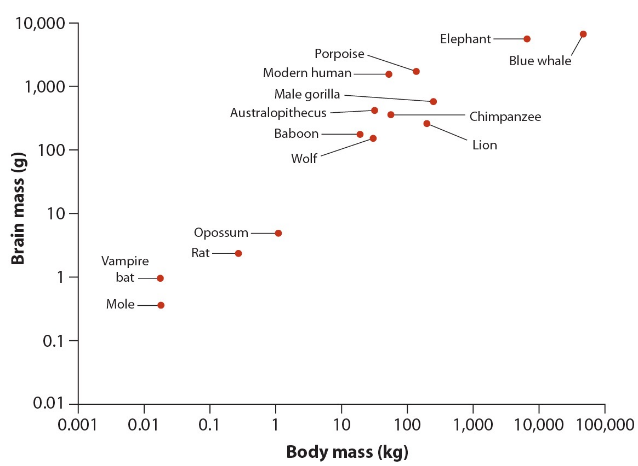
Chapter 16: Correlation between numerical variables

1

Correlation

- When two numerical variables are associated then they are **correlated**



2

Correlation coefficient

- The **correlation coefficient** measures the strength and direction of the association between two numerical variables
  - AKA linear correlation coefficient or Pearson's correlation coefficient
- Correlation coefficient (statistic), r
- Population correlation coefficient (parameter), ρ

3

Correlation coefficient

Measures how deviations in X and Y vary together

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2}\sqrt{\sum_i (Y_i - \bar{Y})^2}}$$

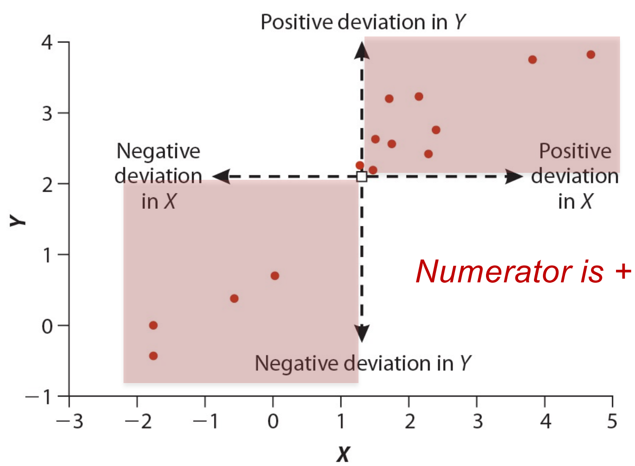


Fig 16.1-1

4

Correlation coefficient

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2}\sqrt{\sum_i (Y_i - \overline{Y})^2}}$$

Square root of sum of squares
(part of standard deviation calculation)

5

Correlation coefficient

- Ranges from -1 to 1

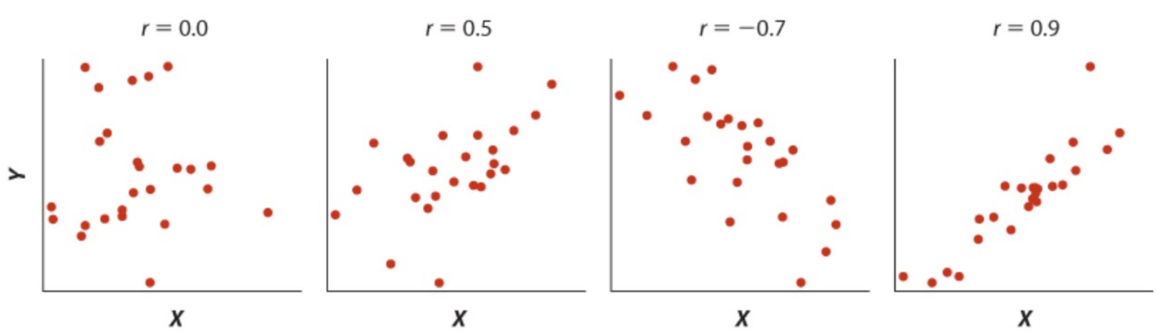


Fig 16.1-2

6

Correlation coefficient

- Ranges from -1 to 1
- Possible that two variables can be strongly associated but have no correlation ($r = 0$)
  - Non-linear association

