# Lecture 25: Linear Regression Part II

Chapter 7.2-7.4

# Questions for Today: Example From Text p.342

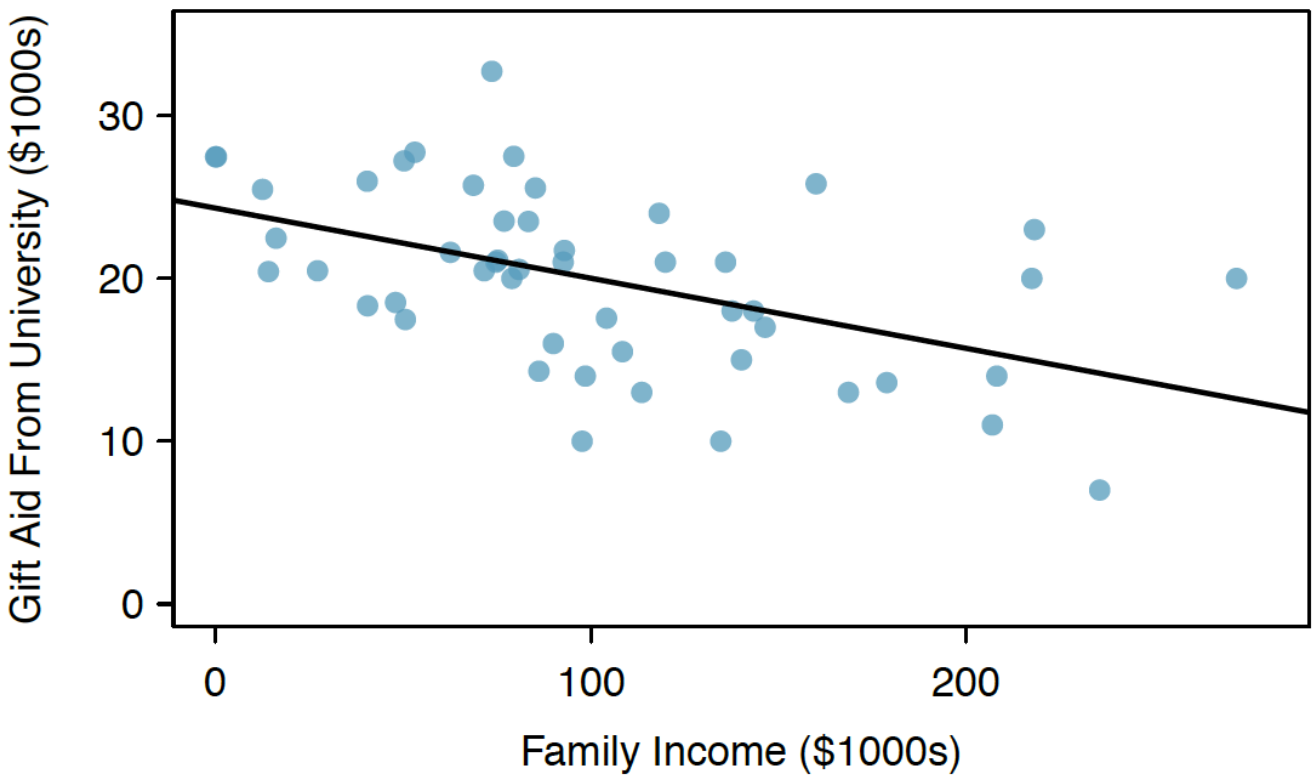
- Data: random sample of 50 students in the 2011 freshman class of Elmhurst College in Illinois.

# Questions for Today: Example From Text p.342

- Data: random sample of 50 students in the 2011 freshman class of Elmhurst College in Illinois.
- Explanatory variable: family income

# Questions for Today: Example From Text p.342

- Data: random sample of 50 students in the 2011 freshman class of Elmhurst College in Illinois.
- Explanatory variable: family income
- Outcome variable: gift aid



# Questions for Today: Example From Text p.342

Using these values,

| | family income in \$1000's (x) | gift aid in \$1000's (y) |
|---|---|---|
| mean | $\overline{x} = 101.8$ | $\overline{y} = 19.94$ |
| sd | $s_x = 63.2$ | $s_y = 5.46$ |
| | | $R = -0.499$ |

