

Практическая работа №3

В этой задаче необходимо проанализировать данные волны мониторинга экономического положения и здоровья населения РФ (данные обследования РМЭЗ НИУ ВШЭ).

В случае, если не удаётся прочесть соответствующий файл, обратитесь к материалу, выложенному вместе с заданием.

Прочитайте данные, выберите столбцы, которые Вам кажутся необходимыми, чтобы описать социально-экономическое положение граждан Российской Федерации. Минимальный набор параметров: зарплата, пол, семейное положение, наличие высшего образования, возраст, тип населенного пункта, длительность рабочей недели.

Из параметра, отвечающего семейному положению, сделать дамми-переменные (с помощью one-hot-encoding): 1) переменная `wed1` имеет значение 1 в случае, если респондент женат, 0 – в противном случае; 2) `wed2=1`, если респондент разведён или вдовец; 3) `wed3 = 1`, если респондент никогда не состоял в браке; 4) если считаете необходимым, введите другие параметры. Следите за мультиколлинеарностью (убедитесь в её отсутствии, оценив вспомогательную регрессию любого параметра (например, зарплаты или одного из параметров `wed`) на эти переменные и используя команду `VIF` для неё).

Из параметра `пол` сделайте переменную `sex`, имеющую значение 1 для мужчин и равную 0 для женщин.

Из параметра, отвечающего типу населённого пункта, создайте одну дамми-переменную `city_status` со значением 1 для города или областного центра, 0 – в противоположном случае.

Введите один параметр `higher_educ`, характеризующий наличие полного высшего образования.

Факторные переменные, «имеющие много значений», такие как: зарплата, длительность рабочей недели и возраст, - необходимо преобразовать в вещественные переменные и нормализовать их: вычесть среднее значение по этой переменной, разделить её значения на стандартное отклонение.

1. Постройте линейную регрессию зарплаты на все параметры, которые Вы выделили из данных мониторинга. Не забудьте оценить коэффициент вздутия дисперсии `VIF`.
2. Поэкспериментируйте с функциями вещественных параметров: используйте логарифм и степени (хотя бы от 0.1 до 2 с шагом 0.1).
3. Выделите наилучшие модели из построенных: по значимости параметров, включённых в зависимости, и по объяснённой с помощью построенных зависимостей разбросу $R^2 - R^2_{adj}$.
4. Сделайте вывод о том, какие индивиды получают наибольшую зарплату.
5. Оцените регрессии для подмножества индивидов, указанных в варианте.

Вариант	Номер волны выборки РМЭЗ	Подмножество для пункта 5
---------	--------------------------	---------------------------

	НИУ ВШЭ	
1.	24	Городские жители, состоящие в браке; разведенные, без высшего образования
2.	22	Женщины, не замужем; женщины, живущие в городе, разведённые
3.	21	Женатые, не из города; разведённые или не вступавшие в брак, с высшим образованием
4.	20	Не женатые мужчины, с высшим образованием; городские жители, состоящие в браке
5.	19	Городские жители, не состоявшие в браке; разведенные женщины, без высшего образования
6.	18	Женщины, не замужем; женщины, живущие в городе, разведённые
7.	17	С высшим образованием, не из города; женщины, с высшим образованием
8.	16	Не вступавшие в брак мужчины, без высшего образования; городские жители, мужчины состоящие в браке
9.	15	Городские жители, не состоящие в браке; разведенные женщины, с высшим образованием
10.	14	Мужчины, разведённые с высшим образованием; женщины, живущие в городе, состоящие в браке
11.	13	Женатые, не из города; мужчины без высшего образования
12.	12	Не состоявшие в браке мужчины, с высшим образованием; городские жители, состоящие в браке, женщины
13.	11	Городские жители, женщины, не состоявшие в браке; разведенные женщины, с высшим образованием
14.	27	Мужчины, женатые, с высшим образованием; мужчины, живущие в городе, разведённые
15.	11	Женатые, не из города; разведенные, с высшим образованием
16.	12	Не вступавшие в брак, без высшего образования; городские жители, состоящие в браке
17.	13	Городские жители, состоящие в браке; разведенные, без высшего образования
18.	14	Женщины, не замужем; женщины, живущие в городе, разведённые
19.	15	Женатые, не из города; разведённые или не вступавшие в брак, с высшим образованием
20.	16	Не женатые мужчины, с высшим образованием; городские жители, состоящие в браке
21.	17	Городские жители, не состоявшие в браке; разведенные женщины, без высшего образования
22.	18	Женщины, не замужем; женщины, живущие в городе, разведённые
23.	19	С высшим образованием, не из города; женщины, с высшим образованием
24.	20	Не вступавшие в брак мужчины, без высшего образования; городские жители, мужчины состоящие в браке
25.	21	Городские жители, не состоящие в браке; разведенные женщины, с высшим образованием
26.	22	Мужчины, разведённые с высшим образованием; женщины,

		живущие в городе, состоящие в браке
27.	23	Женатые, не из города; мужчины без высшего образования
28.	24	Не состоявшие в браке мужчины, с высшим образованием; городские жители, состоящие в браке, женщины
29.	27	Городские жители, женщины, не состоявшие в браке; разведенные женщины, с высшим образованием
30.	19	С высшим образованием, из города; мужчины, с высшим образованием