Independent Dummy Variables (1)

Transform a binary qualitative variable (with non-numerical values) to a dummy variable. For example,

GENDER = 1 if the observation is male

= 0 if it is female

Independent Dummy Variables (2)

Note that

1) the setting is arbitrary. However, it should make the interpretation simple.

2) In general, zero will be given to the reference case. In the example, female is treated as “reference”.

Single Dummy (1)

Example Expenditure function

Male $EXP_i = \beta_1 + \beta_2 INC_i + \varepsilon_i$

Female $EXP_i = \gamma_1 + \gamma_2 INC_i + \varepsilon_i$

$$H_0 : \beta_1 = \gamma_1, \beta_2 = \gamma_2$$

$$H_1 : \beta_1 \neq \gamma_1, \beta_2 \neq \gamma_2$$

Do male and female share the same mean equation?

Single Dummy (2)

Integrate the expenditure functions.

Define a dummy variable

$MALE$=1 for male

$$EXP_i = \beta_1 MALE_i + \gamma_1 (1 - MALE_i)$$
$$+ \beta_2 MALE_i \bullet INC_i$$
$$+ \gamma_2 (1 - MALE_i) \bullet INC_i + \varepsilon_i$$

Single Dummy (3)

Integrate the two data sets.

EXP	MALE	1-MALE	MALE*INC	(1-MALE)*INC
Male	1	0	Male	0
	.	.		.
EXP	1	0	INC	0
Female	0	1	0	Female
	.	.	.	.
EXP	0	1	0	INC

Y X

Single Dummy (4)

- Run OLS on the integrated data
- Number of parameters (K) = 4
- degrees of freedom = $n_M + n_F - 4$

where n_M and n_F are sample size of the male and female samples, respectively

- Do Generalized F-test with

$$F_{cal} \sim F(2, n_M + n_F - 4)$$

Single Dummy (5)

RLS or Wald test is OK.

Accept => male and female share the same intercept and the same slope in the expenditure function

Chow Test is equivalent to the two-run generalized F-test (RLS)

Single Dummy (6)

Chow Test (cont'd)

$$F_{cal} = \frac{SSR_T - (SSR_M + SSR_F)}{SSR_M + SSR_F} \frac{n_M + n_F - 4}{2}$$

$$\sim F(2, n_M + n_F - 4)$$

where SSR_M and SSR_F are the sum of squared residuals from the tow separate runs and SSR_T is generated from OLS run on the stacked data set

Single Dummy (7)

Stacked data set

EXP	1	INC
Male	1	Male
	.	
EXP	1	INC
Female	1	Female
	.	.
EXP	1	INC

Y X

Single Dummy (8)

Chow Test (cont'd)

Note that SSR_M+SSR_F is the same as SSR of the unrestricted model and SSR_T is the SSR of the restricted model.

It is referred to as Chow's Breakpoint test in Eviews.

Single Dummy (9)

Partial Chow Test

Case 1 the slopes are identical. Only the intercept could be different

$$EXP_i = \beta_1 MALE_i + \gamma_1(1 - MALE_i) + \beta_2 INC_i + \varepsilon_i$$

$$H_0 : \beta_1 = \gamma_1$$

$$H_1 : \beta_1 \neq \gamma_1$$

$$F_{cal} \sim F(1, n_M + n_F - 3)$$

Single Dummy (10)

Partial Chow Test

Case 2 the intercepts are identical. Only the slope could be different

$$EXP_i = \beta_1 + \beta_2 MALE_i \bullet INC_i + \gamma_2(1 - MALE_i) \bullet INC_i + \varepsilon_i$$

$$H_0 : \beta_2 = \gamma_2$$

$$H_1 : \beta_2 \neq \gamma_2$$

$$F_{cal} \sim F(1, n_M + n_F - 3)$$

Multi-dummies (1)

Expenditure also depends on Province they live.

Define BKK=1 if the observation is in Bangkok. Otherwise, BKK=0.

<u>Unrestricted Model</u>

$$
\begin{aligned}
EXP_i = {} & \alpha_1 MALE_i BKK_i + \beta_1 MALE_i (1 - BKK_i) \\
& + \gamma_1 (1 - MALE_i) BKK_i \\
& + \delta_1 (1 - MALE_i)(1 - BKK_i) \\
& + \alpha_2 MALE_i \bullet INC_i + \gamma_2 (1 - MALE_i) \bullet INC_i + \varepsilon_i
\end{aligned}
$$

Multi-dummies (2)

<u>Assumption</u>

The slope depends only on the gender not location but the intercept could depend on both gender and province.

