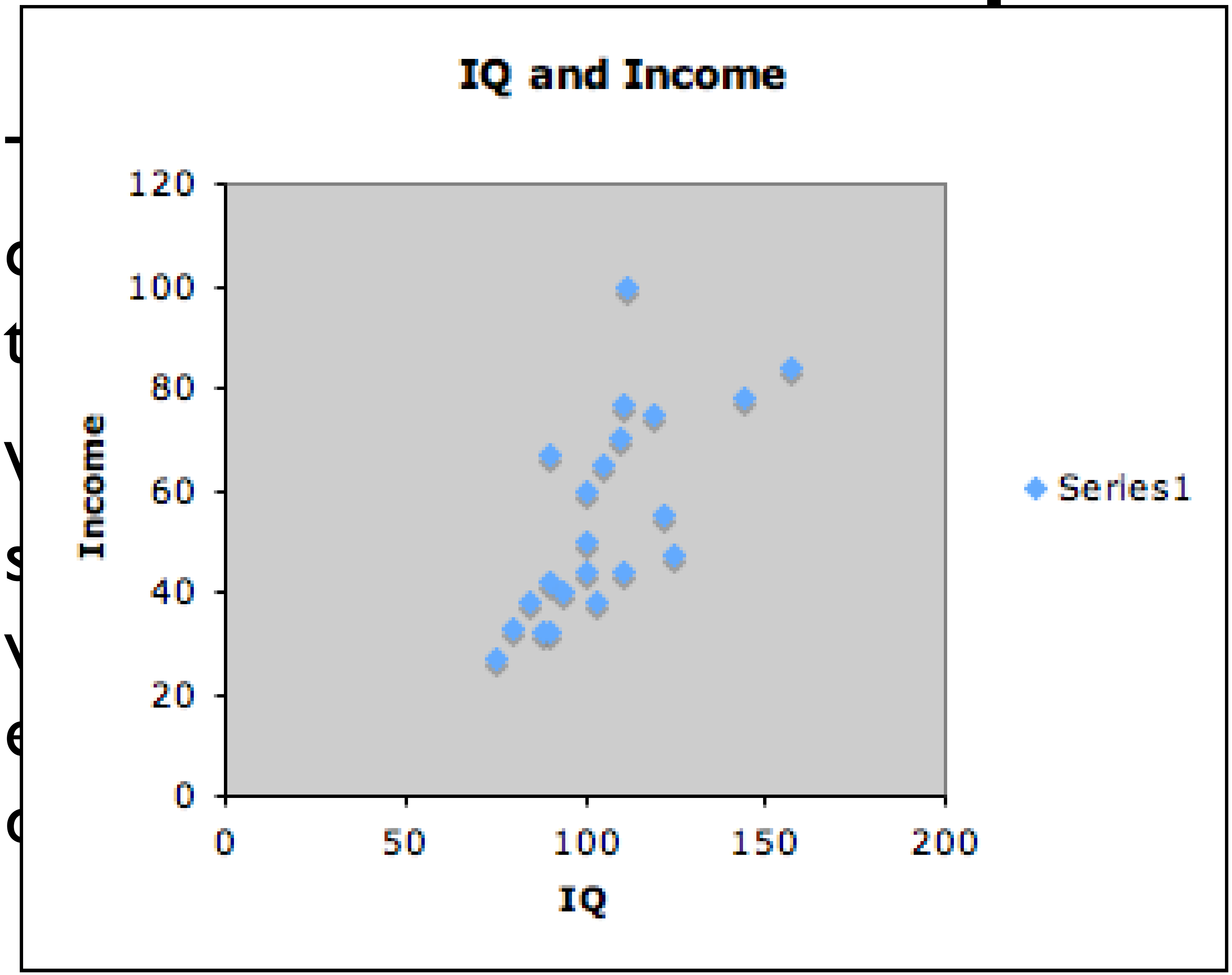
Correlation and Simple Regression

Psychology 3256

Introduction

- All of the procedures we have dealt with so far have looked at differences between means
- You could also look at this as a relationship between the independent and dependent variable
- With a continuous variable the relationship is easy to see

Good ol' scatterplots..



Covariance is a start

$$\mathrm{cov}_{xy} = \frac{\sum (x - \bar{x})(y - \bar{y})}{N - 1}$$

- Measures the degree to which two variables vary together
- if deviations from $\bar{x}$ and $\bar{y}$ go in the same direction you get a positive covariance, otherwise it is negative

We want a measure of association though...

- We will have to standardize covariance, so scales do not matter
- covariance depends on s_x and s_y of course
- Well if it depends on that, why not divide by it?

