Factors:

A factor is a qualitative, or categorical covariate that can take on one or more levels

Examples: degree → ugrad, masters, Ph.D.

breed → italian greyhound, whippet, greyhound

color → red, green, blue

factor = variable

levels = values that the variable can take on.

the levels of a factor are mutually exclusive and exhaustive within the dataset.

each observation takes on exactly one level

The one factor model for y_i:

$$Y_i = b_0 + b_{j(i)} + \varepsilon_i, \quad \varepsilon_i \overset{ind}{\sim} N(0, \sigma^2)$$

$j(i)$ is the level of the i^{th} observation.

Example dataset

i	$color_i$	$j(i)$	y_i
1	red	2	1
2	red	2	3
3	green	1	6
4	green	1	8

⟶

$$Y_1 = b_0 + b_2 + \varepsilon_1$$
$$Y_2 = b_0 + b_2 + \varepsilon_2$$
$$Y_3 = b_0 + b_1 + \varepsilon_3$$
$$Y_4 = b_0 + b_1 + \varepsilon_4$$

$j(i)$ is an integer label for the factor

The model says that observations with factor level k have expected value $b_0 + b_k$

The factor model is a multiple linear model!

Suppose the factor has J levels. Define

$$x_{ij} = \begin{cases} 1 & \text{if } i^{th} \text{ obs has level } j \\ 0 & \text{if } i^{th} \text{ obs does not have level } j \end{cases}$$

Write the model as

$$Y_i = b_0 + \sum_{j=1}^{J} b_j x_{ij} + \varepsilon_i \,, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$$

Back to our color example: $J = 2$ colors (levels)

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + \varepsilon_i$$

$$Y_1 = b_0 + b_1 0 + b_2 1 + \varepsilon_1 = b_0 + b_2 + \varepsilon_1$$

$$Y_2 = b_0 + b_1 0 + b_2 1 + \varepsilon_2 = b_0 + b_2 + \varepsilon_2$$

$$Y_3 = b_0 + b_1 1 + b_2 0 + \varepsilon_3 = b_0 + b_1 + \varepsilon_3$$

$$Y_4 = b_0 + b_1 1 + b_2 0 + \varepsilon_4 = b_0 + b_1 + \varepsilon_4$$

Since the factor model can be expressed as a multiple linear model, everything we have learned so far applies

There is one catch though.

Design matrix

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

What is the rank of X?

rank of $X = 2$

<u>always happens in 1 factor model</u>

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

does not have an inverse, $\text{rank}(X^T X) = 2 < 3$

Cannot use $\hat{\underline{b}} = (X^T X)^{-1} X^T \underline{y}$

Example with $n = 5$ observations, $J = 3$ levels

$$X = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \qquad X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 2 & 2 & 1 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

$\text{rank}(X) = 3$, $\text{rank } X^T X = 3$

however: $(X^T X)\underline{b} = X^T y$ will have a solution

in fact, it has ∞ many solutions.

any of the solutions minimizes

$$RSS(\underline{b}^*) = \sum_{i=1}^{5} \left(y_i - b_0^* - b_1^* x_{i1} - b_2^* x_{i2} \right)^2$$

example: $\underline{y} = (6, 8, 1, 3)^T$. $X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$

$$X^T \underline{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 8 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 18 \\ 4 \\ 14 \end{bmatrix}$$

set $\hat{\underline{b}} = (0, 2, 7)^T \rightarrow X^T X \underline{b} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 18 \\ 4 \\ 14 \end{bmatrix}$

fitted values: $\hat{\underline{y}} = X \hat{\underline{b}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 2 \\ 2 \end{bmatrix}$

$$\text{rss} = (1-2)^2 + (3-2)^2 + (6-7)^2 + (8-7)^2 = 4$$

This solution also works:

$\hat{\underline{b}} = (1, 1, 6)^T \rightarrow X^T X \underline{b} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 6 \end{bmatrix} = \begin{bmatrix} 18 \\ 4 \\ 14 \end{bmatrix}$

fitted values: $\hat{\underline{y}} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 2 \\ 2 \end{bmatrix}$ of course, same rss as well.

