

РГПУ им. А.И. Герцена

К работе допущены _____
Работа выполнена _____
Отчёт сдан _____

Отчет по лабораторной работе №8(часть2)

«Регрессионный анализ: модели и методы»

Работу выполнили: Леонтьева А.В.,
Шадрин А.В.,
Тихонова Э.К.,
Столяренко К.А.

Факультет: ИИТиТО
Группа: 2об-ИВТ-1/20

Санкт-Петербург, 2022 г

Лабораторная работа №8.2

Регрессионный анализ: модели и методы

Выполнил: студент 2 курса РГПУ им. Герцена, ИВТ 1/2, Шадрин А.В.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Оценить качество регрессионной модели, полученной в лабораторной работе №8.1.

- 1.1 Вычислить коэффициент эластичности
- 1.2 Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента
- 1.3 Определить статистическую надежность результатов регрессионного анализа с использованием критерия F-Фишера
- 1.4 Вычислить прогнозное значение результативного признака

Решение:

Для проведения всех расчетов строится вспомогательная таблица:

| № п/п | x | y | x ² | y ² | xy | y' | y-y' | (y-y') ² | A= (y-y')/y |
|------------------|-----------|----------|----------------|----------------|----------|----------|----------|---------------------|--------------|
| 1 | 33 | 13,8 | 1089 | 190,44 | 455,4 | 14,73412 | -0,93412 | 0,872581 | 0,067689899 |
| 2 | 40 | 13,8 | 1600 | 190,44 | 552 | 16,84681 | -3,04681 | 9,283045 | 0,220783264 |
| 3 | 36 | 14 | 1296 | 196 | 504 | 15,63956 | -1,63956 | 2,688152 | 0,117111322 |
| 4 | 60 | 22,5 | 3600 | 506,25 | 1350 | 22,88306 | -0,38306 | 0,146736 | 0,017024966 |
| 5 | 55 | 24 | 3025 | 576 | 1320 | 21,374 | 2,626001 | 6,895884 | 0,109416726 |
| 6 | 80 | 28 | 6400 | 784 | 2240 | 28,91931 | -0,91931 | 0,845139 | 0,032832658 |
| 7 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 8 | 70 | 20,9 | 4900 | 436,81 | 1463 | 25,90119 | -5,00119 | 25,01188 | 0,239291296 |
| 9 | 48 | 22 | 2304 | 484 | 1056 | 19,26131 | 2,73869 | 7,500422 | 0,124485904 |
| 10 | 53 | 21,5 | 2809 | 462,25 | 1139,5 | 20,77037 | 0,729627 | 0,532355 | 0,033936126 |
| 11 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 12 | 75 | 35 | 5625 | 1225 | 2625 | 27,41025 | 7,589749 | 57,60429 | 0,216849964 |
| 13 | 63 | 24 | 3969 | 576 | 1512 | 23,7885 | 0,2115 | 0,044732 | 0,008812515 |
| 14 | 112 | 37,9 | 12544 | 1436,41 | 4244,8 | 38,57732 | -0,67732 | 0,458761 | 0,017871207 |
| 15 | 70 | 27,5 | 4900 | 756,25 | 1925 | 25,90119 | 1,598812 | 2,5562 | 0,058138615 |
| Итого | 985 | 368,9 | 72111 | 9867,85 | 26466,7 | 368,9 | 4,62E-14 | 118,6249 | 1,354650959 |
| Среднее значение | 65,666667 | 24,59333 | 4807,4 | 657,8567 | 1764,447 | 24,59333 | 3,08E-15 | 7,908328 | 0,090310064 |

1.1 Вычисление коэффициента эластичности.

При линейной форме связи средний коэффициент эластичности находится по формуле:

$$\varepsilon = b_1 \times \frac{\bar{X}}{\bar{Y}}$$

\bar{X} и \bar{Y} – средние значения признаков.

$$\varepsilon = 0.318 \times \frac{65.667}{24.593} = 0.806$$

Коэффициент эластичности показывает, что при увеличении общей площади квартиры на 1% её стоимость возрастает на 0.806%.

1.2 Оценка значимости коэффициентов корреляции и регрессии по критерию t-Стьюдента.

