

## О С Н О В Н Ы Е П О Н Я Т И Я (РЕГРЕССИОННЫЙ АНАЛИЗ)

1. Определение математического ожидания дискретной случайной величины,  $N$  – количество элементов генеральной совокупности:

$$\mathbf{E}(X) \equiv \mu_X = \sum_{i=1}^N x_i p_i$$

2. Определение математического ожидания непрерывной случайной величины:

$$\mathbf{E}(X) \equiv \mu_X = \int_{-\infty}^{\infty} x f(x) dx$$

3. Математическое ожидание линейной комбинации случайных величин:

$$\mathbf{E}(c + a \cdot X + b \cdot Y) = c + a \cdot \mathbf{E}(X) + b \cdot \mathbf{E}(Y)$$

4. Определение дисперсии дискретной случайной величины:

$$\mathbf{Var}(X) \equiv \sigma_X^2 \equiv \mathbf{D}(X) = \mathbf{E}[(X - \mu_X)^2] = \sum_{i=1}^N (x_i - \mu_X)^2 p_i$$

5. Определение дисперсии непрерывной случайной величины:

$$\mathbf{Var}(X) \equiv \sigma_X^2 \equiv \mathbf{D}(X) = \mathbf{E}[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx$$

6. Дисперсия линейной комбинации случайных величин:

$$\mathbf{Var}(c + a \cdot X + b \cdot Y) = a^2 \sigma_X^2 + 2ab \sigma_{XY} + b^2 \sigma_Y^2$$

7. Свойство дисперсии (выражение через  $\mathbf{E}(X^2)$ ):

$$\sigma_X^2 = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2$$

8. Определение стандартного отклонения и вариабельности:

$$\sigma_X = \sqrt{\mathbf{Var}(X)}; \quad \nu_X = \frac{\sigma_X}{\mathbf{E}(X)}$$

9. Случайная величина Бернулли: случайная величина, принимающая значение «1» с вероятностью  $p$  и «0» с вероятностью  $(1 - p)$ . Математическое ожидание случайной величины Бернулли:  $p$ ; дисперсия:  $p(1 - p)$ .

10. Математическое ожидание и дисперсия среднего выборочного,  $n$  – количество элементов в выборке:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \mathbf{E}(\bar{X}) = \mathbf{E}(X); \quad \mathbf{Var}(\bar{X}) = \frac{\mathbf{Var}(X)}{n}$$

11. Определение ковариации дискретных случайных величин:

$$\mathbf{Cov}(X; Y) \equiv \sigma_{XY} = \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] = \sum_{i=1}^N \sum_{j=1}^M (x_i - \mu_X)(y_j - \mu_Y) p_{ij}$$

12. Определение ковариации непрерывных случайных величин:

$$\text{Cov}(X; Y) \equiv \sigma_{XY} = \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] = \iint_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)dx dy$$

13. Ковариация линейной комбинации случайных величин:

$$\text{Cov}(c + a \cdot X + b \cdot Y; Z) = a\sigma_{XZ} + b\sigma_{YZ}$$

14. Свойство ковариации (выражение через  $\mathbf{E}(XY)$ ):

$$\sigma_{XY} = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

15. Математическое ожидание произведения случайных величин (выражение через  $\sigma_{XY}$ ):

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y) + \sigma_{XY}$$

16. Определение корреляции:

$$\text{Corr}(X; Y) \equiv \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

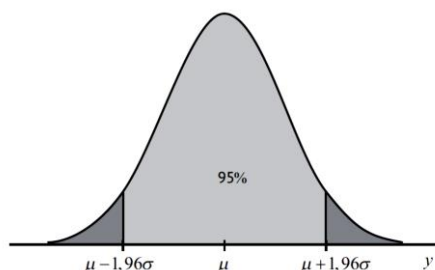
17. Определение независимости случайных величин:

$$\mathbf{E}(F(X)G(Y)) = \mathbf{E}(F(X))\mathbf{E}(G(Y))$$

18. Свойства независимых случайных величин:

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y); \quad \sigma_{XY} = 0$$

19. Квантили нормального распределения:



|     |      |
|-----|------|
| 50% | 0.67 |
| 90% | 1.64 |
| 95% | 1.96 |
| 99% | 2.58 |

20. Несмещенная оценка выборочной дисперсии,  $n$  – количество элементов в выборке:

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

21. Выборочное стандартное отклонение,  $n$  – количество элементов в выборке:

$$s_X = \sqrt{s_X^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

22. Уравнение парной линейной регрессии:  $y_i = \beta_0 + \beta_1 x_i + u_i$ ;  $y_i$  – зависимая переменная;  $x_i$  – независимая переменная, объясняющая переменная, регрессор;  $\beta_0$  – константа линии теоретической регрессии,  $\beta_1$  – угловой коэффициент;  $u_i$  – остаток регрессии.

