ECON 351* -- NOTE 21

Using Dummy Variables to Test for Coefficient Differences

- The **population regression equation (PRE)** for the general multiple linear regression model takes the form:

  $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

  where u_i is an iid (independently and identically distributed) random error term.

- To illustrate the use of dummy variable regressors in testing for coefficient differences, we **use the *simplest possible* multiple regression equation** with only two non-constant regressors:

  $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

1. Definition and Properties of Indicator (Dummy) Variables

♦ **Indicator (or dummy) variables** are ***binary*** **variables** -- i.e., **variables that take *only two values***.

The value **1** indicates **the presence** of some characteristic or attribute.

The value **0** indicates **the absence** of that same characteristic or attribute.

♦ Consider a **two-way partitioning** of a population or sample into **two *mutually exclusive* and *exhaustive* subsets or groups**, denoted as Group 1 and Group 2.

- Let $\mathbf{D1_i}$ be the **Group 1 dummy variable**, defined as follows:

$$D1_i = 1 \text{ if observation i belongs to Group 1 } (\forall\ i \in \text{ Group 1})$$
$$= 0 \text{ if observation i does not belong to Group 1 } (\forall\ i \notin \text{ Group 1}).$$

- Let $\mathbf{D2_i}$ be the **Group 2 dummy variable**, defined as follows:

$$D2_i = 1 \text{ if observation i belongs to Group 2 } (\forall\ i \in \text{ Group 2})$$
$$= 0 \text{ if observation i does not belong to Group 2 } (\forall\ i \notin \text{ Group 2}).$$

♦ **Adding-Up Property of the Indicator Variables $\mathbf{D1_i}$ and $\mathbf{D2_i}$**

For each and every i (population member or sample observation):

if $D1_i = 1$ then $D2_i = 0$

and

if $D2_i = 1$ then $D1_i = 0$.

The definition of the indicator variables $D1_i$ and $D2_i$ thus implies that they satisfy the following **adding-up property**:

$$\mathbf{D1_i + D2_i = 1} \qquad \forall\ \mathbf{i} \qquad \text{i.e., for all i = 1, ..., N.}$$

♦ **Implications of the Adding-Up Property**

1. **Only *one* of the two dummy variables $D1_i$ and $D2_i$** is required to ***completely represent* the *two-way partitioning*** of a population and sample into Group 1 and Group 2.

   - given $D2_i$ values, the adding-up property implies that $D1_i = 1 - D2_i$.

   - given $D1_i$ values, the adding-up property implies that $D2_i = 1 - D1_i$.

2. **<u>General Rule</u>:** A ***categorical* variable with n categories** can be completely represented by **a set of n−1 indicator (dummy) variables**.

   The **general adding-up property** states that

   $$D1_i + D2_i + D3_i + \cdots + Dn_i = 1 \quad \forall \ i.$$

♦ **Examples of Using Dummy Variables to Represent Categorical Variables**

Example 1: The categorical variable gender or sex, $\boldsymbol{sex_i}$, which is coded in Canadian Census data files for individual persons as follows:

$\boldsymbol{sex_i}$ = 1 if person i is ***female***;
= 2 if person i is ***male***.

- Define a set of two (2) gender dummy variables to represent the categorical variable $\boldsymbol{sex_i}$:

$\boldsymbol{female_i}$ = 1 if $\boldsymbol{sex_i}$ = 1, = 0 otherwise;
$\boldsymbol{male_i}$ = 1 if $\boldsymbol{sex_i}$ = 2, = 0 otherwise.

- By definition, the two gender dummy variables satisfy the ***adding-up property***:

$$\text{female}_i + \text{male}_i = 1 \qquad \forall \; i \text{ (for all i).}$$

- **<u>Implication of the adding-up property</u>**: The partitioning of the adult population or sample into ***two*** **mutually exclusive and exhaustive gender categories** can be completely represented by **either *one* of the two gender dummy variables** $\boldsymbol{female_i}$ and $\boldsymbol{male_i}$.

  For example, the male dummy variable $\boldsymbol{male_i}$ can be computed from the female dummy variable $\boldsymbol{female_i}$ as follows:

$$\text{male}_i = 1 - \text{female}_i \qquad \forall \; i.$$

Example 2: Consider a categorical variable $AGEGROUP_i$ defined as follows:

$AGEGROUP_i$ = 1 if person i is 15-19 years of age;
= 2 if person i is 20-24 years of age;
= 3 if person i is 25-34 years of age;
= 4 if person i is 35-44 years of age;
= 5 if person i is 45-54 years of age;
= 6 if person i is 55-64 years of age;
= 7 if person i is 65 years of age or over.

- Define a set of seven (7) age group dummy variables to represent the categorical variable $AGEGROUP_i$.

  $DAGE1_i$ = 1 if $AGEGROUP_i$ = 1, = 0 otherwise;
  $DAGE2_i$ = 1 if $AGEGROUP_i$ = 2, = 0 otherwise;
  $DAGE3_i$ = 1 if $AGEGROUP_i$ = 3, = 0 otherwise;
  $DAGE4_i$ = 1 if $AGEGROUP_i$ = 4, = 0 otherwise;
  $DAGE5_i$ = 1 if $AGEGROUP_i$ = 5, = 0 otherwise;
  $DAGE6_i$ = 1 if $AGEGROUP_i$ = 6, = 0 otherwise;
  $DAGE7_i$ = 1 if $AGEGROUP_i$ = 7, = 0 otherwise.

- By definition, the seven age group dummy variables satisfy the ***adding-up property***:

$$DAGE1_i + DAGE2_i + DAGE3_i + DAGE4_i + DAGE5_i + DAGE6_i + DAGE7_i = 1 \quad \forall \ i.$$

- **Implication of the adding-up property**: The partitioning of the population or sample into ***seven*** **mutually exclusive and exhaustive age groups** can be completely represented by ***any six*** **of the seven age group dummy variables** $DAGE1_i$, $DAGE2_i$, $DAGE3_i$, $DAGE4_i$, $DAGE5_i$, $DAGE6_i$, and $DAGE7_i$.

  For example, the age group dummy variable $DAGE1_i$ can be computed from the other six age group dummy variables as follows:

$$DAGE1_i = 1 - DAGE2_i - DAGE3_i - DAGE4_i - DAGE5_i - DAGE6_i - DAGE7_i \quad \forall \ i.$$

Example 3: The categorical variable Marital Status, $MARSTAT_i$, which is coded in Canadian Census data files for individuals as follows:

$MARSTAT_i$ = 1 if person i is ***single, never married***;
= 2 if person i is ***married with spouse present***;
= 3 if person i is ***widowed***;
= 4 if person i is ***separated***;
= 5 if person i is ***divorced***.