При линейной зависимости, степень тесноты связи между переменными X и Y определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$$

Где σ_x и σ_y – средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{4807.4 - 65.667^2} = 22.254$$

$$\sigma_y = \sqrt{\overline{Y^2} - (\bar{Y})^2} = \sqrt{657.857 - 24.593^2} = 7.283$$

$$r = \frac{1764.447 - 65.667 * 24.593}{22.254 * 7.283} = 0.922$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0.922^2 = 0.850$ показывает, что 85% различий в стоимости квартир объясняется вариацией из общей площади, а 15% - другими, неучтенными факторами

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции.

Выдвигаем нулевую гипотезу H_0 : коэффициент корреляции в генеральной совокупности равен нулю, и изучаемый фактор не оказывает существенного влияния на резульативный признак.

$H_0: r_s = 0$, при $H_1: r_s \neq 0$

Для проверки нулевой гипотезы применим t-критерий Стьюдента.

Найдем расчетное значение t-критерия:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.922}{\sqrt{\frac{1-0.922^2}{15-2}}} = 8.58$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне $\alpha=0.05$ и числе степеней свободы $k=n-2=15-2=13$

Для двусторонней критической области $t_{\text{кр}}=2.16$

Используйте необходимую таблицу.

Сравним $t_{\text{расч}}$ с $t_{\text{кр}}$. Так как $t_{\text{расч}} > t_{\text{кр}}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием t-критерия Стьюдента.

Находится расчетное значение критерия:

$$t_{\text{расч}} = \frac{b_1}{m_{b_1}}$$
$$m_{b_1} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sum(x - \bar{X})^2}} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sigma_x^2 * n}} = \sqrt{\frac{118.625}{(15-2) * 22.254^2 * 15}}$$
$$t_{\text{расч}} = \frac{0.3018}{0.035} = 8.62$$

Критическое значение t также равно 2.16. Так как $t_{\text{расч}} > t_{\text{кр}}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

1.3 Нахождение статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера.

Расчетное (фактическое) значение F-критерия находится по формуле:

$$F_{\text{расч}} = \frac{\sum(y' - \bar{y})^2 / k}{\sum(y' - \bar{y})^2 / (n - k - 1)}$$

где k – число параметров при переменных X.

Если применяется линейное уравнение регрессии, то $F_{\text{расч}}$ упрощается.

$$F_{\text{расч}} = \frac{r^2}{1 - r^2} (n - 2) = \frac{0.85}{1 - 0.85} * 13 = 73.67$$

Так как $F_{\text{расч}} > F_{\text{кр}}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты. Так как $t_{b_1}^2 = t_r^2 = F_{\alpha, k_1, k_2}$

1.4 Вычисление прогнозного значения результативного признака.

Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p)

По условию $x_p = \bar{x} * 1.2 = 65.667 = 78.8$

Тогда прогнозное значение стоимости квартиры составит

$$y'_p = b_0 + b_1 * x_p = 4.7743 + 0.3018 * 78.8 = 28.56$$

Значит при общей площади квартиры 78.8 м^2 возможная стоимость квартиры составляет 25.56 тыс. у.е.

Вывод: в ходе лабораторной работы было оценено качество регрессионной модели. Был вычислен коэффициент эластичности $\Theta = 0.806$. Была оценена значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента, был подтверждён вывод о значимости влияния общей площади на стоимость квартир. Был подтверждён вывод о статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера. Было вычислено прогнозное значение результативного признака $y'_p = 28.56$.

Лабораторная работа №8.2

Регрессионный анализ: модели и методы

Выполнил: студент 2 курса РГПУ им. Герцена, ИВТ 1/2, Столяренко К.А.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Оценить качество регрессионной модели, полученной в лабораторной работе №8.1.

