

Эконометрика. Лекция 11

Модели бинарного выбора

Д. С. Терещенко

НИУ ВШЭ, Санкт-Петербург

2 декабря 2022 г.

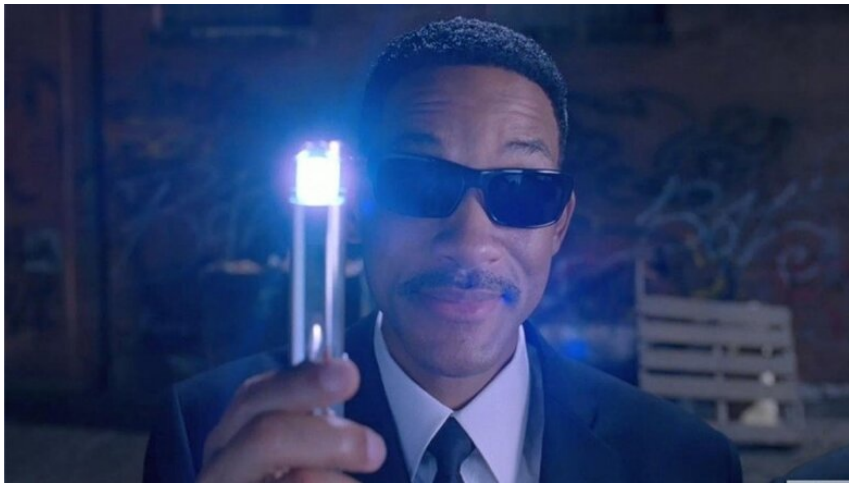
Основные источники

- **SW, глава 11**
- HE, chapter 25

Содержание

- 1 Мотивация
- 2 Постановка модели
- 3 Интерпретация коэффициентов и предельные эффекты
- 4 Модели бинарного выбора на практике

Забудем пока про эндогенность



Содержание

① Мотивация

② Постановка модели

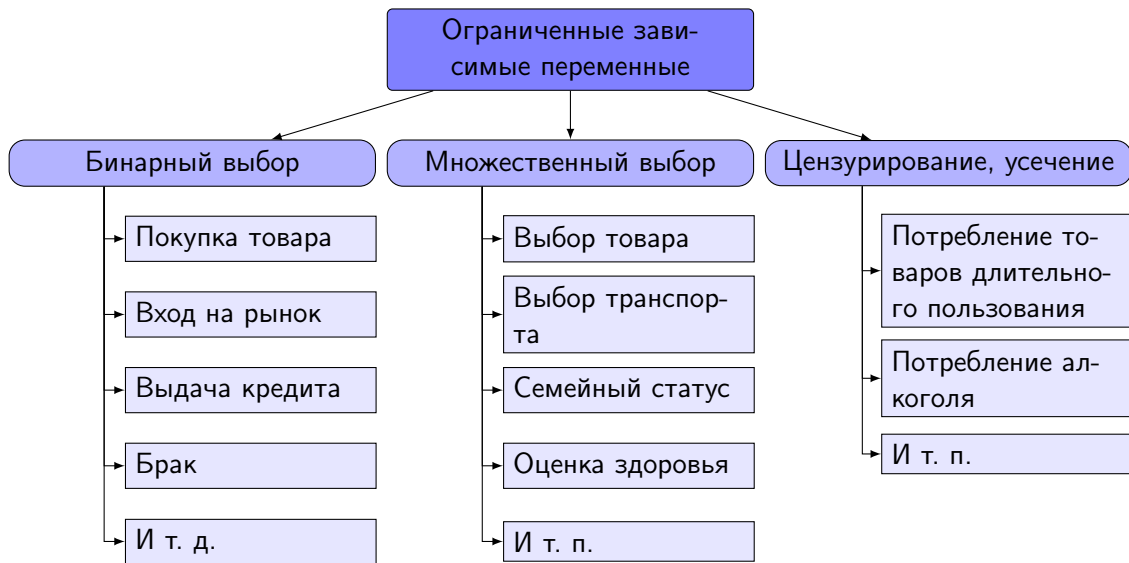
③ Интерпретация коэффициентов и предельные эффекты

④ Модели бинарного выбора на практике

Ограниченные зависимые переменные и линейные модели

- До сих пор мы использовали в качестве зависимых «обычные числовые», непрерывные переменные.
- В случае, когда зависимая переменная не является непрерывной, ее называют **ограниченной** (limited dependent variable).
- Если зависимая переменная является ограниченной, использование для анализа линейных моделей становится затрудненным.

