REGRESSION ANALYSIS

OLS. MULTIPLE REGRESSION ANALYSIS: HOMOGENEITY TESTING. DUMMY (BINARY) VARIABLES

МОДЕЛЬ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ: ТЕСТИРОВАНИЕ ОДНОРОДНОСТИ. ФИКТИВНЫЕ ПЕРЕМЕННЫЕ

TEST CHOW: TEST FOR STRUCTURAL STABILITY

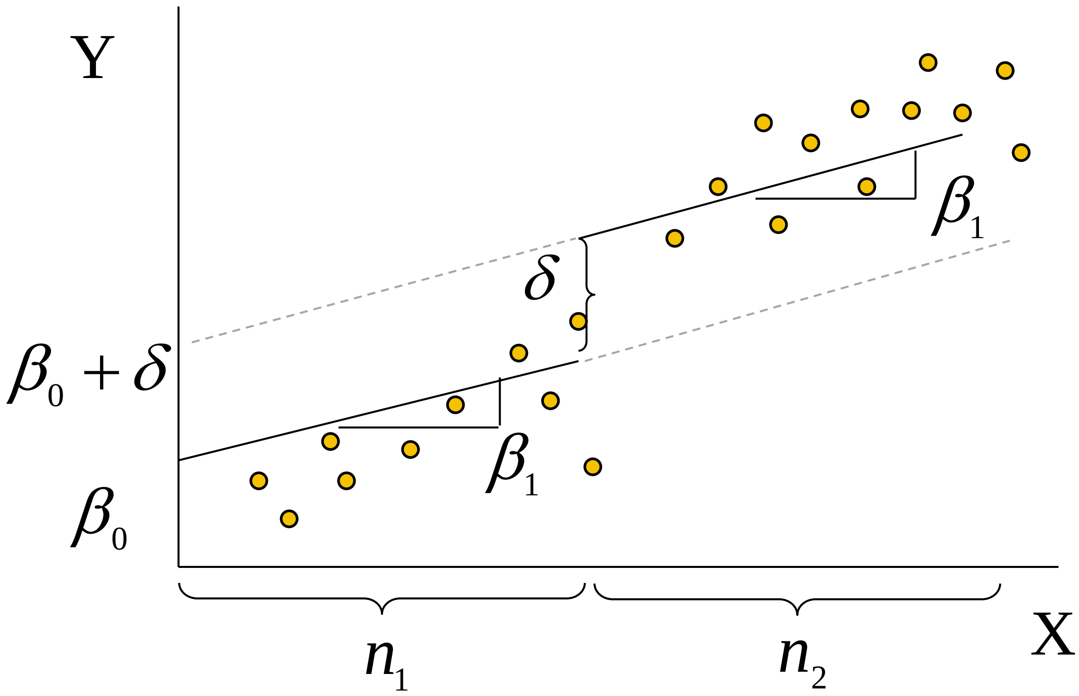
The Chow test is a statistical test of whether the coefficients in two linear regressions on different data sets are equal

Используется для тестирования:

- однородности выборки / homogeneity of the sample
- структурного сдвига / structural break
- наличия в выборке точки разрыва / presence of a point of discontinuity
- для оценки наличия различий во влиянии экзогенных переменных для подвыборок / to determine whether the independent variables have different impacts on different subsamples

