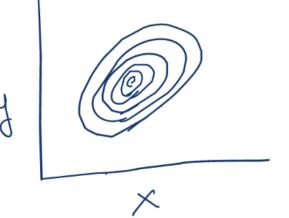
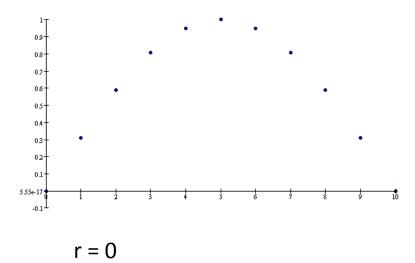
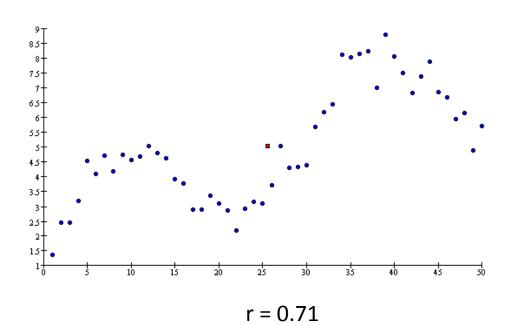
# **Assumptions:**

- Random sample
- Linearity
- Correlation depends on range of values
- Homoscedastic variances
- Bivariate Normal Distribution
  - X is normally distributed
  - Y is normally distributed
  - X and Y have linear relationship