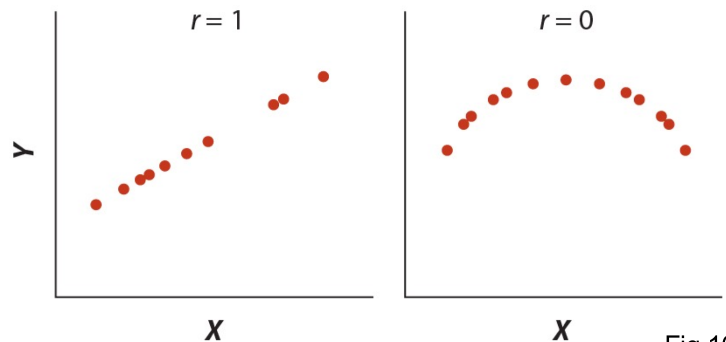


Fig 16.1-3

7

8

Ex 16.1: Flipping the bird

- Mistreated children often become adults that mistreat their young
- Does this happen in other species?
- Nazca boobies in Galapagos
- When adults visit nests of non-related chicks they typically act aggressively toward them
- Is there an association between number of non-relative nest visits and future aggression?

9

TABLE 16.1 Number of non-parent adult visits experienced by boobies as nestlings compared to the number of similar behaviors performed by the same birds when an adult. $n = 24$.

Number of visits	Future aggressive behavior
1	–0.80
7	–0.92
15	–0.80
4	–0.46
11	–0.47
14	–0.46
23	–0.23
14	–0.16
9	–0.23
5	–0.23
4	–0.16
10	–0.10
13	–0.10
13	0.04
14	0.13
12	0.19
13	0.25
9	0.23
8	0.15
18	0.23
22	0.31
22	0.18
23	0.17
31	0.39

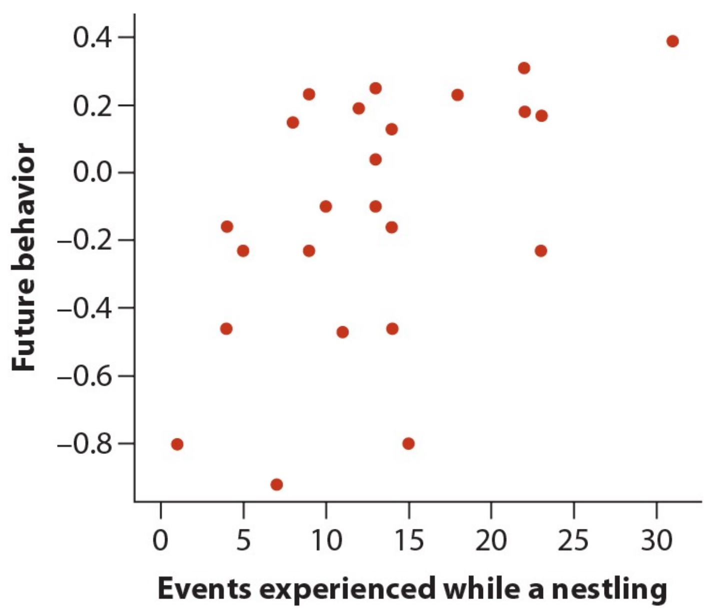


Fig 16.1-4

10

Correlation coefficient

$$r = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2}\sqrt{\sum_i (Y_i - \overline{Y})^2}}$$

$$\sum_i (X_i - \overline{X})(Y_i - \overline{Y}) = 33.086$$

$$\sum_i (X_i - \overline{X})^2 = 1194.625$$

$$\sum_i (Y_i - \overline{Y})^2 = 3.217$$

$$r = \frac{33.086}{\sqrt{1194.625}\sqrt{3.217}} = 0.534$$

11

Standard error of correlation coefficient

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - (0.534)^2}{24 - 2}} = 0.180$$

- But the sampling distribution of r is not normally distributed, so SE_r is not used in calculating the 95% CI

12

Approx. confidence interval of correlation coefficient

- Bit complicated…
- Involves conversion of r that includes natural log, and then back conversion

$$0.166 < \rho < 0.771$$

Fairly broad range,
but does not include zero

Consistent with positive linear correlation between antagonistic events during development and future aggression