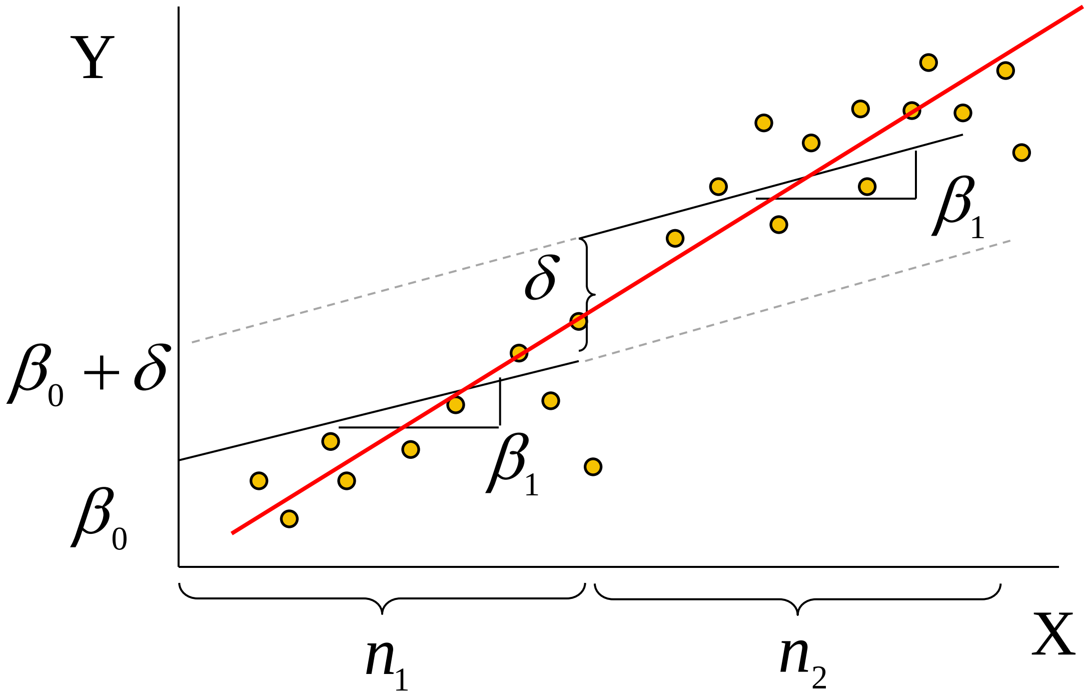
HOMOGENEITY & STRUCTURAL STABILITY

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \qquad n = n_1 + n_2$$



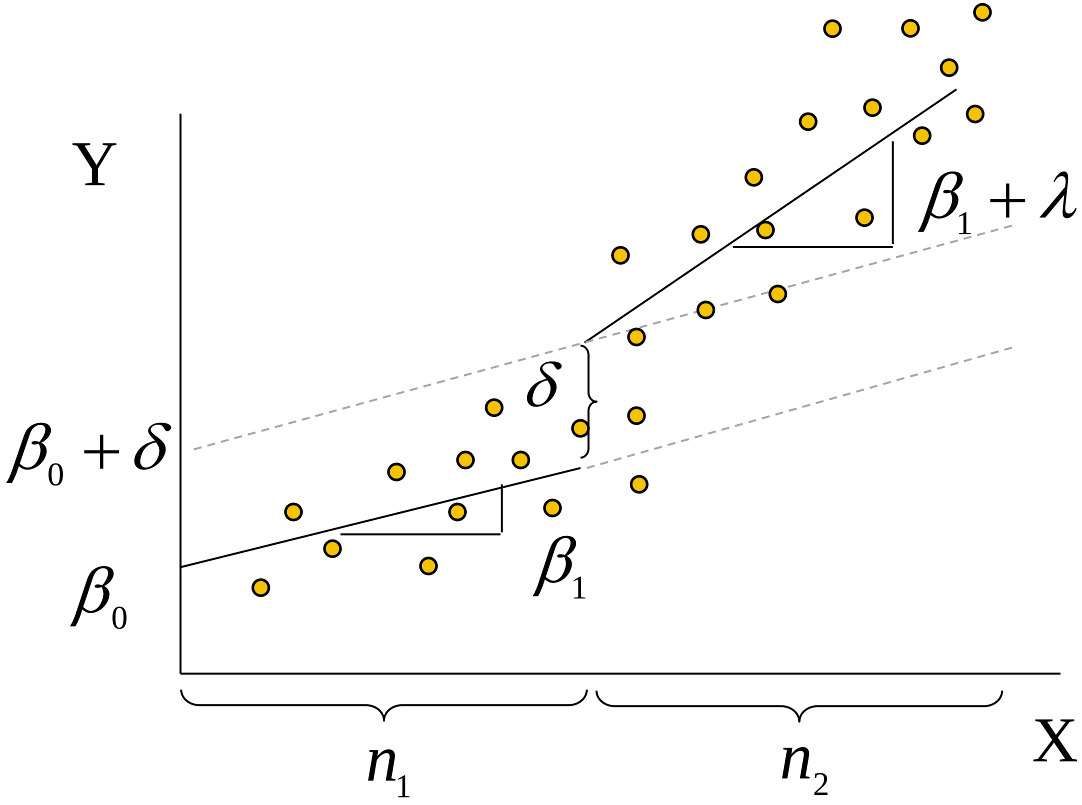
HOMOGENEITY & STRUCTURAL STABILITY

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \qquad n = n_1 + n_2$$



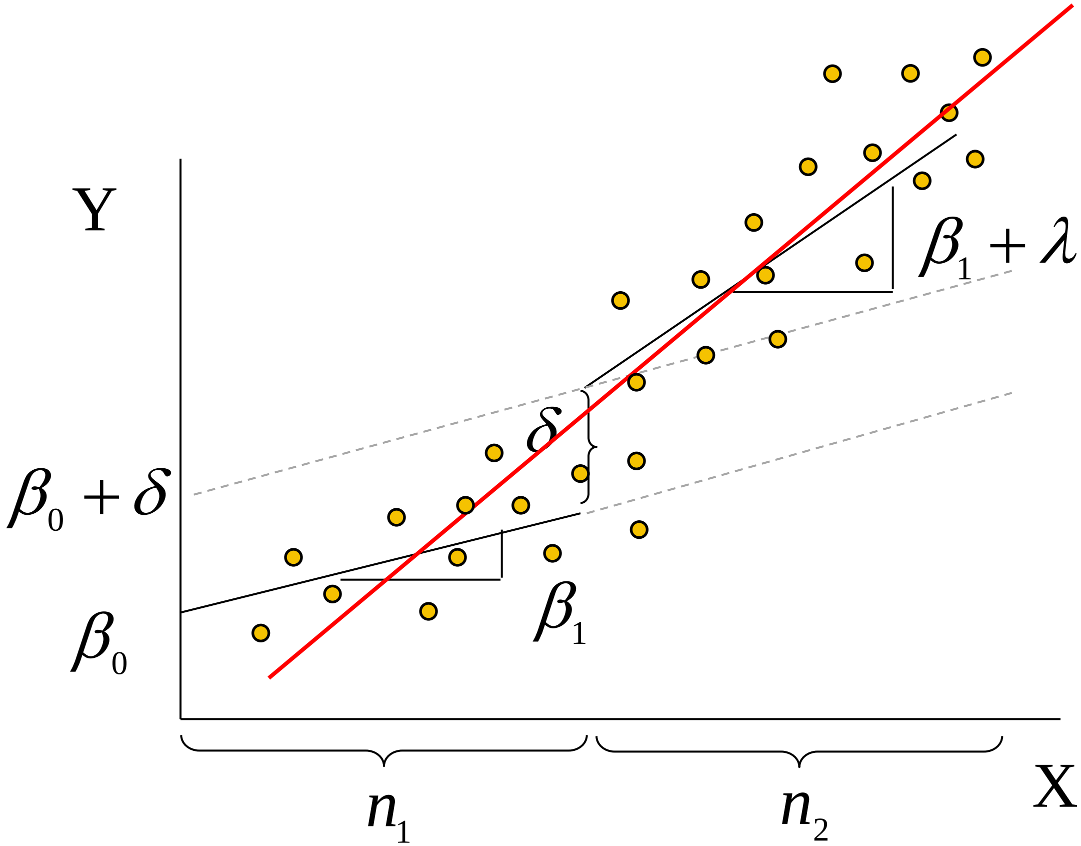
HOMOGENEITY & STRUCTURAL STABILITY

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \qquad n = n_1 + n_2$$



HOMOGENEITY & STRUCTURAL STABILITY

$$Y = \beta_0 + \beta_1 \cdot X + \varepsilon \qquad n = n_1 + n_2$$



TEST CHOW

$$H_0 : RSS_0 = RSS_1 + RSS_2 \Leftrightarrow RSS_0 - (RSS_1 + RSS_2) = 0$$

$$H_1 : RSS_0 > RSS_1 + RSS_2$$

$$F = \frac{\left(RSS_0 - (RSS_1 + RSS_2)\right)/(m+1)}{\left(RSS_1 + RSS_2\right)/\left(n - 2(m+1)\right)} \sim F_{\alpha;m+1;n-2(m+1)}$$

$$RSS_1 + RSS_2 : \left(n_1 - m - 1\right) + \left(n_2 - m - 1\right) =$$

$$= \left(n_1 + n_2\right) - 2m - 2 = n - 2(m+1)$$

$$RSS_0 - \left(RSS_1 + RSS_2\right) : \left(n - m - 1\right) - \left(n - 2(m+1)\right) = m + 1$$

