)=0$. Если регрессоры g, h, x центрировать т нормировать, то получится матрица $\tilde{X}$. Вычислите одну или две главные компоненты (выразите их через вектор-столбцы матрицы $\tilde{X}$), объясняющие не менее 70% общей выборочной дисперсии регрессоров.

Ответ. $(\tilde{X}_1-\tilde{X}_2)/\sqrt{2};\quad \tilde{X}_3$

6) (Демешев, Борзых, 7.13)

Известно, что выборочная корреляция между переменными x и z равна 0.9.

1. Найдите коэффициенты VIF для x и z в регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$.
2. В каких пределах могут лежать коэффициенты VIF для x и z в регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$?

7) **(Демешев, Борзых, 7.11)**

Эконометресса Алевтина перешла от исходных регрессоров к трём главным компонентам, z_1, z_2 и z_3. И далее посчитала коэффициенты вздутия дисперсии, VIF_j, для главных компонент. Чему они оказались равны?

2. Прогнозирование по регрессионной модели

Прогнозирование по регрессионной модели и его точность. Доверительные интервалы для прогнозных значений.

Задачи

1) На основании 5 наблюдений получена МНК оценка уравнения регрессии $\hat{Y}_i = 1.56 + 0.21X_i$ и оценка остаточной дисперсии $\hat{\sigma}_\varepsilon^2 = 0.04$. Матрица наблюдений регрессоров имеет вид: $X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 8 \end{pmatrix}'$.

Построить 95% доверительный интервал для прогноза, если прогнозное значение X=2.

3. Метод максимального правдоподобия. Тесты Вальда, отношения правдоподобия, множителей Лагранжа

Метод максимального правдоподобия. Свойства оценок метода максимального правдоподобия. Соотношение между оценками коэффициентов линейной регрессии, полученными методом максимального правдоподобия и методом наименьших квадратов в случае нормально распределенной случайной составляющей. Свойства оценки дисперсии случайной составляющей, полученной методом максимального правдоподобия. Проверка гипотез с помощью теста Вальда, теста отношения правдоподобия, теста множителей Лагранжа.

Задачи

1) Логарифмическая функция правдоподобия, используемая для оценивания классической линейной регрессионной модели, имеет вид:

1) $\ln L(\beta, \sigma_\varepsilon^2) = \frac{n}{2}\ln\sqrt{2\pi} - \frac{n}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2}\frac{(Y - X\beta)'(Y - X\beta)}{\sigma_\varepsilon^2}$;

2) $\ln L(\beta,\sigma_\varepsilon^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_\varepsilon^2 - \frac{1}{2}\frac{\varepsilon'\varepsilon}{\sigma_\varepsilon^2}$;

3) $\ln L(\beta,\sigma_\varepsilon^2) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(Y - X\beta)'\Sigma^{-1}(Y - X\beta)$;

4) $\ln L(\beta,\sigma_\varepsilon^2) = \frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2}(Y - X\beta)'\Sigma(Y - X\beta)$;

5) ни один из вышеперечисленных.

2) Оценки метода максимального правдоподобия:
1) всегда состоятельные
2) всегда несмещенные
3) всегда имеют нормальное распределение
4) могут быть как смещенными, так и несмещенными
5) могут иметь произвольное асимптотическое распределение
6)

3) Число звонков по товару распределено по закону Пуассона P(X=x) =

$$P(X = x) = \frac{\lambda^x}{x!}e^{\lambda}$$

Произведены замеры в течении рабочего дня, было установлено что количество звонков распределилось следующим образом:

Номер часа от начала распродаж	**1**	**2**	**3**	**4**	**5**
Число звонков	**22**	**44**	**59**	**58**	**47**

Найти параметр λ методом максимального правдоподобия.

Ответ. 46

4) Магнус, Катышев, Пересецкий, № 10.10
Известно, что в модели $Y = X\beta + \varepsilon$
Имеется гетероскедастичность, причем
$\mathrm{var}(\varepsilon_i) = \sigma_1^2, i = 1,\ldots,n_1$, $\mathrm{var}(\varepsilon_i) = \sigma_2^2, i = n_1 + 1,\ldots,n = n_1 + n_2$
$\mathrm{cov}(\varepsilon_i,\varepsilon_j) = 0, i \neq j$.