Ограниченные зависимые переменные: классификация и примеры



Содержание

- 1 Мотивация
- 2 Постановка модели
- 3 Интерпретация коэффициентов и предельные эффекты
- 4 Модели бинарного выбора на практике

Бинарные зависимые переменные

- Сегодня мы говорим про бинарные зависимые переменные.
- Любую такую переменную (решение о покупке, браке, выдаче кредита) можно перекодировать, чтобы она принимала значение 0 или 1.
- Тогда зависимая переменная будет иметь распределения Бернулли.

Регрессия с бинарными зависимыми переменными

- Мы хотим получить модель в привычном для нас виде:

$$Y = \mathbb{E}[Y|X_1, \dots, X_k] + u$$

$$\mathbb{E}[u|X_1, \dots, X_k] = 0, \dots$$

- Можно показать, что:

$$\mathbb{E}(Y) = 0 \times \Pr(Y = 0) + 1 \times \Pr(Y = 1) = \Pr(Y = 1)$$

$$\mathbb{E}(Y|X_1, \dots, X_k) = \Pr(Y = 1|X_1, \dots, X_k) \equiv P(\mathbf{x})$$

- Получается, что регрессия с бинарными зависимыми переменными — моделирование вероятности того, что зависимая переменная равна 1

$$Y = P(\mathbf{x}) + u$$

Регрессия с бинарными зависимыми переменными

$$Y = P(\mathbf{x}) + u$$

- Ошибка в модели выглядит нестандартно и принимает только два значения:

$$u = \begin{cases} 1 - P(\mathbf{x}) & \text{с вероятностью } P(\mathbf{x}) \\ -P(\mathbf{x}) & \text{с вероятностью } 1 - P(\mathbf{x}) \end{cases}$$

- Ошибка гетероскедастична по построению:

$$\text{var}[u|\mathbf{x}] = P(\mathbf{x})(1 - P(\mathbf{x}))$$

Модели бинарного выбора

Линейная вероятностная модель

- Мы моделируем вероятность того, что $Y = 1$.
- Наиболее простой и прямолинейный подход — линейная вероятностная модель (ММЛР с бинарной зависимой переменной):

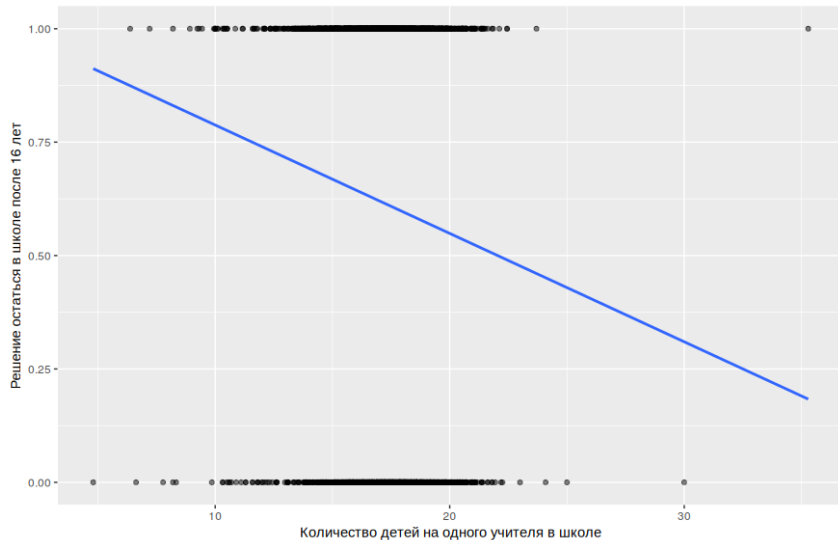
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

$$\Pr(Y = 1 | X_1, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$$

- В этом случае можно оценивать коэффициенты при помощи МНК и проверять гипотезы, как мы уже делали в ММЛР.

Линейная вероятностная модель — пример

Кто видит проблему?



Модели бинарного выбора

Модели линейного индекса

- Мы моделируем вероятность того, что $Y = 1$.
- Более продвинутый и распространенный способ — модели линейного индекса:
 - $\Pr(Y = 1|X_1, \dots, X_k) = G(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$
 - $G(h)$ — функция распределения, $0 \leq G(h) \leq 1$ (на практике часто предполагается симметричность относительно нуля: $G(-h) = 1 - G(h)$).
 - $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ — линейный индекс.
- На практике наиболее распространены:
 - Пробит-модель;
 - Логит-модель.
- Модели линейного индекса более уважительно относятся к границам вероятности, но их оценивание и интерпретация немного сложнее.