DUMMY VARIABLES

Dummy variables are discrete variables taking a value of '0' or '1'. They are often called 'on' 'off' variables, being 'on' when they are 1.

$$D = \begin{cases} 1, \ if \ \ .... \\ \\ 0, \ in\ another\ case \end{cases}$$

Dummy variables can be used either as explanatory variables or as the dependent variable.

When they act as the dependent variable there are specific problems with how the regression is interpreted, however when they act as explanatory variables they can be interpreted in the same way as other variables.

DUMMY VARIABLES

Types of Explanatory Dummy Variable:

- Qualitative dummy variables: i.e. age, sex, education, health etc.;
- Dummy Variable for Single Outlier;
- Seasonal dummy variables: depends on the nature of the data, so quarterly data requires three dummy variables etc.;
- Dummy variables that represent a change in policy for example:

  Intercept dummy variables, that pick up a change in the intercept of the regression;

  Slope dummy variables, that pick up a change in the slope of the regression.

DUMMY VARIABLES

Таким образом, фиктивные переменные можно классифицировать в зависимости от целей, с которыми они используются в регрессионной модели.

Такая классификация находит свое отражение и в терминологических обозначениях фиктивных переменных, заданных при этом одной и той же математической формулой:

- качественные переменные (т.е. не количественные)
- фиктивные переменные
  - аддитивного выброса
  - сезонности
  - изменения линии уровня и наклона тренда
- фиктивные переменные взаимодействия
- бинарные переменные
- дихотомические переменные

QUALITATIVE DUMMY VARIABLES

Качественные фиктивные переменные отвечают за наличие у объекта некоторого качества или признака.

Например, зависимость фактора ежемесячной оплаты труда Y от:

количественных факторов
количества отработанных за месяц дней X1
количества лет стажа X2
количества лет, потраченных на образование X3

качественных факторов
наличия высшего образования X4
пола X5
членства в профсоюзных организациях X6

QUALITATIVE DUMMY VARIABLES

Качественные фиктивные переменные могут быть использованы как в случае пространственных данных, так и в случае временных данных; как в случае микроэконометрических моделей, так и в случае макроэконометрических моделей.

Например, зависимость экспорта Y страны от:
количественных факторов
ВВП X1
обменный курс X2
ПИИ X3
качественных факторов
наличие общей границы со страной-партнером X4
наличие выхода к морским торговым путям X5
вхождение в торговый союз X6

QUALITATIVE DUMMY VARIABLES

Качественные фиктивные переменные могут быть введены как в **аддитивной** форме, так и в **мультипликативной** форме, а также одновременно в обеих формах, в зависимости от того, на что влияет качественный признак в модели – на коэффициент уровня, т.е. коэффициент свободного члена (level, intercept), или коэффициенты угла наклона (slope).

$$Y_t = \beta_0 + \beta_1 \cdot X_t + \beta_2 \cdot D_t + \beta_3 \cdot \left(D_t \cdot X_t\right) + \varepsilon, t = \overline{1,n}$$

$$t = k: \quad Y_t = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) \cdot X_t + \varepsilon$$

$$t \neq k: \quad Y_t = \beta_0 + \beta_1 \cdot X_t + \varepsilon$$

$$D = \begin{cases} 1, \text{при } t = k \\ 0, \text{при } t \neq k \end{cases}$$

QUALITATIVE DUMMY VARIABLES

Фиктивные переменные для моделирования аддитивных выбросов (single outlier), сезонности (seasonal), изменения уровня линии тренда (intercept) или угла наклона тренда (излома) (slope) используются, как правило, в случае моделирования на основе временных данных.