- 1.1 Вычислить коэффициент эластичности
- 1.2 Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента
- 1.3 Определить статистическую надежность результатов регрессионного анализа с использованием критерия F-Фишера
- 1.4 Вычислить прогнозное значение результативного признака

Решение:

Для проведения всех расчетов строится вспомогательная таблица:

| № п/п | x | y | x ² | y ² | xy | y' | y-y' | (y-y') ² | A= (y-y')/y |
|------------------|-----------|----------|----------------|----------------|----------|----------|----------|---------------------|--------------|
| 1 | 33 | 13,8 | 1089 | 190,44 | 455,4 | 14,73412 | -0,93412 | 0,872581 | 0,067689899 |
| 2 | 40 | 13,8 | 1600 | 190,44 | 552 | 16,84681 | -3,04681 | 9,283045 | 0,220783264 |
| 3 | 36 | 14 | 1296 | 196 | 504 | 15,63956 | -1,63956 | 2,688152 | 0,117111322 |
| 4 | 60 | 22,5 | 3600 | 506,25 | 1350 | 22,88306 | -0,38306 | 0,146736 | 0,017024966 |
| 5 | 55 | 24 | 3025 | 576 | 1320 | 21,374 | 2,626001 | 6,895884 | 0,109416726 |
| 6 | 80 | 28 | 6400 | 784 | 2240 | 28,91931 | -0,91931 | 0,845139 | 0,032832658 |
| 7 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 8 | 70 | 20,9 | 4900 | 436,81 | 1463 | 25,90119 | -5,00119 | 25,01188 | 0,239291296 |
| 9 | 48 | 22 | 2304 | 484 | 1056 | 19,26131 | 2,73869 | 7,500422 | 0,124485904 |
| 10 | 53 | 21,5 | 2809 | 462,25 | 1139,5 | 20,77037 | 0,729627 | 0,532355 | 0,033936126 |
| 11 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 12 | 75 | 35 | 5625 | 1225 | 2625 | 27,41025 | 7,589749 | 57,60429 | 0,216849964 |
| 13 | 63 | 24 | 3969 | 576 | 1512 | 23,7885 | 0,2115 | 0,044732 | 0,008812515 |
| 14 | 112 | 37,9 | 12544 | 1436,41 | 4244,8 | 38,57732 | -0,67732 | 0,458761 | 0,017871207 |
| 15 | 70 | 27,5 | 4900 | 756,25 | 1925 | 25,90119 | 1,598812 | 2,5562 | 0,058138615 |
| Итого | 985 | 368,9 | 72111 | 9867,85 | 26466,7 | 368,9 | 4,62E-14 | 118,6249 | 1,354650959 |
| Среднее значение | 65,666667 | 24,59333 | 4807,4 | 657,8567 | 1764,447 | 24,59333 | 3,08E-15 | 7,908328 | 0,090310064 |

1.1 Вычисление коэффициента эластичности.

При линейной форме связи средний коэффициент эластичности находится по формуле:

$$\varepsilon = b_1 \times \frac{\bar{X}}{\bar{Y}}$$

\bar{X} и \bar{Y} – средние значения признаков.

$$\varepsilon = 0.318 \times \frac{65.667}{24.593} = 0.806$$

Коэффициент эластичности показывает, что при увеличении общей площади квартиры на 1% её стоимость возрастает на 0.806%.

1.2 Оценка значимости коэффициентов корреляции и регрессии по критерию t-Стьюдента.

При линейной зависимости, степень тесноты связи между переменными X и Y определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$$

Где σ_x и σ_y – средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{4807.4 - 65.667^2} = 22.254$$

$$\sigma_y = \sqrt{\overline{Y^2} - (\bar{Y})^2} = \sqrt{657.857 - 24.593^2} = 7.283$$

$$r = \frac{1764.447 - 65.667 * 24.593}{22.254 * 7.283} = 0.922$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0.922^2 = 0.850$ показывает, что 85% различий в стоимости квартир объясняется вариацией из общей площади, а 15% - другими, неучтенными факторами

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции.

Выдвигаем нулевую гипотезу H_0 : коэффициент корреляции в генеральной совокупности равен нулю, и изучаемый фактор не оказывает существенного влияния на резульативный признак.

$H_0: r_s = 0$, при $H_1: r_s \neq 0$

Для проверки нулевой гипотезы применим t-критерий Стьюдента.

Найдем расчетное значение t-критерия:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.922}{\sqrt{\frac{1-0.922^2}{15-2}}} = 8.58$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне $\alpha=0.05$ и числе степеней свободы $k=n-2=15-2=13$

Для двусторонней критической области $t_{\text{кр}}=2.16$

Используйте необходимую таблицу.

Сравним $t_{\text{расч}}$ с $t_{\text{кр}}$. Так как $t_{\text{расч}} > t_{\text{кр}}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием t-критерия Стьюдента.