# Point Estimates of Intercept

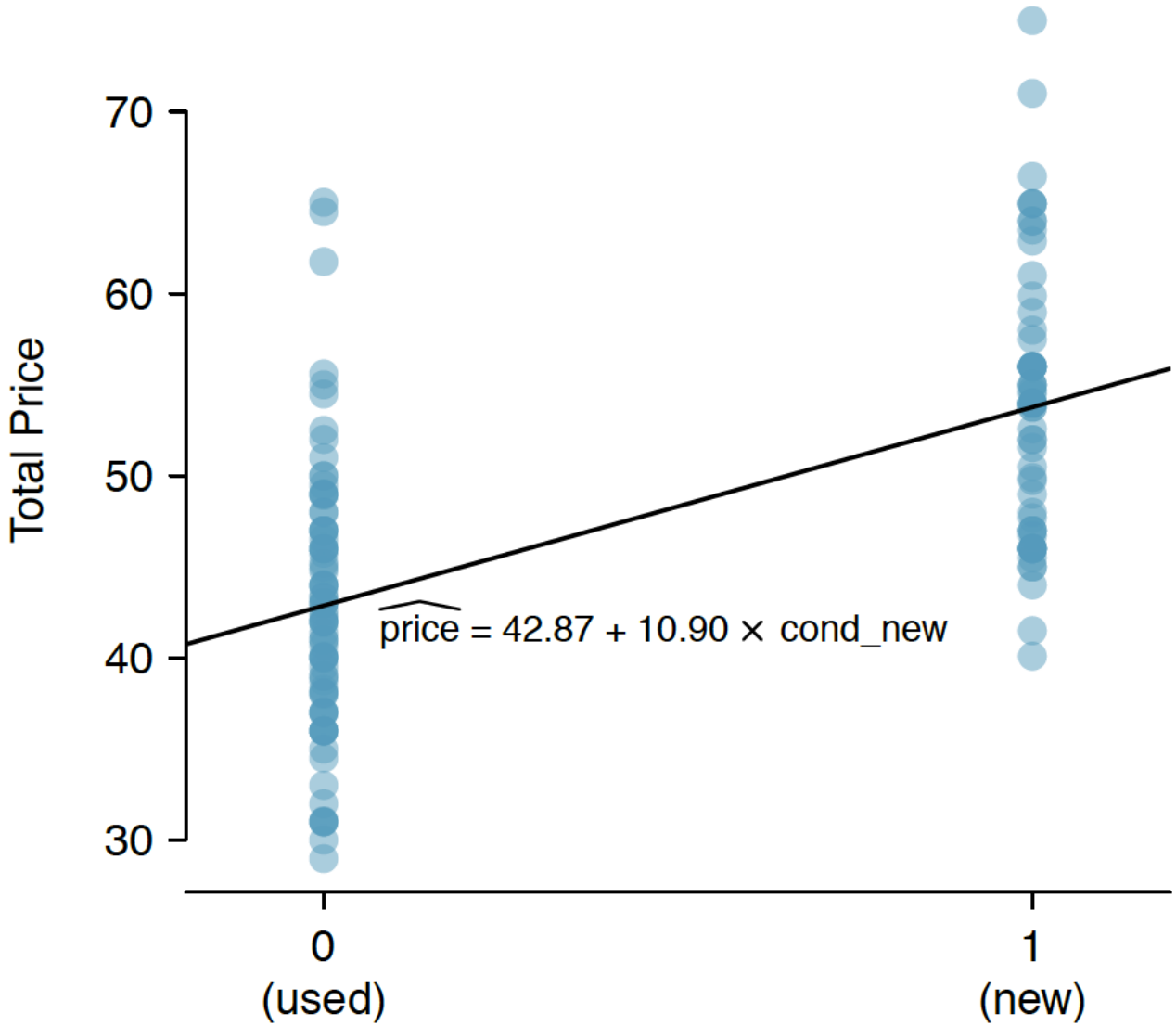
# Point Estimates of Slope

# Extrapolate with Care

Extrapolation: extend the application of a method or conclusion to an unknown situation by assuming that existing trends will continue or similar methods will be applicable.