Пробит и логит

Пробит-модель — стандартная нормальная функция распределения вероятностей:

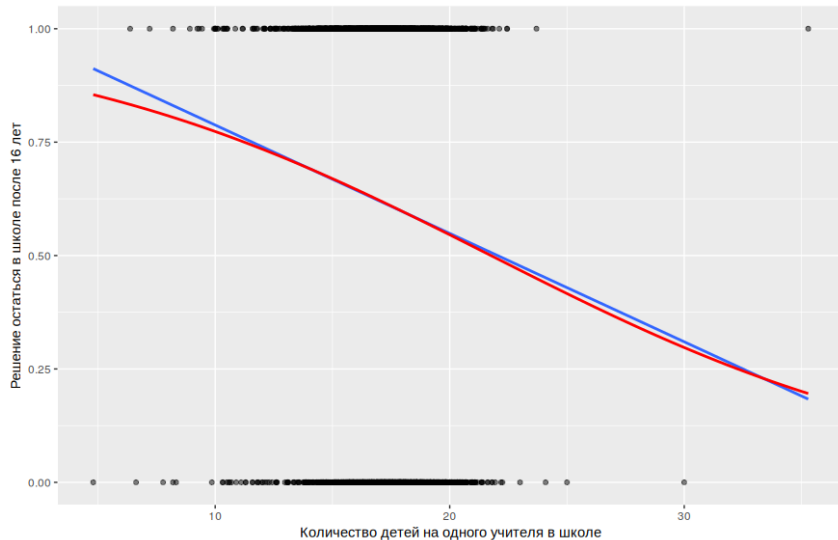
$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$
$$\Phi(h) = \Pr(Z \leq h)$$

Логит-модель — логистическая функция распределения:

$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_{ki}) =$$
$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

Линейная вероятностная модель и модель линейного индекса

Что лучше?



Содержание

- 1 Мотивация
- 2 Постановка модели
- 3 Интерпретация коэффициентов и предельные эффекты**
- 4 Модели бинарного выбора на практике

Интерпретация в моделях бинарного выбора

При интерпретации в моделях бинарного выбора нас могут интересовать:

- условная вероятность $\Pr(Y = 1|X_1, X_2, \dots, X_k)$;
- коэффициенты;
- предельные эффекты.
- ...

Интерпретация в моделях бинарного выбора

Предельные эффекты в общем случае

- Предельный эффект X_1 по определению:

$$ME_{X_1} = \frac{\partial \mathbb{E}[Y|X_1, \dots, X_k]}{\partial X_1} = \frac{\partial \Pr(Y = 1|X_1, \dots, X_k)}{\partial X_1}$$

- Для линейной вероятностной модели:

...

- Для моделей линейного индекса:

$$ME_{X_1} = \beta_1 g(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

где $g(h) = \frac{\partial G(h)}{\partial h}$ — функция плотности распределения G .

Интерпретация в моделях бинарного выбора

Линейная вероятностная модель

$$\Pr(Y = 1|X_1, \dots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

- Самая простая с точки зрения интерпретации.
- Предельный эффект равен коэффициенту:

$$ME_{X_1} = \beta_1$$

Интерпретация в моделях бинарного выбора

Пробит

$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

- Можно интерпретировать только знаки коэффициентов, т. е. говорить о положительной или отрицательной связи.
- Предельный эффект в пробит-модели:

$$ME_{X_1} = \beta_1 \phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

где $\phi(h)$ — функция плотности стандартного нормального распределения.

Интерпретация в моделях бинарного выбора

Логит

$$\Pr(Y = 1|X_1, X_2, \dots, X_k) = \Lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

- Можно интерпретировать знаки коэффициентов, т. е. говорить о положительной или отрицательной связи (но не только).
- Предельный эффект в логит-модели:

$$ME_{X_1} = \beta_1 \lambda(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)$$

где $\lambda(h)$ — функция плотности стандартизированного логистического распределения.

