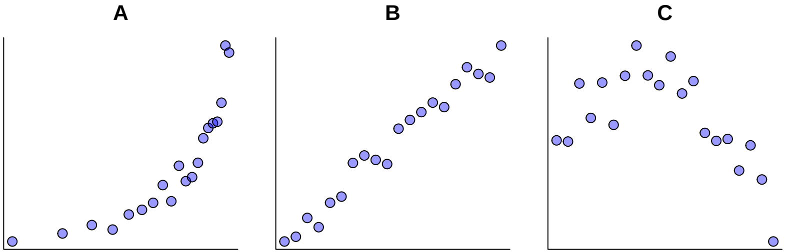
Correlations

Introduction to Quantitative Ecology
Fall 2018
Chris Sutherland
csutherland@umass.edu

Group evaluations

1. We would *not* use the Spearman's rank test to calculate a correlation coefficient - which one?



Group evaluations

2. For her PhD, Eugene is studying the effects of annual temperature on American robin reproductive success (number of eggs hatched). What is the *dependent* variable?

A) American robin

B) Eugene

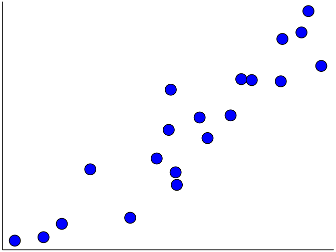
C) Temperature

D) Number of eggs

Group evaluations

3. What is the most likely Pearson's correlation coefficient (r) for this relationship?

A) 0.90
B) 0.09
C) -0.9
D) -0.09



Group evaluations

4. What is the slope in this equation of a straight line?

$$y = mx + c$$

A) y
B) m
C) x
D) c

Group evaluations

5. Which of the following is *polynomial* relationship?

   A) $y = ax + c$
   B) $y = ax + bx^2 + c$
   C) $y = log(ax) + c$
   D) $y = c$

Correlations

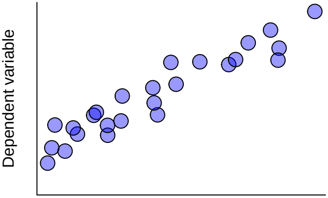
- What are examples of correlations?
- Why would we be interested in correlations?

?

Correlations

Interested in the relationship between two samples

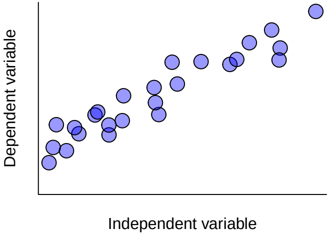
- Dependent variable:
  - data we are interested in explaining
  - Y-axis
  - *response* variable



Correlations

Interested in the relationship between two samples

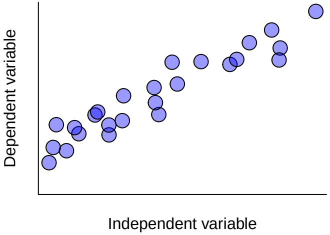
- Dependent variable:
  - data we are interested in explaining
  - Y-axis
  - *response* variable
- Independent variable:
  - data used to describe variation in dependent variable
  - X-axis
  - *explanatory* variable



Correlations

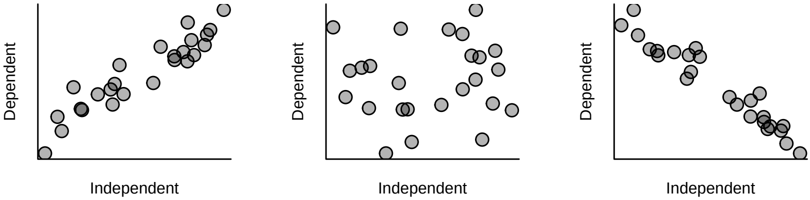
Interested in the relationship between two samples

- Dependent variable:
  - data we are interested in explaining
  - Y-axis
  - *response* variable
- Independent variable:
  - data used to describe variation in dependent variable
  - X-axis
  - *explanatory* variable
- Dealing with *pairs* of values!



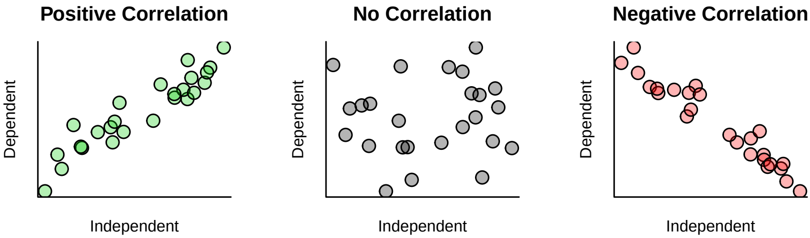
Correlations

What is the sign of the correlation?



