- **Which model is better – with or without interaction?**

- Is interaction term good pred ~~of~~ compared to the model without interaction? Yes
- R^2_{adj} : 58.5% > 58.3% Yes

Debatable — of R^2_{adj}
We can also say the diff is very small and we prefer the simpler model (m w/out int)

- **Fitted Equations:**

$$WT = -128 + 4.00\,HT - 56.16\,Gender + 1.14\,HT{*}Gen$$

Baseline Gender=0 $\quad WT = -128 + 4.00\,HT - \cancel{56.16(0) + 1.14\,HT(0)}$

$$= -128 + 4.00\,HT \rightarrow \text{Female}$$

Gender=1 $\quad WT = -128 + 4.00\,HT - 56.16(1) + 1.14\,HT(1)$

$$= -184.16 + 5.14\,HT \rightarrow \text{Male}$$

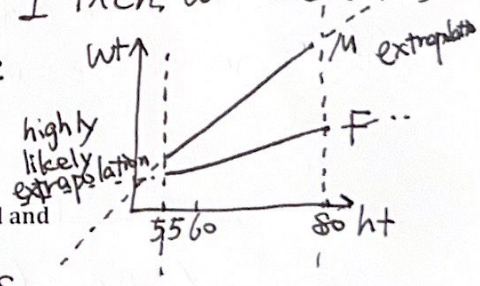
- **Interpretation of Coefficients in the Fitted Equation:**

F y-int = −128 Do not interpret

slope = 4.00 For female, as ht increases by 1 inch, wt increases by 4 pounds on average
(For female, as ht increases by 1 inch, wt is expected (predicted) to increase 4 pounds)

M y-int = −184.16 Do not interpret

slope = 5.14 For male, as ht increases by 1 inch, wt increases by 5.14 pounds on average



- **Sketch the two equations.**
- **Predict weight for Males and Females who are 5ft=60 inches tall and**

F: $WT = -128 + 4.00 \times 60 = 112$ lbs

M: $WT = -184.16 + 5.14 \times 60 = 124.24$ lbs

Models with Categorical Variables with 3 Groups

Suppose we want to predict weight (y) from height (x_1) and race (x_2): White, Black or Hispanic. How can we account for the three groups? **Code race with 2 dummy variables**, i.e.

1 Quant res
1 Quant pred
2 Dummy var

$\left.\right\}$ # predictors = 3 = p

$$x_2 = \begin{cases} 1, & \text{Black} \\ 0, & \text{otherwise} \end{cases}$$

$$x_3 = \begin{cases} 1, & \text{Hispanic} \\ 0, & \text{otherwise} \end{cases}$$

White: $X_2=0,\ X_3=0$

Black: $X_2=1,\ X_3=0$

Hispanic: $X_2=0,\ X_3=1$

NO Interaction Model:

- **How many predictor variables in the model?**

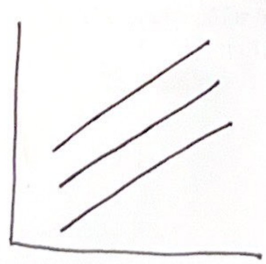
- **Model Equations:** $\underset{\text{WT}}{y} = \alpha + \beta_1 \underset{\text{HT}}{X_1} + \beta_2 \underset{\text{Dummy for B}}{X_2} + \beta_3 \underset{\text{Dummy for H}}{X_3} + \varepsilon$

White $(X_2=X_3=0)$: $y = \alpha + \beta_1 X_1 + \varepsilon$ → Baseline

Black: $(X_2=1, X_3=0)$ $y = (\alpha+\beta_2) + \beta_1 X_1 + \varepsilon$

Hispanic: $(X_2=0, X_3=1)$ $y = (\alpha+\beta_3) + \beta_1 X_1 + \varepsilon$

- **What type of lines does the NO Interaction model allow?**



parallel lines

- **Interpretation of Coefficients in the Full Model:**

α constant y-intercept for White (Baseline)