Интерпретация в моделях бинарного выбора

Логит

$$P(\mathbf{x}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

- Предельный эффект в логит-модели:

$$ME_{X_1} = \beta_1 \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}}{(1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k})^2}$$

- Можно вывести шансы успеха (odds), отношение шансов (odds ratio, OR) и логарифм отношения шансов:

$$odds(X_1, X_2, \dots, X_k) = \frac{P(\mathbf{x})}{1 - P(\mathbf{x})} = \dots = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k}$$

$$OR = \frac{odds(X_1 + 1, \dots)}{odds(X_1, \dots)} = \dots = e^{\beta_1}$$

$$L = \log(OR) = \dots = \beta_1$$

Содержание

- 1 Мотивация
- 2 Постановка модели
- 3 Интерпретация коэффициентов и предельные эффекты
- 4 Модели бинарного выбора на практике

Пробит vs. логит

Историческая справка

HE, p. 804:

The intriguing labels probit and logit have a long history in statistical analysis. The term probit was coined by Chester Bliss in 1934 as a contraction of "probability unit". The logistic function was introduced by Pierre François Verhulst in 1938 as a modified exponential growth model. It is speculated that he used the term logistic as a contrast to logarithmic. In 1944 Joseph Berkson proposed a binary choice model based on the logistic distribution function. He motivated the logistic as a convenient computational approximation to the normal. As his model was an analog of the probit Berkson called his model the logit.

Пробит vs. логит

Какую модель выбрать?

- В теории основная разница в том, что у логистического распределения более толстые хвосты, то есть условная вероятность $P(x)$ приближается к 0 или 1 медленнее в логите, чем в пробите. Т. е. различия в оценках можно наблюдать на хвостах распределения.
- На практике большой разницы в оценках нет.
- Логит-модель проще при расчетах и интерпретации.
- Выбор модели может быть обусловлен либо какими-то теоретическими представлениями о распределении, либо сложившейся исследовательской традицией.

Кратко об оценивании и свойствах оценок

- Наиболее распространенный способ оценивания параметров в пробит- и логит-моделях — метод максимального правдоподобия (maximum likelihood).
- Оценки метода максимального правдоподобия:
 - состоятельные;
 - асимптотически нормальные.
- Также можно использовать нелинейный метод наименьших квадратов.

Примеры

Toivanen, O., & Waterson, M. (2005). Market Structure and Entry: Where's the Beef? The RAND Journal of Economics, 36(3), 680–699.

TABLE 3 Reduced-Form Random-Effects Probit Estimations

Variable	BK	McD
<i>Constant</i>	−2.9876*** (.6152)	−1.4932*** (.5879)
<i>Area</i>	.1042 (.2011)	−.2620 (.2300)
<i>Population</i>	.6335 (.5405)	.5319 (.5164)
<i>Population * Area</i>	−.4029** (.1944)	.0968 (.1954)
<i>BusRate</i>	.2055** (.1009)	.1693* (.0980)
<i>Council Tax</i>	1.3153* (.6933)	−.7576 (.5914)
<i>Unemployment</i>	−1.1163 (2.5784)	.0270 (2.4320)
<i>Youth</i>	.6363 (2.4344)	−.8777 (2.3193)
<i>Pension</i>	−3.4895 (3.3107)	−5.6388* (3.1058)
<i>Wage</i>	.1183 (.1833)	−.1740 (.1788)
<i>Rival Outlets in Neighboring Markets</i>	−.02730 (.0544)	.0470 (.0499)
<i>Own Outlets in Neighboring Markets</i>	−.0966*** (.0168)	.0215 (.0125)
<i>London</i>	.0654 (.2738)	−.3932 (.2453)

TABLE 4 Marginal Effects of Market Structure Indicators

Market Structure Dummies	BK	McD
<i>M1, B0</i>	.1089*** (.0315)	.0413*** (.0150)
<i>M1, B1</i>	.0185 (.218)	.0589** (.0247)
<i>M0, B1</i>	.0528 (.0404)	.0345 (.0294)
<i>M2, B0</i>	.1834*** (.0553)	.1448*** (.0428)
<i>M2, B1</i>	.0442 (.0313)	.1759*** (.0524)
<i>M0, B2</i>	—	.1585** (.0847)
<i>M1, B2</i>	—	.1233 (.0798)
<i>M2, B2</i>	−.0074 (.0299)	.2341** (.1120)
<i>M3, B0</i>	.2011*** (.0935)	−.0037 (.0066)
<i>M3, B <</i>	.0291 (.0281)	.2584*** (.0693)
<i>M <, B3</i>	−.0102 (.0316)	—
ρ^*		

Примеры

Enikolopov, R., Petrova, M., & Zhuravskaya, E. (2011). Media and Political Persuasion: Evidence from Russia. *American Economic Review*, 101(7), 3253–3285.