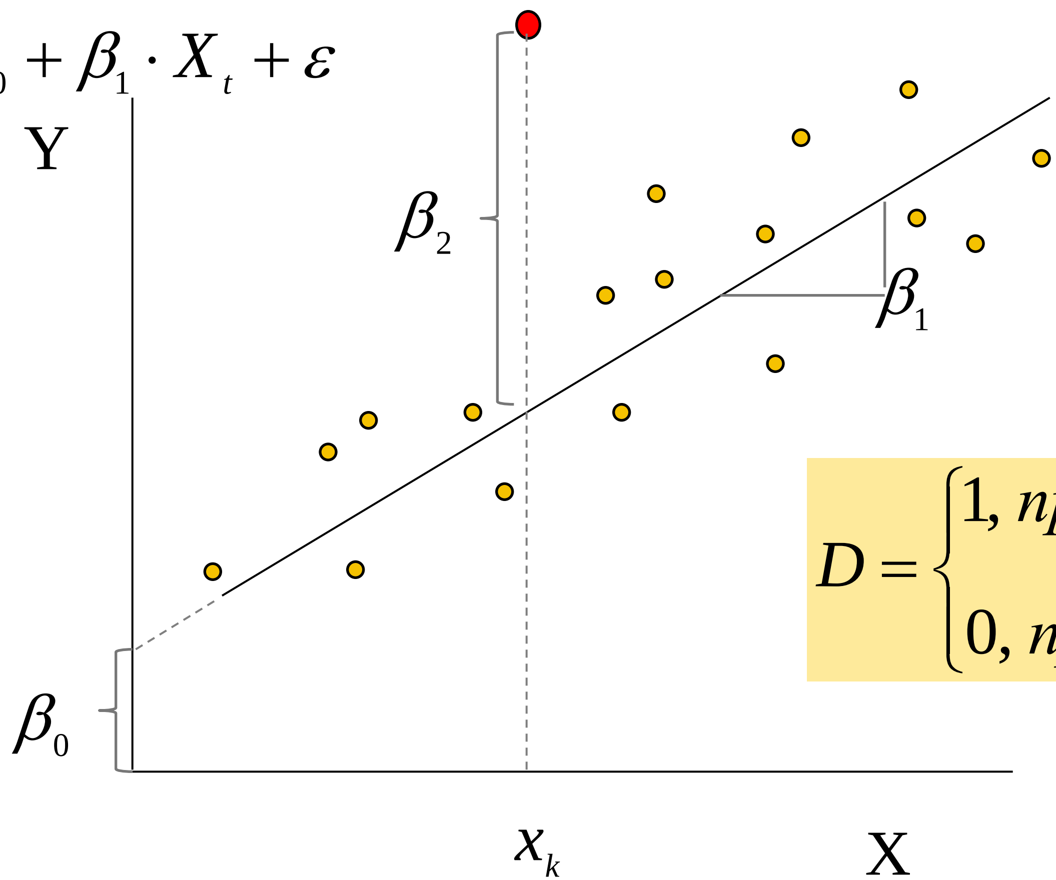
Необходимость введения фиктивных переменных такого рода выдвигается на основе предварительного анализа исходных статистических данных, но главным образом ***на основе анализа случайных отклонений модели.***

DUMMY VARIABLES: SINGLE OUTLIER

$$Y_t = \beta_0 + \beta_1 \cdot X_t + \beta_2 \cdot D_t + \varepsilon, t = \overline{1,n}$$

$$t = k: \quad Y_t = (\beta_0 + \beta_2) + \beta_1 \cdot X_t + \varepsilon$$

$$t \neq k: \quad Y_t = \beta_0 + \beta_1 \cdot X_t + \varepsilon$$



$$D = \begin{cases} 1, npu\ t = k \\ 0, npu\ t \neq k \end{cases}$$

DUMMY VARIABLES: INTERCEPT AND SLOPE OF TREND, SEASONAL

При учете в модели изменения линии уровня тренда в данных, возможно ввести в модель переменную ***TI*** в аддитивной форме.

При учете в модели излома тренда, возможно либо ввести в модели переменную ***TI*** в мультипликативной форме, либо переменную ***TS*** в аддитивной форме.

Моделирование сезонности с помощью соответствующих переменных ***Q*** квартальной или ***M*** помесячной сезонности возможно как в аддитивной, так и в мультипликативной форме.

$$TI = \begin{cases} 1, \text{при } t \geq k \\ 0, \text{при } t < k \end{cases} \qquad TS = \begin{cases} t, \text{при } t \geq k \\ 0, \text{при } t < k \end{cases} \qquad Q_i = \begin{cases} 1, \text{при } Q = i \\ 0, \text{при } Q \neq i \end{cases}$$

“DUMMY VARIABLE TRAP”

Качественный признак может иметь ***не менее двух*** альтернативных значений.

