Тема 9

Корреляционный анализ

Составитель: доцент кафедры отраслевой экономики Ахмедова М.М.

План:

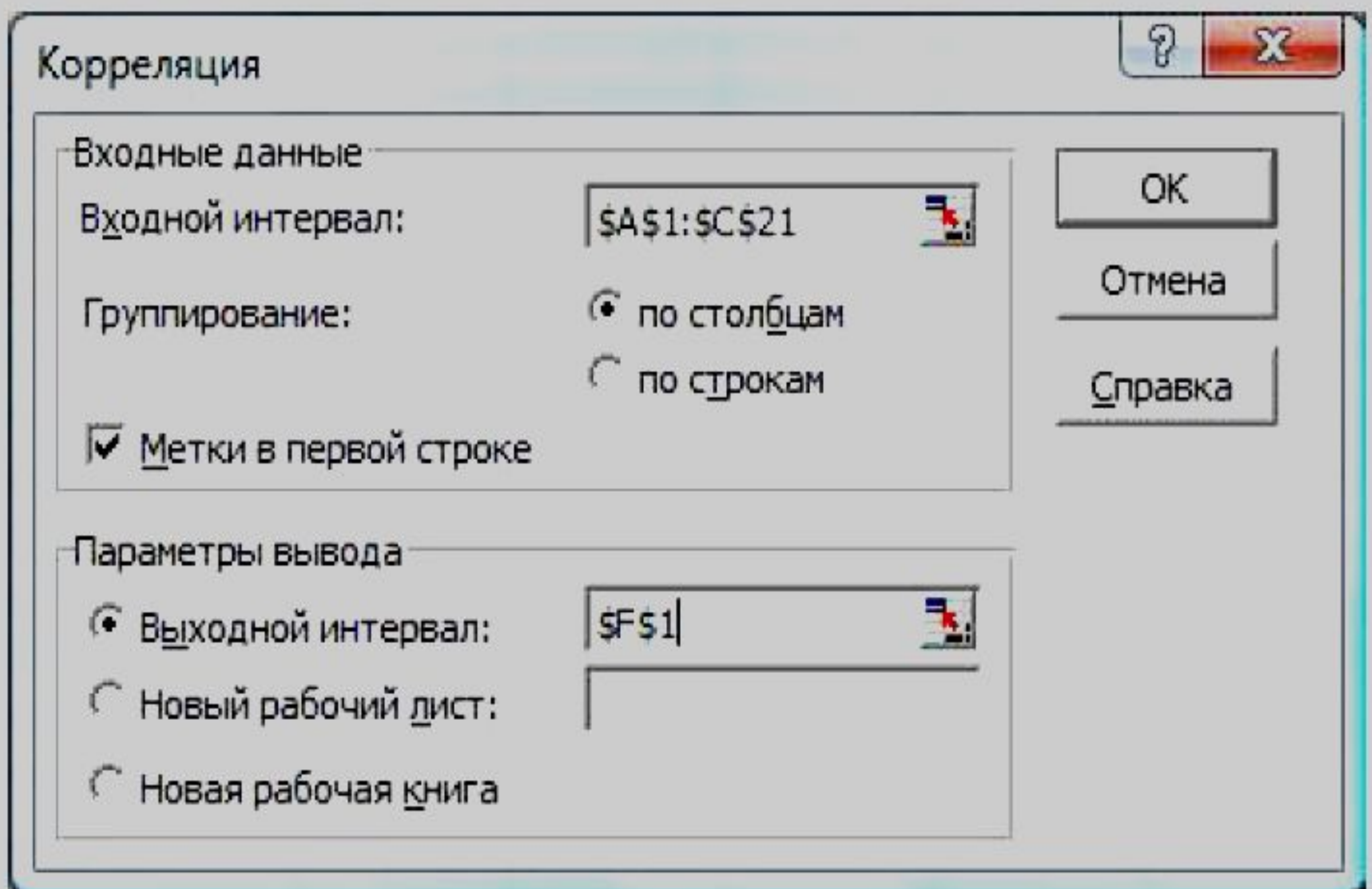
- 1. Сущность корреляционного анализа.
- 2. Однофакторный и многофакторный корреляционный анализ.

1. Виды матрицы и их характеристики

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

1. Виды матрицы и их характеристики

Найдем матрицу парных коэффициентов корреляции (**Сервис→Анализ данных→Корреляция**):



1. Виды матрицы и их характеристики

При линейной зависимости *коэффициент множественной корреляции* можно определить через матрицы парных коэффициентов корреляции:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

1. Виды матрицы и их характеристики

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix}$$

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix}$$

1. Виды матрицы и их характеристики

Этапы проведения комплексного корреляционно-регрессионного анализа

↓

1. Предварительный анализ явлений и выявление причин возникновения взаимосвязей между признаками, характеризующими эти явления

↓

2. Разделение признаков на факторные и результативные, выбор наиболее существенных признаков для их исследования на предмет включения в корреляционно-регрессионные модели

↓

3. Построение матрицы коэффициентов парной корреляции и оценка возможных вариантов группировки признаков корреляционно-регрессионных моделей

↓

4. Предварительная оценка формы уравнения регрессии

↓

5. Решение уравнения регрессии, вычисление коэффициентов регрессии и их смысловая интерпретация

1. Виды матрицы и их характеристики

Этапы проведения комплексного корреляционно-регрессионного анализа

6. Расчет теоретически ожидаемых (воспроизведенных по уравнению регрессии) значений результативного признака

7. Определение и сравнительный анализ дисперсий: общей, факторной и остаточной; оценка тесноты связи между признаками, включенными в регрессионную модель