13

14

Hypothesis testing for correlation

- H_0: $\rho = 0$
- H_A: $\rho \neq 0$

15

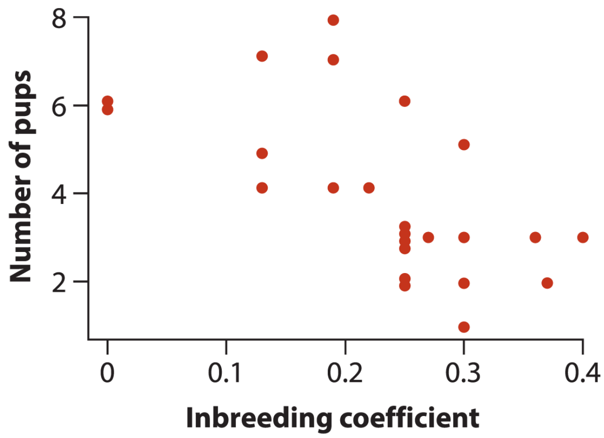
Ex 16.2: What big inbreeding coefficients you have

- Wolves wiped out from Norway and Sweden by 1970, but then area colonized by two individuals in 1980
- ~100 wolves by 2002
- With small number of founders, expect high inbreeding
- Do pairs with high inbreeding coefficients have fewer surviving pups?

16

TABLE 16.2-1 Inbreeding coefficients of litters of mated wolf pairs and the number of pups surviving their first winter. $n = 24$ litters.

Inbreeding coefficient	Number of pups	Inbreeding coefficient	Number of pups
0.00	6	0.25	3
0.00	6	0.25	2
0.13	7	0.25	2
0.13	5	0.25	6
0.13	4	0.27	3
0.19	8	0.30	5
0.19	7	0.30	3
0.19	4	0.30	2
0.22	4	0.30	1
0.25	3	0.36	3
0.25	3	0.37	2
0.25	3	0.40	3



Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

Fig 16.2-1

17

Correlation coefficient

$$r = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2}\sqrt{\sum_i (Y_i - \bar{Y})^2}}$$

$$\sum_i (X_i - \bar{X})(Y_i - \bar{Y}) = -2.690$$

$$\sum_i (X_i - \bar{X})^2 = 0.230$$

$$\sum_i (Y_i - \bar{Y})^2 = 80.958$$

$$r = \frac{-2.690}{\sqrt{0.230}\sqrt{80.958}} = -0.623$$

18

Hypothesis testing for correlation

- H_0: no relationship between inbreeding coefficient and number of pups ($\rho = 0$)
- H_A: Inbreeding coefficients and the number of pups are correlated ($\rho \neq 0$)

19

Test statistic

- Based on Student's t-distribution

$$t = \frac{r}{SE_r} = \frac{-0.623}{0.167} = -3.74$$

$$df = n - 2 = 24 - 2 = 22$$

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - (-0.623)^2}{24 - 2}} = 0.167$$

20

P-value

- $t_{22} = -3.60$
- Calculate P-value with stats table or computer
- $P = 0.002$
- Reject the null hypothesis: there is a significant, negative correlation between inbreeding coefficient and number of surviving pups

21

22

Assumptions

- Random sample from the population
- Bivariate normal distribution
  - Bell-shaped in two dimensions rather than one

