# ECO220Y
# Linear Relationship: Association, Correlation and Linear Regression
## Readings: Chapter 7

Fall 2011

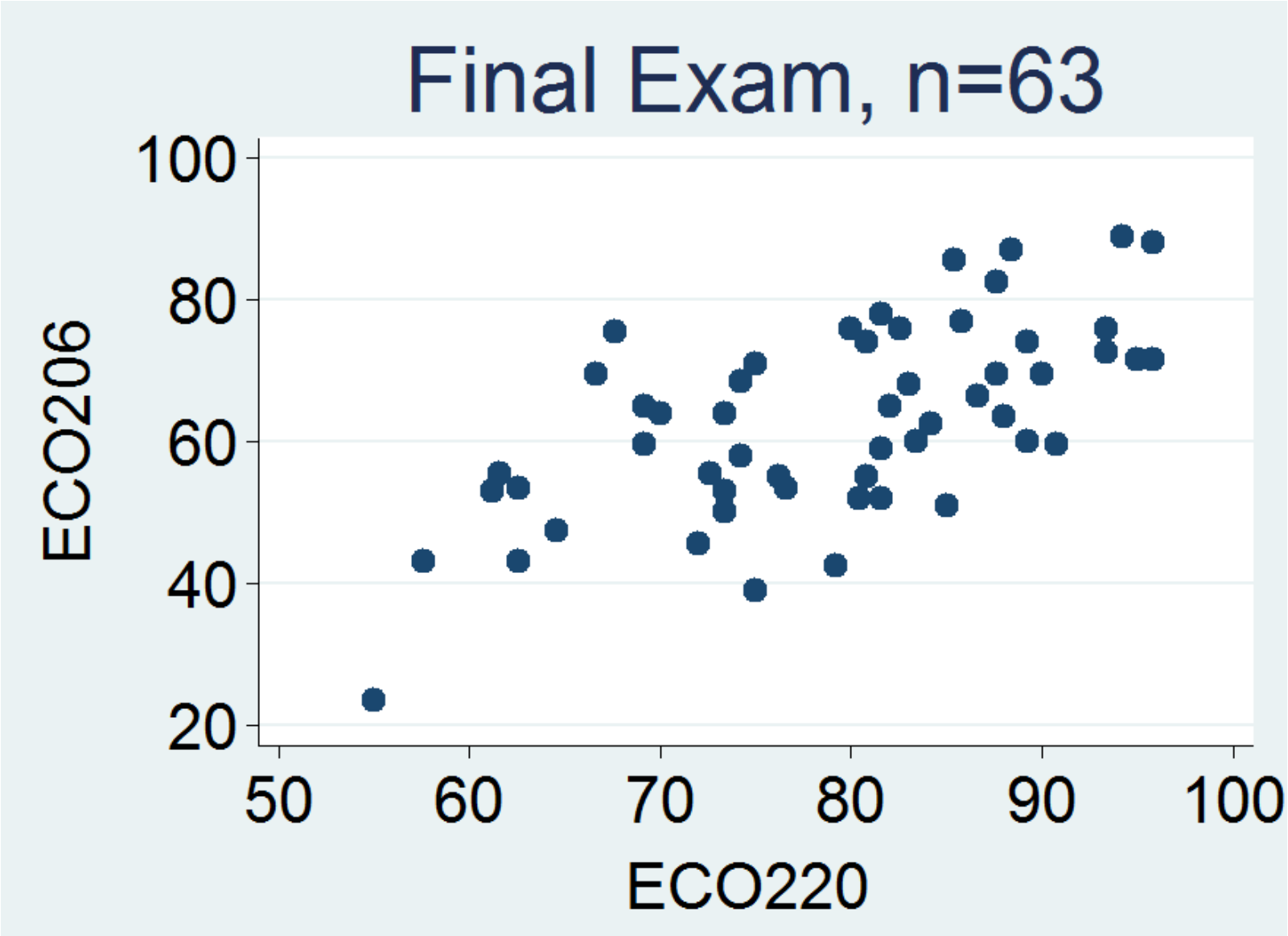
Lecture 3
Part 1 of 2

# Economic Questions...

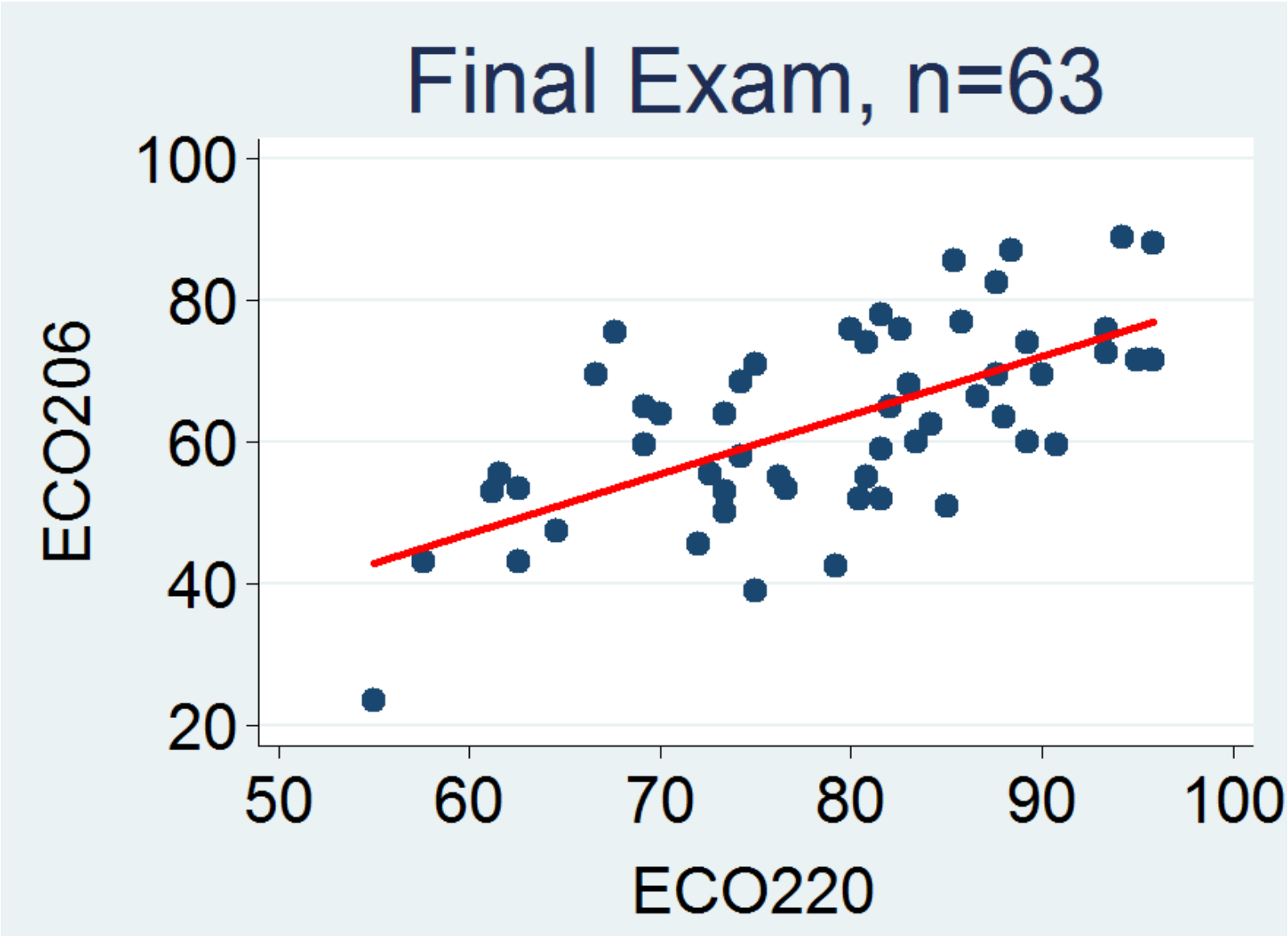
- What is the effect of small class size on test scores?
- What is the effect of advertising on sales?
- What is the effect of education on earnings?
- What is the effect of income tax on labour supply?

