Lecture 6: Linear Regression (Textbook 3.3)

Department of Statistical Science, Cornell University

Other Considerations in the Regression Model

Qualitative Predictors

- Some predictors are not quantitative, but rather qualitative, taking on a discrete set of values.

Other Considerations in the Regression Model

Qualitative Predictors

- Some predictors are not quantitative, but rather qualitative, taking on a discrete set of values.
- These are also referred to as **categorical predictors** or **factor variables**.

Other Considerations in the Regression Model

Qualitative Predictors

- Some predictors are not quantitative, but rather qualitative, taking on a discrete set of values.
- These are also referred to as **categorical predictors** or **factor variables**.
- For example, consider the credit card data, which includes qualitative variables such as gender, student status, marital status, and ethnicity.

Other Considerations in the Regression Model

Qualitative Predictors

- Some predictors are not quantitative, but rather qualitative, taking on a discrete set of values.
- These are also referred to as **categorical predictors** or **factor variables**.
- For example, consider the credit card data, which includes qualitative variables such as gender, student status, marital status, and ethnicity.
- These qualitative variables can take on specific categories, such as male/female, student/non-student, etc.

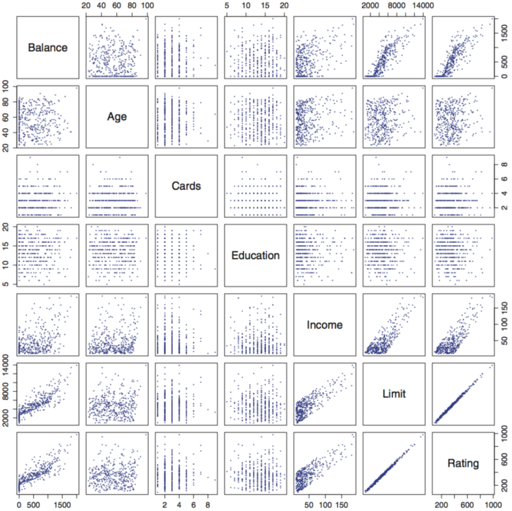
Other Considerations in the Regression Model

Qualitative Predictors

- Some predictors are not quantitative, but rather qualitative, taking on a discrete set of values.
- These are also referred to as **categorical predictors** or **factor variables**.
- For example, consider the credit card data, which includes qualitative variables such as gender, student status, marital status, and ethnicity.
- These qualitative variables can take on specific categories, such as male/female, student/non-student, etc.
- How can we incorporate these qualitative predictors into our regression model?

Credit Card Data

The Credit data set contains information about balance, age, cards, education, income, limit, and rating for a number of potential customers.



Qualitative Predictors

Example: investigate differences in credit card balance between males and females, ignoring the other variables. We create a new **dummy variable**

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male} \end{cases}$$

Resulting model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

Qualitative Predictors

Example: investigate differences in credit card balance between males and females, ignoring the other variables. We create a new **dummy variable**

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male} \end{cases}$$

Resulting model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

Interpretation of β_0 and β_1 ?

Qualitative Predictors

Example: investigate differences in credit card balance between males and females, ignoring the other variables. We create a new **dummy variable**

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male} \end{cases}$$

Resulting model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

β_0 can be interpreted as the average credit card balance among males.
$\beta_0 + \beta_1$ as the average credit card balance among females
β_1 as the average difference in credit card balance between males and females.

Qualitative Predictors

Example: investigate differences in credit card balance between males and females, ignoring the other variables. We create a new **dummy variable**

$$x_i = \begin{cases} 1 & \text{if } i\text{th person is female} \\ 0 & \text{if } i\text{th person is male} \end{cases}$$

Resulting model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is female} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is male.} \end{cases}$$

The decision to code 0 for males and 1 for females is arbitrary and has no effect on the regression ft, but does alter the interpretation of the coefficients.

Qualitative Predictors with More Than Two Levels

With more than two levels, we create additional dummy variables. For example, for the ethnicity variable we create two dummy variables. The first could be

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person is Asian} \\ 0 & \text{if } i\text{th person is not Asian,} \end{cases}$$

and the second could be

$$x_{i2} = \begin{cases} 1 & \text{if } i\text{th person is Caucasian} \\ 0 & \text{if } i\text{th person is not Caucasian.} \end{cases}$$

Qualitative Predictors with More Than Two Levels

Then both of these variables can be used in the regression equation, in order to obtain the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is AA.} \end{cases}$$

There will always be one fewer dummy variable than the number of levels. The level with no dummy variable – African American in this example – is known as the baseline.

