MLR – response variable Y is related to MORE THAN ONE EXPLANATORY VARIABLE

FIRST ORDER MODEL: $\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon}$
Highest exponent of any X = 1
(Order is determined by highest exponent of the INDEPENDENT VARIABLES)

SECOND ORDER MODELS: $\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \beta_3 x_3 + \varepsilon}$
$\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \underbrace{x_1 x_2} + \varepsilon}$

"cross-product" or "Interaction" terms
These are LINEAR models because their β parameters are linear, (not the independent variables).

i.e. these are NOT LINEAR MODELS: $\boldsymbol{y = \beta_0 + \beta_1 x_1 + {\beta_2}^2 x_2 + \varepsilon}$; $\boldsymbol{y = \beta_0 + \beta_0 e^{\beta_1 x_1} + \varepsilon}$

Steps in MLR

1) In general too challenging to scatterplot by hand, go to step 2
2) STATE MODEL (i.e. $\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon}$), AND ASSUMPTIONS:
   a. $x_1, x_2, \ldots, x_k$ are **OBSERVED WITHOUT ERROR** (it is ok if they are RELATED)
   b. y's and errors (ε) are **INDEPENDENTLY DISTRIBUTED** so that:
      mean of $\boldsymbol{y} = \boldsymbol{E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots, \beta_k x_k}$
      in other words, $\boldsymbol{E(\varepsilon)} = 0$
   c. $y's$ (or ε's) have **CONSTANT VARIANCE**, σ^2 for any value of x_1, x_2, x_3.... etc.
   d. Response variables (y) and errors (ε) are **NORMALLY DISTRIBUTED:**
      $\boldsymbol{y \cong N(E(y), \sigma^2)}$ ***or*** $\boldsymbol{\varepsilon \cong N(0, \sigma^2)}$ for every value of $x_1, x_2, \ldots x_k$
3) Use the sample to find the least squares fitted line: $\boldsymbol{\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \widehat{\beta_3} x_3}$
4) Perform RESIDUAL ANALYSIS to check for any violations of the assumptions
5) test whether a a LINEAR relationship exists (F-TEST MODEL)
6) If Linear relationship exists (or in the case of F-Partial test, if parameters:
   a. Calculate r^2 to determine how much variability belongs to regression vs error
   b. Calculate r^2-adjusted to determine how good the model is
      i. If the model is not good, perform **CRITERION TESTS** to determine which parameters to keep
      ii. Go back to step 5 and Test the PARTIAL model vs the FULL model in F-PARTIAL (or F-DROP) test.
   c. If H_0 for FULL MODEL F-TEST is REJECTED (i.e. a Linear relationship exists), use $\boldsymbol{\hat{y}}$ to estimate $\boldsymbol{E(y)}$, or prediction of future value. (Plug numbers found for parameters into equation)

Interpretation of Parameters: (only true if all x's are independent and FIRST ORDER MODEL)

β_0 = y-intercept

β_j , (j=1,k) are PARTIAL SLOPES and represent the CHANGE IN THE AVERAGE VALUE OF y for a UNIT CHANGE in x_j, holding all other x's constant.

Least Squares Method for MLR

SSE = $\sum_{j=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{j=1}^{n}(y_i - (\boldsymbol{\beta_0} + \boldsymbol{\beta_1 x_i} + \boldsymbol{\beta_2 x_{i2}} + \ldots, \boldsymbol{\beta_k x_{ik}}))^2$

$\boldsymbol{y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + \varepsilon}$ → **$Y = X\beta + \varepsilon$** } Matrix notation

$\boldsymbol{\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \ldots \widehat{\beta_k} x_k + \varepsilon}$ → **$\hat{Y} = X\beta$** } Matrix notation

Y is a vector $= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ y_n \end{bmatrix}$ n x 1

$\hat{\beta}$ is a vector $= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ . \\ . \\ \hat{\beta}_k \end{bmatrix}$ (k+1) x1

ε is a vector $= \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ . \\ . \\ \varepsilon_n \end{bmatrix}$ n x 1

x is a DESIGN MATRIX or MATRIX of CONSTANTS:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \ldots & x_{1k} \\ x_{21} & x_{22} & x_{23} & x_{24} & \ldots & x_{2k} \\ x_{31} & x_{32} & x_{33} & x_{34} & \ldots & x_{3k} \\ x_{41} & x_{42} & x_{43} & x_{44} & \ldots & x_{4k} \\ & & \ldots & & & \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} & \ldots & x_{nk} \end{bmatrix}$$ n x (k+1)

Y is the VECTOR of RESPONSE variables

$\boldsymbol{\hat{\beta}}$ is the vector of PARAMETER ESTIMATES

Normal equations can be written as $(X^T X)\hat{\beta} = X^T Y$ → $\boldsymbol{\hat{\beta} = (X^T X)^{-1} (X^T Y)}$

Must solve k+1 normal equations

eg. $\boldsymbol{\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \widehat{\beta_2} x_2 + \varepsilon}$

$Y = \begin{bmatrix} 25 \\ 30 \\ 31 \end{bmatrix}$ $X = \begin{bmatrix} 1 & 6 & 3 \\ 1 & 5 & 3 \\ 1 & 8 & 4 \end{bmatrix}$ $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$ **$Y = X\beta + \varepsilon$**

Transpose = swap rows/columns...