- Define a set of five (5) marital status dummy variables to represent the categorical variable $MARSTAT_i$:

  $mssgl_i$ = 1 if $MARSTAT_i$ = 1, = 0 otherwise;
  $msmar_i$ = 1 if $MARSTAT_i$ = 2, = 0 otherwise;
  $mswid_i$ = 1 if $MARSTAT_i$ = 3, = 0 otherwise;
  $mssep_i$ = 1 if $MARSTAT_i$ = 4, = 0 otherwise;
  $msdiv_i$ = 1 if $MARSTAT_i$ = 5, = 0 otherwise.

- By definition, the five marital status dummy variables satisfy the ***adding-up property***:

$$\text{mssgl}_i + \text{msmar}_i + \text{mswid}_i + \text{mssep}_i + \text{msdiv}_i = 1 \qquad \forall \text{ i (for all i).}$$

- **<u>Implication of the adding-up property</u>**: The partitioning of the adult population or sample into ***five*** **mutually exclusive and exhaustive marital status categories** can be completely represented by ***any four*** **of the five marital status dummy variables** ***$mssgl_i$***, ***$msmar_i$***, ***$mswid_i$***, ***$mssep_i$***, and ***$msdiv_i$***.

  For example, the maritial status dummy variable ***$mssgl_i$*** can be computed from the other four marital status dummy variables as follows:

$$\text{mssgl}_i = 1 - \text{msmar}_i - \text{mswid}_i - \text{mssep}_i - \text{msdiv}_i \qquad \forall \text{ i.}$$

2. The Framework

- Consider a ***sample* or *population* of N observations** that is partitioned into two mutually exclusive and exhaustive subsamples or subsets:

  (1) Group 1 subsample of **N_1 observations**, for which

  $$D1_i = 1 \quad \text{and} \quad D2_i = 0$$

  (2) Group 2 subsample of **N_2 observations**, for which

  $$D2_i = 1 \quad \text{and} \quad D1_i = 0$$

  Note: The total number of observations is **$N = N_1 + N_2$**.

- **<u>Case 1</u>: All regression coefficients are *constant*, or *equal***, across the entire population. In this case, the PRE for the entire population (and hence for all N sample observations) is:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad \forall\ i \text{ (for all i)} \quad i = 1, ..., N = N_1 + N_2 \qquad \textbf{... (1)}$$

- **<u>Case 2</u>: All regression coefficients *differ*, or are *unequal***, between the two groups. In this case, Group 1 and Group 2 have completely different PREs.

  (1) The **Group 1 PRE** is

  $$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_{1i} \qquad \forall\ i \in \text{Group 1} \qquad \textbf{... (2.1)}$$

  where u_{1i} is $NID(0, \sigma_1^2)$.

  (2) The **Group 2 PRE** is

  $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{2i} \qquad \forall\ i \in \text{Group 2} \qquad \textbf{... (2.2)}$$

  where u_{2i} is $NID(0, \sigma_2^2)$.

- **<u>Objective</u>: To test for *pairwise* coefficient *differences* between the Group 1 and Group 2 PRFs.**

  - The ***null* hypothesis of *complete* coefficient equality** is

    $$H_0:\ \alpha_0 = \beta_0 \quad \textit{and} \quad \alpha_1 = \beta_1 \quad \textit{and} \quad \alpha_2 = \beta_2 \qquad \textbf{... (3)}$$

    or, more compactly,

    $$H_0:\ \alpha_j = \beta_j \qquad \forall\ j = 0, 1, 2.$$

    Interpretation of H_0: The null hypothesis H_0 says that ***all* regression coefficients** (including the intercept coefficients) **are *equal* in the Group 1 and Group 2 regression functions**.

  - The ***alternative* hypothesis** is

    $$H_1:\ \alpha_0 \neq \beta_0 \quad \textit{and/or} \quad \alpha_1 \neq \beta_1 \quad \textit{and/or} \quad \alpha_2 \neq \beta_2 \qquad \textbf{... (4)}$$

    or, more compactly,

    $$H_1:\ \alpha_j \neq \beta_j \qquad j = 0, 1, 2.$$

    Interpretation of H_1: The alternative hypothesis H_1 says that ***at least one (some or all)* of the regression coefficients are *unequal (or different)* in the Group 1 and Group 2 regression functions**.

❑ <u>The Restricted and Unrestricted Models</u> Corresponding to H_0 and H_1

♦ The ***unrestricted*** **model** corresponding to the **alternative hypothesis H_1** consists of the Group 1 PRE (2.1) and the Group 2 PRE (2.2):

Group 1 PRE: $Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_{1i}$ $\quad i = 1, ..., N_1$ **... (2.1)**

Group 2 PRE: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{2i}$ $\quad i = 1, ..., N_2$ **... (2.2)**

♦ The ***restricted*** **model** corresponding to the **null hypothesis H_0** is obtained by imposing on the unrestricted model the coefficient restrictions specified by H_0.

1) Set $\alpha_0 = \beta_0$ and $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$ in PRE (2.1): the two PREs (2.1) and (2.2) can then be written as

Group 1 PRE: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{2i}$ $\quad i = 1, ..., N_1$

Group 2 PRE: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{2i}$ $\quad i = 1, ..., N_2$.

2) Assume that the error terms u_{1i} and u_{2i} have the same distribution. Since u_{1i} and u_{2i} both have zero means, this amounts to assuming that **u_{1i} *and* u_{2i}** have the **same variance** -- i.e., that $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

<u>Result</u>: The Group 1 PRE and the Group 2 PRE are *identical*, so that the ***restricted*** **model** for the entire sample of $N = N_1 + N_2$ observations can be written as PRE (1):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad \forall\ i = 1, ..., N = N_1 + N_2 \qquad \textbf{... (1)}$$

❑ Two Approaches to Testing for Inter-Group Coefficient Differences

Both approaches involve using an ***F-test*** to perform a test of

H_0: $\alpha_0 = \beta_0$ *and* $\alpha_1 = \beta_1$ *and* $\alpha_2 = \beta_2$ (3 coefficient restrictions)

against

H_1: $\alpha_0 \neq \beta_0$ *and/or* $\alpha_1 \neq \beta_1$ *and/or* $\alpha_2 \neq \beta_2$

Approach 1: *Separate* Regressions Approach

- does *not* use indicator (or dummy) variables as regressors.