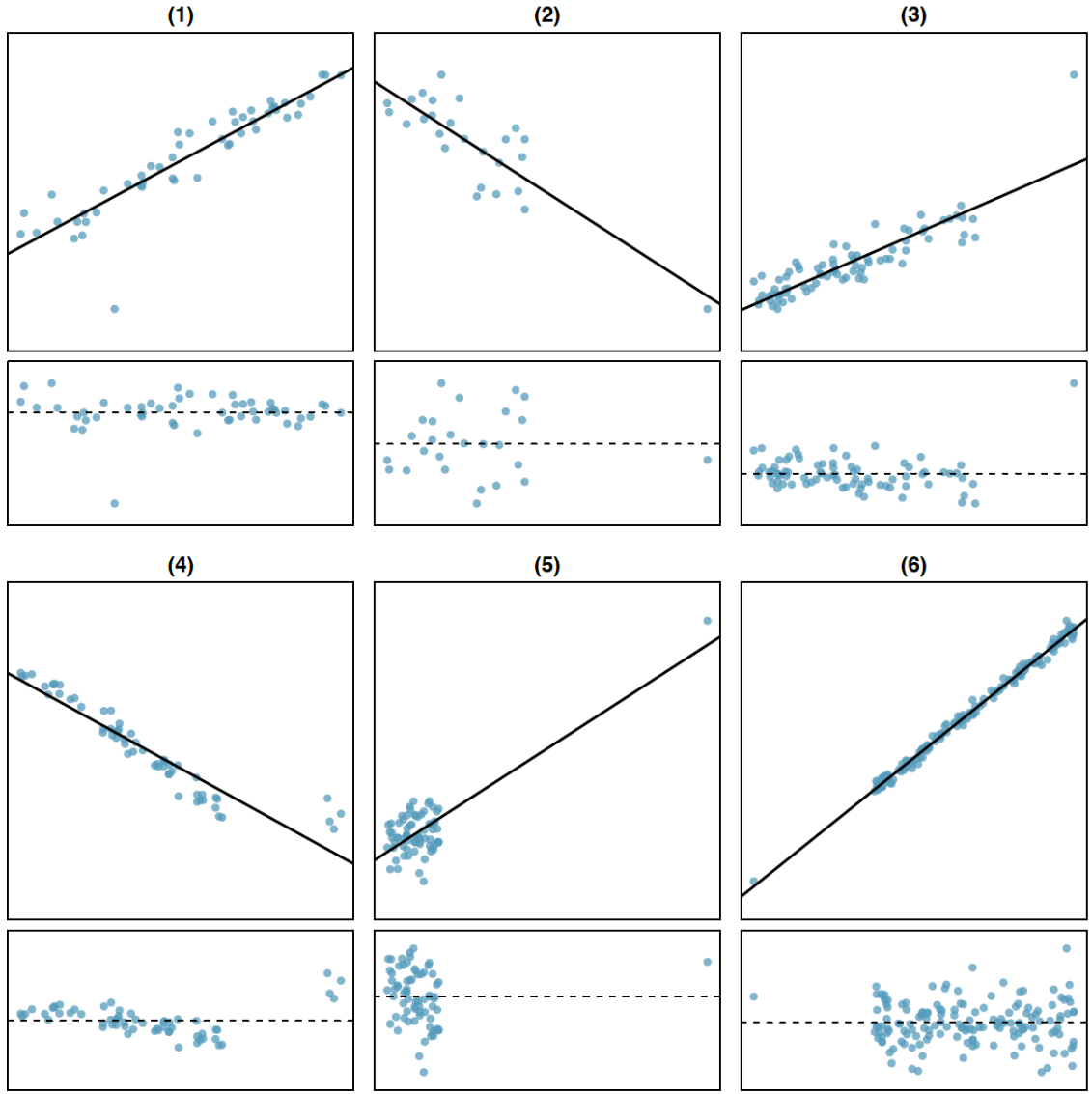
# Categorical Predictor $x$ With Two Levels

# Categorical Predictor $x$ With Two Levels



# Categorical Predictor $x$ With Two Levels

# Types of Outliers in Linear Regression



# Types of Outliers in Linear Regression

Especially in cases 3 and 5, the outliers seem to be pulling the least-squares line towards them.

# Types of Outliers in Linear Regression

Especially in cases 3 and 5, the outliers seem to be pulling the least-squares line towards them.

Points that fall horizontally away from the center of the cloud tend to pull harder on the line, so we call them points with high leverage, i.e. large influence.

