

3. Оценка статистической значимости коэффициентов регрессии и коэффициента корреляции с помощью F и t- тестов

(первая часть РГР)

Пример 2. В таблице приведена динамика экономических показателей России: валовой внутренний продукт РФ (в процентах к предыдущему году) – показатель y (ВВП) и капитальные вложения в основные фонды РФ (в процентах к предыдущему году) – фактор x (КВОФ). Требуется оценить влияние фактора x на результативный признак y .

| Год | ВВП (y,%) | КВОФ (x,%) |
|------------|-------------------------------|--------------------------------|
| 1991 | 95 | 85 |
| 1992 | 85,5 | 60 |
| 1993 | 91,3 | 88 |
| 1994 | 87,3 | 76 |
| 1995 | 95,8 | 90 |
| 1996 | 94 | 82 |
| 1997 | 100,4 | 95 |
| 1998 | 95,1 | 88 |
| 1999 | 104,6 | 105,3 |
| 2000 | 109,9 | 117,4 |
| 2001 | 105 | 108,7 |
| 2002 | 104,3 | 109,9 |

Имеем уравнение ПЛР:

$$Y = 55,9 + 0,45X$$

- где: 0,45 – коэффициент регрессии означает, что при увеличении признака X на 1% изменение результирующего признака Y в среднем составит рост на 0,45%, т.к. имеет место **сильная прямая линейная связь** между темпом роста валового внутреннего продукта РФ (ВВП) - Y и темпом роста капитальных вложений в основные фонды РФ (КВОФ) – X .

3. Оценим качество УПЛР и значимость его коэффициентов

3.1. Для оценки качества линейной регрессии рассчитаем коэффициент детерминации

-алгебраически:

$$R^2 = r^2 = (0,97)^2 = 0,94$$

-статистически

$$R^2 = 0,93$$

Вывод:

Коэффициент детерминации показывает, что 93 % различий в объеме ВВП по годам объясняется вариацией капиталовложений - X , а оставшиеся 7 % объясняются другими неучтенными факторами.

3.2. Охарактеризуем **статистическую надежность** результатов регрессионного анализа с помощью **F-критерия (F-тест)** на уровне значимости 0,05:

- Вычислим алгебраически:

$$F_{\text{факт}} = \frac{R^2}{1 - R^2}(n-2) =$$

- Определим статистически :

$$F_{\text{факт}} =$$

- Сравним $F_{\text{факт}}$ с $F_{\text{табл}}(\alpha; k_1; k_2) = 4,96$

где $k_1=1; k_2=12-1-1=10,$

- Вывод: $F_{\text{факт}} > F_{\text{табл}}$, т.е. **отвергается H_0** о статистической незначимости уравнения регрессии и показателя тесноты связи коэффициента - **В**
- Имеется существенная связь между признаками уравнения с

3.3. Оценим **значимость коэффициентов** регрессии и корреляции с помощью **t-критерия Стьюдента** при уровне значимости 0,05 (с расчетом доверительных интервалов каждого из показателей). Также выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля.

- Оценка значимости параметров (коэффициентов) регрессии и коэффициента корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки ***m***:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r}{m_r}.$$

- **Случайные ошибки** – m - параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_{\text{ост}}^2}{\sum (x - \bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x \sqrt{n}};$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_{\text{ост}}^2 \frac{\sum x^2}{n^2 \sigma_x^2}} = S_{\text{ост}} \frac{\sqrt{\sum x^2}}{n \sigma_x};$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}}.$$

Сравнивая фактическое и критическое (табличное) значения t -статистики, принимаем или отвергаем гипотезу H_0 :

- Если $t_{\text{табл}} < t_{\text{факт}}$, то H_0 отклоняется, т.е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x .
- Если $t_{\text{табл}} > t_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a , b или r_{xy} .

Принимаем уровень значимости 0,05, степень свободы 10.

Тогда для данного примера (слайд 2):

- $t_{\text{табл}} = 2,2281$ (для всех параметров!)

Расчет (с продолжением таблицы данных) фактических статистик дает:

$t_{\text{факт } a} =$

$t_{\text{факт } b} =$

$t_{\text{факт } r} =$

3.4 Далее рассчитываем доверительный интервал для каждого параметра УПЛР

$$\begin{aligned}\gamma_a &= a \pm \Delta_a; & \gamma_{a_{\min}} &= a - \Delta_a; & \gamma_{a_{\max}} &= a + \Delta_a; \\ \gamma_b &= b \pm \Delta_b; & \gamma_{b_{\min}} &= b - \Delta_b; & \gamma_{b_{\max}} &= b + \Delta_b.\end{aligned}$$

определяем предельную ошибку для каждого параметра:

$$\Delta_a = t_{\text{табл}} m_a, \quad \Delta_b = t_{\text{табл}} m_b.$$

**4. Определение
прогнозного значения U
с помощью УПЛР и проверка
ошибки**

В итоге определяется прогнозное значение Y_p путем подстановки в уравнение регрессии $Y_p = a + bX$ соответствующего (прогнозного) значения X_p .

Средняя стандартная ошибка прогноза рассчитывается по формуле:

$$m \hat{y}_p = \sigma_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

$$\text{где } \sigma_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}};$$

- Вычислим алгебраически:

$$F_{\text{факт}} = \frac{R^2}{1-R^2}(n-2) =$$

- Определим статистически :

$$F_{\text{факт}} =$$

- Сравним $F_{\text{факт}}$ с $F_{\text{табл}}(\alpha; k_1; k_2) = 4,96$

где $k_1=1; k_2=12-1-1=10,$

- **Вывод:** $F_{\text{факт}} > F_{\text{табл}}$, т.е. отвергается H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи коэффициента - **В**
- Имеется существенная связь между признаками уравнения с уровнем значимости α

| № | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| коэфф | 1,3 | 1,5 | 1,6 | 1,7 | 1,9 | 2,0 | 2,1 | 2,2 | 2,3 | 2,4 | 2,5 | 2,6 | 2,7 | 2,8 | 2,9 | 3,0 | 3,1 | 3,2 | 3,3 | 3,4 | 3,5 | 3,6 | 3,7 | 3,8 |

Строится доверительный интервал прогноза:

$$\gamma \hat{y}_p = \hat{y}_p \pm \Delta \hat{y}_p; \quad \gamma \hat{y}_{p \min} = \hat{y}_p - \Delta \hat{y}_p; \quad \gamma \hat{y}_{p \max} = \hat{y}_p + \Delta \hat{y}_p,$$

где $\Delta \hat{y}_p = t_{\text{табл}} \cdot m \hat{y}_p$.

Выполненные правильно расчеты, оформленные выводы по прогнозу для приведенного примера (для каждого студента свой прогноз) засчитываются в 1-ю часть РГР.