Test if neither gender nor province has no effect on the expenditure.

$$H_0 : \alpha_1 = \beta_1 = \gamma_1 = \delta_1, \alpha_2 = \gamma_2$$

$$H_1 : \alpha_1 \neq \beta_1 \neq \gamma_1 \neq \delta_1, \alpha_2 \neq \gamma_2$$

Multi-dummies (3)

<u>Restricted Model</u>

$$EXP_i = \alpha_1 + \alpha_2 INC_i + \varepsilon_i$$

Do F-test using

$$F_{cal} \sim F(4, n_M + n_F - 6)$$

Incremental Setting (1)

So far, the above setting of dummy variables is of “switching” type. A dummy variable is used to select the appropriate parameter for each observation. Another setting is “incremental” type.

Incremental Setting (2)

Example Expenditure Function

$$EXP_i = \gamma_1 + \delta_1 MALE_i$$
$$+ \gamma_2 INC_i + \delta_2 MALE_i \bullet INC_i + \varepsilon_i$$

Note that

γ_1, γ_2 are the intercept and slope for female

δ_1 is the intercept deviation for male

δ_2 is the slope deviation for male

Multi-category Variable (1)

Examples

- Color: RED, BLUE, GREEN
- Day-of-Week: Mo,Tu,We,Th,Fr

Question

Price volatility of a certain day depends on whether it is a week-beginning day, a mid-week day or a week-ending day.

Multi-category Variable (2)

Define

STD = 1 if the day is a week-starting day
= 0, otherwise.

MID = 1 if it is a mid-week day
= 0, otherwise.

END = 1 if it is a week-ending day
= 0, otherwise.

Note that STD+MID+END=1 always and one of the dummies could be eliminated

Multi-category Variable (3)

Switching Setting

$$VOL_i = \beta_1 STD_i + \beta_2 STD_i \bullet VAL_i$$
$$+ \gamma_1 MID_i + \gamma_2 MID_i \bullet VAL_i$$
$$+ \delta_1 END_i + \delta_2 END_i \bullet VAL_i + \varepsilon_i$$

Note that

β_1, γ_1, δ_1 are the intercepts for each day category

β_2, $\gamma_{2/}$, δ_2 are the slopes for each day category

Multi-category Variable (4)

Incremental Setting

$$VOL_i = \beta_1 STD_i + \beta_2 STD_i \bullet VAL_i$$
$$+\gamma_1 + \gamma_2 \bullet VAL_i$$
$$+\delta_1 END_i + \delta_2 END_i \bullet VAL_i + \varepsilon_i$$

Note that

$\beta_1+\gamma_1$, γ_1, $\delta_1+\gamma_1$ are the intercepts for each category

$\beta_2+\gamma_2$, $\gamma_{2/}$, $\delta_2+\gamma_2$ are the slopes for each category

Mid-week is used as the reference

Multi-category Variable (5)

To test that there is no difference between categories (F-test or Chi-square test)

Switching Setting

$$H_0 : \beta_1 = \gamma_1 = \delta_1, \beta_2 = \gamma_2 = \delta_2$$
$$H_1 : \beta_1 \neq \gamma_1 \neq \delta_1, \beta_2 \neq \gamma_2 \neq \delta_2$$

Incremental Setting

$$H_0 : \beta_1 = \delta_1 = 0, \beta_2 = \delta_2 = 0$$
$$H_1 : \beta_1 \neq \delta_1 \neq 0, \beta_2 \neq \delta_2 \neq 0$$

Multi-category Variable (6)

$$F_{cal} \sim F(4, n_{STD} + n_{MID} + n_{END} - 6)$$

where $n = n_{STD} + n_{MID} + n_{END}$

Chow Test for Switching Setting

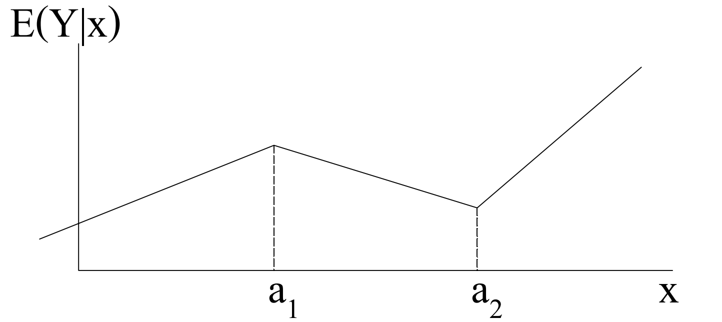
$$F_{cal} = \frac{SSR_T - (SSR_{STD} + SSR_{MID} + SSR_{END})}{SSR_{STD} + SSR_{MID} + SSR_{END}} \frac{n-6}{4}$$

where SSR_{STD}, SSR_{MID} and SSR_{END} are SSR from separate runs and SSR_T is SSR from the stacked data

Piecewise Linear Model (1)

Conditional mean of Y is a piecewise linear function of X

Two kinks take place at $X=a_1$ and $X=a_2$



Piecewise Linear Model (2)

Switching Setting

Define

$D_0 = 1$ if $X < a_1$
$= 0$, otherwise.