TABLE 7—SELF-REPORTED VOTE AND NTV IN 1999, SURVEY DATA

	Opposed by NTV in 1999		Supported by NTV in 1999	
	Vote for Unity in 1999 (centrist, progovernment) IV probit	Probit	Vote for OVR in 1999 (centrist, opposition) IV probit	Probit
<i>Panel A</i>				
Watched NTV in 1999	−0.831 [0.301]***	−0.139 [0.128]	1.180 [0.477]**	0.135 [0.156]
Marginal effect	−0.26 [0.09]***	−0.05 [0.04]	0.25 [0.14]*	0.02 [0.02]
Controls	Yes	Yes	Yes	Yes
Observations	901	901	901	901
Number of subregions	42	42	42	42
χ^2 statistics for the exclusion of NTV1999 in the first stage	34.72		24.79	
	Supported by NTV in 1999			
	Vote for SPS in 1999 (liberal) IV probit	Probit	Vote for Yabloko in 1999 (liberal) IV probit	Probit
<i>Panel B</i>				
Watched NTV in 1999	1.210 [0.405]***	0.272 [0.149]*	0.467 [0.555]	0.039 [0.178]
Marginal effect	0.24 [0.11]**	0.04 [0.02]*	0.06 [0.08]	0.004 [0.02]
Controls	Yes	Yes	Yes	Yes
Observations	901	901	901	901
Number of subregions	42	42	42	42
χ^2 statistics for the exclusion of NTV1999 in the first stage	28.90		30.47	

Примеры

Lipson, S. K., Zhou, S., Wagner III, B., Beck, K., & Eisenberg, D. (2016). Major differences: Variations in undergraduate and graduate student mental health and treatment utilization across academic disciplines. *Journal of College Student Psychotherapy*, 30(1), 23-41.

Table 4. Multivariable Correlates of Mental Health Problems Among Master's Students, Logistic Regressions.

	Dep				Anx				SI				NSSI				Any				Tx			
	OR	p	95% CI		OR	p	95% CI		OR	p	95% CI		OR	p	95% CI		OR	p	95% CI		OR	p	95% CI	
SOC	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
HUM	2.15	.001	1.35	3.42	0.98	.94	0.55	1.73	2.10	.01	1.19	3.72	1.58	.03	1.05	2.37	1.70	.003	1.20	2.41	1.02	.95	0.61	1.68
NAT	1.22	.33	0.82	1.82	0.76	.33	0.43	1.33	1.77	.03	1.07	2.94	1.12	.57	0.76	1.66	1.10	.54	0.81	1.49	1.13	.64	0.67	1.91
ART	1.95	<.001	1.39	2.72	1.44	.12	0.91	2.27	1.61	.05	1.01	2.56	1.32	.13	0.93	1.87	1.80	<.001	1.39	2.33	1.09	.71	0.70	1.70
ENG	1.35	.08	0.97	1.89	0.62	.11	0.34	1.11	1.07	.80	0.66	1.71	1.12	.54	0.77	1.63	1.15	.32	0.88	1.49	0.54	.02	0.33	0.89
BUS	0.90	.55	0.64	1.27	0.71	.10	0.48	1.06	0.49	.003	0.31	0.78	0.47	<.001	0.34	0.66	0.68	.002	0.53	0.87	0.66	.07	0.42	1.03
LAW	0.79	.52	0.40	1.59	1.20	.76	0.38	3.79	0.58	.38	0.17	1.96	0.44	.14	0.15	1.31	0.64	.15	0.34	1.18	0.55	.32	0.17	1.81
SW	0.88	.66	0.49	1.56	1.10	.69	0.69	1.73	0.71	.30	0.36	1.37	0.87	.49	0.59	1.29	0.86	.40	0.60	1.23	2.13	.004	1.27	3.56
PH	0.72	.15	0.47	1.13	0.79	.36	0.48	1.31	0.78	.43	0.43	1.43	0.53	.01	0.33	0.83	0.60	.001	0.44	0.82	1.30	.37	0.74	2.30
NUR	0.82	.47	0.47	1.41	1.05	.88	0.55	2.00	0.42	.06	0.17	1.04	0.37	.01	0.19	0.74	0.66	.05	0.44	1.01	0.93	.87	0.40	2.17
MED	1.62	.05	1.00	2.64	1.35	.31	0.75	2.43	0.98	.97	0.44	2.21	0.68	.23	0.36	1.28	1.06	.77	0.71	1.58	1.13	.73	0.57	2.23
OTH	1.07	.68	0.77	1.50	0.99	.95	0.68	1.43	0.90	.62	0.60	1.36	0.70	.02	0.52	0.94	0.95	.67	0.76	1.20	0.94	.75	0.64	1.38
MULT	1.75	.001	1.24	2.47	1.39	.13	0.91	2.14	1.34	.18	0.87	2.07	0.92	.59	0.67	1.25	1.32	.03	1.02	1.70	0.95	.80	0.64	1.41
Constant	0.17	<.001	0.10	0.29	0.09	<.001	0.04	0.19	0.06	<.001	0.02	0.15	0.35	<.001	0.20	0.60	0.54	.004	0.35	0.82	0.22	<.001	0.10	0.47
N	9,282				6,866				9,275				9,171				9,157				2,346			