## and Answers:

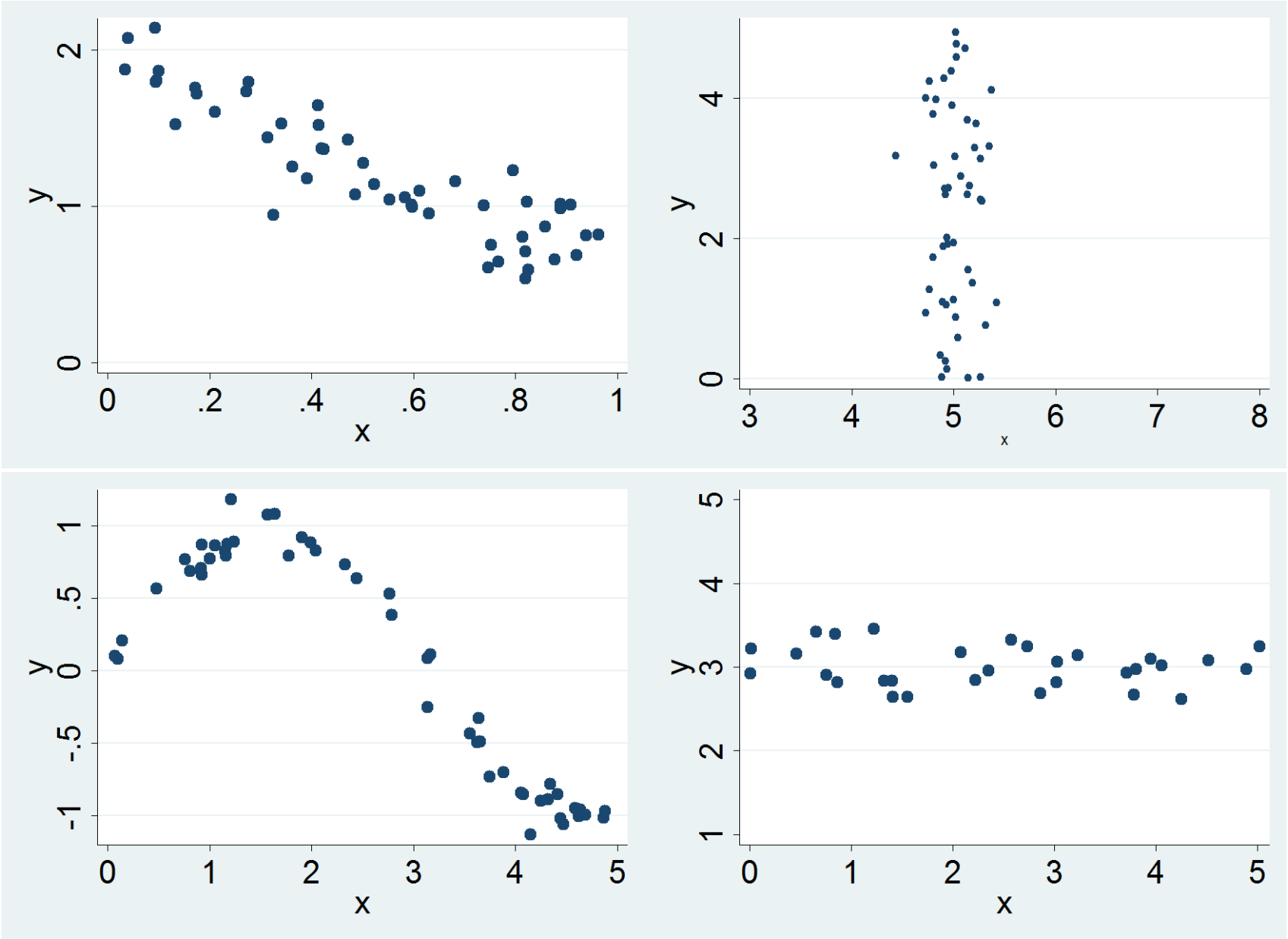
- Reduction in class size by one student leads to improvement in test scores by ?
- Increase in advertising budget by 10% leads to ? increase in sales
- One additional year of education implies a ?% increase in wages
- Income tax cut leads to ?% increase in hours of work




What qualitative statements can we make?




What qualitative statements can we make?


y
0
1
2
0
.2
.4
.6
.8
1
x
y
0
2
4
3
4
5
6
7
8
x
y
-1
-.5
0
.5
1
0
1
2
3
4
5
x
y
1
2
3
4
5
0
1
2
3
4
5
x

We can qualitatively characterize relationship:

