Федеральное государственное бюджетное образовательное учреждение
высшего образования

«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»

**Департамент анализа данных,
принятия решений и финансовых технологий**

Контрольная работа

по дисциплине «Эконометрика»

Выполнила:

студентка группы ПИ19-1

факультета

«Информационных технологий и анализа больших данных»

Воронина К. М.

___________________________________ (Подпись)

Дата выполнения: _______________________________

Дата защиты: ___________________________________

Преподаватель:

доцент, к.т.н., Петросов Д. А.

___________________________________ (Подпись)

2021 г.

Контрольная работа   Воронина Ксения, ПИ19-1

Вариант 2

① Линейная модель парной регрессии является простейшей моделью связи между x_i и y_i.

Модель имеет вид: $y_i = \alpha + \beta x_i + \varepsilon_i$, где

α, β – коэффициенты

x_i – фактор

ε_i – случайное возмущение

Мы можем найти оценки параметров модели, после чего она примет следующий вид: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

$\hat{y}_i$ – оценка реального значения y_i

$y_i - \hat{y}_i = e_i$ – величина ошибки.

Оценки $\hat{\alpha}$ и $\hat{\beta}$ находятся путем минимизации суммы квадратов ошибок e_i^2: $Q = \sum_{i=1}^{n} e_i^2$

$$Q(\hat{\alpha};\hat{\beta}) = \sum_{i=1}^{n} e_i^2 = \sum_i (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \to \min$$

$$\begin{cases} \dfrac{d Q(\hat{\alpha};\hat{\beta})}{d\hat{\alpha}} = 0; \\ \dfrac{d Q(\hat{\alpha};\hat{\beta})}{d\hat{\beta}} = 0, \end{cases}$$

Существует готовая формула:

$$\begin{cases} \hat{\beta} = \dfrac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}, \left(\dfrac{\widehat{cov}(x,y)}{\widehat{var}(x_i)}\right) \\ \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

② Линейная модель множественной регрессии – уравнение связи с несколькими независимыми переменными.

$$y_i = f(x_1, x_2, \dots, x_k)$$ ← факторы

результативный признак

Для построения используются функции следующего типа: 1) Линейная

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

2) Степенная, экспоненциальная, гиперболическая... (рассматриваем в этом вопросе только линейную)

Алгоритм МНК для множ. регрессии:

1) $Q(\hat{\beta}_0; \hat{\beta}_1 \ldots \hat{\beta}_k) = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})^2 = \sum e_i^2 \to min$

Система уравнений:

$$\begin{cases} \sum y = an + b_1 \sum x_1 + b_2 \sum x_2 + \ldots + b_k \sum x_k, \\ \sum y x_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \ldots + b_k \sum x_1 x_k, \\ \ldots \\ \sum y x_k = a \sum x_k + b_1 \sum x_1 x_k + \ldots + b_k \sum x_k^2. \end{cases}$$

$a, b_1, b_2 \ldots$ коэфф. $\beta_0, \beta_1, \beta_2 \ldots$

2) Матричный способ: $\boxed{B = (X^T X)^{-1} X^T Y}$, где $B = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \ldots \\ \hat{\beta}_k \end{pmatrix}$,

$X = (x_{i1}, x_{i2}, \ldots, x_{ik})^T$

$Y = (y_1, y_2, \ldots, y_k)^T$

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T \quad (Y = XB + \varepsilon)$

③ Теорема Гаусса - Маркова и ее предпосылки

Предпосылки регрессионного анализа:

1) $M(\varepsilon_i) = 0, \; i = 1 \ldots n$

2) Дисперсия постоянна

$D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2, \; i = 1 \ldots n$

3) Случайные величины должны быть статистически независимы: $cov(\varepsilon_i, \varepsilon_j) = 0, \ (i \neq j)$

4) Ошибки регрессии нормально распределены:

$\varepsilon_i \sim N(0, \sigma^2), \ i = 1 \dots n$

5) Случайные отклонения должны быть стат. независимы от объясняющих переменных:

$cov(x_{ji}, \varepsilon_i) = 0$.

$\Downarrow$

Если условия выполнены, то МНК дает лучшие оценки:
① Св-во несмещенности оценки (→ к М)
② Св-во эффективности (имеют наименьшую дисперсию)
③ Св-во состоятельности (при достаточно большом „n" оценки близки к параметрам)

④ Задача

1) Провести корреляционный анализ:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left(n\sum x_i^2 - (\sum x_i)^2\right)\left(n\sum y_i^2 - (\sum y_i)^2\right)}} = \frac{18 \cdot 1217949 - 16859 \cdot 1280,48}{\sqrt{(18 \cdot 26409831 - 16859^2) \cdot (18 \cdot \cancel{1280,48}\ 91137,66 - (\cancel{91137,66}\ 1280,48)^2)}}$$

$= 0,83282$ — Вывод: высокая прямая связь

2) Построить модель линейной парной регрессии

$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i \quad , \quad \hat{\beta} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{18637,2}{10619504,28} = 0,0018$$

$$\hat{\alpha} = 71,14 - 0,0018 \cdot 936,61 = 69,5$$

$$\boxed{\hat{y}_i = 69,5 + 0,0018 x_i}$$ < Расчёт в Excel. >

3) Рассчитать коэффициент детерминации:

$$R^2 = \frac{ESS}{TSS} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{32,71}{47,15} = 0,69 \longrightarrow$$

Вывод: $R^2 = 0,69$ – в целом, модель линейной регрессии даёт хорошее описание