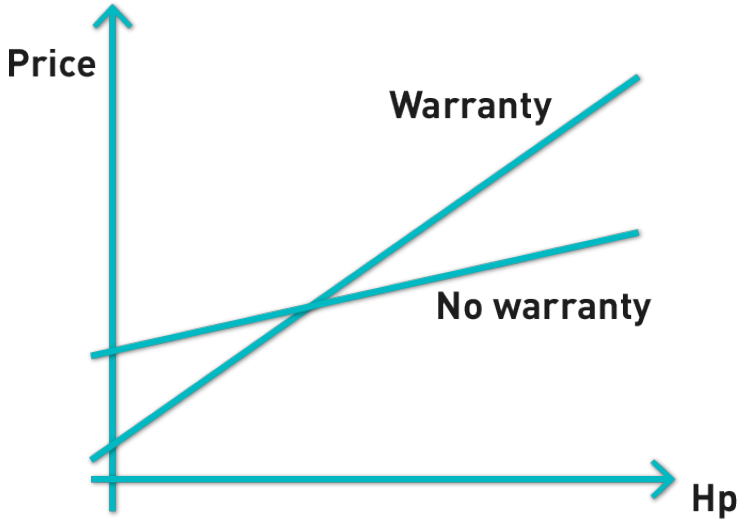
Example: Indicator and Metric Variable

Model of car price as a function of horsepower and a dummy for warranty

- $\text{Price} = \beta_0 + \beta_1 \,\text{hp} + \beta_2 \,\text{warranty} + \beta_3 \,\text{hp} \times \text{warranty} + u$
- If we don't include interaction term, β_2 will act as intercept shift.
  - Two lines with same slope, but different intercepts to fit each group as closely as possible.
- Include interaction term.
  - Slope of price with respect to horsepower:
    - $\delta\text{price}/\delta\text{hp} = \beta_1 + \beta_3 \,\text{warranty}$
  - Takes on two values:
    - β_1 for cars with no warranty
    - $\beta_1 + \beta_3$ for cars with a warranty

Price of Car as Function of Horsepower and Warranty



Different Slopes

- Interaction term allows both intercepts and slopes to vary independently.
- Interaction coefficient is the difference between the two slopes.
- OLS will fit the best line to warranty group and no warranty group independently.
- Hypothesis: Do the two groups have the same slope?
  - Null hypothesis: Groups have same slope, $\beta_3 = 0$.
  - Standard *t*-test for coefficient included with R regression output.

Testing for Equal Lines

- Do groups have same slopes and same intercepts?
  - Hypothesis: $\beta_2 = \beta_3 = 0$
- Run *F*-test between full model and restricted model.
- Full model (contains full set of interactions):
  - $\text{Price} = \beta_0 + \beta_1 \,\text{hp} + \beta_2 \,\text{warranty} + \beta_3 \,\text{hp} \times \text{warranty} + u$
- Restricted model (same regression for both groups)
  - $\text{Price} = \beta_0 + \beta_1 \,\text{hp} + u$
- Compute *F*-statistic to see if saturated model yields statistically significant improvement in model fit.