23. Метод наименьших квадратов (МНК) находит коэффициенты  $\beta_0$  и  $\beta_1$ , минимизирующие сумму квадратов остатков:  $\sum_{i=1}^n (u_i)^2$

24. Условия, которым удовлетворяют остатки парной регрессии, построенной МНК:

$$\sum_{i=1}^n u_i = 0; \quad \sum_{i=1}^n x_i u_i = 0$$

25. Коэффициенты парной линейной регрессии, построенной МНК:

$$\beta_1 = \frac{\sigma_{XY}}{\sigma_X^2}; \quad \beta_0 = E(Y) - \beta_1 E(X)$$

26. Определение объясненной суммы квадратов остатков,  $n$  – количество наблюдений:

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = (n - 1)\sigma_{\hat{Y}}^2$$

27. Выражение объясненной суммы квадратов остатков через  $\sigma_{XY}$  и  $\sigma_X$ :

$$ESS = (n - 1) \left( \frac{\sigma_{XY}}{\sigma_X} \right)^2$$

28. Определение полной суммы квадратов остатков,  $n$  – количество наблюдений:

$$TSS = \sum_{i=1}^n (y_i - \bar{y})^2 = (n - 1)\sigma_Y^2$$

29. Остаточная сумма квадратов,  $n$  – количество наблюдений:

$$RSS \equiv SSR = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (u_i)^2$$

30. Стандартная ошибка множественной регрессии ( $n$  – количество наблюдений,  $k$  – количество параметров модели):

$$SER = \sqrt{\frac{RSS}{n - k}}$$

31. Определение коэффициента детерминации:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

32. Свойства коэффициента детерминации (выражение через  $\sigma_{XY}$ ):

$$R = \sqrt{\frac{ESS}{TSS}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

33. Скорректированный коэффициент детерминации ( $n$  – количество наблюдений,  $k$  – кол-во параметров модели):

$$R_{adj}^2 = 1 - \frac{RSS/(n - k)}{TSS/(n - 1)}$$

34.  $F$ -статистика теста Фишера ( $n$  – количество наблюдений,  $k$  – кол-во параметров модели):

$$F = \frac{ESS/(k - 1)}{RSS/(n - k)}$$

35. Статистика теста Чоу ( $RSS_R$  – остатки общей модели,  $RSS_{UR}$  – сумма остатков частичных моделей,  $n$  – количество наблюдений,  $k$  – количество параметров общей модели):

$$F = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(n - 2k)}$$

36.  $F$ -статистика теста, сравнивающего две модели ( $RSS_1$  – остатки первой модели,  $RSS_2 < RSS_1$  – остатки второй модели,  $n$  – количество наблюдений,  $k_1$  – количество параметров первой модели,  $k_2$  – количество параметров второй модели):

$$F = \frac{(RSS_1 - RSS_2)/(k_2 - k_1)}{RSS_2/(n - k_2)}; \quad k_2 > k_1$$

37. Гетероскедастичность случайных ошибок регрессии – зависимость дисперсии ошибок регрессии от объясняющих переменных.
38. В каком случае возникает смещение оценки коэффициента регрессии из-за пропущенной переменной? В случае если пропущенная переменная влияет на зависимую переменную и коррелирует с другими объясняющими переменными.
39. Два типа возможных ошибок при статистической проверке гипотез:  
*ошибка 1 рода*, когда отвергается фактически истинная нулевая гипотеза;  
*ошибка 2 рода*, когда НЕ отвергается фактически ложная нулевая гипотеза.
40. Статистический размер теста: вероятность совершить ошибку первого рода;  
Статистическая мощность теста: вероятность НЕ совершить ошибку второго рода.