# Statistics with Spa  ows

Lecture 12

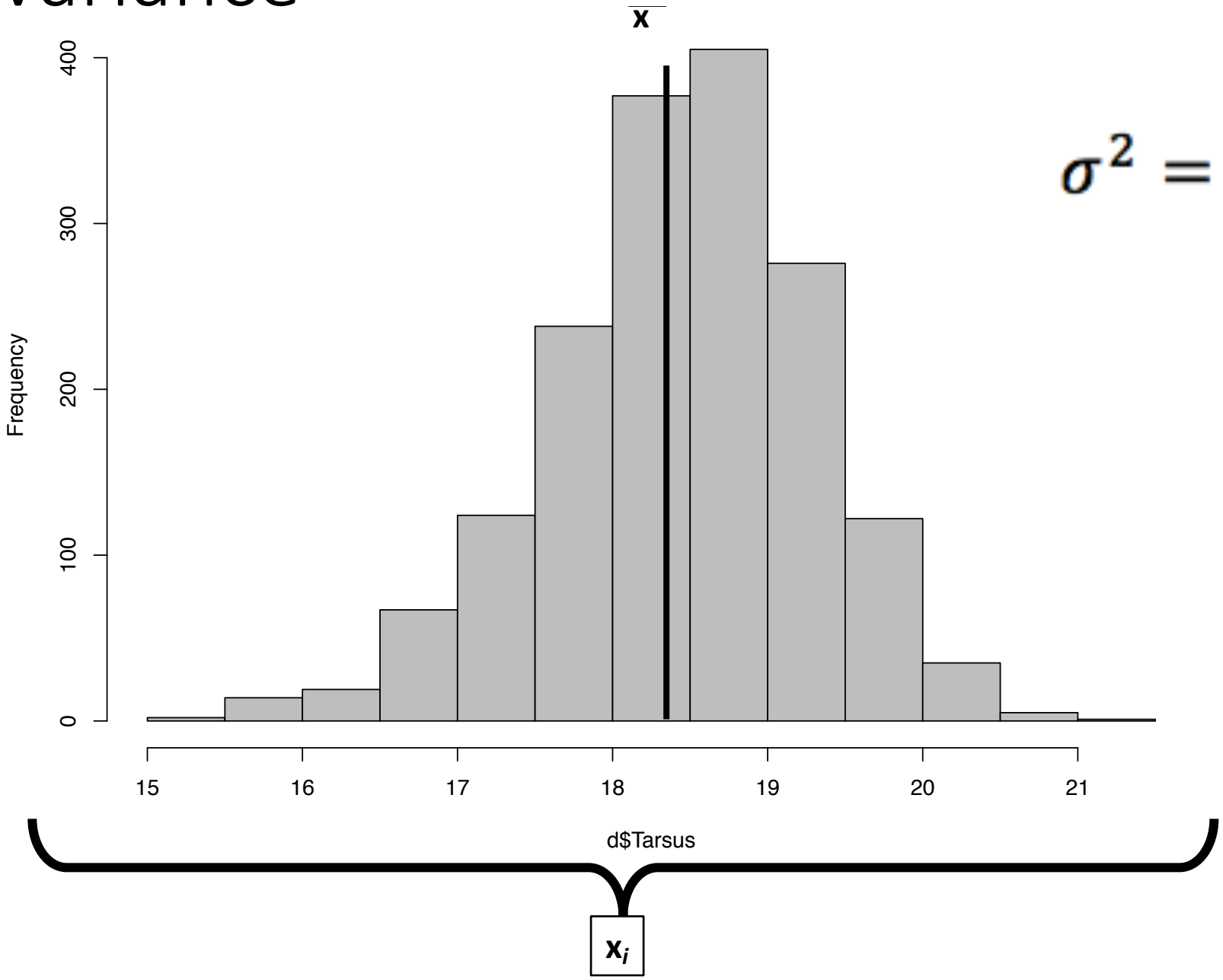
Julia Schroeder

Julia.schroeder@imperial.ac.uk

# Outline

- Check-up on variance
- Covariance and correlation