В предположении нормальности вектора ошибок постройте тест отношения правдоподобия (LR-test) для проверки гипотезы $H_0 : \sigma_1^2 = \sigma_2^2$.

5) Борзых, Демешев, № 5.19, 5.20

5.19 Исследователь Вениамин пытается понять, как логарифм количества решённых им по эконометрике задач зависит от количества съеденных им пирожков. Для этого он собрал 100 наблюдений. Первые 50 наблюдений — относятся к пирожкам с мясом, а последние 50 наблюдений — к пирожкам с повидлом. Вениамин считает, что ожидаемое количество решённых задач не зависит от начинки пирожков, а только от их количества, т.е. $y_i = \beta x_i + u_i$. Однако он полагает, что для пирожков с мясом — $u_i \sim \mathcal{N}(0; \sigma_M^2)$, а для пирожков с повидлом — $u_i \sim \mathcal{N}(0; \sigma_J^2)$.

1. Выпишите логарифмическую функцию правдоподобия.
2. Выпишите условия первого порядка для оценки β, σ_M^2, σ_J^2.

5.20 После долгих изысканий Вениамин пришёл к выводу, что $\beta = 0$, т.е. что логарифм количества решенных им по эконометрике за вечер задач имеет нормальное распределение y_i с математическим ожиданием ноль. Однако он по прежнему уверен, что дисперсия y_i зависит от того, какие пирожки он ел в этом вечер. Для пирожков с повидлом $y_i \sim \mathcal{N}(0; \sigma_J^2)$, а для пирожков с мясом — $y_i \sim \mathcal{N}(0; \sigma_M^2)$. Всего 100 наблюдений. Первые 50 вечеров относятся к пирожкам с мясом, последние 50 вечеров — к пирожкам с повидлом:

$$\sum_{i=1}^{50} y_i = 10, \ \sum_{i=1}^{50} y_i^2 = 100, \ \sum_{i=51}^{100} y_i = -10, \ \sum_{i=51}^{100} y_i^2 = 300$$

1. Найдите оценки σ_M^2, σ_J^2, которые получит Вениамин.

4. Гетероскедастичность

Нарушение гипотезы о гомоскедастичности. Экономические причины гетероскедастичности. Последствия гетероскедастичности для оценок коэффициентов регрессии методом наименьших квадратов и проверки статистических гипотез. Поведение графика остатков регрессии, как признак гетероскедастичности. Тесты Голдфелда-Квандта (Goldfeld-Quandt), Глейзера (Glejser Бройша-Пагана (Breusch-Pagan). Методы борьбы с гетероскедастичностью. Робастные стандартные ошибки в форме Уайта (White). Взвешенный метод наименьших квадратов. Обобщенный метод наименьших квадратов

Задачи

1) При гетероскедастичности возмущений нарушается условие теоремы Гаусса – Маркова ______________________________

б) (2) Оценки МНК коэффициентов в этом случае останутся

1) BEST 2) LINEAR 3) UNBIASED 4) ESTIMATOR 5) ничего из перечисленного

2) По данным с 1946 г. по 1975 г. Hanushek и Jackson оценили коэффициенты уравнений регрессии (в скобках указаны оценки стандартных отклонений)

$$\hat{C}_t = \underset{(2.73)}{26.19} + \underset{(0.006)}{0.6248}\, GNP_t - \underset{(0.0736)}{0.4398}\, D_t$$

$$\left(\frac{\hat{C}}{GNP}\right)_t = \underset{(2.22)}{25.92}\,\frac{1}{GNP_t} + \underset{(0.00597)}{0.6246} - \underset{(0.0597)}{0.4315}\left(\frac{D}{GNP}\right)_t,$$

где C – агрегированные частные потребительские расходы,

GNP – ВНР,

D – национальные расходы на оборону.

С какой целью оценили второе уравнение? Какое при этом было сделано предположение о дисперсии ошибок?

Ответ. $\sigma_i^2 \sim GNP_i^2$

3) По данным для 45 стран исследователь оценил зависимость инвестиций I от государственных расходов G и валового внутреннего продукта Y (все переменные измеряются в миллиардах долларов США):

$$\hat{I} = \underset{(7.79)}{18.1} - \underset{(0.14)}{1.07}\, G + \underset{(0.02)}{0.36}\, Y, \quad R^2 = 0.98$$

Исследователь упорядочил наблюдения по увеличению Y и оценил регрессии снова для 15 стран с наименьшим Y и 15 стран с наибольшим Y. Величины RSS для этих регрессий равны 421 и 3219 соответственно. Протестируйте модель на наличие пропорциональной формы гетероскедастичности.