Approach 2: *Pooled* Regression Approach

- uses indicator (or dummy) variables as regressors.

3. Approach 1: Separate Regressions Approach

❑ Step 1: Under H_0, estimate equation (1) by OLS on the full sample of N observations.

The ***restricted*** **OLS-SRE** is

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{u}_i \qquad i = 1, ..., N = N_1 + N_2 \qquad \textbf{... (4.1)}$$

The ***restricted*** **residual sum-of-squares** is

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with } df_0 = df_R = N - K_0 \qquad \textbf{... (4.2)}$$

- **Step 2: Under H_1, *separately* estimate by OLS *(1)* PRE (2.1) on the subsample of N_1 observations for Group 1 and *(2)* PRE (2.2) on the subsample of N_2 observations for Group 2.**

1. The **Group 1 OLS-SRE** for the subsample of N_1 observations for Group 1 is

$$Y_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \hat{\alpha}_2 X_{2i} + \hat{u}_{1i} \qquad i = 1, ..., N_1 \qquad \textbf{... (5.1)}$$

The **residual sum-of-squares** for the Group 1 OLS-SRE is

$$RSS_{(1)} = \sum_{i=1}^{N_1} \hat{u}_{1i}^2 \qquad \text{with } df_{(1)} = N_1 - K_0 = N_1 - 3 \qquad \textbf{... (5.2)}$$

2. The **Group 2 OLS-SRE** for the subsample of N_2 observations for Group 2 is

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{u}_{2i} \qquad i = 1, ..., N_2 \qquad \textbf{... (6.1)}$$

The **residual sum-of-squares** for the Group 2 OLS-SRE is

$$RSS_{(2)} = \sum_{i=1}^{N_2} \hat{u}_{2i}^2 \qquad \text{with } df_{(2)} = N_2 - K_0 = N_2 - 3 \qquad \textbf{... (6.2)}$$

3. The ***unrestricted*** **residual sum-of-squares** is

$$RSS_1 = RSS_U = RSS_{(1)} + RSS_{(2)} = \sum_{i=1}^{N_1} \hat{u}_{1i}^2 + \sum_{i=1}^{N_2} \hat{u}_{2i}^2 \qquad \textbf{... (7.1)}$$

$$\text{with } df_1 = df_U = df_{(1)} + df_{(2)} = N - 2K_0 = N - 6 \qquad \textbf{... (7.2)}$$

Calculation of $df_1 = df_U$:

$$\begin{aligned} df_1 &= df_{(1)} + df_{(2)} \\ &= N_1 - K_0 + N_2 - K_0 \\ &= N_1 + N_2 - 2K_0 \\ &= N - 2K_0 = N - 2(3) = N - 6. \end{aligned}$$

- **<u>Step 3</u>: Compute the *sample value* of the F-statistic** for testing H_0 against H_1.

$$F_{SR} = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{\left[RSS_0 - \left(RSS_{(1)} + RSS_{(2)}\right)\right]/(df_0 - df_1)}{(RSS_1 + RSS_2)/df_1} \qquad \textbf{... (8)}$$

- The ***restricted* RSS under H_0** is

$$RSS_0 = RSS_R \qquad \text{with} \qquad df_0 = df_R = N - K_0.$$

- The ***unrestricted* RSS under H_1** is

$$RSS_1 = RSS_U = RSS_{(1)} + RSS_{(2)} = \sum_{i=1}^{N_1} \hat{u}_{1i}^2 + \sum_{i=1}^{N_2} \hat{u}_{2i}^2$$

with $df_1 = df_U = df_{(1)} + df_{(2)} = N - 2K_0$.

- The difference **($RSS_0 - RSS_1$)** in the numerator of F is

$$RSS_0 - RSS_1 = RSS_0 - \left(RSS_{(1)} + RSS_{(2)}\right) = \sum_{i=1}^{N} \tilde{u}_i^2 - \left(\sum_{i=1}^{N_1} \hat{u}_{1i}^2 + \sum_{i=1}^{N_2} \hat{u}_{2i}^2\right).$$

- The ***numerator* degrees of freedom** equal

$$df_{num} = df_0 - df_1 = (N - K_0) - (N - 2K_0) = N - K_0 - N + 2K_0 = K_0.$$

- The ***denominator* degrees of freedom** equal

$$df_{den} = df_1 = N - 2K_0.$$

- The ***sample value*** **of the F-statistic for H_0 against H_1** is

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1}$$

$$= \frac{(RSS_0 - RSS_1)/K_0}{RSS_1/(N - 2K_0)} \qquad \textbf{... (9)}$$

$$= \frac{\left[RSS_0 - (RSS_{(1)} + RSS_{(2)})\right]/K_0}{(RSS_1 + RSS_2)/(N - 2K_0)}$$

❑ **<u>Step 4</u>: Apply the conventional *decision rule*.**

- ***Null distribution*** **of F_0:** $F_0 \sim F(K_0, N - 2K_0)$ under H_0.

- ***Decision Rule:*** At the 100α percent significance level

1. ***reject*** **H_0** if $F_0 \geq F_\alpha(K_0,\, N - 2K_0)$ *or* p-value for $F_0 \leq \alpha$;

2. ***retain*** **H_0** if $F_0 < F_\alpha(K_0,\, N - 2K_0)$ *or* p-value for $F_0 > \alpha$.

4. Approach 2: Pooled (Full-Interaction) Regression Approach

- **<u>Strategy</u>:** Estimate a **single pooled regression equation** that incorporates the full set of coefficient differences between the PREs for Group 1 and Group 2.