The Pearson r

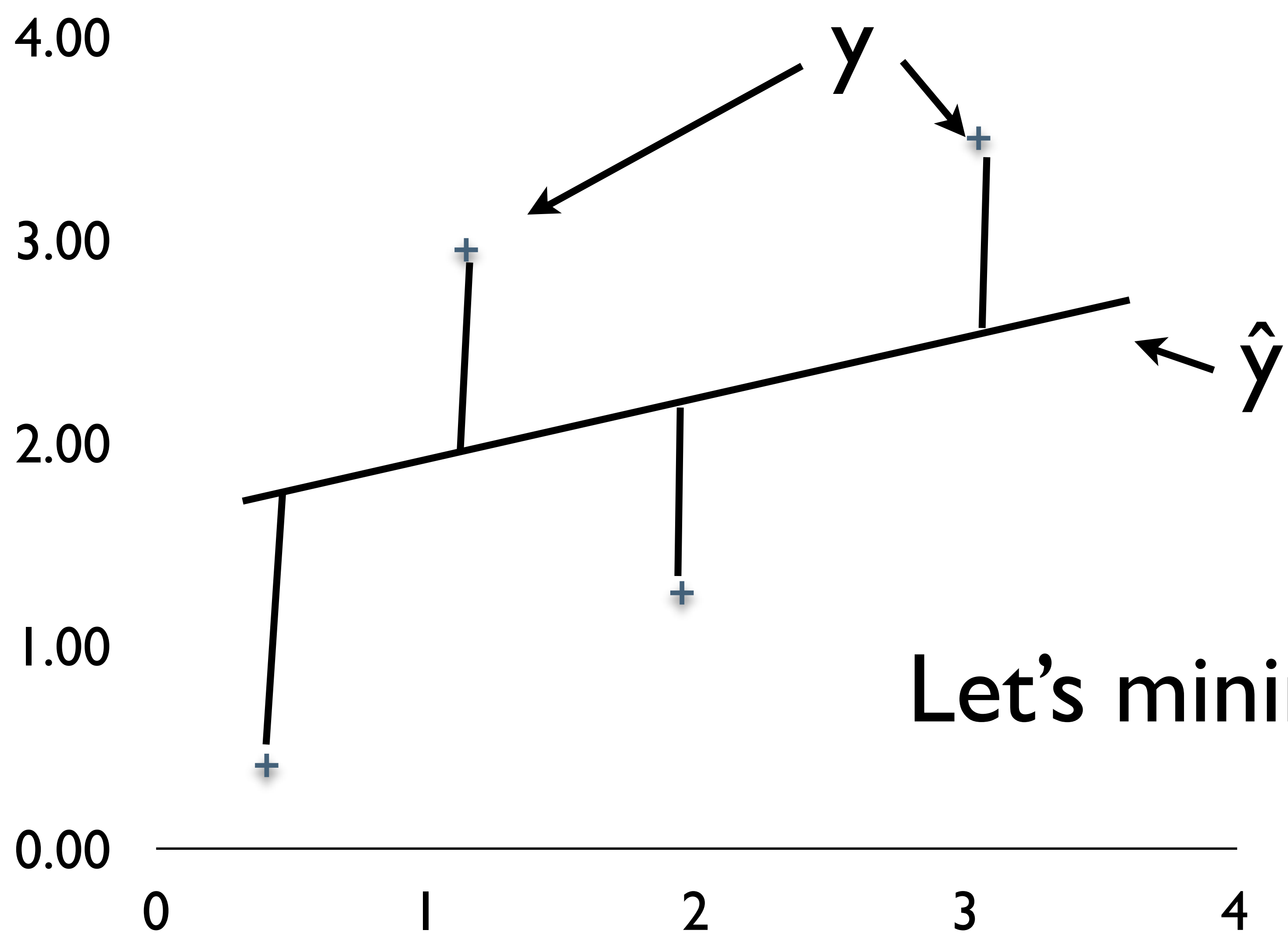
$$r = \frac{\mathrm{cov}_{xy}}{s_x s_y}$$

- if $|\mathrm{cov}_{xy}| = s_x s_y$ then $|r| = 1$ and you haave a perfect relationship
- the sign of r only shows the direction
- $-1 <= r <= 1$
- straight lines only eh

It would be cool if..

- we should be able to predict y from x
- Basically just by drawing a line through a scatterplot
- The most common approach is to use a least squares regression line

Here is the idea



Let's minimize $\Sigma\ (y-\hat{y})^2$

You end up with a prediction equation

$$\hat{y} = a + bx$$

Predicted y

Intercept

Slope

finding b and a

$$b = \frac{\mathrm{cov}_{xy}}{s_x^2}$$

$$a = \overline{y} - b\overline{x}$$

Interpretation

- a, the intercept, is where x=0, not always that meaningful
- linear relationships only
  - look at the residuals (e)
  - $cov_{xe} = 0$
- don't go outside the range