4) По данным для 20 стран были оценены коэффициенты уравнения регрессии

$$\hat{Y}_i = \underset{t=4.79}{111.78} - \underset{t=-2.53}{0.0042}\, X_{2i} - \underset{t=-1.71}{0.4898}\, X_{3i} \quad R^2 = 0.492$$

где Y – младенческая смертность (количество в расчете на тысячу рожденных живыми),

X_2 – GNP в расчете на душу населения,

X_3 – процент имеющих начальное образование в определенной возрастной группе.

При проведении теста Уайта была оценена регрессия

$$e_i^2 = \underset{t=4.86}{4987} - \underset{t=-0.59}{0.4718}\, X_{2i} - \underset{t=-2.45}{0.8442}\, X_{3i}$$

$$+ \underset{t=1.27}{0.00001}\, X_{2i}^2 + \underset{t=1.62}{0.4435}\, X_{3i}^2 + \underset{t=0.35}{0.0026}\, X_{2i} X_{3i} \quad R^2 = 0.649$$

Проведя тест Уайта, проверьте гипотезу об отсутствии гетероскедастичности.

5) Предположим, что для модели парной регрессии $\sigma_{ui}^2 = \sigma^2 X_i^4$.

Как избавиться от проблемы гетероскедастичности ошибок?

6) Тестом, который не только позволяет выявить наличие гетероскедастичности, но и указать способ оценивания параметров σ_i^2, является

1) тест Уайта 2) тест Глейзера 3) тест Рамсея 4) тест Хаусмана

7) Оценки метода наименьших квадратов коэффициентов регрессии : $Y = \beta_1 + \beta_2 X_2 + ... + \beta_k X_k + u$ останутся несмещенными при нарушении условий теоремы Гаусса – Маркова

1) $\text{var}(u_i) = \sigma^2$ при всех i

2) $\text{cov}(u_i; u_j) = 0$; i ≠ j,

3) состоящих во включении в модель лишнего объясняющего фактора Z,

4) состоящих в невключении в модель необходимого фактора

8) FGLS-оценка коэффициентов линейной регрессионной модели

$$Y = X\beta + \varepsilon, \quad \hat{V}[\varepsilon] = \Sigma$$

имеет вид:

1) $(X'X)^{-1}(X'Y)$
2) $(X'\hat{\Sigma}^{-1}X)^{-1}(X'Y)$
3) $(X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}Y)$
4) $(X'\hat{\Sigma}^{-1}X)^{-1}(X'\hat{\Sigma}^{-1}Y)$
5) нет верного ответа

9) (Д.А.Борзых, Б.Б.Демешев, задача 8.26)

Пусть $Y_i = \beta X_i + \varepsilon_i$, где $E(\varepsilon_i) = 0$ и известно, что оценка для параметра $\tilde{\beta} = \dfrac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i}$ является наилучшей (в смысле минимума дисперсии) среди всех линейных несмещенных оценок параметра β. Чему равна в этом случае матрица ковариаций вектора ε с точностью до пропорциональности?

Задачник Борзых и Демешев

8.2 В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \sigma^2 x_i^2$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

8.3 В модели $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ присутствует гетероскедастичность вида $\mathrm{Var}(\varepsilon_i) = \lambda |x_i|$. Как надо преобразовать исходные регрессоры и зависимую переменную, чтобы устранить гетероскедастичность?

8.4 Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на x_i^2 гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

8.5 Известно, что после деления каждого уравнения регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ на $\sqrt{x_i}$ гетероскедастичность ошибок была устранена. Какой вид имела дисперсия ошибок, $\mathrm{Var}(\varepsilon_i)$?

8.7 По наблюдениям $x = (1, 2, 3)'$, $y = (2, -1, 3)'$ оценивается модель $y = \beta_1 + \beta_2 x + \varepsilon$. Ошибки ε гетероскедастичны и известно, что $\mathrm{Var}(\varepsilon_i) = \sigma^2 \cdot x_i^2$.

1. Найдите оценки $\hat{\beta}_{ols}$ с помощью МНК и их ковариационную матрицу.
2. Найдите оценки $\hat{\beta}_{gls}$ с помощью обобщенного МНК и их ковариационную матрицу.