Находится расчетное значение критерия:

$$t_{\text{расч}} = \frac{b_1}{m_{b_1}}$$
$$m_{b_1} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sum(x - \bar{X})^2}} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sigma_x^2 * n}} = \sqrt{\frac{118.625}{(15-2) * 22.254^2 * 15}}$$
$$t_{\text{расч}} = \frac{0.3018}{0.035} = 8.62$$

Критическое значение t также равно 2.16. Так как $t_{\text{расч}} > t_{\text{кр}}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

1.3 Нахождение статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера.

Расчетное (фактическое) значение F-критерия находится по формуле:

$$F_{\text{расч}} = \frac{\sum(y' - \bar{y})^2 / k}{\sum(y' - \bar{y})^2 / (n - k - 1)}$$

где k – число параметров при переменных X.

Если применяется линейное уравнение регрессии, то $F_{\text{расч}}$ упрощается.

$$F_{\text{расч}} = \frac{r^2}{1 - r^2} (n - 2) = \frac{0.85}{1 - 0.85} * 13 = 73.67$$

Так как $F_{\text{расч}} > F_{\text{кр}}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты. Так как $t_{b_1}^2 = t_r^2 = F_{\alpha, k_1, k_2}$

1.4 Вычисление прогнозного значения результативного признака.

Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p)

По условию $x_p = \bar{x} * 1.2 = 65.667 = 78.8$

Тогда прогнозное значение стоимости квартиры составит

$$y'_p = b_0 + b_1 * x_p = 4.7743 + 0.3018 * 78.8 = 28.56$$

Значит при общей площади квартиры 78.8 м² возможная стоимость квартиры составляет 25.56 тыс. у.е.

Вывод: в ходе лабораторной работы было оценено качество регрессионной модели. Был вычислен коэффициент эластичности $\mathcal{E} = 0.806$. Была оценена значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента, был подтверждён вывод о значимости влияния общей площади на стоимость квартир. Был подтверждён вывод о статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера. Было вычислено прогнозное значение результативного признака $y'_p = 28.56$.

Лабораторная работа №8.2

Регрессионный анализ: модели и методы

Выполнил: студентка 2 курса РГПУ им. Герцена, ИВТ 1/2, Леонтьева А.В.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Оценить качество регрессионной модели, полученной в лабораторной работе №8.1.

- 1.1 Вычислить коэффициент эластичности
- 1.2 Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента
- 1.3 Определить статистическую надежность результатов регрессионного анализа с использованием критерия F-Фишера
- 1.4 Вычислить прогнозное значение результативного признака

Решение:

Для проведения всех расчетов строится вспомогательная таблица:

| № п/п | x | y | x ² | y ² | xy | y' | y-y' | (y-y') ² | A= (y-y')/y |
|------------------|-----------|----------|----------------|----------------|----------|----------|----------|---------------------|--------------|
| 1 | 33 | 13,8 | 1089 | 190,44 | 455,4 | 14,73412 | -0,93412 | 0,872581 | 0,067689899 |
| 2 | 40 | 13,8 | 1600 | 190,44 | 552 | 16,84681 | -3,04681 | 9,283045 | 0,220783264 |
| 3 | 36 | 14 | 1296 | 196 | 504 | 15,63956 | -1,63956 | 2,688152 | 0,117111322 |
| 4 | 60 | 22,5 | 3600 | 506,25 | 1350 | 22,88306 | -0,38306 | 0,146736 | 0,017024966 |
| 5 | 55 | 24 | 3025 | 576 | 1320 | 21,374 | 2,626001 | 6,895884 | 0,109416726 |
| 6 | 80 | 28 | 6400 | 784 | 2240 | 28,91931 | -0,91931 | 0,845139 | 0,032832658 |
| 7 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 8 | 70 | 20,9 | 4900 | 436,81 | 1463 | 25,90119 | -5,00119 | 25,01188 | 0,239291296 |
| 9 | 48 | 22 | 2304 | 484 | 1056 | 19,26131 | 2,73869 | 7,500422 | 0,124485904 |
| 10 | 53 | 21,5 | 2809 | 462,25 | 1139,5 | 20,77037 | 0,729627 | 0,532355 | 0,033936126 |
| 11 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 12 | 75 | 35 | 5625 | 1225 | 2625 | 27,41025 | 7,589749 | 57,60429 | 0,216849964 |
| 13 | 63 | 24 | 3969 | 576 | 1512 | 23,7885 | 0,2115 | 0,044732 | 0,008812515 |
| 14 | 112 | 37,9 | 12544 | 1436,41 | 4244,8 | 38,57732 | -0,67732 | 0,458761 | 0,017871207 |
| 15 | 70 | 27,5 | 4900 | 756,25 | 1925 | 25,90119 | 1,598812 | 2,5562 | 0,058138615 |
| Итого | 985 | 368,9 | 72111 | 9867,85 | 26466,7 | 368,9 | 4,62E-14 | 118,6249 | 1,354650959 |
| Среднее значение | 65,666667 | 24,59333 | 4807,4 | 657,8567 | 1764,447 | 24,59333 | 3,08E-15 | 7,908328 | 0,090310064 |