# Simple Linear Regression Regression Table

eBay price of old vs new Mario Kart using $n = 141$. On page 355:

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 42.87 | 0.81 | 52.67 | 0.0000 |
| `cond_new` | 10.90 | 1.26 | 8.66 | 0.0000 |
| | | | | $df = 139$ |

# Simple Linear Regression Regression Table

eBay price of old vs new Mario Kart using $n = 141$. On page 355:

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | 42.87 | 0.81 | 52.67 | 0.0000 |
| `cond_new` | 10.90 | 1.26 | 8.66 | 0.0000 |
| | | | | $df = 139$ |

where

- degrees of freedom
  $df = n - k - 1 = n - (k + 1) = 141 - 2 = 139$
- $k$ is the # of predictors in the model
- $k + 1$ is the # of parameters in the model: $\beta_0$ and $\beta_1$

# Confidence Interval and Hypothesis Test for $\beta_1$

Looking at t-table, for $df = 139$, $t^*_{df=139} = 1.98$, so a 95% confidence interval for $\beta_1$ is

$$\begin{aligned} b_1 \pm 1.98 \times SE_{b_1} &= 10.90 \pm 1.98 \times 1.26 \\ &= (8.40, 13.39) \end{aligned}$$

# Confidence Interval and Hypothesis Test for $\beta_1$

Looking at t-table, for $df = 139$, $t^*_{df=139} = 1.98$, so a 95% confidence interval for $\beta_1$ is

$$\begin{aligned} b_1 \pm 1.98 \times SE_{b_1} &= 10.90 \pm 1.98 \times 1.26 \\ &= (8.40, 13.39) \end{aligned}$$

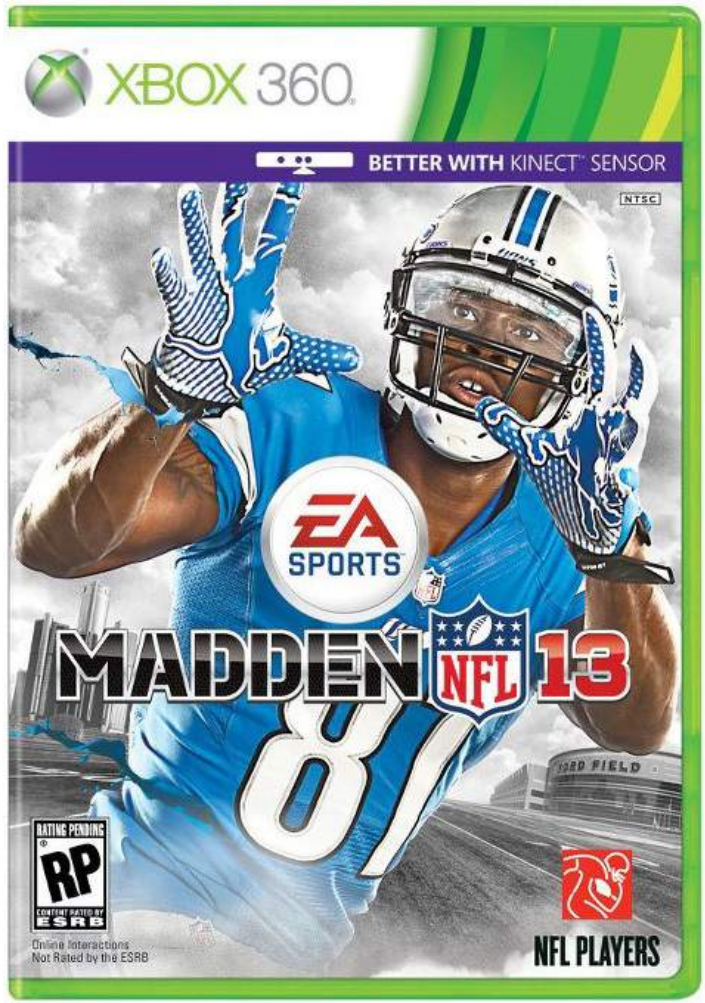
The p-value for the two-sided hypothesis test of

$$\begin{aligned} & H_0 : \beta_1 = 0 \\ \text{vs} \quad & H_A : \beta_1 \neq 0 \end{aligned}$$

is essentially 0, so we reject the null hypothesis and declare that there is an association between `price` and `cond_new`.

# Concept: Regression to the Mean

The Madden Curse. Many NFL players who feature on the cover of the video game Madden end up having subpar subsequent years, leading many to believe there is a curse.



# Concept: Regression to the Mean

Regression to the mean is the phenomenon that if a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement.

# Concept: Regression to the Mean

Regression to the mean is the phenomenon that if a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement.

Madden is selecting players who had exceptional seasons the previous year: the exceptional performance by the players who appear on the cover is not sustainable.

# Concept: Regression to the Mean

Regression to the mean is the phenomenon that if a variable is extreme on its first measurement, it will tend to be closer to the average on its second measurement.

Madden is selecting players who had exceptional seasons the previous year: the exceptional performance by the players who appear on the cover is not sustainable.

So while it looks like a curse, it is just players reverting back to their "mean" level of performance.