Workspace 'Workspace' in 'lec7'

Page 1 (row 1, column 1)

Module 3: Linear Regression

Lecture 7

Jan 25th, 2023

Some Important Questions

MICHIGAN STATE UNIVERSITY

- Is at least one of the predictors useful in predicting the response?

$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$

- Do all the predictors help to explain Y, or is only a subset of the of the predictors useful?

2^p, Forward / Backward. Shrinkage. F test. BIC, AIC...

- Given a set of predictor values, what response value should we predict, and how accurate is our prediction?

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_2 X_2 + \hat{\beta}_5 X_5$

Confidence Interval and Prediction

MICHIGAN STATE UNIVERSITY

- ① Confidence interval for a fixed x: a random interval which can cover **f(x)** with probability 95%
- ② Prediction interval for a fixed x : a random interval which can cover a **new** random responds Y with probability 95%

fixed

random

$$\boxed{Y = f(x) + \varepsilon}$$

f

$\mathrm{Var}(\varepsilon) = \sigma^2 \leftarrow$ irreducible variance

special case of ①

C.I. for β_1

$$x = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad x^T\beta = (0\ 1\ 0\ 0)\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_x \end{pmatrix} = \beta_1$$

$\beta_2 \quad x = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$

Confidence Interval and Prediction

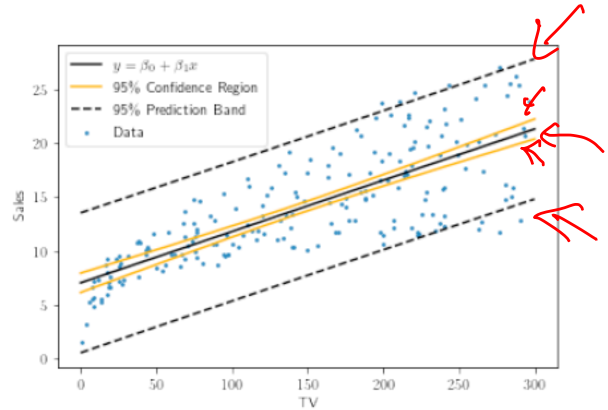
MICHIGAN STATE UNIVERSITY

- Confidence interval for a fixed x: a random interval which can cover **f(x)** with probability 95%
- Prediction interval for a fixed x : a random interval which can cover a **new** random responds Y with probability 95%

$Y = f(x) + \epsilon$

Confidence Interval
The range likely to contain the population parameter (mean, standard deviation) of interest.

Prediction Interval
The range that likely contains the value of the dependent variable for a single new observation given specific values of the independent variables.



Other Considerations

MICHIGAN STATE UNIVERSITY

Quantitative Predictors

age

hp

TV

2 4 6

Qualitative

Gender ⇒ # ?

Color

Countries

Blue Red Green

0 2 1

Qualitative Predictors

- Some predictors are not categorical predictors or factor variables.
- For example: gender, student (student status), status (marital status), and ethnicity (Caucasian, African American (AA) or Asian).

Qualitative Predictors

Example: investigate differences in credit card balance between males and females, ignoring the other variables.

Qualitative Predictors

MICHIGAN STATE UNIVERSITY

Example: investigate differences in credit card balance between males and females, ignoring the other variables.

We create a new 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?

Credit card data

MICHIGAN STATE UNIVERSITY

Female = $\hat{\beta}_0 + \hat{\beta}_1$

male = $\hat{\beta}_0$

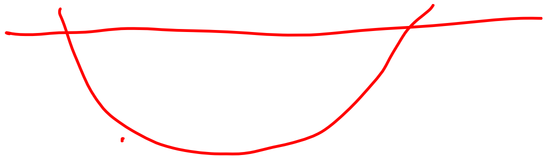
	Coefficient	Std. Error	t-statistic	p-value
Intercept	509.80	33.13	15.389	< 0.0001
gender[Female]	19.73	46.05	0.429	0.6690

More than two levels?

MICHIGAN STATE UNIVERSITY

For the ethnicity: Asian, Caucasian, African American

1 — 2 — 3



$$X = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} = 0$$

2

$$\text{Asian} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad C = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\cancel{AA = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}}$$

$$X \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\hat{\beta} = \boxed{(X^T X)^{-1}} X^T Y$$

More than two levels?

MICHIGAN STATE UNIVERSITY

For the ethnicity: Asian, Caucasian, African American

We will create two dummy variables.

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

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

AS.n. C, AA

$$\begin{pmatrix}1\\0\end{pmatrix} \quad \begin{pmatrix}0\\1\end{pmatrix} \quad \begin{vmatrix}0\\0\end{vmatrix}$$

More than two levels?

MICHIGAN STATE UNIVERSITY

Then we have the following 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.