Qualitative Predictors with More Than Two Levels

Then both of these variables can be used in the regression equation, in order to obtain the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i = \begin{cases} \beta_0 + \beta_1 + \epsilon_i & \text{if } i\text{th person is Asian} \\ \beta_0 + \beta_2 + \epsilon_i & \text{if } i\text{th person is Caucasian} \\ \beta_0 + \epsilon_i & \text{if } i\text{th person is AA.} \end{cases}$$

There will always be one fewer dummy variable than the number of levels. The level with no dummy variable – African American in this example – is known as the baseline.

Interpretation of β_0, β_1 and β_2 ?

Credit Card Data

	Coefficient	Std. error	t-statistic	p-value
Intercept	531.00	46.32	11.464	< 0.0001
ethnicity[Asian]	−18.69	65.02	−0.287	0.7740
ethnicity[Caucasian]	−12.50	56.68	−0.221	0.8260

Interpretation: The Asian category will have 18.69 less debt than the African American category, and that the Caucasian category will have 12.50 less debt than the African American category.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.
- It makes restrictive assumptions, often violated in practice.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.
- It makes restrictive assumptions, often violated in practice.
- Two key assumptions are **additivity** and **linearity**.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.
- It makes restrictive assumptions, often violated in practice.
- Two key assumptions are **additivity** and **linearity**.
- **Additivity**: The association between a predictor X_j and the response Y does not depend on the other predictors.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.
- It makes restrictive assumptions, often violated in practice.
- Two key assumptions are **additivity** and **linearity**.
- **Additivity**: The association between a predictor X_j and the response Y does not depend on the other predictors.
- **Linearity**: The change in Y associated with a one-unit change in X_j is constant, regardless of the value of X_j.

Assumptions of the Standard Linear Regression Model

Additivity and Linearity Assumptions

- The standard linear regression model provides interpretable results and works well on many real-world problems.
- It makes restrictive assumptions, often violated in practice.
- Two key assumptions are **additivity** and **linearity**.
- **Additivity**: The association between a predictor X_j and the response Y does not depend on the other predictors.
- **Linearity**: The change in Y associated with a one-unit change in X_j is constant, regardless of the value of X_j.
- In later chapters, we explore methods that relax these assumptions.

Extensions of the Linear Model

Removing the additive assumption: **interactions** and **nonlinearity**.

Extensions of the Linear Model

Removing the additive assumption: **interactions** and **nonlinearity**.

Interactions:

Extensions of the Linear Model

Removing the additive assumption: **interactions** and **nonlinearity**.

Interactions:

- Consider the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

  Regardless of the value of X_2, a one-unit increase in X_1 will lead to a β_1-unit increase in Y.

Extensions of the Linear Model

Removing the additive assumption: **interactions** and **nonlinearity**.

Interactions:

- Consider the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

  Regardless of the value of X_2, a one-unit increase in X_1 will lead to a β_1-unit increase in Y.

- Consider the model with **interaction** terms

$$\begin{aligned} Y &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon \\ &= \beta_0 + (\beta_1 + \beta_3 X_2) X_1 + \beta_2 X_2 + \epsilon \\ &= \beta_0 + \tilde{\beta}_1 X_1 + \beta_2 X_2 + \epsilon, \end{aligned}$$

  where $\tilde{\beta}_1 = \beta_1 + \beta_3 X_2$. Since $\tilde{\beta}_1$ changes with X_2, the effect of X_1 on Y is no longer constant: adjusting X_2 will change the impact of X_1 on Y.

Interaction Effects in the Advertising Data

Synergy and Interaction Effects

- In our previous analysis, both TV and radio advertising were associated with sales.

Interaction Effects in the Advertising Data

Synergy and Interaction Effects

- In our previous analysis, both TV and radio advertising were associated with sales.
- The linear model assumed that the effect of one medium is independent of the other.

Interaction Effects in the Advertising Data

Synergy and Interaction Effects

- In our previous analysis, both TV and radio advertising were associated with sales.
- The linear model assumed that the effect of one medium is independent of the other.
- However, this may not be correct. Spending on radio may increase the effectiveness of TV ads.