- **<u>Derivation of Pooled Regression Equation</u>**

Group 1 PRE: $Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_{1i}$ $\quad i = 1, ..., N_1$ **... (2.1)**

Group 2 PRE: $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_{2i}$ $\quad i = 1, ..., N_2$ **... (2.2)**

1. **Multiply** equation (2.1) by $D1_i$ and equation (2.2) by $D2_i$:

$$D1_i Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + D1_i u_i \qquad (10.1)$$

$$D2_i Y_i = \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + D2_i u_i \qquad (10.2)$$

Note: We again assume that $u_{1i} = u_{2i} = u_i$ -- i.e., that u_{i1} and u_{i2} have identical distributions, so that they have equal variances $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

2. **Combine equations (10.1) and (10.2)** by ***adding them together*** for each observation i = 1, ..., N.

$$\begin{aligned} D1_i Y_i + D2_i Y_i &= \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + D1_i u_i \\ &\quad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + D2_i u_i \\ &= \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} \\ &\quad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} \\ &\quad + D1_i u_i + D2_i u_i \end{aligned}$$

or

$$\begin{aligned} (D1_i + D2_i) Y_i &= \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} \\ &\quad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} \qquad (11) \\ &\quad + (D1_i + D2_i) u_i \end{aligned}$$

3. Use the *adding-up property* to set $D1_i + D2_i = 1$ for all i in equation (11):

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \quad \forall\ i \qquad (12.0)$$

- ***Result:*** This equation (12.0) is the ***pooled full-interaction PRE*** **corresponding to the alternative hypothesis H_1**:

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \quad \forall\ i \qquad \mathbf{(12.0)}$$

- **Characteristics of equation (12.0):**

1) Equation (12.0) has ***no intercept* coefficient**.

2) Equation (12.0) **contains the *full set of regression coefficients*** for both the Group 1 PRF (α_0, α_1, and α_2) and the Group 2 PRF (β_0, β_1, and β_2).

Estimation of the pooled full-interaction regression equation (12.0) thus yields coefficient estimates of both the Group 1 PRF and the Group 2 PRF.

3) Both the **Group 1 and Group 2 PREs can be obtained from equation (12.0)**.

- **Group 1 PRE** is obtained by setting $D1_i = 1$ and $D2_i = 0$ in (12.0):

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + u_i \qquad \forall\ i \text{ such that } D1_i = 1$$

- **Group 2 PRE** is obtained by setting $D2_i = 1$ and $D1_i = 0$ in (12.0):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad \forall\ i \text{ such that } D2_i = 1$$

❑ **<u>Two Alternative Forms</u> of the *Pooled* Full-Interaction PRE (12.0):**

1. **Group 1 selected as *base group***: set $D1_i = 1 - D2_i$ in equation (12.0);
2. **Group 2 selected as *base group***: set $D2_i = 1 - D1_i$ in equation (12.0).

1. <u>Group 1 selected as base group</u>: Set $D1_i = 1 - D2_i$ in equation (12.0):

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \qquad \textbf{...(12.0)}$$

$$\begin{aligned} Y_i &= \alpha_0(1 - D2_i) + \alpha_1(1 - D2_i)X_{1i} + \alpha_2(1 - D2_i)X_{2i} \\ &\quad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \\ &= \alpha_0 - \alpha_0 D2_i + \alpha_1 X_{1i} - \alpha_1 D2_i X_{1i} + \alpha_2 X_{2i} - \alpha_2 D2_i X_{2i} \\ &\quad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \end{aligned}$$

Re-arrange the terms on the right-hand side of the above equation:

$$\begin{aligned} Y_i &= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} - \alpha_0 D2_i - \alpha_1 D2_i X_{1i} - \alpha_2 D2_i X_{2i} \\ &\qquad + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \\ &= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \beta_0 D2_i - \alpha_0 D2_i \\ &\qquad + \beta_1 D2_i X_{1i} - \alpha_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} - \alpha_2 D2_i X_{2i} + u_i \end{aligned}$$

Collect like terms in the above equation:

$$\begin{aligned} Y_i &= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \beta_0 D2_i - \alpha_0 D2_i \\ &\qquad + \beta_1 D2_i X_{1i} - \alpha_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} - \alpha_2 D2_i X_{2i} + u_i \\ &= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + (\beta_0 - \alpha_0) D2_i \\ &\qquad + (\beta_1 - \alpha_1) D2_i X_{1i} + (\beta_2 - \alpha_2) D2_i X_{2i} + u_i \end{aligned}$$

Define the coefficients of $D2_i$, $D2_iX_{1i}$, and $D2_iX_{2i}$ as

$$\gamma_0 = \beta_0 - \alpha_0; \quad \gamma_1 = \beta_1 - \alpha_1; \quad \gamma_2 = \beta_2 - \alpha_2.$$

Finally, the pooled full-interaction regression equation with Group 1 as the base group can be written as:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + (\beta_0 - \alpha_0)D2_i + (\beta_1 - \alpha_1)D2_i X_{1i} + (\beta_2 - \alpha_2)D2_i X_{2i} + u_i$$

$$= \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \gamma_0 D2_i + \gamma_1 D2_i X_{1i} + \gamma_2 D2_i X_{2i} + u_i$$

- ***Result:*** The **pooled full-interaction regression equation with *Group 1* as the *base group*** can be written as:

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \gamma_0 D2_i + \gamma_1 D2_i X_{1i} + \gamma_2 D2_i X_{2i} + u_i \qquad \textbf{(12.1)}$$

where $\gamma_0 = \beta_0 - \alpha_0; \quad \gamma_1 = \beta_1 - \alpha_1; \quad \gamma_2 = \beta_2 - \alpha_2$.

Interpretation of the Coefficients of PRE (12.1):

1) Equation (12.1) contains an ***intercept* coefficient** – specifically, the **Group 1 *intercept* coefficient α_0**.

2) The ***slope* coefficients of the *regressors* X_{1i} *and* X_{2i}** are the **Group 1 *slope* coefficients α_1 *and* α_2**.

3) The **coefficient γ_0 of the Group 2 *dummy variable* $D2_i$** is the difference between the Group 2 intercept coefficient β_1 and the Group 1 intercept coefficient α_1 – i.e., $\boldsymbol{\gamma_0 = \beta_0 - \alpha_0}$.

4) The **coefficient γ_1 of the *interaction term* $D2_iX_{1i}$** is the difference between the Group 2 slope coefficient for X_{1i} (β_1) and the corresponding Group 1 slope coefficient for X_{1i} (α_1) – i.e., $\boldsymbol{\gamma_1 = \beta_1 - \alpha_1}$.

5) The **coefficient γ_2 of the *interaction term* $D2_iX_{2i}$** is the difference between the Group 2 slope coefficient for X_{2i} (β_2) and the corresponding Group 1 slope coefficient for X_{2i} (α_2) – i.e., $\boldsymbol{\gamma_2 = \beta_2 - \alpha_2}$.

