### Степени свободы для сумм квадратов

• Число степеней свободы некоторой величины — это разность между числом ее исследований (наблюдений) и числом констант, которые установлены (найдены) в результате этих исследований, независимо друг от друга

### Применительно к SST, SSR и SSE

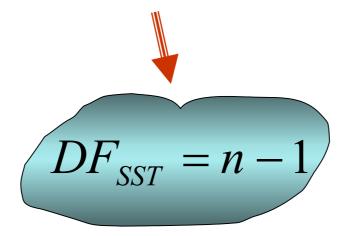
**Степень свободы** — число, которое показывает, сколько независимых элементов информации, полученных из элементов  $y_1, y_2, \dots, y_n$ , необходимо для расчета данной суммы квадратов

#### SST

$$y_1 - \overline{y}$$
,  $y_2 - \overline{y}$ , ...,  $y_n - \overline{y}$ 

$$\sum_{i=1}^{n} (y_i - \overline{y}) = 0 \qquad \sum_{i=1}^{n-1} (y_i - \overline{y}) \neq 0$$

n-1 разность независима

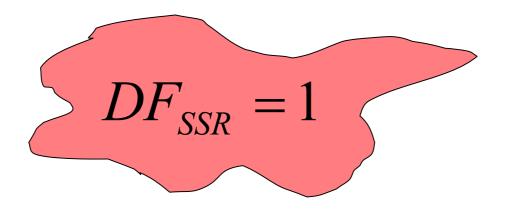


### SSR

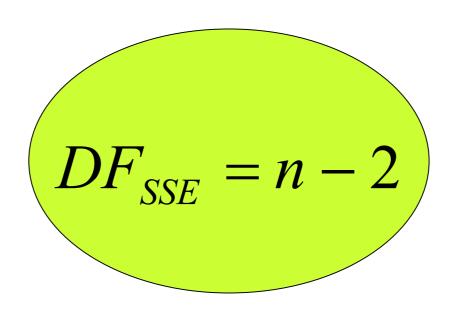
$$\widehat{a}_{\scriptscriptstyle 1}$$

$$\widehat{y}_i - \overline{y} = \widehat{a}_0 + \widehat{a}_1 x_i - \widehat{a}_0 - \widehat{a}_1 \overline{x} = \widehat{a}_1 (x_i - \overline{x})$$

$$\sum_{i=1}^{n} (\widehat{y}_{i} - \overline{y})^{2} = \widehat{a}_{1}^{2} \sum_{i=1}^{n} (x_{i} - \overline{x})^{2}$$



### SSE



### Определение

• Средний квадрат – сумма квадратов, поделенная на соответствующее число степеней свободы

### Средний квадрат ошибок

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{n-2}$$

## Средний квадрат, объясняющий регрессию

$$MSR = \frac{\sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2}{1} = SSR$$

# •Проверка адекватности модели и ее параметров

•Проверка адекватности модели по *F*-критерию Фишера

$$F = \frac{MSR}{MSE}$$

$$v_1 = 1 \qquad v_2 = n - 2$$

$$\hat{a}_1 = 0$$



$$\hat{y}_{i} = \hat{a}_{0} + \hat{a}_{1}x_{i} = \bar{y} - \hat{a}_{1}\bar{x} + \hat{a}_{1}x_{i} = \bar{y} + \hat{a}_{1}(x_{i} - \bar{x}) = \bar{y}$$

Гипотеза

$$H_0: \widehat{a}_1 = 0$$

### Этапы проверки

1. Рассчитываем величину F-критерия

$$F_{pacy.} = \frac{MSR}{MSE}$$

### 2. Задаем уровень значимости СС

3. Находим табличное (критическое)

значение 
$$F$$
-критерия  $\left(F_{(1,n-2,\alpha)}\right)$  с  $v_1=1$ 

и  $v_2 = n - 2$  степенями свободы и уровнем

значимости О

4. Если  $F_{pacy} > F_{(1,n-2,\alpha)}$  , то гипотеза

$$H_0: \widehat{a}_1 = 0$$

отвергается, если ,  $F_{pac^{q}} \leq F_{(1,n-2,\alpha)}$  то гипотеза принимается

### Рассчитаем F-статистику Фишера

$$MSR = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{1} = \frac{274,7935}{1} = 274,7935$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-2} = \frac{10,70652}{8-2} = 1,78442$$

$$F_{pacu.} = \frac{MSR}{MSE} = \frac{274,7935}{1,78442} = 153,996$$

### Для

$$v_1 = 1$$
  $v_2 = n - 2$ 

$$\alpha = 0.05$$
  $F_{(1,6,0.05)} = 5.99$ 

$$F_{pacy.} = 153,996 > F_{(1,6,0.05)} = 5,99$$

## Проверка гипотезы о значимости коэффициента корреляции

Гипотеза

$$H_0: r_{yx} = 0$$

### Используем критерий Стьюдента

$$t_{pacy.} = r_{yx} \cdot \sqrt{\frac{n-2}{1-r_{yx}^2}}$$

ИЛИ

$$t_{pacy.} = r_{yx} \cdot \sqrt{\frac{n-2}{1-R^2}}$$

$$\nu = n-2$$
 - число степеней свободы

Для заданных v и  $\alpha$  находим табличное значение  $t_{(v,\alpha)}$ 

Если 
$$\left|t_{pacu.}\right| > t_{(v,\alpha)}$$
 , то гипотеза  $H_0: r_{yx} = 0$  отвергается, если  $\left|t_{pacu.}\right| \leq t_{(v,\alpha)}$  - то принимается

