

Лабораторная работа №4.

Оценивание и верификация моделей линейной регрессии.

Часть I.

В таблице (см. файл **lab_4_DATA_2024.xls**) содержатся данные о работниках: **ЗАРПЛАТА**, **СТАЖ РАБОТЫ**, **ВРЕМЯ**, **ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ**, и **ПОЛ РАБОТНИКА**.

Выберите из файла **lab_4_DATA_2024.xls** данные n наблюдений, где объем выборки n приведен в таблице **ВАРИАНТЫ ЗАДАНИЙ**.

Далее следует оценивать модели по выборке из n наблюдений.

1. Найдите оценки параметров линейной модели зависимости **ЗАРПЛАТЫ** от фактора **Z** (см. **ВАРИАНТЫ ЗАДАНИЙ**).
2. Проверьте на уровне значимости α , обладает ли построенная модель высокой объясняющей способностью, или же модель низкого качества.
3. Найдите оценки параметров линейной модели зависимости **ЗАРПЛАТЫ** от **СТАЖА РАБОТЫ** и **ВРЕМЕНИ, ЗАТРАЧЕННОГО НА ОБРАЗОВАНИЕ**.
4. На сколько в среднем изменится **ЗАРПЛАТА**, если **СТАЖ РАБОТЫ** увеличить на 1 год, а **ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ**, останется неизменным?
5. Какую в среднем прибавку к **ЗАРПЛАТЕ** дает дополнительный год обучения?
6. Проверьте на уровне значимости α , обладает ли построенная двухфакторная модель высокой объясняющей способностью, или же модель низкого качества.
7. С надежностью γ постройте доверительные интервалы для теоретических коэффициентов регрессии модели.

Часть II.

1. Найдите оценки параметров линейной модели зависимости **ЗАРПЛАТЫ** от **СТАЖА РАБОТЫ**, **ВРЕМЕНИ, ЗАТРАЧЕННОГО НА ОБРАЗОВАНИЕ**, и **ПОЛА РАБОТНИКА**.
2. Проверьте на уровне значимости α , обладает ли построенная трехфакторная модель высокой объясняющей способностью, или же эта модель низкого качества.
3. Существенно ли на уровне значимости α различаются зарплаты мужчины и женщины при прочих равных условиях (т.е. равных стаже и числу лет, затраченных на образование)?
6. Проверьте на уровне значимости α , отражает ли полученное при оценке регрессии значение коэффициента при факторе **СТАЖ РАБОТЫ** истинную зависимость, или же оно появилось случайно?
7. Проверьте на уровне значимости α , отражает ли полученное при оценке регрессии значение коэффициента при факторе **ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ**, истинную зависимость, или же оно появилось случайно?
8. С надежностью γ найдите интервальную оценку прогнозной **ЗАРПЛАТЫ** работника с указанной в варианте гендерной принадлежностью, если его **СТАЖ РАБОТЫ** составляет a лет, а **ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ** - b лет.
9. Как в среднем изменится **ЗАРПЛАТА**, если **СТАЖ РАБОТЫ** увеличится на 2 года, **ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ**, останется неизменным?
10. Можно ли считать, что при начислении зарплаты имеет место дискриминация по гендерному признаку?

ВАРИАНТЫ ЗАДАНИЙ.

Вариант	<i>n</i>	Наименование показателя <i>Z</i>	γ	α	Пол, а, b
<u>1</u>	500	СТАЖ РАБОТЫ (лет)	0,99	0,01	1,1,8
<u>2</u>	525	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,98	0,02	0,10,9
<u>3</u>	450	ПОЛ (1-жен, 0- муж)	0,97	0,03	0,2,10
<u>4</u>	490	СТАЖ РАБОТЫ (лет)	0,975	0,025	1,11,12
<u>5</u>	515	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,96	0,04	0,4,12
<u>6</u>	480	ПОЛ (1-жен, 0- муж)	0,955	0,045	1,7,10
<u>7</u>	490	СТАЖ РАБОТЫ (лет)	0,95	0,05	0,8,13
<u>8</u>	525	ПОЛ (1-жен, 0- муж) (лет)	0,94	0,06	1,4,14
<u>9</u>	460	СТАЖ РАБОТЫ (лет)	0,935	0,065	0,6,12
<u>10</u>	510	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,93	0,07	1,11,12
<u>11</u>	520	ПОЛ (1-жен, 0- муж)	0,925	0,075	0,4,12
<u>12</u>	435	СТАЖ РАБОТЫ (лет)	0,92	0,08	1,7,12
<u>13</u>	485	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,915	0,085	0,12,10
<u>14</u>	505	ПОЛ (1-жен, 0- муж)	0,91	0,09	1,4,10
<u>15</u>	495	СТАЖ РАБОТЫ (лет)	0,99	0,01	1,5,13
<u>16</u>	475	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,985	0,015	0,4,15
<u>17</u>	505	ПОЛ (1-жен, 0- муж)	0,98	0,02	1,7,11
<u>18</u>	485	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,975	0,025	1,5,10
<u>19</u>	535	СТАЖ РАБОТЫ (лет)	0,97	0,03	0,14,15
<u>20</u>	530	ПОЛ (1-жен, 0- муж)	0,965	0,035	0,14,18
<u>21</u>	480	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,96	0,04	0,14,13
<u>22</u>	495	СТАЖ РАБОТЫ (лет)	0,955	0,045	1,4,12
<u>23</u>	510	ПОЛ (1-жен, 0- муж)	0,93	0,07	0,2,15

<u>24</u>	466	СТАЖ РАБОТЫ (лет)	0,985	0,015	1,14,12
<u>25</u>	499	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,955	0,015	1,5,16
<u>26</u>	509	ПОЛ (1-жен, 0- муж)	0,975	0,02	0,11,14
<u>27</u>	469	СТАЖ РАБОТЫ (лет)	0,935	0,035	0,6,10
<u>28</u>	511	ПОЛ (1-жен, 0- муж)	0,925	0,06	1,3,16
<u>29</u>	529	СТАЖ РАБОТЫ (лет)	0,965	0,04	1,12,13
<u>30</u>	433	ВРЕМЯ, ЗАТРАЧЕННОЕ НА ОБРАЗОВАНИЕ (лет)	0,945	0,025	0,5,11
<u>31</u>	489	ПОЛ (1-жен, 0- муж)	0,97	0,075	1,7,9

Замечание.

Примеры оценивания линейных моделей методом наименьших квадратов на Питоне

См.

https://www.simplilearn.com/tutorials/scikit-learn-tutorial/sklearn-linear-regression-with-examples#what_is_sklearn_