Simple Linear Regression

Christine P. Chai

Department of Statistical Science
Duke University

February 23, 2016

Data Example

Name	Midterm	Final
Alice	84	90
Bob	71	68
Carol	92	89
Dick	86	93
Emily	65	77
Frank	75	?
⋮	⋮	⋮

- Frank had an emergency and could not take the final.
- Unfortunately, there is no time for a make-up exam.
- How to predict his score on final exam using the midterm grades?

- Expected score on final $= \alpha + \beta \times$ midterm
- α, β are coefficients to be determined

Data Example

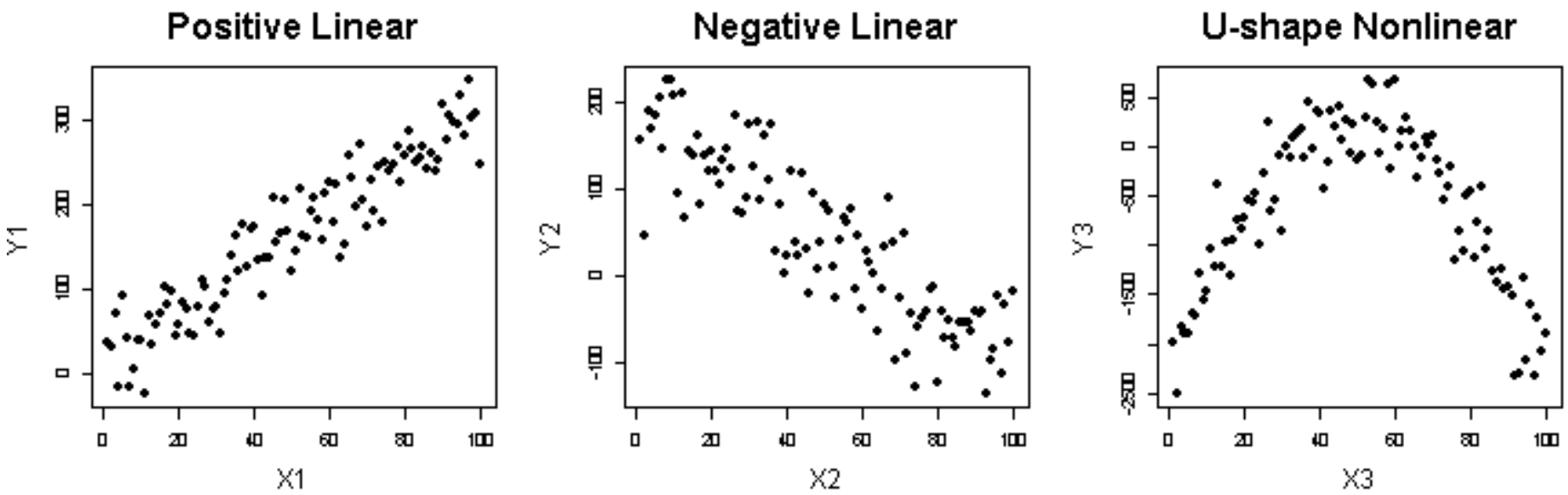
Name	Midterm	Final
Alice	84	90
Bob	71	68
Carol	92	89
Dick	86	93
Emily	65	77
Frank	75	?
⋮	⋮	⋮

- Frank had an emergency and could not take the final.
- Unfortunately, there is no time for a make-up exam.
- How to predict his score on final exam using the midterm grades?