2. **<u>Group 2 selected as base group</u>**: Set $D2_i = 1 - D1_i$ in equation (12.0):

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \qquad \textbf{…(12.0)}$$

$$\begin{aligned} Y_i &= \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} \\ &\quad + \beta_0(1 - D1_i) + \beta_1(1 - D1_i)X_{1i} + \beta_2(1 - D1_i)X_{2i} + u_i \\ &= \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} \\ &\quad + \beta_0 - \beta_0 D1_i + \beta_1 X_{1i} - \beta_1 D1_i X_{1i} + \beta_2 X_{2i} - \beta_2 D1_i X_{2i} + u_i \end{aligned}$$

Re-arrange the terms on the right-hand side of the above equation:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} \\ &\qquad - \beta_0 D1_i - \beta_1 D1_i X_{1i} - \beta_2 D1_i X_{2i} + u_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha_0 D1_i - \beta_0 D1_i \\ &\qquad + \alpha_1 D1_i X_{1i} - \beta_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} - \beta_2 D1_i X_{2i} + u_i \end{aligned}$$

Collect like terms in the above equation:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \alpha_0 D1_i - \beta_0 D1_i \\ &\qquad + \alpha_1 D1_i X_{1i} - \beta_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} - \beta_2 D1_i X_{2i} + u_i \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (\alpha_0 - \beta_0) D1_i \\ &\qquad + (\alpha_1 - \beta_1) D1_i X_{1i} + (\alpha_2 - \beta_2) D1_i X_{2i} + u_i \end{aligned}$$

Define the coefficients of $D1_i$, $D1_i X_{2i}$, and $D1_i X_{3i}$ as

$$\delta_0 = \alpha_0 - \beta_0; \quad \delta_1 = \alpha_1 - \beta_1; \quad \delta_2 = \alpha_2 - \beta_2 .$$

Finally, the pooled full-interaction regression equation with Group 2 as the base group can be written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + (\alpha_0 - \beta_0)D1_i + (\alpha_1 - \beta_1)D1_i X_{1i} + (\alpha_2 - \beta_2)D1_i X_{2i} + u_i$$

$$= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \delta_0 D1_i + \delta_1 D1_i X_{1i} + \delta_2 D1_i X_{2i} + u_i$$

♦ ***Result:*** The **pooled full-interaction regression equation with Group 2 as the base group** can be written as:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \delta_0 D1_i + \delta_1 D1_i X_{1i} + \delta_2 D1_i X_{2i} + u_i \qquad \textbf{(12.2)}$$

where $\delta_0 = \alpha_0 - \beta_0$; $\delta_1 = \alpha_1 - \beta_1$; $\delta_2 = \alpha_2 - \beta_2$.

Interpretation of the Coefficients of PRE (12.2):

1) Equation (12.2) contains an ***intercept*** **coefficient** – specifically, the **Group 2 *intercept* coefficient β_0**.

2) The ***slope*** **coefficients of the *regressors* X_{1i} *and* X_{2i}** are the **Group 2 *slope* coefficients β_1 *and* β_2**.

3) The **coefficient δ_0 of the Group 1 *dummy variable* $D1_i$** is the difference between the Group 1 intercept coefficient α_0 and the Group 2 intercept coefficient β_0 – i.e., $\boldsymbol{\delta_0 = \alpha_0 - \beta_0}$.

4) The **coefficient δ_1 of the *interaction term* $D1_iX_{1i}$** is the difference between the Group 1 slope coefficient for X_{1i} (α_1) and the corresponding Group 2 slope coefficient for X_{1i} (β_1) – i.e., $\boldsymbol{\delta_1 = \alpha_1 - \beta_1}$.

5) The **coefficient δ_2 of the *interaction term* $D1_iX_{2i}$** is the difference between the Group 1 slope coefficient for X_{2i} (α_2) and the corresponding Group 2 slope coefficient for X_{2i} (β_2) – i.e., $\boldsymbol{\delta_2 = \alpha_2 - \beta_2}$.

❑ **<u>Properties of Pooled Regression Equations</u> (12.0), (12.1) and (12.2)**

There are ***three* different but equivalent ways** of writing the pooled full-interaction regression equation:

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \quad \textbf{...(12.0)}$$

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \gamma_0 D2_i + \gamma_1 D2_i X_{1i} + \gamma_2 D2_i X_{2i} + u_i \quad \textbf{(12.1)}$$

where $\gamma_0 = \beta_0 - \alpha_0$; $\gamma_1 = \beta_1 - \alpha_1$; $\gamma_2 = \beta_2 - \alpha_2$.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \delta_0 D1_i + \delta_1 D1_i X_{1i} + \delta_2 D1_i X_{2i} + u_i \quad \textbf{(12.2)}$$

where $\delta_0 = \alpha_0 - \beta_0$; $\delta_1 = \alpha_1 - \beta_1$; $\delta_2 = \alpha_2 - \beta_2$.

Note: The **γ_j coefficients in pooled regression (12.1) are *equal in magnitude* but *opposite in sign* to the δ_j coefficients in pooled regression (12.2)**.

$$\gamma_j = \beta_j - \alpha_j = -\delta_j \qquad \text{or} \qquad \delta_j = \alpha_j - \beta_j = -\gamma_j \qquad j = 0, 2, 3.$$

1. **Equations (12.0), (12.1), and (12.2) are *observationally equivalent***: OLS estimation of equations (12.0), (12.1), and (12.2) yield ***identical* values** of

   RSS $\equiv$ residual sum-of-squares

   $\hat{\sigma}^2 \equiv$ the estimator of the error variance σ^2

   $R^2 \equiv$ the coefficient of determination.

2. The ***unrestricted* RSS** from OLS estimation of equations (12.0), (12.1), and (12.2) equals

   $$RSS_1 = RSS_U = RSS_{(1)} + RSS_{(2)} \qquad \text{with} \qquad df_1 = df_U = N - 2K_0.$$

   That is, the RSS from OLS estimation of pooled equations (12.0), (12.1), and (12.2) equals the sum of the RSS values from separate OLS estimation of the Group 1 and Group 2 regression equations.