# Variance

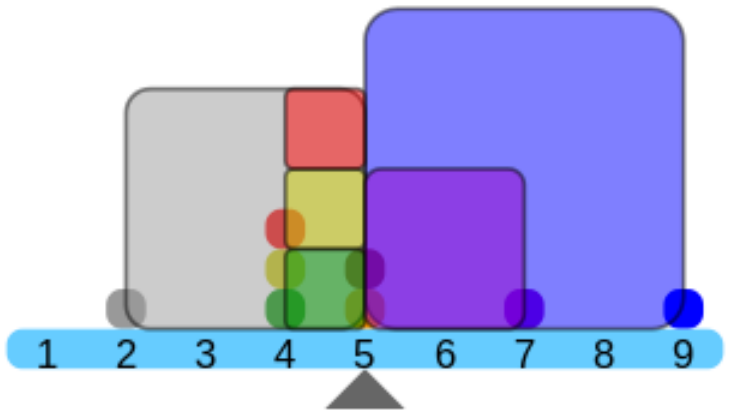



$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

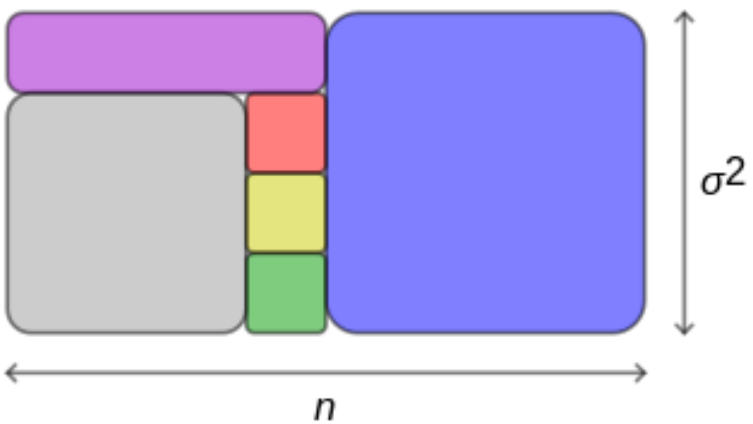
1.


2.


3.




4.