$D_1 = 1$ if $a_1 < X < a_2$
$= 0$, otherwise.

$D_2 = 1$ if $X > a_2$
$= 0$, otherwise.

Piecewise Linear Model (3)

Switching Setting

$$Y_i = \beta_1 D_{0i} + \beta_2 X_i D_{0i} + (\beta_1 + \beta_2 a_1) D_{1i} + \beta_3 (X_i - a_1) D_{1i} + \{\beta_1 + \beta_2 a_1 + \beta_3 (a_2 - a_1)\} D_{2i} + \beta_4 (X_i - a_2) D_{2i} + \varepsilon_i$$

β_1 is the intercept

$\beta_2, \beta_3, \beta_4$ are the slope of each section

Piecewise Linear Model (4)

Re-arrange

$$Y_i = \beta_1 \underbrace{(D_{0i} + D_{1i} + D_{2i})}_{X_1} + \beta_2 \underbrace{(X_i D_{0i} + a_1 D_{1i} + a_1 D_{2i})}_{X_2} + \beta_3 \underbrace{\{(X_i - a_1) D_{1i} + (a_2 - a_1) D_{2i}\}}_{X_3} + \beta_4 \underbrace{(X_i - a_2) D_{2i}}_{X_4} + \varepsilon_i$$

Piecewise Linear Model (5)

Incremental Setting

Define

$D_1 = 1$ if $X > a_1$
$= 0$, otherwise.

$D_2 = 1$ if $X > a_2$
$= 0$, otherwise.

Piecewise Linear Model (6)

Incremental Setting (more simple)

$$Y_i = \beta_1 + \beta_2 X_i + \delta_1 (X_i - a_1) D_{1i} + \delta_2 (X_i - a_2) D_{2i} + \varepsilon_i$$

β_1 is the intercept

β_2 is the slope of the first section

δ_1, δ_2 are the incremental of the slope

Piecewise Linear Model (7)

To test whether the function is single piece (slope is constant for all X)

Switching Setting

$$H_0 : \beta_2 = \beta_3 = \beta_4$$

$$H_1 : \beta_2 \neq \beta_3 \neq \beta_4$$

Incremental Setting

$$H_0 : \delta_1 = \delta_2 = 0$$

$$H_1 : \delta_1 \neq \delta_2 \neq 0$$

Non-linear Approximation to a Piecewise Linear Model (1)

In general, the locations of kinks (a_1,a_2) are unknown. How can we estimate them?

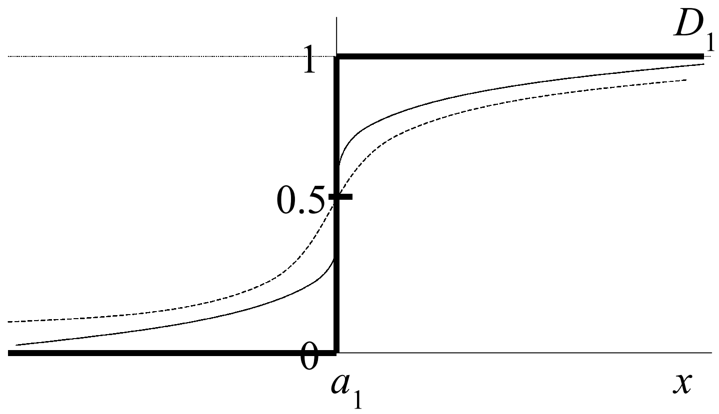
Logistic Transformation of dummy variables for Incremental Setting

Approx. continuous function of X_i for D_{1i} is

$$\frac{1}{1 + e^{-M(X_i - a_1)}}$$

if M is a large positive value

Non-linear Approximation to a Piecewise Linear Model (2)



Non-linear Approximation to a Piecewise Linear Model (3)

Approx. non-linear regression

$$Y_i = \beta_1 + \beta_2 X_i + \delta_1 \frac{X_i - a_1}{1 + e^{-M(X_i - a_1)}}$$

$$+ \delta_2 \frac{X_i - a_2}{1 + e^{-M(X_i - a_2)}} + \varepsilon_i$$

Apply Non-linear LS. Another approx. is a polynomial regression.