$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 \\ 6 & 5 & 8 \\ 3 & 3 & 4 \end{bmatrix}$

$\mathbf{X^T X} = \begin{bmatrix} 1 & 1 & 1 \\ 6 & 5 & 8 \\ 3 & 3 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & 6 & 3 \\ 1 & 5 & 3 \\ 1 & 8 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 19 & 10 \\ 19 & 125 & 65 \\ 10 & 65 & 34 \end{bmatrix}$

3 x 3   3 x 3 → 3 x 3

Multiply and SUM values :

1x1 + 1x1 + 1x1 = 3

1x6 + 1x5 + 1x8 = 19

1x3 + 1x3 + 1x4 = 10

To calculate INVERSE $(X^TX)^{-1}$, use IDENTITY MATRIX:

$$\left[\begin{array}{ccc|ccc} a_{11} & a_{12} & a_{13} & 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 1 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 1 \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & 1 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & 1 & b_{31} & b_{32} & b_{33} \end{array}\right]$$

$$\mathbf{X^TX} = \begin{bmatrix} 3 & 19 & 10 \\ 19 & 125 & 65 \\ 10 & 65 & 34 \end{bmatrix} \qquad (\mathbf{X^TX})^{-1} = \begin{bmatrix} 25 & 4 & -15 \\ 4 & 2 & -5 \\ -15 & -5 & 14 \end{bmatrix}$$

$$\mathbf{X^TY} = \begin{bmatrix} 1 & 1 & 1 \\ 6 & 5 & 8 \\ 3 & 3 & 4 \end{bmatrix} \times \begin{bmatrix} 25 \\ 30 \\ 31 \end{bmatrix} = \begin{bmatrix} 86 \\ 548 \\ 289 \end{bmatrix}$$

$$1x25 + 1x30 + 1x31 = 86$$
$$6x25 + 5x30 + 8x31 = 548$$
$$3x35 + 3x30 + 4x31 = 289$$

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^TX})^{-1}(\mathbf{X^TY}) = \begin{bmatrix} 25 & 4 & -15 \\ 4 & 2 & -5 \\ -15 & -5 & 14 \end{bmatrix} \times \begin{bmatrix} 86 \\ 548 \\ 289 \end{bmatrix} = \begin{bmatrix} 7 \\ -5 \\ 16 \end{bmatrix}$$

$$\widehat{\beta_0} = 7 \qquad \widehat{\beta_1} = -5 \qquad \widehat{\beta_2} = 16$$

$$\hat{\boldsymbol{\beta}}^{\mathbf{T}} = [\,7 \quad -5 \quad 16\,]$$

regression line: $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \varepsilon \rightarrow \hat{y} = 7 - 5x_1 + 16x_2 + \varepsilon$

SSR = $\hat{\boldsymbol{\beta}}^{\mathbf{T}}(\mathbf{X^TY}) = [\,7 \quad -5 \quad 16\,] \times \begin{bmatrix} 86 \\ 548 \\ 289 \end{bmatrix}$ = (7x86) + (-5x548) + (16x289) = **2486**

FITTED LEAST SQUARES REGRESSION EQUATION (know this!): $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$

AKA: LEAST SQUARES FITTED EQUATION or LEAST SQUARES REGRESSION EQUATION or LEAST SQUARES PREDICTION EQUATION

For a model with INTERACTION TERMS: I.E.

$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}x_2 + \widehat{\beta_3}x_1x_2 + \varepsilon$

where $x_2 = 1 \text{ if type } A$

$x_2 = 0 \text{ if type } B$

The regressions are as follows:

For type A: $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}(1) + \widehat{\beta_3}x_1(1) + \varepsilon$ **or** $\hat{y} = (\widehat{\beta_0} + \widehat{\beta_2}) + (\widehat{\beta_1} + \widehat{\beta_3})x_1 + \varepsilon$

For type B: $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \widehat{\beta_2}(0) + \widehat{\beta_3}x_1(0) + \varepsilon$ **or** $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}x_1 + \varepsilon$

Therefore: $\widehat{\beta_0}$ → **the INTERCEPT of line A**

Therefore: $\widehat{\beta_1}$ → **the SLOPE of line A**

Therefore: $\widehat{\beta_2} = (\widehat{\beta_0} + \widehat{\beta_2}) - \widehat{\beta_0}$ → **the difference in *y-INTERCEPTS* between type A and B lines**

