CONTINUED FROM LAST CLASS.

- LET $r^2_{y3\cdot2}$ = PARTIAL COEFFICIENT OF DETERMINATION BETWEEN y AND x_3, AFTER REMOVING THE INFLUENCE OF x_2.

THEN PARTIAL CORRELATION COEFFICIENT BETWEEN y AND x_3, HOLDING x_2 CONSTANT $= \sqrt{r^2_{y3\cdot2}} = r_{y3\cdot2}$.

(NOTE: R^2 = PLAIN COEFFICIENT OF DETERMINATION
$r^2_{yxx} \Rightarrow$ PARTIAL " ——————— ")

- WE CAN FIND THE REGRESSION OF y ON x_2.

THE COEFFICIENT OF DETERMINATION IS:

$$R^2_{y\cdot2} = \frac{\text{EXPLAINED SS DUE TO } x_2}{\text{TOTAL SUM OF SQUARE}}$$

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i}$$

OR $\hat{y}_i = \hat{\beta}_2 x_{2i}$.

$$\hat{y}_i = \hat{Y}_i - \bar{Y}$$
$$= (\hat{\beta}_1 - \hat{\beta}_1) + \hat{\beta}_2 (X_{2i} - \bar{X}_i)$$

(a) $1 - R^2_{y\cdot2}$ = PART OF VARIATION IN Y THAT REMAINED UNEXPLAINED, WHICH IS DUE TO THE FACT THAT x_3 AND OTHER FACTORS ARE NOT INCLUDED IN THE REGRESSION ABOVE.

- THEN WE ADD x_3.

FROM ABOVE $\hat{y}_{\cdot2} = \hat{\beta}_2 x_2 \rightarrow$ WHERE

$$R^2_{y\cdot2} = \frac{\text{EXPLAINED SS DUE TO } x_2}{\text{TOTAL SUM OF SQUARE}}$$

NOW $\hat{y}_{\cdot23} = \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 \rightarrow$ WHERE

$$R^2_{y\cdot23} = \frac{\text{EXPLAINED SS DUE TO } x_2 \text{ AND } x_3}{\text{TOTAL SUM OF SQUARE}}$$

(b) DEFINE $1 - R^2_{y\cdot23}$ = PART OF VARIATION THAT REMAIN UNEXPLAINED BECAUSE OTHER VARIABLES ARE LEFT OUT.

THEREFORE (a) − (b) $= (1 - R^2_{y\cdot2}) - (1 - R^2_{y\cdot23})$

= PART OF VARIATION THAT CAN BE EXPLAINED BY x_3. (AFTER REMOVING THE INFLUENCE OF x_2).

IN RELATIVE TERMS:

$$r^2_{y3\cdot2} = \frac{(1 - R^2_{y\cdot2}) - (1 - R^2_{y\cdot23})}{(1 - R^2_{y\cdot2})}$$

$= r^2_{13\cdot2}$ IN GUJARATI, WHERE $y = 1$

= PARTIAL COEFFICIENT OF DETERMINATION BETWEEN y AND x_3 AFTER REMOVING THE INFLUENCE OF x_2.

IF WE PUT IN THE EXPRESSION FOR THE EXPLAINED SS AND THE FORMULA FOR $\hat{\beta}_2$ AND $\hat{\beta}_3$, IT CAN BE SHOWN THAT THE PARTIAL COEFFICIENT OF DETERMINATION CAN BE EXPRESSED IN TERMS OF SIMPLE CORRELATION COEFFICIENTS.

$$\boxed{r^2_{y3\cdot 2} = \frac{(r_{y3} - r_{y2} r_{23})^2}{(1 - r^2_{y2})(1 - r^2_{23})}}$$

WHERE
$$r_{y2} = \frac{\sum_{i=1}^{n} x_{2i} y_i}{\sqrt{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} y_i^2}}$$

$$r_{y3} = \frac{\sum_{i=1}^{n} x_{3i} y_i}{\sqrt{\sum_{i=1}^{n} x_{3i}^2 \sum_{i=1}^{n} y_i^2}}$$

$$r_{23} = \frac{\sum_{i=1}^{n} x_{2i} x_{3i}}{\sqrt{\sum_{i=1}^{n} x_{2i}^2 \sum_{i=1}^{n} x_{3i}^2}}$$