Correlations

What is the sign of the correlation?



Correlations

Lets play a game!

http://www.istics.net/Correlations/

Correlations

We need a way to

- quantify correlations/relationships
- assess whether correlations/relationships are *significant*

Correlations

We need a *test*!

1. Spearman's rank test (r_s)
   - determine the strength of the link between 2 samples
   - data are *not* normally distributed
   - realtionship not linear
     - but still exhibits a positive of negative trend
     - is not u-shaped or n-shaped
   - use the *ranks* of values
   - correlation strenght ranges from -1 to 1
     - -1: perfect *negative* correlation
     - 1: perfect *positive* correlation
     - 0: no correlation

Correlations

We need a *test*!

2. Pearson's product moment (r)
    - determine the strength of the link between 2 samples
    - data *are* normally distributed
    - relationship assumed to be linear
        - positive of negative trend
        - not u- or n-shaped
    - use actual values
    - correlation strenght ranges from -1 to 1
        - -1: perfect *negative* correlation
        - 1: perfect *positive* correlation
        - 0: no correlation

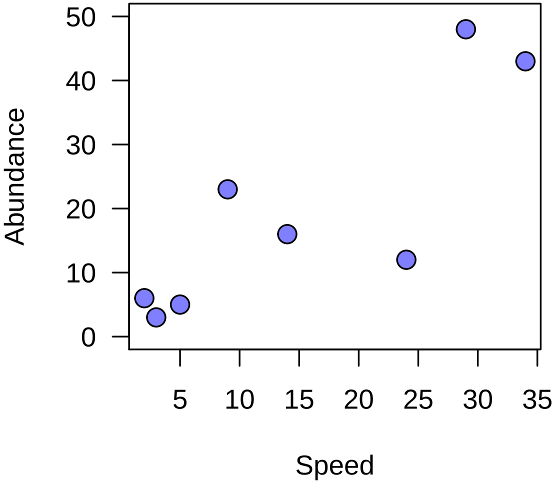
Correlations

1. Spearman's rank test (r_s)
   - determine the strength of the link between 2 samples
   - data are *not* normally distributed
   - not neccesarily linear
   - use the *ranks* of values
2. Pearson's product moment (r)
   - determine the strength of the link between 2 samples
   - data *are* normally distributed
   - assumed to be linear linear
   - use actual values

- For both, the *correlation coefficient* ranges from -1 to 1
  - -1: perfect *negative* correlation
  - 1: perfect *positive* correlation
  - 0: no correlation

Example: the Mayfly data

```
head(mayfly)
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
```



Spearman's rank test (r_s) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

Spearman's rank test (r_s) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

Spearman's rank test (r_s) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

- H_1 (the *alternative* hypothesis):

Spearman's rank test (r_s) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

- H_1 (the *alternative* hypothesis):

There is a positive correlation between stream speed and mayfly abundance

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- programs (`Excel` and `R`) will do the math for us
- <u>BUT</u> we should be aware of what's going on!

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- programs (`Excel` and `R`) will do the math for us
- <u>BUT</u> we should be aware of what's going on!

```
mayfly
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
7    29        48
8    34        43
```

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- first, calculate the *ranks* of the values: speed

```
mayfly
  Speed Abundance Speed.rank
1     2         6          1
2     3         3          2
3     5         5          3
4     9        23          4
5    14        16          5
6    24        12          6
7    29        48          7
8    34        43          8
```

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- first, calculate the *ranks* of the values: abundance

```
mayfly
  Speed Abundance Speed.rank Abundance.rank
1     2         6          1              3
2     3         3          2              1
3     5         5          3              2
4     9        23          4              6
5    14        16          5              5
6    24        12          6              4
7    29        48          7              8
8    34        43          8              7
```

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- then, calculate the difference in the *ranks*: D

```
mayfly
  Speed Abundance Speed.rank Abundance.rank Diff
1     2         6          1              3    2
2     3         3          2              1   -1
3     5         5          3              2   -1
4     9        23          4              6    2
5    14        16          5              5    0
6    24        12          6              4   -2
7    29        48          7              8    1
8    34        43          8              7   -1
```

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- then, square the difference in the *ranks*: D^2

```
mayfly
  Speed Abundance Speed.rank Abundance.rank Diff Diff.sq
1     2         6          1              3    2       4
2     3         3          2              1   -1       1
3     5         5          3              2   -1       1
4     9        23          4              6    2       4
5    14        16          5              5    0       0
6    24        12          6              4   -2       4
7    29        48          7              8    1       1
8    34        43          8              7   -1       1
```