Dummy Variable Trap или Ловушка фиктивных переменных - ситуация, когда в модель для описания качественного признака вводится количество фиктивных переменных, равное количеству альтернативных значений. В таком случае сумма таких фиктивных переменных дает число 1 (единицу), фактически константу, которой в модели уже соответствует коэффициент свободного члена. Это приводит к нарушению требуемых свойств получаемых оценок коэффициентов.

DUMMY VARIABLES AND TEST CHOW

Although the Chow test is usually used to test for a structural break, an alternative test involving the dummy variables can also be used.

Use the F-test for nested models from lecture 5 (!).

It involves running two regressions, one with the dummy variables (unrestricted or full model) and collecting the RSS.

The other regression excludes the dummy variables (restricted model) and collect this RSS.

Use the F-test formula to produce the F-statistic and compare with the critical values, the null hypothesis being that the regression is structurally stable.

EXAMPLE: DUMMY

i	Y=Salary, euro per week	X1=Work hours, per week	X2=sex, woman or man	
1	60,5	27	1	man
2	62,5	32	0	woman
3	64,5	34	0	woman
4	57,5	26	1	man
5	56,5	24	1	man
6	68,0	32	1	man
7	62,5	32	0	woman
8	67,0	34	1	man
9	60,0	32	0	woman
10	58,0	28	1	man

EXAMPLE: DUMMY

$$Y = 21{,}805 + 1{,}248 \cdot X_1 + 3{,}868 \cdot X_2 + \varepsilon \qquad R^2 = 0{,}853$$

$$(S)\ (6{,}495)\quad (0{,}198)\quad (1{,}359)$$

$$(t)\ (3{,}357)\quad (6{,}301)\quad (2{,}847)$$

$$(P)\ (0{,}012)\quad (0{,}000)\quad (0{,}025)$$

$$Y = 34{,}026 + 0{,}919 \cdot X_1 + \varepsilon \qquad R^2 = 0{,}684$$

$$(S)\ (6{,}697)\quad (0{,}221)$$

$$(t)\ (5{,}081)\quad (4{,}158)$$

$$(P)\ (0{,}001)\quad (0{,}003)$$

EXAMPLE: DUMMY

$$Y = 21{,}805 + 1{,}248 \cdot X_1 + 3{,}868 \cdot X_2 + \varepsilon \qquad R^2 = 0{,}853$$

$$(S) \quad (6{,}495) \quad (0{,}198) \quad (1{,}359)$$

$$(t) \quad (3{,}357) \quad (6{,}301) \quad (2{,}847)$$

$$(P) \quad (0{,}012) \quad (0{,}000) \quad (0{,}025)$$

$$Y = 34{,}026 + 0{,}919 \cdot X_1 + \varepsilon \qquad R^2 = 0{,}684$$

$$(S) \quad (6{,}697) \quad (0{,}221)$$

$$(t) \quad (5{,}081) \quad (4{,}158)$$

$$(P) \quad (0{,}001) \quad (0{,}003)$$

EXAMPLE: DUMMY

$R^2_{FULL} = 0,853$

$R^2_{REDUCED} = 0,684$

$$H_0 : R^2_A = R^2_B,\ R^2_A - R^2_B = 0 \quad or \quad H_0 : \beta_2 = 0$$

$$H_1 : R^2_A > R^2_B,\ R^2_A - R^2_B > 0$$

$$F = \frac{\left(R^2_A - R^2_B\right)/k}{\left(1 - R^2_A\right)/\left(n - m - 1\right)} \sim F_{\alpha;k;n-m-1}\ (k = m - l)$$

$$F = \frac{\left(0,853 - 0,684\right)/1}{\left(1 - 0,853\right)/\left(10 - 2 - 1\right)} =$$

$$= 8,048 \sim F_{\alpha;k;n-m-1} = F_{0,05;1;7} = 5,591\ or\ P = 0,0252$$

EXAMPLE: DUMMY

$$Y = 21{,}805 + 1{,}248 \cdot X_1 + 3{,}868 \cdot X_2 + \varepsilon \qquad R^2 = 0{,}853$$