Therefore: $\widehat{\beta_3} = (\widehat{\beta_1} + \widehat{\beta_3}) - \widehat{\beta_1}$ → **the difference in *SLOPES* between type A and B lines**

(crossproduct interaction terms are always diff in slope)

ANOVA

Source of variation	df	SS	MS	F
Regression	k	SSR	MSR = SSR/df	MSR/MSE
Error	n-(k+1)	SSE	MSE = SSE/df	
Total	n-1			

k = number of parameters (not including β_0)
n = number of measurements in the sample

F-TEST: FULL MODEL – testing the LINEAR RELATIONSHIP significance

H_0: $\beta_1 = \beta_2 = \beta_k = 0$ H_a: At least one of the β's is $\neq 0$

The degrees of freedom for **FULL F-TEST** are **$F_\alpha(df_{SSR}, df_{SSE})$.**

Rejection Region: When **F-TEST** ≥ **F-TABLE** **(df_{SSR}, df_{SSE})**

If we reject H_0 we can conclude that at the α*100 level of significance **there is evidence that the model is significant** – i.e. that there is a **linear relationship between y and at least one of the x's**.

F-TEST: PARTIAL MODEL – testing the significance of VARIABLE CONTRIBUTION TO THE MODEL

H_0: $\beta_3 = \beta_4 = 0$ (β's dropped) H_a: At least one of the β's is $\neq 0$

The degrees of freedom for **PARTIAL F-TEST are ($df_{SSR(REDUCED)}$, $df_{SSE(FULL)}$).**

Rejection Region: When **F-PARTIAL** ≥ **F-TABLE** (α,(**$df_{SSR(REDUCED)}$**, **$df_{SSE(FULL)}$**))

If we reject H_0 we can conclude that at the α*100 level of significance **there is evidence that the PARAMETERS** in question (dropped in the reduced model) **contribute to the model**.

T-TEST: Test for SPECIFIC PARAMETER CONTRIBUTION to the MODEL

For β_k contribution to the model: **H_0: $\beta_k = 0$ H_a: $\beta_k \neq 0$**

Rejection Region: When |t-TEST| > $t\text{-TABLE}_{(\alpha/2, n-(k+1))}$
If we reject H_0 we can conclude that at the α*100 level of significance **there is evidence that the PARAMETER** tested **contributes to the model**.

r^2 and r^2-adjusted
r^2 – explains what percentage of the variation in the data is due to **REGRESSION (vs error)**

r^2-adjusted – tests the **fitness of the model -** if lower than r^2 and/or if prior t-test shows that parameters do not contribute, we can conclude the model isn't very good. If r^2-adjusted is close to 1 then the model is likely pretty good.

METHODS FOR CHOOSING MODEL PARAMETERS

MAXIMUM r^2 CRITERION

1) Calculate r^2 for every model version. Take model with 2nd highest r^2 (i.e. full is usually highest – take next highest), unless there is a HUGE jump, in which case, take FULL
2) If two models have similar r^2 then calculate r^2–adjusted and take the one with the higher value. (r^2-adjusted DROPS if unimportant vars are in the model)
3) since r^2 = SSR/TSS, you can quickly determine that the model with the largest SSR will have the highest r^2

MINIMUM MSE CRITERION

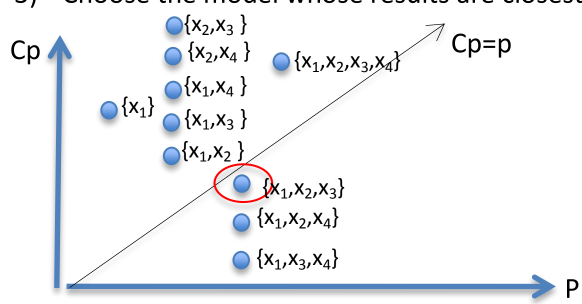
Min MSE = MAX r^2 → TSS = SSR+SSE Therefore the one with the max r^2 will have the minimum SSE (when I have high SSR, I will have low SSE)
Equivalent to MAX r^2

MALLOWS Cp CRITERION

Measures difference of a fitted regression line ($\hat{y}$ from y) from the TRUE model, along with RANDOM ERROR

$Cp = \frac{SSE_p}{MSE_f} - (n - 2p)$ where $\boldsymbol{p}$ is the **number of parameters in the model, including β_0**
and SSE_p is the SSE of the model with p parameters.

1) Calculate Cp for every possible model combination (arduous!)
2) **Plot $\boldsymbol{p}$ vs $\boldsymbol{Cp}$**
3) Choose the model whose results are closest to the $Cp = p$ line (the best fitting)



(Not all combinations shown…)

RESIDUAL ANALYSIS

Plot of y_i vs e_i → tests violations of INDEPENDENCE
Plot of x_i vs e_i → tests violations of CONSTANT VARIANCE
Histogram of errors → tests violations of NORMALITY