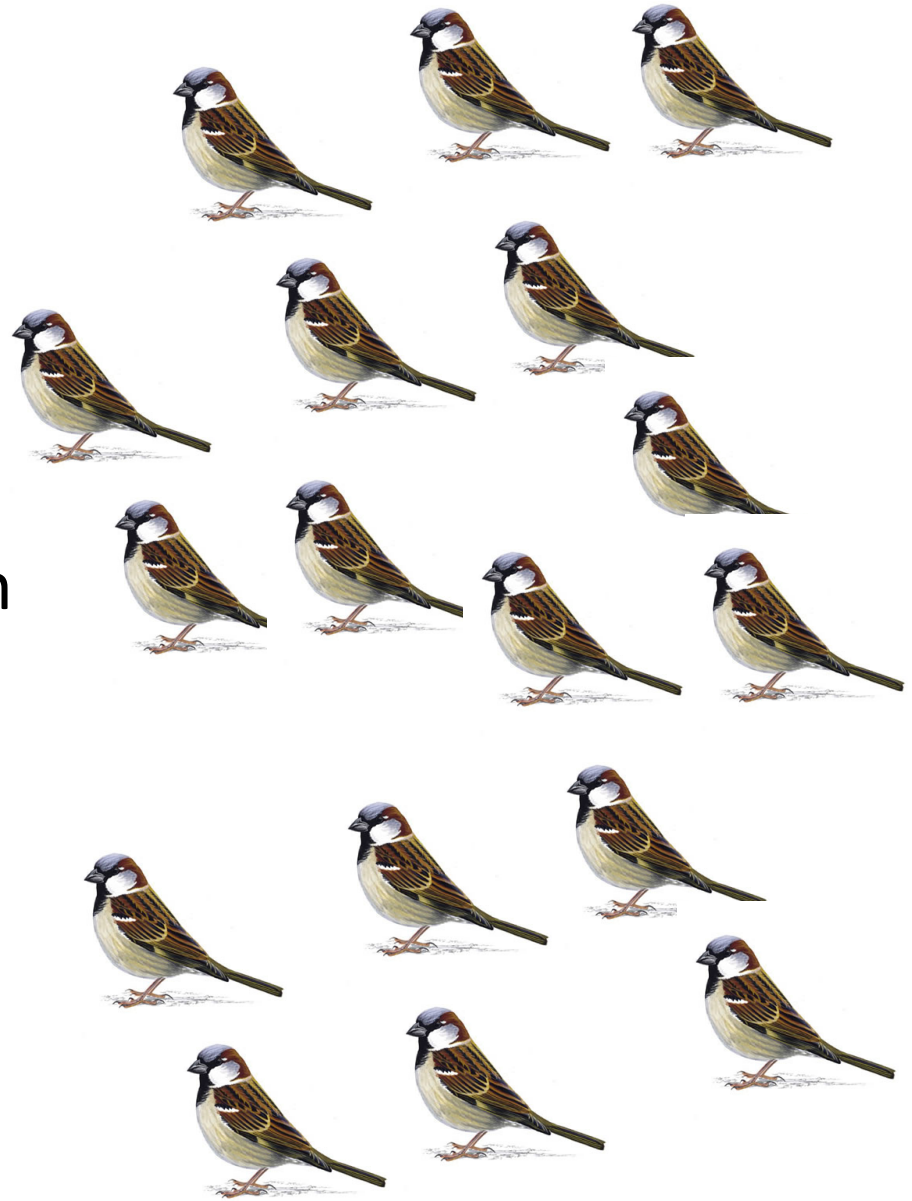

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

# VARIANCE

It’s a Sum of Squares

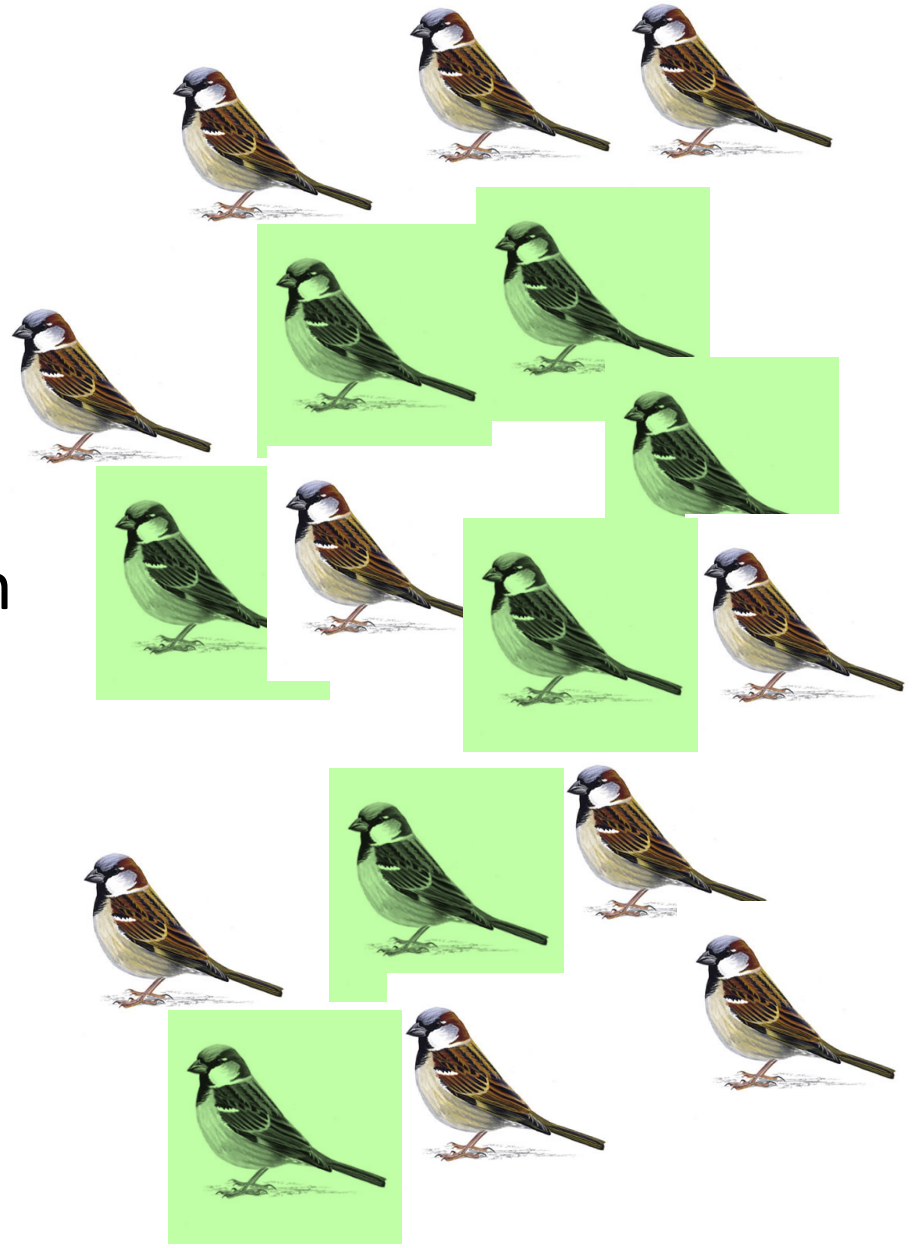
# Covariance

- How two variables change together
- Population:  joint probability distribution



# Covariance

- How two variables change together
- Population:  joint probability distribution
- Sample: covariance estimate



# Covariance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

# Covariance

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

$$Cov_{x,y} = \frac{\sum(xy) - n\bar{x}\bar{y}}{n-1}$$

# Covariance vs Correlation

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

$$Cor\ \rho_{x,y} = \frac{\sum(xy) - n\bar{x}\bar{y}}{n-1} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$$

# Covariance vs Correlation

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

- Ok, why do we need two versions of this?

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

$$Cor\ \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$$

# Covariance vs Correlation

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

- Ok, why do we need two versions of this?

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

The correlation coefficient is the covariance divided by the product of the standard deviations

$$Cor\ \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$$

# Covariance vs Correlation

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

- Ok, why do we need two versions of this?

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

$$Cor\ \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$$

The correlation coefficient is the covariance divided by the product of the standard deviations

It is the *standardized* version of the covariance

# Covariance vs Correlation

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

- Ok, why do we need two versions of this?

$$Cov_{x,y} = \frac{\sum(x - \bar{x})\sum(y - \bar{y})}{n-1}$$

The correlation coefficient is the covariance divided by the product of the standard deviations