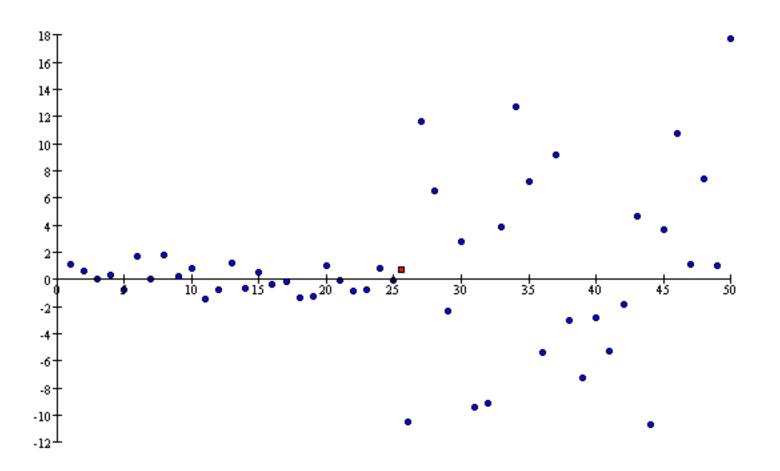


### Non-linearity:

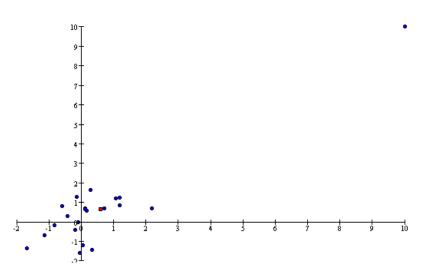


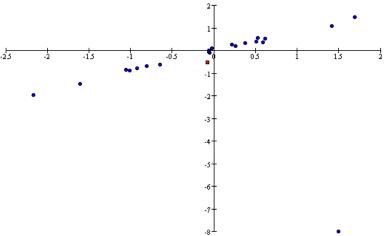


#### **Heteroscedasticity:**

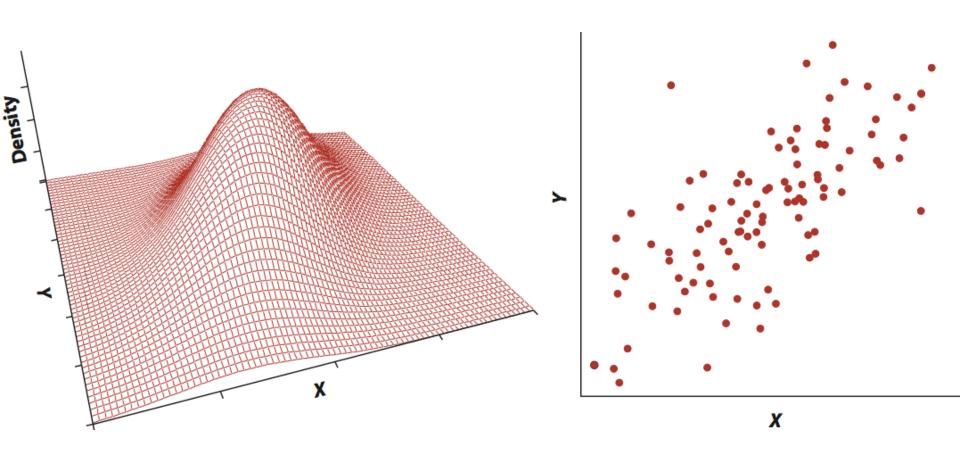


#### **Outliers:**





#### This is what **bivariate** looks like:



If data are not bivariate or are not linearly related try **transformation** of data.

If data are heteroscedastic or have outliers... try a **non-parametric method**.....

 Remember: non-parametric methods are more conservative (they have less power) than parametric

# Spearman's rank

### assumes:

- \* random sample
- \* linear relationship

Spearman's rank correlation:

 Measures strength and direction of linear association between the **ranks** of two variables

Two variables are ranked separately

Parameter: ρ<sub>s</sub>; sample estimate: r<sub>s</sub>

#### Spearman's rank correlation:

Test for correlation in the normal way....

Step 1: declare null and alternate

 $H_0$ : Zero correlation ( $\rho_s$ =0)

 $H_A$ : Some correlation ( $\rho_s \neq 0$ )

**Step 2: test statistic** 

$$r_{s} = \frac{\sum (R - \overline{R})(S - \overline{S})}{\sqrt{\sum (R - \overline{R})^{2}} \sqrt{\sum (S - \overline{S})^{2}}}$$

Step 3:State α/P-value/Critical value

Table G

**Step 4: State conclusion** 

### If n > 100:

$$t = r_s - \rho_s$$

$$SE_{r(s)}$$

where:

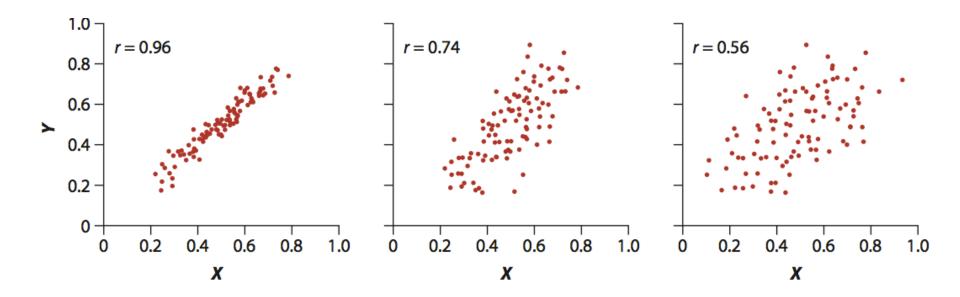
$$SE_{r_s} = \sqrt{\frac{1 - r_s^2}{n - 2}}$$

- t is ~t-distributed with n 2 degrees of freedom
- Tricky part: reject null hypothesis if

  - $t \ge t_{0.05(2),n-2}$   $t \le -t_{0.05(2),n-2}$

#### **Attentuation**

- measurement error weakens correlation

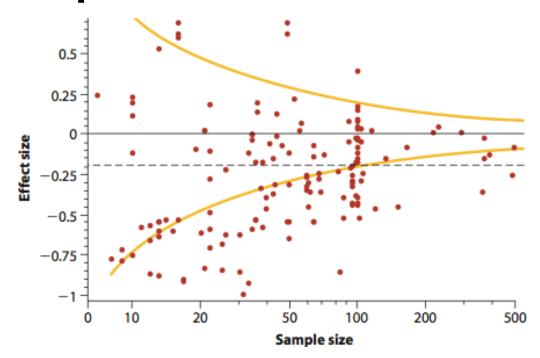


## **Publication Bias**

## Papers that:

- Reject null
- Have large effect

### tend to be published



http://www.badscience.net/about-dr-ben-goldacre/