More than two levels?

	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

Interactions

In the advertising data, we assume the effect on sales of increasing one medium is independent of other media.

$$\widehat{\texttt{sales}} = \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times \texttt{newspaper}$$

But suppose that spending money on radio advertising increases the effectiveness of TV advertising. How to model it?

Extensions of the Linear Model

MICHIGAN STATE UNIVERSITY

Interactions

In the advertising data, we assume the effect on sales of increasing one medium is independent of other media.

$$\widehat{\texttt{sales}} = \beta_0 + \beta_1 \times \texttt{TV} + \beta_2 \times \texttt{radio} + \beta_3 \times \texttt{newspaper}$$

But suppose that spending money on radio advertising increases the effectiveness of TV advertising. How to model it?

x_3 $\quad$ $\text{radio}^2 \cdot TV$

This is called an interaction effect.

$\boxed{x_1 \cdot x_2}$ $\quad$ $\sin TV \ldots$

$$\begin{aligned}
\texttt{sales} &= \beta_0 + \beta_1 \times \overset{x_1}{\texttt{TV}} + \beta_2 \times \overset{x_2}{\texttt{radio}} + \beta_3 \times \boxed{(\cancel{\texttt{radio} \times \texttt{TV}})} + \epsilon \\
&= \beta_0 + \underbrace{\boxed{(\beta_1 + \beta_3 \times \texttt{radio})}}_{\alpha_1} \times \texttt{TV} + \beta_2 \times \texttt{radio} + \epsilon.
\end{aligned}$$

Interpretation

MICHIGAN STATE UNIVERSITY

	Coefficient	Std. Error	t-statistic	p-value
β_0 Intercept	6.7502	0.248	27.23	< 0.0001
β_1 TV	0.0191	0.002	12.70	< 0.0001
β_2 radio	0.0289	0.009	3.24	0.0014
β_3 TV×radio	0.0011	0.000	20.73	< 0.0001

The p-value for the interaction term TV × radio is extremely low, indicating that there is strong evidence for $H_A : \beta_3 \neq 0$.

6%

The R2 for the interaction model is 96.8%, compared to only 89.7% for the model that predicts sales using TV and radio without an interaction term.

This means that (96.8 − 89.7)/(100 − 89.7) = 69% of the variability in sales that remains after fitting the additive model has been explained by the interaction term.

Interpretation

MICHIGAN STATE UNIVERSITY

	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

The coefficient estimates in the table suggest that an increase in TV advertising of \$1, 000 is associated with increased sales of

$$(\hat{\beta}_1 + \hat{\beta}_3 \times \texttt{radio}) \times 1000 = 19 + 1.1 \times \texttt{radio} \text{ units.}$$

An increase in radio advertising of \$1, 000 will be associated with an increase in sales of

$$(\hat{\beta}_2 + \hat{\beta}_3 \times \texttt{TV}) \times 1000 = 29 + 1.1 \times \texttt{TV} \text{ units.}$$

Hierarchy

Sometimes it is the case that an interaction term has a very small p-value, but the associated main effects (in this case, TV and radio) do not.

The hierarchy principle:

If we include the interaction term, we should include the main effects no matter what!

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1 x_2)$$

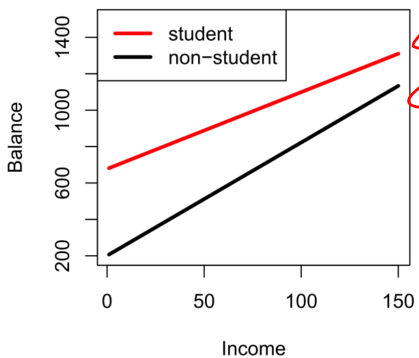
Qualitative and quantitative variables?

Consider the Credit data set, and suppose that we wish to predict balance using income (quantitative) and student (qualitative). With interactions, we have

student$_i$ student$_i$

$$\texttt{balance}_i \approx \beta_0 + \beta_1 \times \texttt{income}_i + \begin{cases} \beta_2 + \beta_3 \times \texttt{income}_i & \text{if student} \\ 0 & \text{if not student} \end{cases}$$