8. Общая оценка качества модели, отсев несущественных (или включение дополнительных) факторов, построение модели, т.е. повторение п. 1—7

9. Статистическая оценка достоверности параметров уравнения регрессии, построение доверительных границ для теоретически ожидаемых по уравнению регрессии значений функции

10. Практические выводы из анализа

1. Виды матрицы и их характеристики

Теснота связи изучаемых явлений

для линейной регрессии

Линейный коэффициент парной корреляции

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x,y)}{\sigma_x \cdot \sigma_y}$$

$$\left(-1 \le r_{xy} \le 1\right)$$

для нелинейной регрессии

Индекс корреляции

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{\sum\left(y - \hat{y}_x\right)^2}{\sum\left(y - \overline{y}\right)^2}}$$

$$\left(0 \le \rho_{xy} \le 1\right)$$

1. Виды матрицы и их характеристики

Проверка значимости коэффициента корреляции

Коэффициенты корреляции как статистические величины подвергаются в анализе оценке на достоверность. Это объясняется тем, что любая совокупность наблюдений представляет собой некоторую выборку, следовательно, значение любого показателя, вычисленное на основе выборки, не может рассматриваться как истинное, а является только более или менее точной его оценкой. В связи с этим возникает необходимость проверки существенности (значимости) показателей.

1. Виды матрицы и их характеристики

Проверка значимости коэффициента корреляции

Для *оценки значимости* коэффициента корреляции используют *t-критерий Стьюдента (t-статистику)*, который применяется при *t*-распределении, отличном от нормального. При этом выдвигается и проверяется нулевая гипотеза (H_0) о равенстве r_{yx} нулю, т.е. $H_0: r_{yx} = 0$. Если нулевая гипотеза отвергается, то коэффициент корреляции признается значимым, а связь между переменными существенной.

1. Виды матрицы и их характеристики

Частные коэффициенты (или индексы) корреляции, измеряющие влияние на y фактора x_i, при элиминировании (исключении влияния) других факторов, можно определить по формуле

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 \ldots x_i \ldots x_m}}{1 - R^2_{yx_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m}}},$$

1. Виды матрицы и их характеристики

Частные коэффициенты корреляции характеризуют тесноту связи между двумя показателями за вычетом влияния других показателей. Порядок частного коэффициента корреляции равен числу показателей, влияние которых исключается. Коэффициенты линейной парной корреляции можно считать частными коэффициентами нулевого порядка. Коэффициенты частной корреляции можно определить, используя рекуррентную формулу:

1. Виды матрицы и их характеристики

Частные коэффициенты корреляции

$$r_{yx_k \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_m} =$$

$$= \frac{r_{yx_k \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_{m-1}} - r_{yx_m \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_{m-1}} r_{x_k x_m \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_{m-1}}\right) \cdot \left(1 - r^2_{x_k x_m \cdot x_1 \ldots x_{k-1} x_{k+1} \ldots x_{m-1}}\right)}}$$

1. Виды матрицы и их характеристики

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде.

1. Виды матрицы и их характеристики

Частные коэффициенты корреляции

Частные коэффициенты первого порядка:

$$r_{yx_1\cdot x_2}=\frac{r_{yx_1}-r_{yx_2}r_{x_1x_2}}{\sqrt{\left(1-r_{yx_2}^2\right)\left(1-r_{x_1x_2}^2\right)}},$$

$$r_{yx_2\cdot x_1}=\frac{r_{yx_2}-r_{yx_1}r_{x_2x_1}}{\sqrt{\left(1-r_{yx_1}^2\right)\left(1-r_{x_2x_1}^2\right)}},$$

$$r_{x_1x_2\cdot y}=\frac{r_{x_1x_2}-r_{x_1y}r_{x_2y}}{\sqrt{\left(1-r_{x_1y}^2\right)\left(1-r_{x_2y}^2\right)}}.$$

1. Виды матрицы и их характеристики

Частные коэффициенты корреляции

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \quad r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}.$$

1. Виды матрицы и их характеристики

Формула расчета *t*-критерия Стьюдента

***t*-критерий Стьюдента (*t*-статистика)**

$$t_{\text{расч}} = r_{yx}\sqrt{\frac{n-k-1}{1-r_{yx}^2}},$$

где k — число факторных признаков, включенных в модель

Значение t-критерия сравнивают с табличным $t_{\alpha\,\gamma}$, где α — заданный уровень значимости (обычно принимается равным 0,05 или 0,01); $\gamma = (n - k - 1)$ — число степеней свободы.

1. Виды матрицы и их характеристики

Если выполняется неравенство $t > t_{\alpha,\gamma}$, то значение коэффициента корреляции признается значимым, т.е. нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается и делается вывод о том, что между исследуемыми переменными есть тесная статистическая взаимосвязь

1. Виды матрицы и их характеристики

Зная линейный коэффициент корреляции, можно определить *парный коэффициент детерминации*, он представляет собой r_{yx}^2.

Парный коэффициент детерминации r_{yx}^2

показывает, какая доля вариации переменной *y* учтена в модели и обусловлена влиянием на нее переменной *x*

2.Методы оценки коэффициентов и их расчеты

Для расчета доверительного интервала определяем *предельную ошибку* Δ для каждого показателя:

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

Формулы для расчета *доверительных интервалов* имеют следующий вид:

$$\gamma_a = a \pm \Delta_a; \qquad \gamma_{a_{\min}} = a - \Delta_a; \quad \gamma_{a_{\max}} = a + \Delta_a;$$

$$\gamma_b = b \pm \Delta_b; \qquad \gamma_{b_{\min}} = b - \Delta_b; \quad \gamma_{b_{\max}} = b + \Delta_b;$$

2.Методы оценки коэффициентов и их расчеты

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.