$$\text{rss} = (1-2)^2 + (3-2)^2 + (6-7)^2 + (8-7)^2 = 4$$

$\hat{\underline{b}} = (0, 2, 7)$ and $\hat{\underline{b}} = (1, 1, 6)$ both minimize rss.

any $\hat{\underline{b}}$ such that $\hat{b}_0 + \hat{b}_1 = 2$, $\hat{b}_0 + \hat{b}_2 = 7$ will minimize rss.

normal equations have infinitely many solutions.

Consequence: can't interpret individual coefficients.

what does b_0 mean if $\hat{b}_0 = 0$ no better than $\hat{b}_0 = 1$?

can interpret quantities that are the same no matter which solution we pick.

i.e. $b_0 + b_1 \longrightarrow E(Y_i)$ when $j(i) = 1$

$b_0 + b_2 \longrightarrow E(Y_i)$ when $j(i) = 2$

these are examples of estimable functions

A linear combination $\underline{c}^T \underline{b}$ (i.e. $b_0 + b_1$)

is an estimable function if we can find a vector $\underline{d}$ such that $E(\underline{d}^T \underline{Y}) = \underline{c}^T \underline{b}$

example: $\underline{c} = (1, 1, 0)$, $\underline{c}^T \underline{b} = b_0 + b_1$

$\underline{d} = (1, 0, 0, 0)$, $E(\underline{d}^T \underline{Y}) = E(Y_1) = b_0 + b_1$

$\underline{c} = (1, 0, 0)$, $\underline{c}^T \underline{b} = b_0$

can't find $\underline{d}$ such that $E(\underline{d}^T \underline{Y}) = b_0$

$\underline{c} = (0, 1, -1)$, $\underline{c}^T \underline{b} = b_1 - b_2$

$\underline{d} = (1, 0, -1, 0)$, $E(\underline{d}^T \underline{Y}) = E(Y_1 - Y_3) = b_1 - b_2$

estimable functions are the quantities we can interpret.

How do we find a solution to $(X^T X)\underline{b} = X^T \underline{y}$ when $(X^T X)^{-1}$ does not exist?

<u>Dropped column method</u>

Simply remove one column of X, so resulting matrix is full rank

$$X = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \quad X_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \quad X_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \quad X_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

these all produce a full rank design matrix.

Dropping a column is equivalent to setting a regression coefficient to 0:

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ 0 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_0 + b_2 \\ b_0 + b_2 \\ b_0 \\ b_0 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_0 + b_2 \\ b_0 + b_2 \\ b_0 \\ b_0 \end{bmatrix}$$

interpretation: $b_0 = E(Y_3)$ $\quad b_2 = E(Y_1) - E(Y_3)$ $\quad$ <u>R default</u>

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ 0 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_0 \\ b_0 + b_1 \\ b_0 + b_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 + b_1 \\ b_0 + b_1 \\ b_0 \\ b_0 \end{bmatrix}$$

Interpretation: $b_0 = E(Y_1)$ $\quad b_1 = E(Y_3) - E(Y_1)$ $\quad$ <u>SAS default</u>

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_2 \\ b_2 \\ b_1 \\ b_1 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_1 \\ b_2 \\ b_2 \end{bmatrix}$$

Interpretation: $b_2 = E(Y_1)$ $\quad b_1 = E(Y_3)$ $\quad$ <u>in R</u> add -1 to formula

Testing in factor models

if we use a dropped column solution, testing is exactly the same as before. X_0 is full rank, so

$$\underset{\sim}{\hat{B}} = \underbrace{(X_0^T X_0)^{-1}}_{J \times J} \underbrace{X_0^T \underline{Y}}_{J \times 1} \sim N\left(\underline{b},\ \sigma^2 (X_0^T X_0)^{-1}\right)$$

can do t-tests for individual components of $\underline{b}$

$$\boxed{\text{Have to be careful about interpretation!!!}}$$

$$X_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad X_0 \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_0 \\ b_0 + b_1 \\ b_0 + b_1 \end{bmatrix}$$

$H_0 : b_0 = 0 \longrightarrow$ is the EV for red 0?