❑ <u>F-Tests of Complete Coefficient Equality</u> – Three Equivalent Tests

<u>TEST 1</u>: Use OLS estimates of pooled equation (12.0)

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \quad \forall\ i = 1, \ldots, N \qquad \ldots(12.0)$$

to perform an **F-test** of

H_0: $\alpha_j = \beta_j \quad \forall\ j = 0, 1, 2$

H_1: $\alpha_j \neq \beta_j \quad j = 0, 1, 2.$

♦ ***Restricted*** **OLS-SRE** corresponding to H_0 is obtained by OLS estimation of the restricted equation (1) on the full sample of $N = N_1 + N_2$ observations:

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{u}_i \qquad \forall\ i = 1, \ldots, N = N_1 + N_2 \qquad \ldots (1)$$

Yields ***restricted*** **RSS**

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with } df_0 = df_R = N - K_0 = N - 3.$$

♦ ***Unrestricted*** **OLS-SRE** corresponding to H_1 is obtained by OLS estimation of equation (12.0):

$$Y_i = \hat{\alpha}_0 D1_i + \hat{\alpha}_1 D1_i X_{1i} + \hat{\alpha}_2 D1_i X_{2i} + \hat{\beta}_0 D2_i + \hat{\beta}_1 D2_i X_{1i} + \hat{\beta}_2 D2_i X_{2i} + \hat{u}_i \quad \forall\ i = 1, \ldots, N$$

Yields ***unrestricted*** **RSS**

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with } df_1 = df_U = N - K = N - 2K_0 = N - 6.$$

<u>TEST 2</u>: Use OLS estimates of pooled equation (12.1)

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \gamma_0 D2_i + \gamma_1 D2_i X_{1i} + \gamma_2 D2_i X_{2i} + u_i \qquad \textbf{(12.1)}$$

$$\gamma_0 = \beta_0 - \alpha_0; \quad \gamma_1 = \beta_1 - \alpha_1; \quad \gamma_2 = \beta_2 - \alpha_2.$$

to perform an **F-test** of

H_0: $\gamma_j = 0 \quad \forall\ j = 0, 1, 2 \quad \Rightarrow \quad \gamma_j = \beta_j - \alpha_j = 0 \quad \forall\ j = 0, 1, 2$

H_1: $\gamma_j \neq 0 \quad j = 0, 1, 2 \quad \Rightarrow \quad \gamma_j = \beta_j - \alpha_j \neq 0 \quad j = 0, 1, 2.$

- ***Restricted*** **OLS-SRE** corresponding to H_0 is obtained by OLS estimation of equation (12.1) with $\boldsymbol{\gamma_0 = \gamma_1 = \gamma_2 = 0}$ on the full sample of $N = N_1 + N_2$ observations:

$$Y_i = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_{1i} + \tilde{\alpha}_2 X_{2i} + \tilde{u}_i \qquad \forall\ i = 1, ..., N = N_1 + N_2$$

Yields ***restricted*** **RSS**

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with} \quad df_0 = df_R = N - K_0 = N - 3.$$

- ***Unrestricted*** **OLS-SRE** corresponding to H_1 is obtained by OLS estimation of equation (12.1):

$$Y_i = \hat{\alpha}_0 + \hat{\alpha}_1 X_{1i} + \hat{\alpha}_2 X_{2i} + \hat{\gamma}_0 D2_i + \hat{\gamma}_1 D2_i X_{1i} + \hat{\gamma}_2 D2_i X_{2i} + \hat{u}_i$$

$$\hat{\gamma}_0 = \hat{\beta}_0 - \hat{\alpha}_0; \quad \hat{\gamma}_1 = \hat{\beta}_1 - \hat{\alpha}_1; \quad \hat{\gamma}_2 = \hat{\beta}_2 - \hat{\alpha}_2.$$

Yields ***unrestricted*** **RSS**

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with} \quad df_1 = df_U = N - K = N - 2K_0 = N - 6.$$

<u>TEST 3</u>: Use OLS estimates of pooled equation (12.2)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \delta_0 D1_i + \delta_1 D1_i X_{1i} + \delta_2 D1_i X_{2i} + u_i \qquad \textbf{(12.2)}$$

$$\delta_0 = \alpha_0 - \beta_0; \quad \delta_1 = \alpha_1 - \beta_1; \quad \delta_2 = \alpha_2 - \beta_2.$$

to perform an **F-test** of

H_0: $\delta_j = 0 \quad \forall\ j = 0, 1, 2 \quad \Rightarrow \quad \delta_j = \alpha_j - \beta_j = 0\ \forall\ j = 0, 1, 2$

H_1: $\delta_j \neq 0 \quad j = 0, 1, 2 \quad \Rightarrow \quad \delta_j = \alpha_j - \beta_j \neq 0 \quad j = 0, 1, 2.$

- ***Restricted* OLS-SRE** corresponding to H_0 is obtained by OLS estimation of equation (12.2) with $\boldsymbol{\delta_0 = \delta_1 = \delta_2 = 0}$ on the full sample of $N = N_1 + N_2$ observations:

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{u}_i \qquad \forall\ i = 1, ..., N = N_1 + N_2$$

Yields ***restricted* RSS**

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with} \quad df_0 = df_R = N - K_0 = N - 3.$$

- ***Unrestricted* OLS-SRE** corresponding to H_1 is obtained by OLS estimation of equation (12.2):

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\delta}_0 D1_i + \hat{\delta}_1 D1_i X_{1i} + \hat{\delta}_2 D1_i X_{2i} + \hat{u}_i$$

$$\hat{\delta}_0 = \hat{\alpha}_0 - \hat{\beta}_0; \quad \hat{\delta}_1 = \hat{\alpha}_1 - \hat{\beta}_1; \quad \hat{\delta}_2 = \hat{\alpha}_2 - \hat{\beta}_2.$$

Yields ***unrestricted* RSS**

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with} \quad df_1 = df_U = N - K = N - 2K_0 = N - 6.$$

❑ Equivalence of Three F-Tests

The **three F-tests** based on pooled full-interaction regression equations (12.0), (12.1), and (12.2) ***are equivalent to each other*** **and** ***to the F-test computed using the separate regressions approach***.

$$F(12.0) = F(12.1) = F(12.2) = F_{SR} \sim F(K_0, N - 2K_0) \text{ under } H_0.$$

where all four F-statistics F(12.0) = F(12.1) = F(12.2) = F_{SR} = F_0 and

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} = \frac{(RSS_0 - RSS_1)/K_0}{RSS_1/(N - 2K_0)}.$$

❑ Meaning/Interpretation of Joint F-Tests of Complete Coefficient Equality

Meaning of TEST 1: Uses OLS estimates of pooled equation (12.0)

$$Y_i = \alpha_0 D1_i + \alpha_1 D1_i X_{1i} + \alpha_2 D1_i X_{2i} + \beta_0 D2_i + \beta_1 D2_i X_{1i} + \beta_2 D2_i X_{2i} + u_i \qquad \textbf{(12.0)}$$

to perform an **F-test** of H_0: $\alpha_j = \beta_j \;\forall\; j = 0, 1, 2$ versus H_1: $\alpha_j \neq \beta_j$ for j = 0, 1, 2.