Spearman's rank test (r_s) - the statistical test

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- so now we have all the pieces!
  - n: number of pairs of observations (each has an x and a y)
  - D: difference in ranks between variables
  - D^2: the difference squared

Spearman's rank test (r_s) - in practice

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

```
mayfly
  Speed Abundance Speed.rank Abundance.rank Diff Diff.sq
1     2         6          1              3    2       4
2     3         3          2              1   -1       1
3     5         5          3              2   -1       1
4     9        23          4              6    2       4
5    14        16          5              5    0       0
6    24        12          6              4   -2       4
7    29        48          7              8    1       1
8    34        43          8              7   -1       1
```

Spearman's rank test (r_s) - in practice

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

$$r_s = 0.81$$

What does this tell us?

Spearman's rank test (r_s) - in practice

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

$$r_s = 0.81$$

Two useful pieces of information:

1. Sign of the correlation
   - positive value means *positive* correlation

Spearman's rank test (r_s) - in practice

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

$$r_s = 0.81$$

Two useful pieces of information:

1. Sign of the correlation
   - positive value means *positive* correlation
2. Strength of the correlation
   - r ranges from -1 to 1
   - $r_s = 0.81$ is *strongly* positive
   - but is this *significant*?

Spearman's rank test (r_s) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

Spearman's rank test (r_s) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

If the test statistic is <u>greater than</u> the critical value:

- *reject* the null hypothesis
- there *is* a statistically significant positive correlation

Spearman's rank test (r_s) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

If the test statistic is greater than the critical value:

- *reject* the null hypothesis
- there *is* a statistically significant positive correlation

If the test statistic is less than the critical value:

- *fail to reject* the null hypothesis
- there is no evidence of a statistically significant positive correlation

Spearman's rank test (r_s) - mayfly conclusions

- *significance level* (typically 5%)
- *sample size* (number of pairs = 8)
- *critical value* (0.738)
- *test stasistic* ($r_s = 0.81$)
- the *test stasistic* is ?? than the *critical value*

No. of pairs, *n*	Significance level 5%	2%	1%
5	1.000	1.000	-
6	0.886	0.943	1.000
7	0.786	0.893	0.929
8	0.738	0.833	0.881
9	0.683	0.783	0.833
10	0.648	0.746	0.794
12	0.591	0.712	0.777
14	0.544	0.645	0.715
16	0.506	0.601	0.665
18	0.475	0.564	0.625
20	0.450	0.534	0.591
22	0.428	0.508	0.562
24	0.409	0.485	0.537
26	0.392	0.465	0.515
28	0.377	0.448	0.496
30	0.364	0.432	0.478

Spearman's rank test (r_s) - mayfly conclusions

So, $r_s > 0.738$:

- reject the null hypothesis
- accept the alternative hypothesis

There *is* a statistically significant positive correlation between stream flow and mayfly abundance!

Pearson's product moment (r)

2. Pearson's product moment (r)

Pearson's product moment (r)

2. Pearson's product moment (r)
   - determine the strength of the link between 2 samples
   - data *are* normally distributed
   - relationship assumed to be linear
   - use actual values

Pearson's product moment (r)

2. Pearson's product moment (r)
   - determine the strength of the link between 2 samples
   - data *are* normally distributed
   - relationship assumed to be linear
   - use actual values
   - -1: perfect *negative* correlation
   - 1: perfect *positive* correlation
   - 0: no correlation

Pearson's product moment (r)

2. Pearson's product moment (r)
   - determine the strength of the link between 2 samples
   - data *are* normally distributed
   - relationship assumed to be linear
   - use actual values
   - -1: perfect *negative* correlation
   - 1: perfect *positive* correlation
   - 0: no correlation

Pearson's product moment (r)

The linear assumption:

- relationship between two variables can be described by the equation of a straight line:

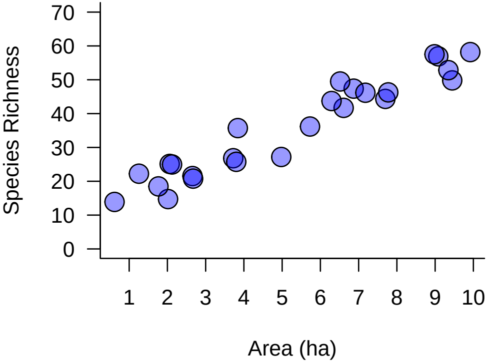
$$y = mx + c$$

- y: the dependent/response variable
- x: the independent/explanatory variable
- m: the *slope* of the relationship
- c: the *intercept*