8.8 Для линейной регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ была выполнена сортировка наблюдений по возрастанию переменной x. Исходная модель оценивалась по разным частям выборки:

Выборка	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	RSS
$i = 1, \ldots, 30$	1.21	1.89	2.74	48.69
$i = 1, \ldots, 11$	1.39	2.27	2.36	10.28
$i = 12, \ldots, 19$	0.75	2.23	3.19	5.31
$i = 20, \ldots, 30$	1.56	1.06	2.29	14.51

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

8.12 Рассмотрим линейную регрессию $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ по 50 наблюдениям. При оценивании с помощью МНК были получены результаты: $\hat{\beta}_1 = 1.21$, $\hat{\beta}_2 = 1.11$, $\hat{\beta}_3 = 3.15$, $R^2 = 0.72$.

Оценена также вспомогательная регрессия: $\hat{\varepsilon}_i^2 = \delta_1 + \delta_2 x_i + \delta_3 z_i + \delta_4 x_i^2 + \delta_5 z_i^2 + \delta_6 x_i z_i + u_i$. Результаты оценивания следующие: $\hat{\delta}_1 = 1.50$, $\hat{\delta}_2 = -2.18$, $\hat{\delta}_3 = 0.23$, $\hat{\delta}_4 = 1.87$, $\hat{\delta}_5 = -0.56$, $\hat{\delta}_6 = -0.09$, $R^2_{aux} = 0.36$

Известно, что ошибки в модели являются независимыми нормальными случайными величинами с нулевым математическим ожиданием. Протестируйте ошибки на гетероскедастичность на уровне значимости 5%.

8.15 Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.16 Рассматривается модель $y_t = \beta_1 + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.17 Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.18 Рассматривается модель $y_t = \beta_1 x_t + \varepsilon_t$, где ошибки ε_t — независимые случайные величины с $\mathbb{E}(\varepsilon_t) = 0$ и $\mathrm{Var}(\varepsilon_t) = t^2$. Найдите наиболее эффективную оценку неизвестного параметра β_1 в классе линейных по y и несмещённых оценок.

8.23 Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$, в которой ошибки ε_i независимы и имеют нормальное распределение $N(0, \sigma^2)$. Для $n = 200$ наблюдений найдите

1. вероятность того, что статистика Уайта окажется больше 10;
2. ожидаемое значение статистики Уайта;
3. дисперсию статистики Уайта.

8.24 Найдите число коэффициентов во вспомогательной регрессии, необходимой для выполнения теста Уайта, если число коэффициентов в исходной регрессии равно k, включая свободный член.

8.25 По 35 наблюдениям сотрудники НИИ оценили уравнение регрессии $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$ и рассчитали остатки ε_i. После того они приступили к диагностике возможных недостатков модели, обнаружили гетероскедастичность и решили её побороть.

1. Самый младший научный сотрудник выдвинул предположение, что стандартное отклонение случайной составляющей может быть выражено так: $\sigma_{\varepsilon,i} = ax_i$, где a — неизвестный коэффициент. Каким образом нужно преобразовать исходное уравнение регрессии, чтобы избавиться от гетероскедастичности?
2. Профессор решил перепроверить результаты и оценил регрессию:
$$\hat{e}_i^2 = -0.3 + 0.08x_i - 0.01x_i^2, R^2 = 0.15.$$
Свидетельствует ли полученный профессором результат о наличии гетероскедастичности?

5. Бинарные объясняемые переменные. Логит и пробит модели

Бинарные объясняемые переменные. Модель линейной вероятности. Логит и Пробит модели, их оценивание. Интерпретация результатов оценивания моделей с бинарными объясняемыми переменными. ROC – кривая.

Задачи

Демидова, Малахов

Задача 11.1.

Исследователя интересует зависимость вероятности найти работу от уровня образования индивидуума. Введя в качестве зависимой переменную EMP, равную 1 для работающих и 0 для неработающих и S – количество лет обучения в качестве объясняющей, он оценил логит – модель:

$$P\{EMP_i = 1\} = \frac{1}{1+\exp\{-Z_i\}}, \quad Z_i = \underset{(0.242)}{-1.006} + \underset{(0.018)}{0.148}\, S,$$

Оцените предельный эффект объясняющего фактора для среднего значения переменной $S = 13.5$.