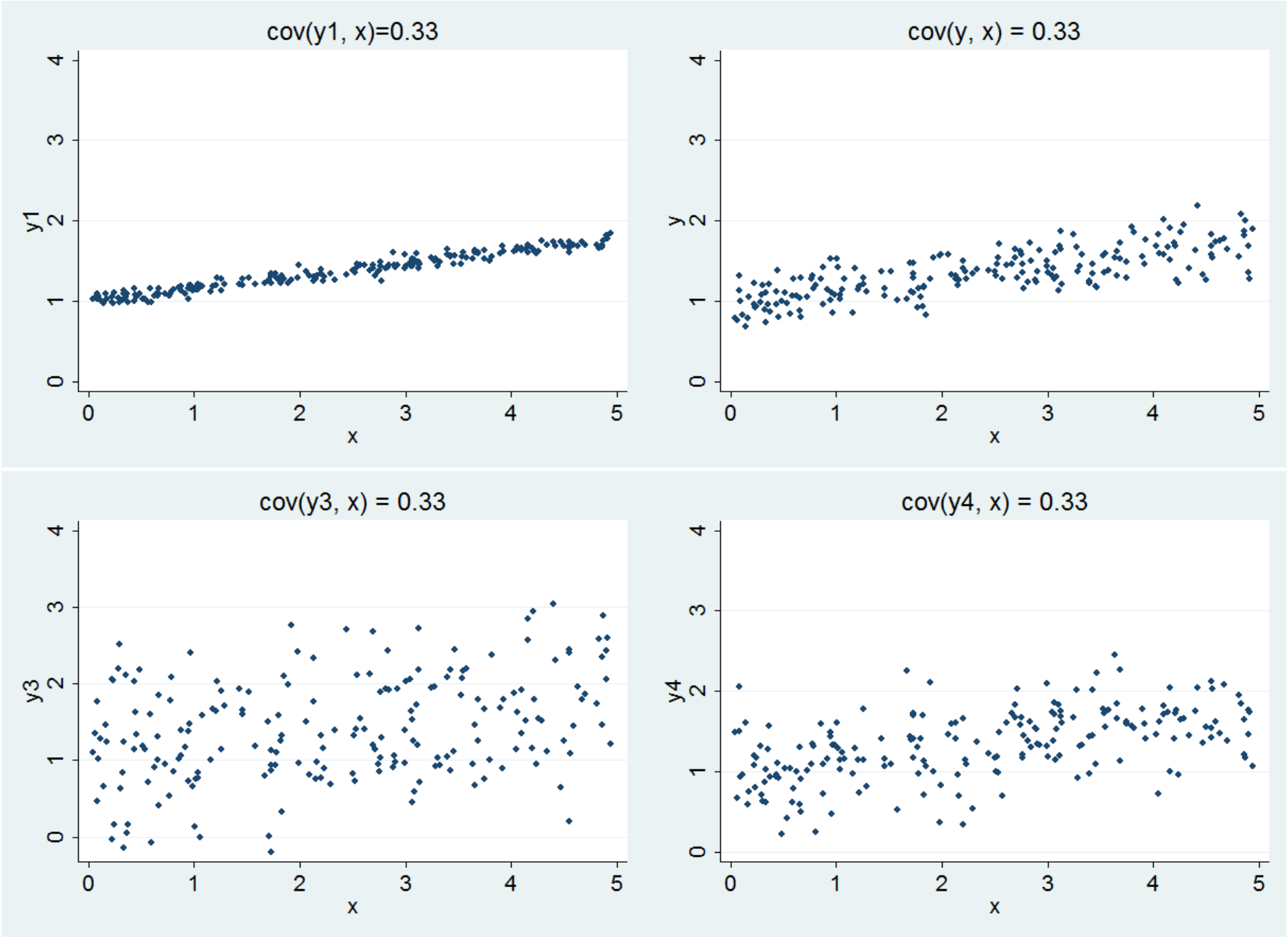
- Direction: Positive ($\nearrow$) or Negative ($\swarrow$) Relationship
- Strength: Strong, Weak, Zero/No Relationship
- Linearity: Linear or Non-Linear Relationship

We can also compute statistics to characterize relationship quantitatively

# Covariance

| | Population | Sample |
|---|---|---|
| Number of observations | $N$ | $n$ |
| Covariance | $\sigma_{xy}$ | $s_{xy}$ |
| How to find | $\frac{\sum_{i=1}^{N}(x_i-\mu_x)(y_i-\mu_y)}{N}$ | $\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n-1}$ |

What are the units of measurement?




How does relationship between x and y differ on these graphs?

# Correlation

| | Population | Sample |
|---|---|---|
| Covariance | $\sigma_{xy}$ | $s_{xy}$ |
| Standard Deviations | $\sigma_x, \sigma_y$ | $s_x, s_y$ |
| How to find | $\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$ | $r = \frac{s_{xy}}{s_x s_y}$ |

What determines the sign?
What are the units of measurement?