$$(S) \quad (6{,}495) \quad (0{,}198) \quad (1{,}359)$$

$$(t) \quad (3{,}357) \quad (6{,}301) \quad (2{,}847)$$

$$(P) \quad (0{,}012) \quad (0{,}000) \quad (0{,}025)$$

$$Y = 16{,}333 + 1{,}417 \cdot X_1 + 9{,}541 \cdot X_2 - 0{,}175 \cdot \left(X_1 X_2\right) + \varepsilon \qquad R^2 = 0{,}854$$

$$(S) \quad (34{,}594) \quad (1{,}064) \quad (35{,}155) \quad (1{,}086)$$

$$(t) \quad (0{,}472) \quad (1{,}331) \quad (0{,}271) \quad (-0{,}161)$$

$$(P) \quad (0{,}654) \quad (0{,}231) \quad (0{,}795) \quad (0{,}877)$$

EXAMPLE: DUMMY

$R^2_{FULL} = 0,854$ $R^2_{REDUCED} = 0,853$

$$H_0 : R^2_A = R^2_B,\; R^2_A - R^2_B = 0 \quad or \quad H_0 : \beta_2 = 0$$

$$H_1 : R^2_A > R^2_B,\; R^2_A - R^2_B > 0$$

$$F = \frac{\left(R^2_A - R^2_B\right)/k}{\left(1 - R^2_A\right)/\left(n - m - 1\right)} \sim F_{\alpha;k;n-m-1} \; (k = m - l)$$

$$F = \frac{\left(0,854 - 0,853\right)/1}{\left(1 - 0,854\right)/\left(10 - 2 - 1\right)} =$$

$$= 0,048 \sim F_{\alpha;k;n-m-1} = F_{0,05;1;7} = 5,591 \; or \; P = 0,8328$$

EXAMPLE: TEST CHOW

i	Y=Salary, euro per week	X1=Work hours, per week	X2=sex, woman or man	
1	62,5	32	0	woman
2	64,5	34	0	woman
3	62,5	32	0	woman
4	60,0	32	0	woman
5	58,0	28	1	man
6	57,5	26	1	man
7	56,5	24	1	man
8	68,0	32	1	man
9	67,0	34	1	man
10	60,5	27	1	man

EXAMPLE : TEST CHOW

$$Y = 34{,}026 + 0{,}919 \cdot X_1 + \varepsilon \qquad R^2 = 0{,}684$$

$$(P) \quad (0{,}001) \qquad (0{,}003) \qquad RSS_0 = 44{,}1665$$

$$Only \quad WOMAN \quad n_1 = 4$$

$$Y = 16{,}333 + 1{,}417 \cdot X_1 + \varepsilon \qquad R^2 = 0{,}591$$

$$(P) \quad (0{,}608) \qquad (0{,}231) \qquad RSS_1 = 4{,}1667$$

$$Only \quad MAN \quad n_1 = 6$$

$$Y = 25{,}874 + 1{,}241 \cdot X_1 + \varepsilon \qquad R^2 = 0{,}872$$

$$(P) \quad (0{,}019) \qquad (0{,}007) \qquad RSS_2 = 16{,}2133$$

TEST CHOW

$$H_0 : RSS_0 = RSS_1 + RSS_2 \Leftrightarrow RSS_0 - (RSS_1 + RSS_2) = 0$$

$$H_1 : RSS_0 > RSS_1 + RSS_2$$

$$F = \frac{\left(RSS_0 - (RSS_1 + RSS_2)\right)/(m+1)}{\left(RSS_1 + RSS_2\right)/\left(n - 2(m+1)\right)} \sim F_{\alpha;m+1;n-2(m+1)}$$

$$F = \frac{\left(44{,}1665 - (4{,}1667 + 16{,}2133)\right)/(1+1)}{\left(4{,}1667 + 16{,}2133\right)/\left(10 - 2(1+1)\right)} =$$

$$= 3{,}501 \sim F_{\alpha;m+1;n-2(m+1)} = F_{0,05;2;6} = 5{,}143 \; or \; P = 0{,}098$$