Задача 11.2.

По наблюдениям для 570 индивидуумов оценена зависимость получения школьником аттестата от обобщенной оценки результатов тестов X. Переменная Y принимает значение 1, если аттестат был получен и 0 в противном случае.

Оцененные модели имеют следующий вид:

Логит: $P\{Y_i = 1\} = \dfrac{1}{1+\exp\{-Z_i\}}, \quad Z_i = \underset{(0.865)}{-5.004} + \underset{(0.021)}{0.1666}\, X,$

Пробит: $P\{Y_i = 1\} = F(Z_i)$, $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$, $Z = -\underset{(0.083)}{2.7} + \underset{(0.0117)}{0.53} X$

Дайте экономическую интерпретацию полученным результатам для логит и пробит моделей. Найдите предельный эффект объясняющего фактора в точке $\overline{X} = 50.15$.

Задача 11.3.

Из 750 обратившихся за ссудой в банк 250 было в ней отказано. Для оценки вероятности получения ссуды были оценены линейная и пробит модели:

$Y = 0.5 + 1.5X$,

$P\{Y_i = 1\} = F(Z_i)$, $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$

$Z = 0.45 + 3X$

где $Y_i = 1$ для получивших ссуду и 0 иначе, X – доход просителя.

По пробит модели найти предельный эффект дохода в среднем.

Задача 11.4.

Для того, чтобы определить, эффективна ли новая методика преподавания микроэкономики, провели следующий эксперимент: протестировали всех студентов по микроэкономике в конце первого и второго семестра. Часть студентов во втором семестре обучали по новой методике, часть по старой. После этого в качестве объясняющей выбрали переменную Y, равную 1, если результат студента улучшился и 0 в противном случае, а в качестве объясняющих переменных X_1 – результаты теста в первом семестре, X_2 – средний балл по остальным предметам, D – равную 1, если студент обучался по новой методике и 0, если по старой.

По имеющимся данным оценили пробит- модель:

$P\{Y_i = 1\} = F(Z_i)$, $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$

$Z = -7.452 + 0.052X_1 + 1.626X_2 + 1.426D$.

Найдите предельный эффект переменной D при средних значениях $\overline{X}_1 = 21.938$, $\overline{X}_2 = 3.117$ (разность вероятностей улучшения результата при D = 1 и D = 0).

Задание 11.3.

Для анализа аудитории, использующей Интернет для учебы по данным для 1314 индивидов были оценены линейная и пробит модели (последняя с предельными эффектами), в которых intlear = 1 при использовании индивидом Интернета для учебы и 0 в противном случае, male = 1 для мужчин и 0 для женщин, income – заработная плата индивида по основному месту работы, age – возраст. В скобках указаны соответственно t или z - статистики. В чем состоят недостатки линейной модели? Дайте интерпретацию полученным результатам.

$$INTLEAR = -\underset{(18.48)}{0.78} - \underset{(-11.45)}{.013}\ AGE - \underset{(-0.96)}{4.53 \cdot 10^{-10}}\ INCOME - \underset{(-3.05)}{0.073} MALE$$

$$P\{INTLEAR_i = 1\} = F(Z_i),\ F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

$$Z = \underset{(7.25)}{1.01} - \underset{(-10.82)}{0.044} AGE - \underset{(-0.98)}{1.51 \cdot 10^{-9}}\ INCOME - \underset{(-3.08)}{0.23} MALE$$

Marginal effects after probit

variable	dy/dx	Std. Err.	z	P>z	[95% C.I.]	
age	-.0147444	.00133	-11.08	0.000	-.017353	-.012136
income	-5.07e-10	.00000	-0.98	0.328	-1.5e-09	5.1e-10
male*	-.0783057	.02536	-3.09	0.002	-.128015	-.028597

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Борзых, Демешев, задача 6.5.

При оценке логит модели $\mathbb{P}(y_i = 1) = \Lambda(\beta_1 + \beta_2 x_i)$ по 500 наблюдениям оказалось, что $\hat{\beta}_1 = 0.7$ и $\hat{\beta}_2 = 3$. Оценка ковариационной матрицы коэффициентов имеет вид

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

1. Проверьте гипотезу о незначимости коэффициента $\hat{\beta}_2$.
2. Найдите предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$ при $x_i = -0.5$.
3. Найдите максимальный предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$.