AND SINCE THE SQUARE ROOT OF r^2_{y32} = $r_{y3\cdot 2}$ AND EQUAL TO

$$\boxed{r_{y3\cdot 2} = \frac{r_{y3} - r_{y2} r_{23}}{\sqrt{(1 - r^2_{y2})(1 - r^2_{23})}}}$$

$\Rightarrow$ PARTIAL CORRELATION BETWEEN y AND x_3, HOLDING THE INFLUENCE OF x_2 CONSTANT.

IN A SIMILAR MANNER,

$r_{y2\cdot 3}$ CAN BE DERIVED IN SIMILAR FASHION...

$$\boxed{r_{y2\cdot 3} = \frac{r_{y2} - r_{y3} r_{23}}{\sqrt{(1 - r^2_{y3})(1 - r^2_{23})}}}$$

WHERE $r_{y2\cdot 3}$ = PARTIAL CORRELATION COEFFICIENT OF y ON x_2, AFTER REMOVING THE INFLUENCE OF x_3.

SUMMARY GENERAL LINEAR MODEL (GLM)

MODEL $Y = X\beta + U$

ASSUMPTIONS ① $E(U) = 0$

② $E(UU') = \sigma^2_U I_n$ (HOMOSKEDASTIC

③ X IS NON STOCHASTIC (A SET OF FIXED NUMBER) (DISTURBANCE)

④ X HAS RANK = k, $n > k$

⑤ THE MODEL IS CORRECTLY SPECIFIED.

OLS ESTIMATOR

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$E(\hat{\beta}) = \beta \qquad \text{(UNBIASNESS)}$$

$$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$$

WE ESTIMATE σ_u^2 BY

$$s^2 = \frac{e'e}{n-k} = \frac{\sum_{i=1}^{n} e_i^2}{n-k}$$

• GIVEN X_{n+1}, A VECTOR OF OBSERVATION FOR $X_1, X_2, \ldots, X_k$ AT TIME $(n+1)$ OUTSIDE THE SAMPLE PERIOD, THE BEST UNBIASED PREDICTOR FOR THE MEAN VALUE OF Y_{n+1} OR $E(Y_{n+1} | X_{n+1})$ IS

$$\hat{Y}_{n+1} = X'_{n+1}\hat{\beta}$$

FORECAST ERROR $= Y_{n+1} - \hat{Y}_{n+1}$

VARIANCE OF FORECAST ERROR $= \sigma_F^2 = \sigma_u^2 + \sigma^2_{\hat{Y}_{n+1}}$

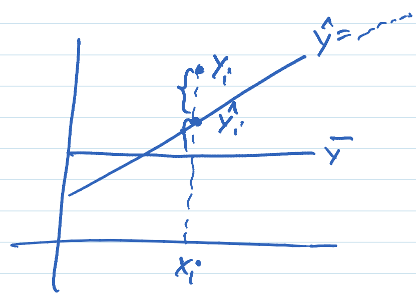
COEFFICIENT OF DETERMINATION (R^2)

FROM $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \ldots + \hat{\beta}_k X_{ki}$.

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki} \qquad \text{(IN DEVIATION FORM)}$$

$$R^2 = \frac{\text{EXPLAINED SS}}{\text{TOTAL SS}} = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2}$$

$$= \frac{\sum (\hat{\beta}_2 x_{2i} + \ldots + \hat{\beta}_k x_{ki})}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$



AND FROM $y_i^2 = \hat{y}_i^2 + e_i^2$

$$\therefore R^2 = \frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}.$$

$\sum_{i=1}^{n} y_i^2$ $\qquad$ $\sum_{i=1}^{n} y_i^2$

$\boxed{\text{REMARK: ONE SHOULD NOT COMPARE } R^2 \text{ OF THE MODELS WITH DIFFERENT } k.}$

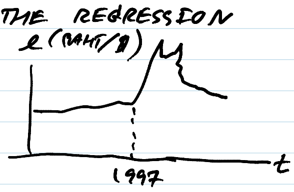
k: # OF EXPLANATORY VARIABLES INCLUDING INTERCEPT TERM