$$Cor\ \rho_{x,y} = \frac{Cov_{x,y}}{\sigma_x \sigma_y}$$

It is the *standardized* version of the covariance

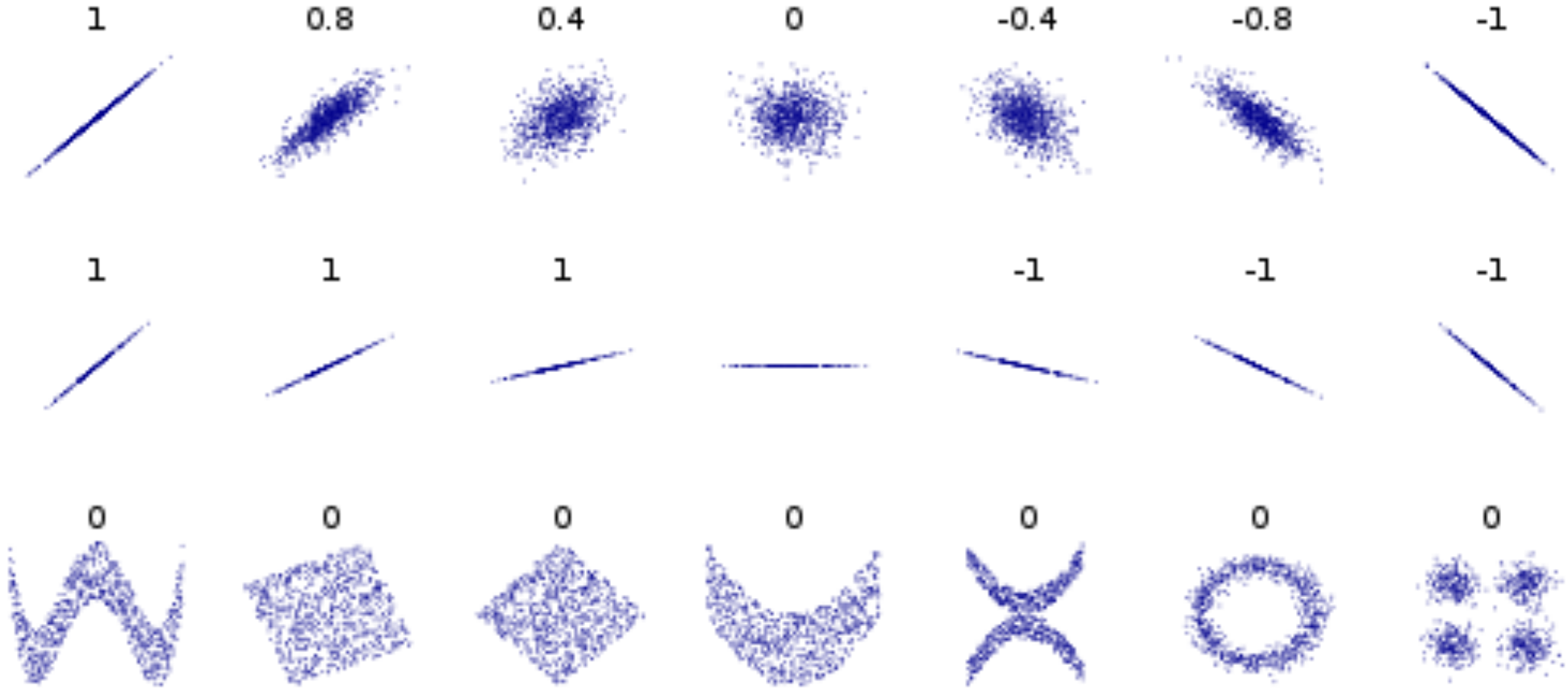
R =Pearson's correlation coefficient

# Correlation coefficient

- Standardized, between -1 and 1.

# Correlation coefficients

- Standardized, between -1 and 1.



# Correlation coefficients

- Standardized, between -1 and 1.

- Guessthecorrelation.com

# Covariance vs Correlation – do it now!

- Calculate covariance of tarsus and mass
- Calculate correlation of tarsus and mass

- Divide Tarsus by 10 (-> cm)
- Calculate covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance?
- What is the function for correlation?
- Covariance of tarsus and mass
- Correlation of tarsus and mass

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance? `cov(d$Mass,d$Tarsus, use="complete.obs")`
- What is the function for correlation?
- Covariance of tarsus and mass
- Correlation of tarsus and mass

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance? `cov(d$Mass,d$Tarsus, use="complete.obs")`
- What is the function for correlation? `cor(d$Mass,d$Tarsus, use="complete.obs")`
- Covariance of tarsus and mass
- Correlation of tarsus and mass

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance?
- What is the function for correlation?
- Covariance of tarsus and mass
- Correlation of tarsus and mass

```
cov(d$Mass,d$Tarsus, use="complete.obs")
cor(d$Mass,d$Tarsus, use="complete.obs")
[1] 0.87
```

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance?
- What is the function for correlation?
- Covariance of tarsus and mass
- Correlation of tarsus and mass

```
cov(d$Mass,d$Tarsus, use="complete.obs")
cor(d$Mass,d$Tarsus, use="complete.obs")
```

```
[1] 0.87
[1] 0.48
```

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

# Covariance vs Correlation – do it now!

- What is the function for covariance?
- What is the function for correlation?
- Covariance of tarsus and mass
- Correlation of tarsus and mass

```
cov(d$Mass,d$Tarsus, use="complete.obs")
cor(d$Mass,d$Tarsus, use="complete.obs")
```

```
[1] 0.87
[1] 0.48
```

- Divide Tarsus by 10 (-> cm)
- Covariance and correlation again

```
[1] 0.087
[1] 0.48
```

# Learning aims

- Covariation is unit-dependent
- Correlation is always between -1 and 1, scaled

- Correlation has no response and explanatory variable (compare with LM)

- We used Spearman correlation – there are others (NP stats)
- Guessthecorrelation.com