# Correlation

- Coefficient of correlation is always between -1 and 1.
- Value close to -1 implies strong negative **linear** relationship.
- Value close to 1 implies strong positive **linear** relationship.
- Value close to 0 implies no **linear** relationship.
- Correlation does not measure the strength of non-linear relationship.
- A strong correlation does not imply that $X$ causes an increase/decrease in $Y$ or $Y$ causes an increase/decrease in $X$.

# From Covariance to Correlation

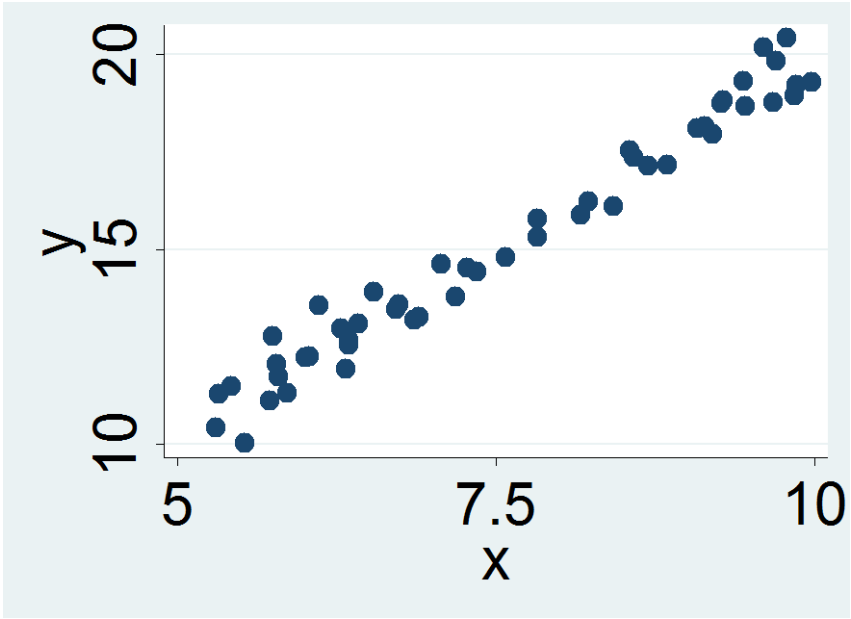
| | X | Y | W |
|---|---|---|---|
| X | 2.32 | | |
| Y | 1.01 | 2.27 | |
| W | -0.27 | 1.71 | 1.86 |

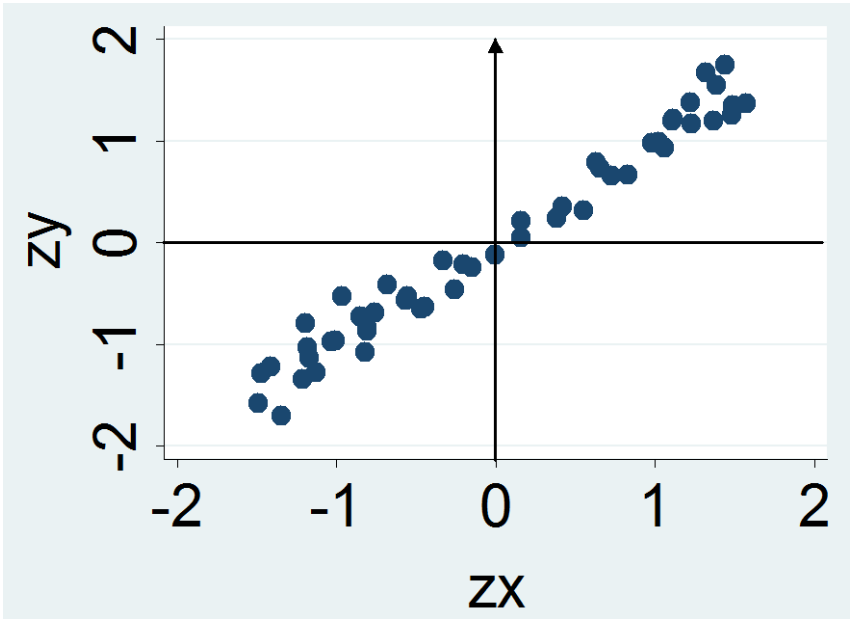
| | X | Y | W |
|---|---|---|---|
| X | 1.00 | | |
| Y | 0.44 | 1.00 | |
| W | -0.13 | 0.83 | 1.00 |