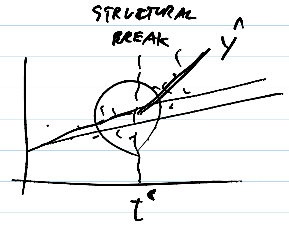
29 MAR 2012

HYPOTHESIS TESTING FOR GENERAL LINEAR MODEL

THIS IS CONSIDERED AS "STATISTICAL INFERENCE"

1. TEST ON INDIVIDUAL PARAMETER (INDIVIDUAL TEST) ✓
2. TEST ON OVERALL INFLUENCE OF EXPLANATORY VARIABLES (JOINT TEST)
3. TEST ON ADDITIONAL VARIABLE
4. TEST ON LINEAR RESTRICTION OF PARAMETER
   - TWO OR MORE COEFFICIENTS ARE EQUAL (ex: $\beta_2 = \beta_3 = \beta_4$)
   - COEFFICIENTS SATISFY CERTAIN RESTRICTIONS (ex: $\beta_2 + \beta_3 = 1$)
5. TEST FOR EQUALITY / STABILITY OF THE REGRESSION EQUATION. (STABILITY TEST)





RELATED TO THE USE OF DUMMY VARIABLE(S)

TEST ON INDIVIDUAL PARAMETER (INDIVIDUAL TEST)

FROM MODEL $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \cdots + \hat{\beta}_k X_{ki}$

(NOTE: IN SIMPLE LINEAR REGRESSION: $Y_i = \hat{\beta}_1 + \hat{\beta}_{2i}$.

$H_0: \beta_2 = 0 \Rightarrow X_2$ HAS NO INFLUENCE ON Y

$H_0: \beta_3 = 0 \Rightarrow X_3$ HAS NO INFLUENCE ON Y

$\vdots$

$H_0: \beta_k = 0 \Rightarrow X_k$ HAS NO INFLUENCE ON Y

EX: CHILD MORTALITY (CM)
PER CAPITA GNP (X_2)
FEMALE LITERACY RATE (X_3)

PER CAPITA GNP (X_2)

FEMALE LITERACY RATE (X_3)

$n = 64$ (CROSS SECTIONAL DATA)

$$CM_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

WE EXPECT THAT COEFFICIENT OF X_2 TO BE NEGATIVE
AND COEFFICIENT OF X_3 TO BE NEGATIVE.

THE RESULTS

$$\hat{CM}_i = 263.64 - \underline{0.0056}\, X_2 - \underline{2.2316}\, X_3 .$$

S.E	(11.59)	(0.0019)	(0.2099)
$\hat{t}$	(22.74)	(−2.82)	(−10.63)
p-value	(0.000)	(0.0065)	(0.000)

$R^2 = 0.7077$ $\bar{R}^2 = 0.6981$

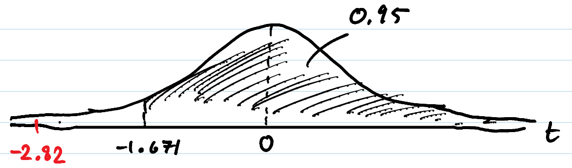
TEST ON INDIVIDUAL COEFFICIENTS

$H_0: \beta_2 = 0$ (X_2 HAS NO INFLUENCE ON CM)

$H_1: \beta_2 < 0$ (X_2 HAS A NEGATIVE INFLUENCE ON CM)

$$\hat{t} = \frac{\hat{\beta}_2 - \beta_2}{S_{\hat{\beta}_2}} = -2.82$$

$t_{.05,\, \underbrace{64-3}_{=61}} = 1.671$ (TRY TO FIND THIS NUMBER IN t-TABLE)



IF THE NULL HYPOTHESIS WERE TRUE, THE PROBABILITY (OR CHANCE) OF OBSERVING A VALUE OF $\hat{t}$ BEING LESS THAN −1.671 IS ONLY 0.05. BUT WE STILL OBSERVE IT!