1.1 Вычисление коэффициента эластичности.

При линейной форме связи средний коэффициент эластичности находится по формуле:

$$\varepsilon = b_1 \times \frac{\bar{X}}{\bar{Y}}$$

\bar{X} и \bar{Y} – средние значения признаков.

$$\varepsilon = 0.318 \times \frac{65.667}{24.593} = 0.806$$

Коэффициент эластичности показывает, что при увеличении общей площади квартиры на 1% её стоимость возрастает на 0.806%.

1.2 Оценка значимости коэффициентов корреляции и регрессии по критерию t-Стьюдента.

При линейной зависимости, степень тесноты связи между переменными X и Y определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$$

Где σ_x и σ_y – средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{4807.4 - 65.667^2} = 22.254$$

$$\sigma_y = \sqrt{\overline{Y^2} - (\bar{Y})^2} = \sqrt{657.857 - 24.593^2} = 7.283$$

$$r = \frac{1764.447 - 65.667 * 24.593}{22.254 * 7.283} = 0.922$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0.922^2 = 0.850$ показывает, что 85% различий в стоимости квартир объясняется вариацией из общей площади, а 15% - другими, неучтенными факторами

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции.

Выдвигаем нулевую гипотезу H_0 : коэффициент корреляции в генеральной совокупности равен нулю, и изучаемый фактор не оказывает существенного влияния на резульативный признак.

$H_0: r_s = 0$, при $H_1: r_s \neq 0$

Для проверки нулевой гипотезы применим t-критерий Стьюдента.

Найдем расчетное значение t-критерия:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.922}{\sqrt{\frac{1-0.922^2}{15-2}}} = 8.58$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне $\alpha=0.05$ и числе степеней свободы $k=n-2=15-2=13$

Для двусторонней критической области $t_{\text{кр}}=2.16$

Используйте необходимую таблицу.

Сравним $t_{\text{расч}}$ с $t_{\text{кр}}$. Так как $t_{\text{расч}} > t_{\text{кр}}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием t-критерия Стьюдента.

Находится расчетное значение критерия:

$$t_{\text{расч}} = \frac{b_1}{m_{b_1}}$$
$$m_{b_1} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sum(x - \bar{X})^2}} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sigma_x^2 * n}} = \sqrt{\frac{118.625}{(15-2) * 22.254^2 * 15}}$$
$$t_{\text{расч}} = \frac{0.3018}{0.035} = 8.62$$

Критическое значение t также равно 2.16. Так как $t_{\text{расч}} > t_{\text{кр}}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

1.3 Нахождение статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера.

Расчетное (фактическое) значение F-критерия находится по формуле:

$$F_{\text{расч}} = \frac{\sum(y' - \bar{y})^2 / k}{\sum(y' - \bar{y})^2 / (n - k - 1)}$$

где k – число параметров при переменных X.

Если применяется линейное уравнение регрессии, то $F_{\text{расч}}$ упрощается.

$$F_{\text{расч}} = \frac{r^2}{1 - r^2} (n - 2) = \frac{0.85}{1 - 0.85} * 13 = 73.67$$

Так как $F_{\text{расч}} > F_{\text{кр}}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты. Так как $t_{b_1}^2 = t_r^2 = F_{\alpha, k_1, k_2}$

1.4 Вычисление прогнозного значения результативного признака.

Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p)

По условию $x_p = \bar{x} * 1.2 = 65.667 = 78.8$

Тогда прогнозное значение стоимости квартиры составит

$$y'_p = b_0 + b_1 * x_p = 4.7743 + 0.3018 * 78.8 = 28.56$$

Значит при общей площади квартиры 78.8 м² возможная стоимость квартиры составляет 28.56 тыс. у.е.

Вывод: в ходе лабораторной работы было оценено качество регрессионной модели. Был вычислен коэффициент эластичности $\varepsilon = 0.806$. Была оценена значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента, был подтверждён вывод о значимости влияния общей площади на стоимость квартир. Был подтверждён вывод о статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера. Было вычислено прогнозное значение результативного признака $y'_p = 28.56$.

Лабораторная работа №8.2

Регрессионный анализ: модели и методы

Выполнил: студентка 2 курса РГПУ им. Герцена, ИВТ 1/2, Тихонова Э.К.

Оборудование: ПК, Excel, Word

Задание 1

Постановка задачи:

Оценить качество регрессионной модели, полученной в лабораторной работе №8.1.

- 1.1 Вычислить коэффициент эластичности
- 1.2 Оценить значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента
- 1.3 Определить статистическую надежность результатов регрессионного анализа с использованием критерия F-Фишера
- 1.4 Вычислить прогнозное значение результативного признака

Решение:

Для проведения всех расчетов строится вспомогательная таблица:

| № п/п | x | y | x ² | y ² | xy | y' | y-y' | (y-y') ² | A= (y-y')/y |
|------------------|-----------|----------|----------------|----------------|----------|----------|----------|---------------------|--------------|
| 1 | 33 | 13,8 | 1089 | 190,44 | 455,4 | 14,73412 | -0,93412 | 0,872581 | 0,067689899 |
| 2 | 40 | 13,8 | 1600 | 190,44 | 552 | 16,84681 | -3,04681 | 9,283045 | 0,220783264 |
| 3 | 36 | 14 | 1296 | 196 | 504 | 15,63956 | -1,63956 | 2,688152 | 0,117111322 |
| 4 | 60 | 22,5 | 3600 | 506,25 | 1350 | 22,88306 | -0,38306 | 0,146736 | 0,017024966 |
| 5 | 55 | 24 | 3025 | 576 | 1320 | 21,374 | 2,626001 | 6,895884 | 0,109416726 |
| 6 | 80 | 28 | 6400 | 784 | 2240 | 28,91931 | -0,91931 | 0,845139 | 0,032832658 |
| 7 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 8 | 70 | 20,9 | 4900 | 436,81 | 1463 | 25,90119 | -5,00119 | 25,01188 | 0,239291296 |
| 9 | 48 | 22 | 2304 | 484 | 1056 | 19,26131 | 2,73869 | 7,500422 | 0,124485904 |
| 10 | 53 | 21,5 | 2809 | 462,25 | 1139,5 | 20,77037 | 0,729627 | 0,532355 | 0,033936126 |
| 11 | 95 | 32 | 9025 | 1024 | 3040 | 33,4465 | -1,4465 | 2,092374 | 0,045203248 |
| 12 | 75 | 35 | 5625 | 1225 | 2625 | 27,41025 | 7,589749 | 57,60429 | 0,216849964 |
| 13 | 63 | 24 | 3969 | 576 | 1512 | 23,7885 | 0,2115 | 0,044732 | 0,008812515 |
| 14 | 112 | 37,9 | 12544 | 1436,41 | 4244,8 | 38,57732 | -0,67732 | 0,458761 | 0,017871207 |
| 15 | 70 | 27,5 | 4900 | 756,25 | 1925 | 25,90119 | 1,598812 | 2,5562 | 0,058138615 |
| Итого | 985 | 368,9 | 72111 | 9867,85 | 26466,7 | 368,9 | 4,62E-14 | 118,6249 | 1,354650959 |
| Среднее значение | 65,666667 | 24,59333 | 4807,4 | 657,8567 | 1764,447 | 24,59333 | 3,08E-15 | 7,908328 | 0,090310064 |

1.1 Вычисление коэффициента эластичности.