Note. CI = confidence interval; OR = odds ratio; Dep = Depression; Anxiety (Anx) excludes 2013; SI = suicidal ideation; NSSI = nonsuicidal self-injury; Any = Any Mental Health Problem; Treatment Utilization (Tx) is among students with at least one apparent mental health (MH) problem; SOC = Social Sciences; HUM = Humanities; NAT = Natural Sciences; ART = Art & Design; ENG = Engineering; BUS = Business; LAW = Law; SW = Social Work; PH = Public Health; NUR = Nursing; MED = Medicine; OTH = Other; MULT = Multidisciplinary. All models control for survey year and students' age, gender, citizenship, race/ethnicity, and parental education. Reference category is *Social Sciences*.

Примеры

Kouvonen, A., Kivimäki, M., Virtanen, M., Pentti, J., & Vahtera, J. (2005). Work stress, smoking status, and smoking intensity: an observational study of 46 190 employees. *Journal of Epidemiology & Community Health*, 59(1), 63-69.

Table 2 Relation of smoking status with job strain and effort-reward imbalance: adjusted odds ratios (ORs) and their 95% confidence intervals (CIs)

	Women			Men			p for sex interaction
	Number	OR*	(95% CI)	Number	OR*	(95% CI)	
Job control (component of job strain)							0.776
High job control	11033	1.00		2739	1.00		
Intermediate job control	12428	0.94	(0.87 to 1.01)	2925	0.98	(0.86 to 1.11)	
Low job control	11454	0.96	(0.89 to 1.04)	2490	1.01	(0.88 to 1.17)	
Job demands (component of job strain)							0.616
Low job demands	8547	1.00		2403	1.00		
Intermediate job demands	13825	0.92	(0.85 to 0.99)	3431	0.97	(0.86 to 1.10)	
High job demands	12452	1.01	(0.94 to 1.09)	2310	0.99	(0.87 to 1.14)	
Job strain							0.083
Low strain	8316	1.00		2173	1.00		
Active jobs	10284	1.00	(0.92 to 1.09)	2398	0.87	(0.76 to 1.01)	
Passive jobs	7387	1.02	(0.93 to 1.11)	2021	0.88	(0.75 to 1.02)	
High strain	8792	1.04	(0.96 to 1.13)	1548	1.00	(0.86 to 1.17)	
Effort (component of effort-reward imbalance)							0.019
High effort	12908	1.00		2169	1.00		
Intermediate effort	18037	0.88	(0.83 to 0.94)	4363	1.09	(0.96 to 1.24)	
Low effort	3860	0.83	(0.75 to 0.92)	1605	0.95	(0.81 to 1.11)	
Rewards (component of effort-reward imbalance)							0.292
High rewards	10236	1.00		2405	1.00		
Intermediate rewards	10379	1.03	(0.96 to 1.11)	2355	1.05	(0.91 to 1.21)	
Low rewards	8534	1.20	(1.11 to 1.30)	2501	1.15	(1.00 to 1.33)	
Effort-reward imbalance							0.129
Low imbalance	9044	1.00		2652	1.00		
Intermediate imbalance	11156	1.13	(1.05 to 1.22)	2441	1.01	(0.88 to 1.15)	
High imbalance	8831	1.28	(1.19 to 1.39)	2152	1.13	(0.98 to 1.29)	

*Adjusted for age, basic education, occupational status, type of employment, and marital status.