- Expected score on final $= \alpha + \beta \times$ midterm
- α, β are coefficients to be determined

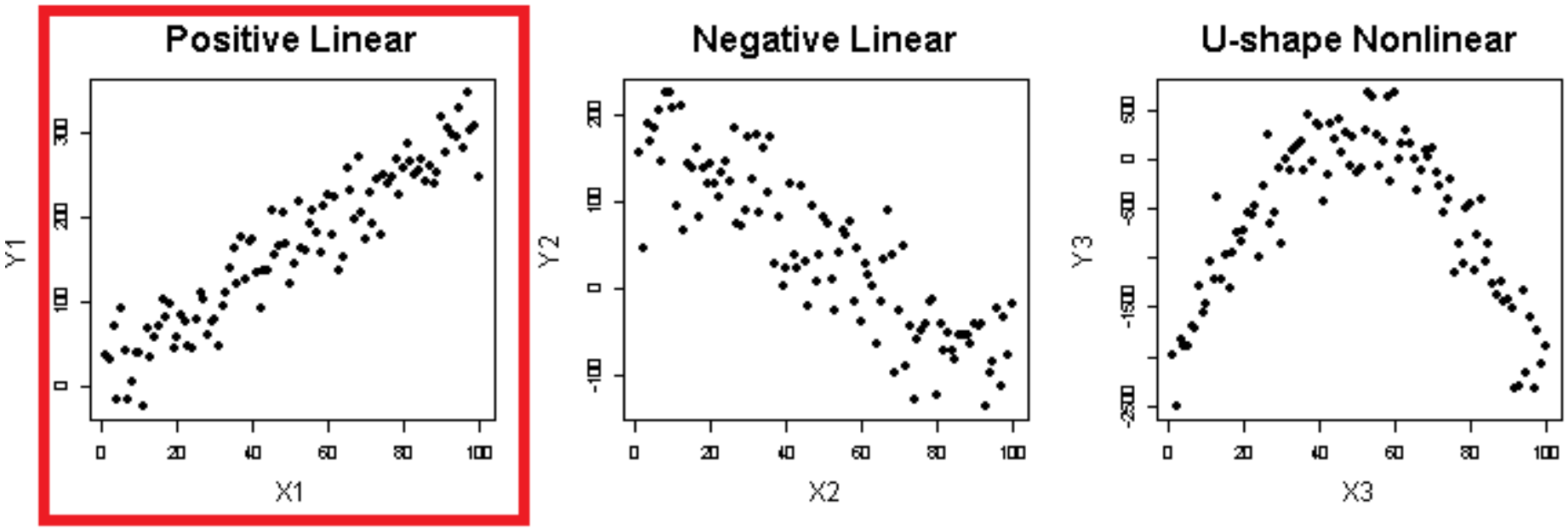
Simple Linear Regression

- **Linear** regression: $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- Predict Y (response variable) from X (explanatory variable)
- Error term ϵ is independent of X, Y, and the coefficients α, β
- Which trend do the midterm-final data look like?



Simple Linear Regression

- **Linear** regression: $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- Predict Y (response variable) from X (explanatory variable)
- Error term ϵ is independent of X, Y, and the coefficients α, β
- Which trend do the midterm-final data look like?



Applications

- $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- Use the R function `lm` to determine the coefficients
- In the midterm-final dataset, $\alpha = 22.13$ and $\beta = 0.770$
- Frank's expected final score $= 22.13 + 0.770 \times 75 = 79.88 \approx 80$

- For every point increase in the midterm score, the final score **is expected to** increase by $\beta = 0.770$ points
- $\alpha = 22.13$ serves as an intercept – the value of Y when $X = 0$
- α may or may not have a meaning

- Do you think it is possible to score 0 points on the midterm exam?

Applications

- $Y = \alpha + \beta X + \epsilon$, with $\epsilon \sim N(0, \sigma^2)$
- Use the R function `lm` to determine the coefficients
- In the midterm-final dataset, $\alpha = 22.13$ and $\beta = 0.770$
- Frank's expected final score $= 22.13 + 0.770 \times 75 = 79.88 \approx 80$

- For every point increase in the midterm score, the final score **is expected to** increase by $\beta = 0.770$ points
- $\alpha = 22.13$ serves as an intercept – the value of Y when $X = 0$
- α may or may not have a meaning

- Do you think it is possible to score 0 points on the midterm exam?