# From Covariance to Correlation

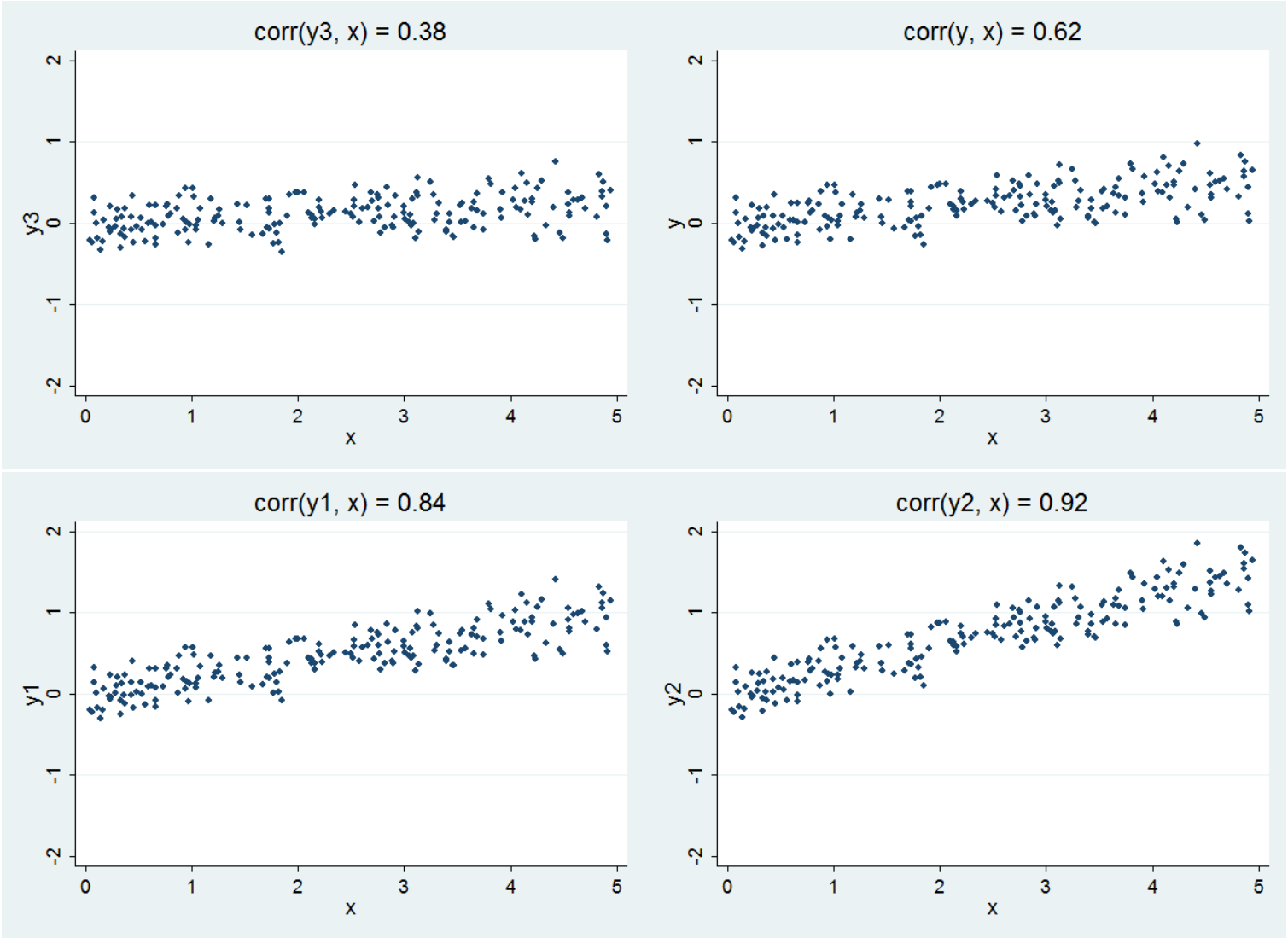
| | X | Y | W |
|---|---|---|---|
| X | $\frac{s_x^2}{s_x s_x}$ | | |
| Y | $\frac{s_{xy}}{s_x s_y}$ | $\frac{s_y^2}{s_y s_y}$ | |
| W | $\frac{s_{xw}}{s_x s_w}$ | $\frac{s_{yw}}{s_y s_w}$ | $\frac{s_w^2}{s_w s_w}$ |

# "Standardized" Scatter Plot



# Next Lecture




What feature of these scatter plots is not captured by correlation?