Interaction Effects in the Advertising Data

Synergy and Interaction Effects

- In our previous analysis, both TV and radio advertising were associated with sales.
- The linear model assumed that the effect of one medium is independent of the other.
- However, this may not be correct. Spending on radio may increase the effectiveness of TV ads.
- $$\begin{aligned} sales &= \beta_0 + \beta_1 TV + \beta_2 radio + \beta_3 (TV \times radio) + \epsilon \\ &= \beta_0 + (\beta_1 + \beta_3 radio) \times TV + \beta_2 radio + \epsilon. \end{aligned}$$

Advertising Data

	Coefficient	Std. Error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Advertising Data

	Coefficient	Std. Error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Interpretation: an increase in TV advertising of $1,000 is associated with increased sales of $(\beta_1 + \beta_3 \textit{radio}) \times 1000 = 19 + 1.1 \times \textit{radio}$ units.

Advertising Data

	Coefficient	Std. Error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Interpretation: an increase in TV advertising of $1,000 is associated with increased sales of $(\beta_1 + \beta_3 \textit{radio}) \times 1000 = 19 + 1.1 \times \textit{radio}$ units.

Interpretation of $\beta_1, \beta_2, \beta_3$?

Advertising Data

	Coefficient	Std. Error	t-statistic	p-value
Intercept	6.7502	0.248	27.23	< 0.0001
TV	0.0191	0.002	12.70	< 0.0001
radio	0.0289	0.009	3.24	0.0014
TV×radio	0.0011	0.000	20.73	< 0.0001

Interpretation: an increase in TV advertising of $1,000 is associated with increased sales of $(\beta_1 + \beta_3 \textit{radio}) \times 1000 = 19 + 1.1 \times \textit{radio}$ units.

Interpretation of $\beta_1, \beta_2, \beta_3$? Read pages 89-90 of the textbook

Hierarchy

The **hierarchy principle**:

Hierarchy

The **hierarchy principle**:

- If we include an interaction $X_1 \times X_2$ in a model, we should also include the main effects X_1 and X_2, even if the p-values associated with their coefficients are not significant.

Hierarchy

The **hierarchy principle**:

- If we include an interaction $X_1 \times X_2$ in a model, we should also include the main effects X_1 and X_2, even if the p-values associated with their coefficients are not significant.

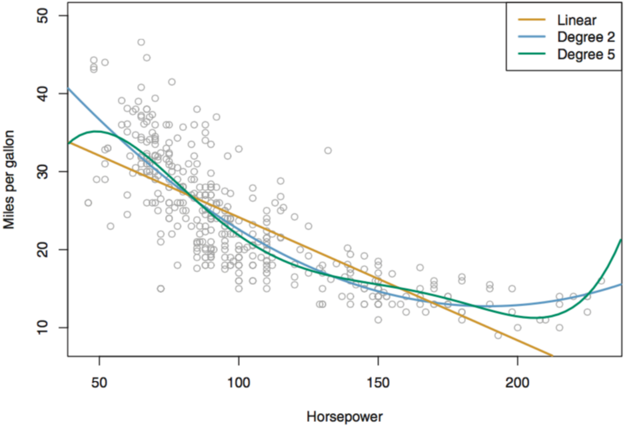
- The rationale for this principle is that interactions are hard to interpret in a model without main effects.

Hierarchy

The **hierarchy principle**:

- If we include an interaction $X_1 \times X_2$ in a model, we should also include the main effects X_1 and X_2, even if the p-values associated with their coefficients are not significant.
- The rationale for this principle is that interactions are hard to interpret in a model without main effects.
- Specifically, the interaction terms also contain main effects, if the model has no main effect terms.

Non-linear Relationships



For a number of cars, mpg and horsepower are shown. The linear regression (orange); the linear regression fit for a model that includes horsepower^2 (blue); the linear regression fit for a model that includes all polynomials of horsepower up to fifth-degree (green).

Non-linear Relationships

The figure suggests that

$$mpg = \beta_0 + \beta_1 horsepower + \beta_2 horsepower^2 + \epsilon,$$

may provide a better fit.

	Coefficient	Std. Error	t-statistic	p-value
Intercept	56.9001	1.8004	31.6	< 0.0001
horsepower	-0.4662	0.0311	-15.0	< 0.0001
horsepower^2	0.0012	0.0001	10.1	< 0.0001

Some general comments:

- A simple approach for incorporating non-linear associations in a linear model is to include transformed versions of the predictors in the model.
- **It is still a linear model!** Can be fitted by least squared with $X_1 = horsepower$, and $X_2 = horsepower^2$.