Correlation pipeline and example

Testing for no correlation:

Step 1: declare null and alternate hypotheses

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

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

Testing for no correlation:

Step 1: declare null and alternate

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

H_A: Some correlation ($\rho\neq0$)

Step 2: test statistic

$$t = \frac{r - \rho}{SE_r}$$

Testing for no correlation:

Step 1: declare null and alternate

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

H_A: Some correlation ($\rho\neq0$)

Step 2: test statistic

$$t = \frac{r}{SE_r}$$

Step 3: P-value/Critical value

- Null distribution has a **sampling distribution** of Student's t-distribution with d.o.f. = *n - 2*
  - *Why n - 2? Use two summaries of data,* $\overline{X}$ *and* $\overline{Y}$
- Use table C to determine critical value

Testing for no correlation:

Step 1: declare null and alternate

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

H_A: Some correlation ($\rho\neq0$)

Step 2: test statistic

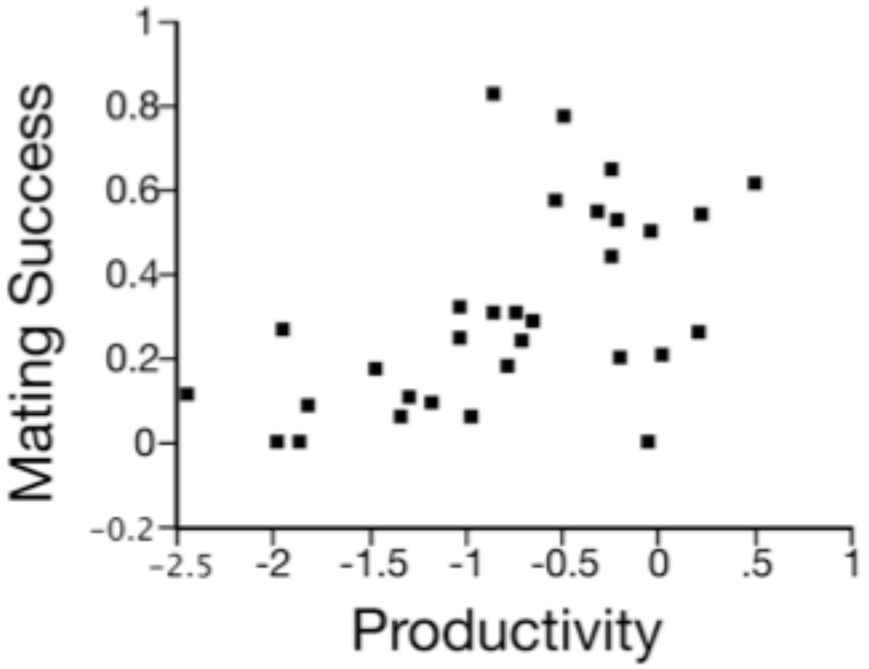
$$t = \frac{r}{SE_r}$$

Step 3: P-value/Critical value

Step 4: State conclusion and Confidence interval

- *Correlation?*
- *CI needs to use transformation since SE_r is not normally distributed*

- Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*



- Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*

H_0: Mating success and productivity are not related

H_A: Mating success and productivity are related

Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*

H_0: Mating success and productivity are not related

H_A: Mating success and productivity are related

X is productivity, Y is the mating success

$\sum X = -24.228$ $\sum Y = 9.498$

$\sum X^2 = 35.1808$ $\sum Y^2 = 4.5391$

$\sum XY = -4.62741$ $n = 31$

Shortcuts

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \left(\sum X_i Y_i\right) - \frac{\sum X_i \sum Y_i}{n}$$

$$\sum (X - \overline{X})^2 = \sum (X_i^2) - \frac{(\sum X_i)^2}{n}$$

$$\sum (Y - \overline{Y})^2 = \sum (Y_i^2) - \frac{(\sum Y_i)^2}{n}$$

Find r:

$$\sum (X_i - \overline{X})(Y_i - \overline{Y}) = \left(\sum X_i Y_i\right) - \frac{\sum X_i \sum Y_i}{n}$$

= 2.796

$$\sum (X - \overline{X})^2 = \sum (X_i^2) - \frac{(\sum X_i)^2}{n}$$

= 16.245

$$\sum (Y - \overline{Y})^2 = \sum (Y_i^2) - \frac{(\sum Y_i)^2}{n}$$

= 1.6289

$$r = \frac{2.796}{\sqrt{(16.245)(1.6289)}} = 0.535$$

$$SE_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{0.7045}{29}} = 0.1558$$

$$t = \frac{0.5435}{01558} = 3.49$$

Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*

H_0: Mating success and productivity are not related

H_A: Mating success and productivity are related

t = 3.49

df = 29

This is greater than $t_{0.05(2),\ 29} = 2.045$, so we can reject the null hypothesis and say that productivity and male mating success are correlated

Example: Are the effects of new mutations on mating success and productivity correlated? Data from various visible mutations in *Drosophila melanogaster*

H_0: Mating success and productivity are not related

H_A: Mating success and productivity are related

t = 3.49

df = 29

This is greater than $t_{0.05(2),\ 29} = 2.045$, so we can reject the null hypothesis and say that productivity and male mating success are correlated ($\rho \neq 0$).

The 95% confidence interval for this parameter is:

$0.22 < \rho < 0.747$

* practice at home since it is not normally distributed and you need to convert it!*