β_1 coeff of HT slope for all groups

β_2 coeff of ~~Q~~Black Change in y-intercept from W to B

β_3 coeff of Hispanic Change in y-intercept from W to His

- **How can we test whether:**
  - The intercept for Blacks ($\alpha+\beta_2$) is significantly different than Whites (α)?

    t-test $H_0: \beta_2=0$ $H_a: \beta_2 \neq 0$

  - The intercept for Hispanics ($\alpha+\beta_3$) is significantly different than Whites (α)?

    t-test $H_0: \beta_3=0$ $H_a: \beta_3 \neq 0$

  - The intercept for Blacks ($\alpha+\beta_2$) is significantly different than Hispanics ($\alpha+\beta_3$)?

    ~~t-test~~ $H_0: \beta_2 = \beta_3$ $H_a: \beta_2 \neq \beta_3$

    NOT on output

    Not required for this class

X_1X_2 , X_1X_3 , $\cancel{X_2X_3}$

Interaction Model: Include *all* interactions between height (x_1) and dummy variables (x_2 and x_3).

- **~~Full~~ WT Model:**

HT, Dummy B, Dummy H, Int btwn HT & B, Int btwn HT & H

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \varepsilon$$

- **How many predictor variables in the model?**

$p = 5$

- **Model Equations:**

~~Full mode~~

White [Baseline] $y = \alpha + \beta_1 X_1 + \varepsilon$
$(X_2 = X_3 = 0)$

Block $(X_2 = 1, X_3 = 0)$

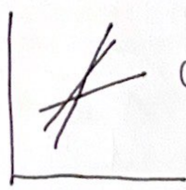
$$y = \alpha + \beta_1 X_1 + \beta_2 (1) + \cancel{\beta_3 (0)} + \beta_4 X_1 (1) + \cancel{\beta_5 X_1 (0)} + \varepsilon$$
$$= (\alpha + \beta_2) + (\beta_1 + \beta_4) X_1 + \varepsilon$$

Hisp $(X_2 = 0, X_3 = 1)$

$$y = \alpha + \beta_1 X_1 + \cancel{\beta_2 (0)} + \beta_3 (1) + \cancel{\beta_4 X_1 (0)} + \beta_5 X_1 (1) + \varepsilon$$
$$= (\alpha + \beta_3) + (\beta_1 + \beta_5) X_1 + \varepsilon$$

- **What type of lines does the Interaction model allow?**

non-parallel lines



(one possibility)

- **Interpretation of Coefficients in the Full Model:**

α	constant	y-intercept for W
β_1	coeff HT	slope for W
β_2	coeff dummy B	change in y-int from W to B
β_3	coeff dummy H	- - - - - - - - H
β_4	coeff interaction btwn HT & B	change in slope from W to B
β_5	coeff interaction btwn HT & H	- - - - - - - H

- **How can we test whether:**
  - The intercept for Blacks ($\alpha + \beta_2$) is significantly different than Whites (α)?

    t-test: $H_0: \beta_2 = 0$ $\quad H_a: \beta_2 \neq 0$
  - The slope for Hispanics ($\beta_1 + \beta_5$) is significantly different than Whites (β_1)?

    t-test: $H_0: \beta_5 = 0$ $\quad H_a: \beta_5 \neq 0$
  - The slope for Blacks ($\beta_1 + \beta_4$) is significantly different than Hispanics ($\beta_1 + \beta_5$)?

    $H_0: \beta_4 = \beta_5$ $\quad H_a: \beta_4 \neq \beta_5$

- **How do we enter the data into the computer?**
  It depends on the program but you need to know if it creates dummy variables for you HOW it was done. And you can always create the columns of 0's and 1's yourself!

Q: If β_4 sig diff from zero but β_2 NOT sig diff from zero, do we eliminate X_2 from the model?

A: NO, if higher-order term is sig, we must include lower-order terms.
X_1X_2 sig $\rightarrow$ include X_1 and X_2 regardless of their p-vals.