

Chapter 8: Regression Models with Qualitative Predictors

- Some predictors may be binary (e.g., male/female) or otherwise categorical (e.g., small/medium/large).
- These typically enter the regression model through indicator variables (dummy variables), which take on values
- For a predictor with c categories, we employ
- Why not an indicator variable for each category?

Example: Table 8.2 (Insurance innovation data)

Y = Time until innovation adapted (in months)

X_1 = size of firm (continuous)

X_2 =

Model:

Mean response for mutual firms:

Mean response for stock firms:

Same

- **Why not fit two separate regressions, one for stock firms and one for mutual firms?**
- **Our model assumes same**
- **It's better to estimate these with the total data set.**
- **Inference for β_0 and β_2 will be more precise when we use all the data (more observations) to fit the model.**
- **We may fit our model with least squares as usual.**

Example (insurance innovation data). Fitted model:

- **Interpretation of b_2 :**
- **95% CI for β_2 :**
- **t-test for**

Predictors with Several Categories

- Suppose a predictor X has four categories:
 X = shirt size (S, M, L, XL) of customer
 Y = amount spent on clothes by customer during store visit
- Why not use a single predictor X defined as

Then for small size:

For medium:

For large:

For XL:

- Note the spacing between mean response functions is
- Defining $c - 1 = 3$ indicator variables here allows more

Then for small size:

For medium:

For large:

For XL:

- **We can estimate the differences in mean response between the different categories by estimating**

Example (Shirt Data): Fitted Model:

Interpretation of b_1 :