При линейной форме связи средний коэффициент эластичности находится по формуле:

$$\varepsilon = b_1 \times \frac{\bar{X}}{\bar{Y}}$$

\bar{X} и \bar{Y} – средние значения признаков.

$$\varepsilon = 0.318 \times \frac{65.667}{24.593} = 0.806$$

Коэффициент эластичности показывает, что при увеличении общей площади квартиры на 1% её стоимость возрастает на 0.806%.

1.2 Оценка значимости коэффициентов корреляции и регрессии по критерию t-Стьюдента.

При линейной зависимости, степень тесноты связи между переменными X и Y определяется с помощью коэффициента корреляции:

$$r = \frac{\overline{XY} - \bar{X}\bar{Y}}{\sigma_x \sigma_y}$$

Где σ_x и σ_y – средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - (\bar{X})^2} = \sqrt{4807.4 - 65.667^2} = 22.254$$

$$\sigma_y = \sqrt{\overline{Y^2} - (\bar{Y})^2} = \sqrt{657.857 - 24.593^2} = 7.283$$

$$r = \frac{1764.447 - 65.667 * 24.593}{22.254 * 7.283} = 0.922$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0.922^2 = 0.850$ показывает, что 85% различий в стоимости квартир объясняется вариацией из общей площади, а 15% - другими, неучтенными факторами

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции.

Выдвигаем нулевую гипотезу H_0 : коэффициент корреляции в генеральной совокупности равен нулю, и изучаемый фактор не оказывает существенного влияния на резульативный признак.

$H_0: r_s = 0$, при $H_1: r_s \neq 0$

Для проверки нулевой гипотезы применим t-критерий Стьюдента.

Найдем расчетное значение t-критерия:

$$t_{\text{расч}} = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.922}{\sqrt{\frac{1-0.922^2}{15-2}}} = 8.58$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне $\alpha=0.05$ и числе степеней свободы $k=n-2=15-2=13$

Для двусторонней критической области $t_{\text{кр}}=2.16$

Используйте необходимую таблицу.

Сравним $t_{\text{расч}}$ с $t_{\text{кр}}$. Так как $t_{\text{расч}} > t_{\text{кр}}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием t-критерия Стьюдента.

Находится расчетное значение критерия:

$$t_{\text{расч}} = \frac{b_1}{m_{b_1}}$$
$$m_{b_1} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sum(x - \bar{X})^2}} = \sqrt{\frac{\sum(y - y')^2}{(n-2) * \sigma_x^2 * n}} = \sqrt{\frac{118.625}{(15-2) * 22.254^2 * 15}}$$
$$t_{\text{расч}} = \frac{0.3018}{0.035} = 8.62$$

Критическое значение t также равно 2.16. Так как $t_{\text{расч}} > t_{\text{кр}}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

1.3 Нахождение статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера.

Расчетное (фактическое) значение F-критерия находится по формуле:

$$F_{\text{расч}} = \frac{\sum(y' - \bar{y})^2 / k}{\sum(y' - \bar{y})^2 / (n - k - 1)}$$

где k – число параметров при переменных X.

Если применяется линейное уравнение регрессии, то $F_{\text{расч}}$ упрощается.

$$F_{\text{расч}} = \frac{r^2}{1 - r^2} (n - 2) = \frac{0.85}{1 - 0.85} * 13 = 73.67$$

Так как $F_{\text{расч}} > F_{\text{кр}}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты. Так как $t_{b_1}^2 = t_r^2 = F_{\alpha, k_1, k_2}$

1.4 Вычисление прогнозного значения результативного признака.

Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p)

По условию $x_p = \bar{x} * 1.2 = 65.667 = 78.8$

Тогда прогнозное значение стоимости квартиры составит

$$y'_p = b_0 + b_1 * x_p = 4.7743 + 0.3018 * 78.8 = 28.56$$

Значит при общей площади квартиры 78.8 м^2 возможная стоимость квартиры составляет 25.56 тыс. у.е.

Вывод: в ходе лабораторной работы было оценено качество регрессионной модели. Был вычислен коэффициент эластичности $\Theta = 0.806$. Была оценена значимость коэффициентов корреляции и регрессии по критерию t-Стьюдента, был подтверждён вывод о значимости влияния общей площади на стоимость квартир. Был подтверждён вывод о статистической надежности результатов регрессионного анализа с использованием критерия F-Фишера. Было вычислено прогнозное значение результативного признака $y'_p = 28.56$.