The **population regression functions** (PRFs) **for Group 1 and Group 2** implied by pooled regression equation (12.0) give the conditional mean value of Y_i for each group for any given values of the explanatory variables X_{1i} and X_{2i}.

- The **Group 1 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.0) is:

$$E\left(Y_i \mid X_{1i}, X_{2i}, D1_i = 1, D2_i = 0\right) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i}$$

- The **Group 2 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.0) is:

$$E\left(Y_i \mid X_{1i}, X_{2i}, D1_i = 0, D2_i = 1\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

The **null hypothesis of complete coefficient equality** – H_0: $\alpha_j = \beta_j \;\forall\; j = 0, 1, 2$ – means that

$$\begin{aligned} E\left(Y_i \mid X_{1i}, X_{2i}, D1_i = 1, D2_i = 0\right) &= E\left(Y_i \mid X_{1i}, X_{2i}, D1_i = 0, D2_i = 1\right) \\ &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \end{aligned}$$

i.e., the **Group 1 conditional mean value of Y** ***equals*** the **Group 2 conditional mean value of Y** for any given values of X_1 and X_2.

<u>Meaning of TEST 2</u>: Uses OLS estimates of pooled equation (12.1)

$$Y_i = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \gamma_0 D2_i + \gamma_1 D2_i X_{1i} + \gamma_2 D2_i X_{2i} + u_i \qquad \textbf{(12.1)}$$

to perform an **F-test** of H_0: $\gamma_j = 0 \;\forall$ j = 0, 1, 2 versus H_1: $\gamma_j \neq 0$ j = 0, 1, 2 where $\gamma_j = \beta_j - \alpha_j \neq 0$ for j = 0, 1, 2.

The **population regression functions** (PRFs) **for Group 1 and Group 2** implied by pooled regression equation (12.1) give the **conditional mean value of Y_i for each group** for any given values of the explanatory variables X_{1i} and X_{2i}.

- The **Group 1 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.1) is:

$$E\left(Y_i \mid X_{1i}, X_{2i}, D2_i = 0\right) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i}$$

- The **Group 2 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.1) is:

$$E\left(Y_i \mid X_{1i}, X_{2i}, D2_i = 1\right) = (\alpha_0 + \gamma_0) + (\alpha_1 + \gamma_1)X_{1i} + (\alpha_2 + \gamma_2)X_{2i}$$

The **null hypothesis of complete coefficient equality** – H_0: $\gamma_j = 0 \;\forall$ j = 0, 1, 2 – means that

$$E\left(Y_i \mid X_{1i}, X_{2i}, D2_i = 0\right) = E\left(Y_i \mid X_{1i}, X_{2i}, D2_i = 1\right) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i}$$

i.e., the **Group 1 conditional mean value of Y** ***equals*** the **Group 2 conditional mean value of Y** for any given values of X_1 and X_2.

<u>Meaning of TEST 3</u>: Uses OLS estimates of pooled equation (12.2)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \delta_0 D1_i + \delta_1 D1_i X_{1i} + \delta_2 D1_i X_{2i} + u_i \qquad \textbf{(12.2)}$$

to perform an **F-test** of H_0: $\delta_j = 0 \ \forall$ j = 0, 1, 2 versus H_1: $\delta_j \neq 0$ j = 0, 1, 2 where $\delta_j = \alpha_j - \beta_j \neq 0$ for j = 0, 1, 2.

The **population regression functions** (PRFs) **for Group 1 and Group 2** implied by pooled regression equation (12.1) give the **conditional mean value of Y_i for each group** for any given values of the explanatory variables X_{1i} and X_{2i}.

- The **Group 1 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.2) is:

$$E(Y_i \mid X_{1i}, X_{2i}, D1_i = 1) = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)X_{1i} + (\beta_2 + \delta_2)X_{2i}$$

- The **Group 2 conditional mean value of Y_i** for given values of X_{1i} and X_{2i} implied by equation (12.2) is:

$$E(Y_i \mid X_{1i}, X_{2i}, D1_i = 0) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

The null hypothesis of complete coefficient equality – H_0: $\delta_j = 0 \ \forall$ j = 0, 1, 2 – means that

$$E(Y_i \mid X_{1i}, X_{2i}, D1_i = 1) = E(Y_i \mid X_{1i}, X_{2i}, D1_i = 0) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

i.e., the **Group 1 conditional mean value of Y *equals*** the **Group 2 conditional mean value of Y** for any given values of X_1 and X_2.

❑ **<u>Advantages of Pooled Regression Approach</u> (Approach 2)**

1. Approach 2 is more *informative*.

- It permits **t-tests of individual coefficient differences** between the Group 1 and Group 2 regression functions.

  This advantage is particularly evident when a base group is selected for the pooled full-interaction regression equation.

2. Approach 2 is more *flexible*.

- **Approach 2** can be used **to test for coefficient equality between any subset of regression coefficients** in the Group 1 and Group 2 PRFs.

  Approach 1 can only test the hypothesis of **complete coefficient equality** between the Group 1 and Group 2 regression functions.

- ***Illustrations of Approach 2:*** Tests for equality of subsets of coefficients.

♦ **<u>Test 1</u>: Equality of *all* (both) slope coefficients**.