$H_0 : b_2 = 0 \longrightarrow$ are the EVs for red and green equal?

if $b_2 = 0 \Rightarrow X_0 \underline{b} = \begin{bmatrix} b_0 \\ b_0 \\ b_0 \\ b_0 \end{bmatrix}$ makes sense

<u>exercises:</u>

if we use $X_0 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}$ what is interpretation of $H_0 : b_0 = 0$ and $H_0 : b_1 = 0$?

if we use $X_0 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$ what are the interpretations of $H_0 : b_1 = 0$ and $H_0 : b_2 = 0$?

F - tests

model $Y_i = b_0 + b_{j(i)} + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$H_0 : b_1 = b_2 = \cdots = b_J$

interpretation: all levels have the same EV.

Can't do a t-test here.

To set up the F test define two models that capture the null and alternative hypotheses.

Reduced Model (null) : $Y_i = b_0 + b_1 + \varepsilon_i$

Full Model (alternative) : $Y_i = b_0 + b_{j(i)} + \varepsilon_i$

$$X_r = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{bmatrix} \qquad \text{rank}(X_r) = 1, \quad mdf_0 = 1$$

X_f has entries $x_{i0} = 1$
$n \times (J+1)$

$$x_{ij} = \begin{cases} 1 & \text{if obs. } i \text{ has level } j \\ 0 & \text{if not} \end{cases} \quad j \geq 1$$

$$\Rightarrow x_{i0} = \sum_{j=1}^{J} x_{ij} \Rightarrow \text{rank}(X_f) = J < J+1, \; mdf_1 = J.$$

get RSS_0 and RSS_1

$$F = \frac{(RSS_0 - RSS_1)/(J-1)}{RSS_1/(n-J)} \sim F_{J-1, n-J}$$

Under the null, this has an F distribution with J-1 and n-J degrees of freedom

Generalized inverse method (optional if we have time)

G is a generalized inverse of A if $AGA = A$

if $(X^TX)^-$ is a Gen inv. of X^TX,

then $\boxed{\hat{\underline{b}} = (X^TX)^- X^T \underline{y}}$ is a solution of $\boxed{(X^TX)\underline{b} = X^T\underline{y}}$

why? write $\underline{y} = \hat{\underline{y}} + \hat{\underline{e}}$, where $\hat{\underline{y}} = P\underline{y}$, $\hat{\underline{e}} = (I-P)\underline{y}$.

$\hat{\underline{y}} \in \text{span}(X) \implies \hat{\underline{y}} = X\underline{v}$ for some $\underline{v}$.

left: $(X^TX)\hat{\underline{b}} = (X^TX)(X^TX)^- X^T\underline{y} = (X^TX)(X^TX)^- X^T(X\underline{v} + \hat{\underline{e}})$

$= \boxed{(X^TX)\underline{v}} + (X^TX)(X^TX)^- X^T\hat{\underline{e}} \leftarrow 0$

right: $X^T\underline{y} = \boxed{X^TX\underline{v}} + X^T\hat{\underline{e}} \leftarrow 0$

Example: $X^TX = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$ $\quad (X^TX)^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$

$$\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \checkmark$$

$$(X^TX)^- X^T y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} 18 \\ 4 \\ 14 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 7 \end{bmatrix}$$

one of our solutions above.

Generalized inverses provide solutions that cannot be obtained by the "dropped columns" method.

<u>example</u>: $\underset{\sim}{\hat{b}} = [4.5 \;\; -2.5 \;\; 2.5]^T$ default solution in JMP.

no coefficient $= 0$ $\longrightarrow$ not a dropped column solution.

$b_1 + b_2 = 0$ "sum to zero" constraint.