### Рассчитаем *t*-статистику Стьюдента

$$t_{pacu.} = r_{yx} \cdot \sqrt{\frac{n-2}{1-r_{yx}^2}} = 0,981 \cdot \sqrt{\frac{8-2}{1-0,9625}} = 12,41$$

При 
$$\alpha$$
=0,05 и  $v$ = $n$ -2= $8$ -2= $6$ 

$$t_{\text{magn.}} = t_{(6,0.05)} = 1,943$$

Так как

$$t_{pacy.} = 12,41 > t_{(6,0.05)} = 1,943,$$

то коэффициент корреляции значим

## • Критерии оценки качества линейной модели

### 1. Средняя ошибка прогноза

$$ME = \overline{u} = \frac{1}{n} \sum_{i=1}^{n} u_i$$

$$ME \rightarrow 0$$
 при  $n \rightarrow \infty$ 

### 2. Дисперсия и стандартное отклонение ошибок

$$var(u) = \frac{1}{n} \sum_{i=1}^{n} (u_i - \overline{u})^2 = \sigma_u^2$$

$$\sigma = \sqrt{\operatorname{var}(u)} = \sqrt{\frac{1}{n}} \sum_{i=1}^{n} (u_i - \overline{u})^2$$

### 3. Среднее абсолютное отклонение

$$MAD(u) = \frac{1}{n} \sum_{i=1}^{n} |u_i - \overline{u}|$$

### 4. Средний квадрат ошибки

$$MSE(u) = \frac{1}{n} \sum_{i=1}^{n} u_i^2$$

(сумма квадратов ошибок)

$$SSE = \sum_{i=1}^{n} u_i^2$$

### 5. Средняя абсолютная процентная ошибка

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|u_i|}{y_i} 100\%$$

## 6. Средняя процентная ошибка аппроксимации

$$MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{u_i}{y_i} 100\%$$

### 7. Средняя абсолютная ошибка

$$MAE(u) = \frac{1}{n} \sum_{i=1}^{n} |u_i|$$

## Рассчитаем среднюю процентную ошибку аппроксимации

$$MPE = \frac{1}{n} \sum_{i=1}^{n} \frac{u_i}{y_i} \cdot 100\% =$$

$$=\frac{-0.28386}{8}\cdot100\%=-3.55\%$$

Несмещенная оценка истинного значения  $\sigma_u^2$ :

а) для парной зависимости

$$\widehat{\sigma}_{u}^{2} = \frac{\sum_{i=1}^{n} u_{i}^{2}}{n-2}$$

### б) для многофакторной модели

$$\widehat{\sigma}_{u}^{2} = \frac{\sum_{i=1}^{n} u_{i}^{2}}{n-k}$$

$$n-2$$
 и  $n-k$  – число степеней свободы

### • Условия применения метода наименьших квадратов

$$1. M(u_i) = 0$$

Факторы, не включенные в модель, не влияют систематически на математическое ожидание

#### 2. Гомоскедастичность

$$\operatorname{var}(u_i) = \sigma_u^2, i = \overline{1,n}$$

## 3. Случайная величина u распределена по нормальному закону

$$u \sim N(0,\sigma_u^2)$$

## 4. Между случайными величинами $\mathcal{U}$ отсутствует автокорреляция

$$cov(u_i, u_j) = M\{[u_i - M(u_i)] \cdot [u_j - M(u_j)]\} =$$

$$=M(u_iu_j)=0, i\neq j$$

5. Значения независимых переменных  $X_i$ 

не зависят от значений случайных

величин  $u_i$ 

$$cov(u_i, x_i) = M\{[u_i - M(u_i)] \cdot [x_i - M(x_i)]\} =$$

$$=M\{u_{i}[x_{i}-M(x_{i})]\}=$$

$$= M(u_i x_i) - M(u_i) \cdot M(x_i) = M(u_i x_i) = 0$$

$$M(u_i) = 0$$
  $M(x_i) = const$ 

6. Независимые переменные не должны быть мультиколлинеарными

7. Регрессионная модель построена (специфицирована) правильно: правильно подобрана линия регрессии, в нее включены только существенно влияющие переменные, модель адекватна экономическому процессу

8. Количество наблюдений больше числа независимых переменных (perpeccopos) (n > k)

9. Регрессоры – детерминированные величины

## 10. Набор переменных X не содержит ошибок наблюдений

11.Набор переменных X содержит все существенно влияющие на регрессант факторы

12. Коэффициенты регрессии – детерминированные величины, неизменные во времени и пространстве

## • <u>Распределение независимой</u> <u>переменной</u> у

Единственный источник изменения *у* – случайная величина *и* 



распределение переменной y зависит от распределения случайной величины u, т.е.

## является нормальным

$$M(y_i) = M(a_0 + a_1x_i + u_i) = M(a_0 + a_1x_i) + M(u_i)$$

$$M(u_i) = 0$$

 $a_0, a_1$  - constants

 $X_i$  — детерминированная величина

$$M(y_i) = M(a_0 + a_1 x_i) = a_0 + a_1 x_i$$

$$D(y_i) = M[y_i - M(y_i)]^2 = \sigma_u^2$$

## Здесь и далее в расчетах используется

несмещенная оценка дисперсии!

$$\widehat{\sigma}_{u}^{2}$$