THEREFORE WE SHOULD REJECT THE NULL HYPOTHESIS AND CONCLUDE THAT X_2 HAS A NEGATIVE INFLUENCE ON THE DEPENDENT VARIABLE (CM).

IN SIMILAR MANNER, WE COULD PERFORM THE TEST ON β_i WHERE $i = 1, 2, 3, \ldots, k$.

② $\boxed{\text{TESTING OVERALL SIGNIFICANCE}}$ (JOINT TEST)

Ex: $CM_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$

$H_0: \beta_2 = \beta_3 = 0$ (BOTH ARE "SIMULTANEOUSLY" EQUAL TO ZERO)

$H_1:$ OTHERWISE

WE USE ANOVA TABLE TO GIVE US F-STATISTIC.

SOURCE OF VARIATION	SUM OF SQUARE	MEAN SUM OF SQUARE
REGRESSION (EXPLAINED)	$\sum_{i=1}^{n} \hat{y}_i^2 \left[= \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2\right]$	$\sum_{i=1}^{n} \hat{y}_i^2 / (k-1)$
RESIDUAL (UNEXPLAINED)	$\sum_{i=1}^{n} e_i^2 \left[= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2\right]$	$\sum_{i=1}^{n} e_i^2 / (n-k)$
TOTOL SUM OF SQUARE	$\sum_{i=1}^{n} y_i^2 \left[= \sum_{i=1}^{n} (Y_i - \bar{Y})^2\right]$	$\sum_{i=1}^{n} y_i^2 / (n-1)$

$$\boxed{\hat{F} = \frac{\sum_{i=1}^{n} \hat{y}_i^2 / (k-1)}{\sum_{i=1}^{n} e_i^2 / (n-k)}} \sim F_{k-1,\, n-k}$$

FOLLOWS OR EXHIBITS

• VERIFY THIS LINK BETWEEN $\hat{F}$ AND R^2

HINT: $\underbrace{\sum_{i=1}^{n} y_i^2}_{(TSS)} = \underbrace{\sum_{i=1}^{n} \hat{y}_i^2}_{(ESS)} + \underbrace{\sum_{i=1}^{n} e_i^2}_{(RSS)}$

$$1 = \underbrace{\frac{\sum_{i=1}^{n} \hat{y}_i^2}{\sum_{i=1}^{n} y_i^2}}_{R^2} + \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

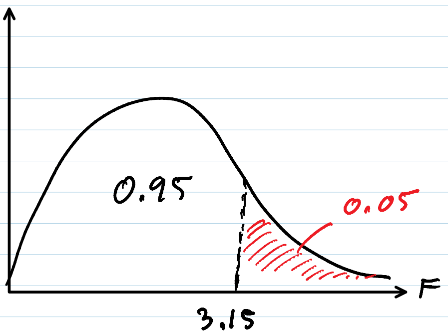
$$\therefore R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} y_i^2}$$

WE CAN EXPRESS $\hat{F}$ IN TERM OF R^2:

$$\boxed{\hat{F} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}}$$

FROM THE CHILD MORTALITY EXAMPLE

SINCE $R^2 = 0.7077 \Rightarrow \hat{F} = \frac{0.7077/(3-1)}{(1-0.7077)/(64-3)} = 73.84.$



FROM F TABLE, $F_{\underset{\alpha}{0.05},\, \underset{(k-1)}{(3-1)},\, \underset{(n-k)}{(64-3)}} = 3.15$

IF THE NULL HYPOTHESIS WERE TRUE, CHANCE OF OBSERVING $\hat{F}$ GREATER THAN THE VALUE OF 3.15 IS ONLY 0.05. BUT WE OBSERVE $\hat{F} = 73.84$, THEREFORE H_0 SHOULD NOT BE TRUE AND WE REJECT H_0. WE CONCLUDE THAT X_2 AND X_3 "JOINTLY" INFLUENCE THE DEPENDENT VARIABLE (CM).

NOTE H_0: $\beta_2 = \beta_3 = 0$

H_1: $\beta_2 \neq 0$ OR $\beta_3 \neq 0$ OR β_2 AND $\beta_3 \neq 0$ $\equiv$ OTHER WISE.