H_0: $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$ in pooled PRE (12.0)
$\gamma_1 = 0$ and $\gamma_2 = 0$ in pooled PRE (12.1)
$\delta_1 = 0$ and $\delta_2 = 0$ in pooled PRE (12.2)

H_1: $\alpha_1 \neq \beta_1$ and/or $\alpha_2 \neq \beta_2$ in pooled PRE (12.0)
$\gamma_1 \neq 0$ and/or $\gamma_2 \neq 0$ in pooled PRE (12.1)
$\delta_1 \neq 0$ and/or $\delta_2 \neq 0$ in pooled PRE (12.2)

- ***Restricted* OLS-SRE** corresponding to H_0 is any one of the following three OLS sample regression equations, for which $\mathbf{K_0 = 4}$:

$$Y_i = \tilde{\alpha}_0 D1_i + \tilde{\beta}_0 D2_i + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{u}_i \qquad \text{from (12.0)}$$

$$Y_i = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_{1i} + \tilde{\alpha}_2 X_{2i} + \tilde{\gamma}_0 D2_i + \tilde{u}_i \qquad \text{from (12.1)}$$

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{\delta}_0 D1_i + \tilde{u}_i \qquad \text{from (12.2)}$$

- The ***restricted* RSS under H_0** is

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with } df_0 = df_R = N - K_0 = N - 4.$$

- The ***unrestricted* RSS under H_1** is

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with } df_1 = df_U = N - K = N - 6.$$

- The ***numerator* degrees of freedom** equal

$$df_{num} = df_0 - df_1 = (N - K_0) - (N - K) = K - K_0 = 6 - 4 = 2.$$

- The ***denominator* degrees of freedom** equal

$$df_{den} = df_1 = N - K = N - 6.$$

♦ **Test 2: Equality of a subset of slope coefficients, e.g. the coefficient of X_{2i}.**

H_0:	$\alpha_2 = \beta_2$	in pooled PRE (12.0)
	$\gamma_2 = 0$	in pooled PRE (12.1)
	$\delta_2 = 0$	in pooled PRE (12.2)
H_1:	$\alpha_2 \neq \beta_2$	in pooled PRE (12.0)
	$\gamma_2 \neq 0$	in pooled PRE (12.1)
	$\delta_2 \neq 0$	in pooled PRE (12.2)

- ***Restricted* OLS-SRE** corresponding to H_0 is any one of the following three OLS sample regression equations, for which $\mathbf{K_0 = 5}$:

$$Y_i = \tilde{\alpha}_0 D1_i + \tilde{\beta}_0 D2_i + \tilde{\alpha}_1 D1_i X_{1i} + \tilde{\beta}_1 D2_i X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{u}_i \qquad \text{from (12.0)}$$

$$Y_i = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_{1i} + \tilde{\alpha}_2 X_{2i} + \tilde{\gamma}_0 D2_i + \tilde{\gamma}_1 D2_i X_{1i} + \tilde{u}_i \qquad \text{from (12.1)}$$

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{\delta}_0 D1_i + \tilde{\delta}_1 D1_i X_{1i} + \tilde{u}_i \qquad \text{from (12.2)}$$

- The ***restricted* RSS under $\mathbf{H_0}$** is

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with } df_0 = df_R = N - K_0 = N - 5.$$

- The ***unrestricted* RSS under $\mathbf{H_1}$** is

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with } df_1 = df_U = N - K = N - 6.$$

- The ***numerator* degrees of freedom** equal

$$df_{num} = df_0 - df_1 = (N - K_0) - (N - K) = K - K_0 = 6 - 5 = 1.$$

- The ***denominator* degrees of freedom** equal

$$df_{den} = df_1 = N - K = N - 6.$$

♦ **<u>Test 3</u>: Equality of *intercept* coefficients only**.

H_0:	$\alpha_0 = \beta_0$	in pooled PRE (12.0)
	$\gamma_0 = 0$	in pooled PRE (12.1)
	$\delta_0 = 0$	in pooled PRE (12.2)
H_1:	$\alpha_0 \neq \beta_0$	in pooled PRE (12.0)
	$\gamma_0 \neq 0$	in pooled PRE (12.1)
	$\delta_0 \neq 0$	in pooled PRE (12.2)

- ***Restricted* OLS-SRE** corresponding to H_0 is any one of the following three OLS sample regression equations, for which $\mathbf{K_0 = 5}$:

$$Y_i = \tilde{\beta}_0 + \tilde{\alpha}_1 D1_i X_{1i} + \tilde{\beta}_1 D2_i X_{1i} + \tilde{\alpha}_2 D1_i X_{2i} + \tilde{\beta}_2 D2_i X_{2i} + \tilde{u}_i \qquad \text{from (12.0)}$$

$$Y_i = \tilde{\alpha}_0 + \tilde{\alpha}_1 X_{1i} + \tilde{\alpha}_2 X_{2i} + \tilde{\gamma}_1 D2_i X_{1i} + \tilde{\gamma}_2 D2_i X_{2i} + \tilde{u}_i \qquad \text{from (12.1)}$$

$$Y_i = \tilde{\beta}_0 + \tilde{\beta}_1 X_{1i} + \tilde{\beta}_2 X_{2i} + \tilde{\delta}_1 D1_i X_{1i} + \tilde{\delta}_2 D1_i X_{2i} + \tilde{u}_i \qquad \text{from (12.2)}$$

- The ***restricted* RSS under $\mathbf{H_0}$** is

$$RSS_0 = RSS_R = \sum_{i=1}^{N} \tilde{u}_i^2 \qquad \text{with} \quad df_0 = df_R = N - K_0 = N - 5.$$

- The ***unrestricted* RSS under $\mathbf{H_1}$** is

$$RSS_1 = RSS_U = \sum_{i=1}^{N} \hat{u}_i^2 \qquad \text{with} \quad df_1 = df_U = N - K = N - 6.$$

- The ***numerator* degrees of freedom** equal

$$df_{num} = df_0 - df_1 = (N - K_0) - (N - K) = K - K_0 = 6 - 5 = 1.$$

- The ***denominator* degrees of freedom** equal

$$df_{den} = df_1 = N - K = N - 6.$$

❑ Computing Test 1, Test 2 and Test 3

For each test, compute the *sample value* of the general F-statistic, and **then apply the conventional *decision rule***:

♦ ***Sample value* of general F-statistic** is computed as:

$$F_0 = \frac{(RSS_0 - RSS_1)/(df_0 - df_1)}{RSS_1/df_1} \sim F(df_0 - df_1, df_1) \text{ under } H_0$$

or

$$F_0 = \frac{(RSS_0 - RSS_1)/(K - K_0)}{RSS_1/(N-K)} \sim F(K - K_0, N - K) \text{ under } H_0.$$

♦ ***Decision Rule:***

(1) **If $F_0 \geq F_\alpha(K-K_0, N-K)$**, or **if the *p-value* for $F_0 \leq \alpha$**, ***reject*** the coefficient restrictions specified by the ***null*** **hypothesis H_0** at the ***100α%*** significance level;

(2) **If $F_0 < F_\alpha(K-K_0, N-K)$**, or **if the *p-value* for $F_0 > \alpha$**, ***retain (do not reject)*** the coefficient restrictions specified by the ***null*** **hypothesis H_0** at the ***100α%*** significance level.