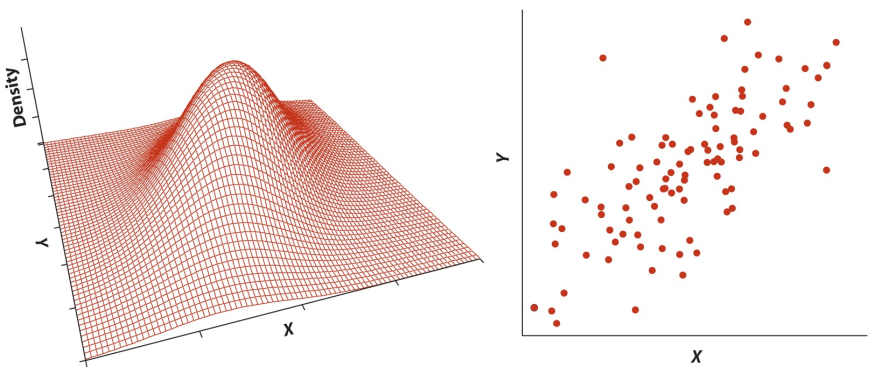


Fig 16.3-1

23

Deviations from bivariate normality

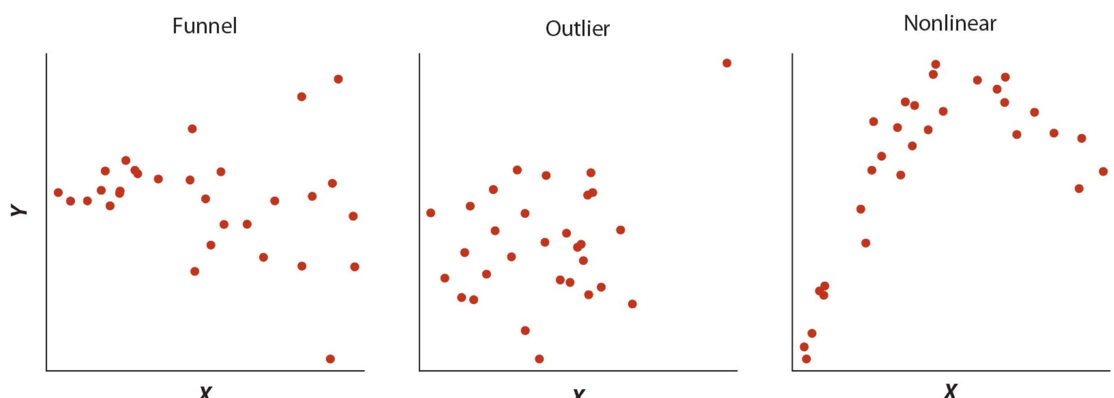


Fig 16.3-2

- Options?
  - Transform data (both variables same way)
  - Nonparametric test (Spearman's rank correlation); skipping

24

Correlation coefficient depends on range

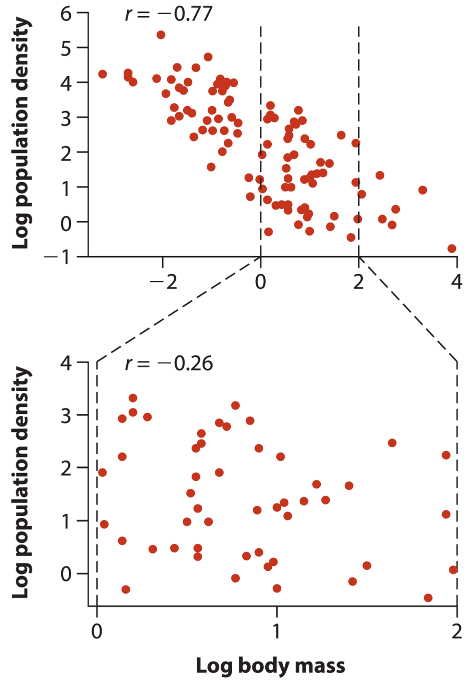
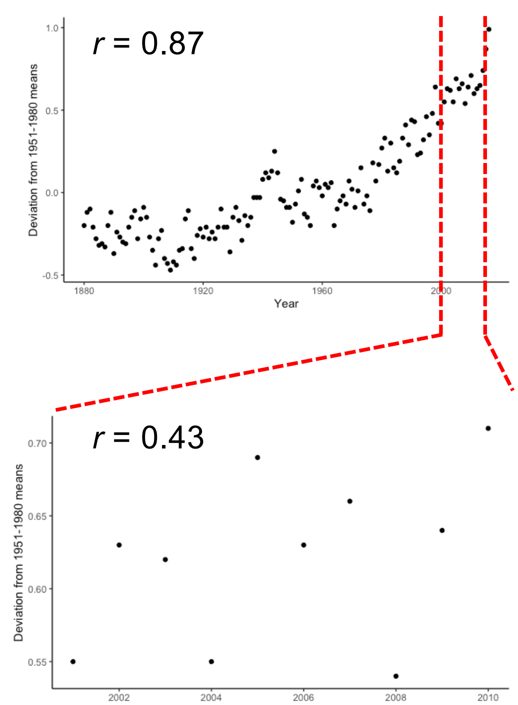


Fig 16.4-1

Whitlock & Schluter, *The Analysis of Biological Data*, 3e © 2020 W. H. Freeman and Company

25

Correlation coefficient depends on range



26

Notes

- Skipping sections 16.5 and 16.6

27