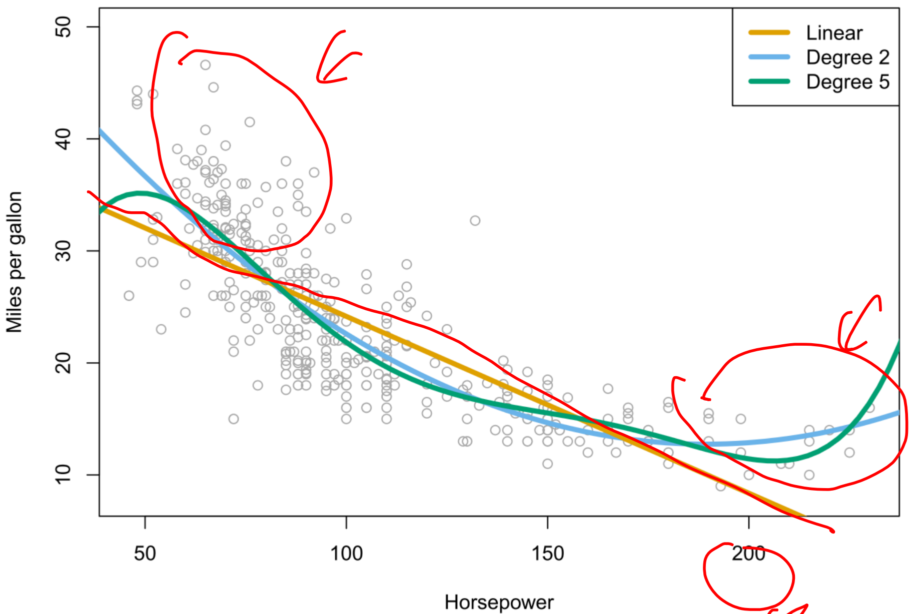
$$= \begin{cases} \boxed{(\beta_0 + \beta_2)} + (\beta_1 + \beta_3) \times \texttt{income}_i & \text{if student} \\ \beta_0 + \beta_1 \times \texttt{income}_i & \text{if not student} \end{cases}$$



Non-linear effects of predictors

MICHIGAN STATE UNIVERSITY

polynomial regression on `Auto` data

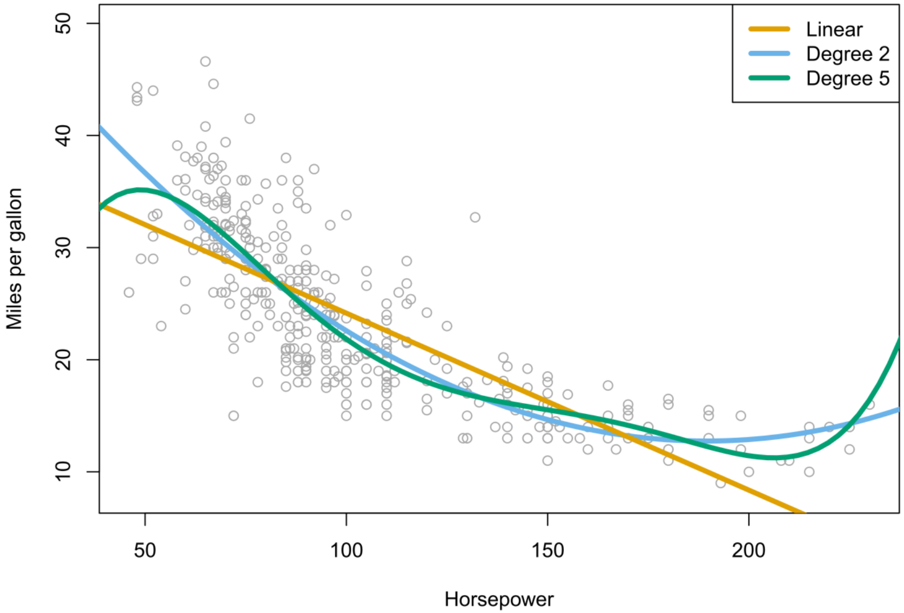


$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$

$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \beta_4 X^4 + \beta_5 X^5 + \epsilon$

Non-linear effects of predictors

polynomial regression on `Auto` data



$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{horsepower} + \beta_2 \times \texttt{horsepower}^2 + \epsilon$$

Non-linear effects of predictors

MICHIGAN STATE UNIVERSITY

$$\texttt{mpg} = \beta_0 + \beta_1 \times \texttt{horsepower} + \beta_2 \times \texttt{horsepower}^2 + \epsilon$$

	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
$\texttt{horsepower}^2$	0.0012	0.0001	10.1	< 0.0001

Bonus Quiz 8

MICHIGAN STATE UNIVERSITY

Assume we have the following model

$$\text{Income} = \beta_0 + \beta_1 X + \epsilon = \beta_0 + \beta_1 \frac{X^*}{100} + \epsilon$$

Here X is height with unit of meter.

$$= \beta_0 + \left(\frac{\beta_1}{100}\right) X^* + \epsilon$$

Now we can have another model

$$\text{Income} = \beta_0 + \beta_1^* X^* + \epsilon \quad = \beta_0 + \beta_1^* X^* + \epsilon$$

$$\Rightarrow \beta_1^* = \frac{\beta_1}{100}.$$

Here X* is also height with unit of centimeter. What is the relationship between β_1 and β_1^*?