Pearson's product moment (r) - an example

Relationship between area & species richness (by hand/eye):

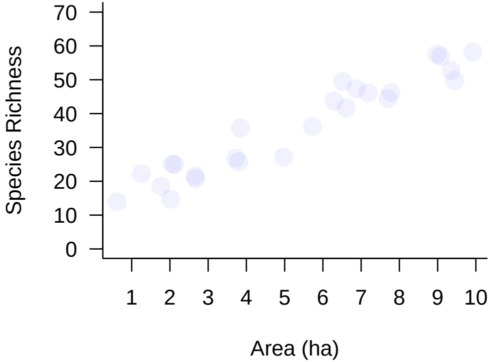
$$species = m \times area + c$$



Pearson's product moment (r) - an example

Relationship between area & species richness (by hand/eye):

$$species = m \times area + c$$



Pearson's product moment (r)

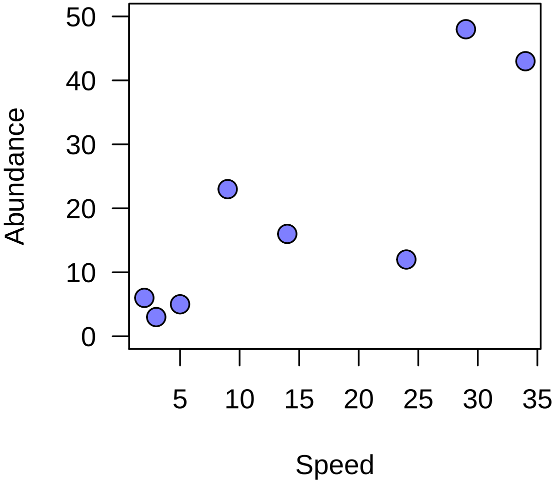
2. Pearson's product moment (r)

- calculates the strength of the correlation
  - similar to Spearman's coefficient
- *also* the equation of the best fit line
  - calculates the slope
  - calculates the intercepts

Pearson's product moment calculates strength of correlation & parameters of linear relationship between two variables.

Example: the Mayfly data

```r
head(mayfly)
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
```



Pearson's product moment (r) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

Pearson's product moment (r) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

Pearson's product moment (r) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

- H_1 (the *alternative* hypothesis):

Pearson's product moment (r) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

- H_1 (the *alternative* hypothesis):

There is a positive correlation between stream speed and mayfly abundance

Pearson's product moment (r) - the hypothesis

Before conducting any statistical test, we need to state the hypotheses!

- H_0 (the *null* hypothesis):

There is no correlation between stream speed and mayfly abundance

- H_1 (the *alternative* hypothesis):

There is a positive correlation between stream speed and mayfly abundance

NOTE: the hypotheses remain the same, but the assumption about the data are different. The assumptions determine the appropriate model!

Pearson's product moment (r) - the straight line

Equation of a straight line:

$$y = mx + c$$

Pearson's product moment (r) - the slope

Equation of a straight line:

$$y = mx + c$$

Calculating the slope:

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

Pearson's product moment (r) - the intercept

Equation of a straight line:

$$y = mx + c$$

Calculating the slope:

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

Calculating the intercept:

$$c = \bar{y} - m\bar{x}$$

Pearson's product moment (r) - the linear equation

Equation of a straight line:

$$y = mx + c$$

- programs (`Excel` and `R`) will do the math for us
- BUT we should be aware of what's going on!
  - calculating the *slope* from the data
  - calculating the *intercept* from the data

Pearson's product moment (r) - the statistical test

Equation of a straight line: $y = mx + c$

- but, is the linear relationship *significant*?
- we can calculate Pearson's correlation coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

Pearson's product moment (r) - the statistical test

Equation of a straight line: $y = mx + c$

- but, is the linear relationship *significant*?
- we can calculate Pearson's correlation coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

```
mayfly
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
7    29        48
8    34        43
```

Pearson's product moment (r) - the statistical test

Equation of a straight line: $y = mx + c$

- but, is the linear relationship *significant*?
- we can calculate Pearson's correlation coefficient

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$r = 0.84$$

Pearson's product moment (r) - in practice

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

$$r = 0.84$$

Again we have two useful pieces of information:

1. Sign of the correlation
   - positive value means *positive* correlation
2. Strength of the correlation
   - r ranges from -1 to 1
   - $r_s = 0.84$ is *strongly* positive
   - but is this *significant*?

Pearson's product moment (r) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

Pearson's product moment (r) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

If the test statistic is greater than the critical value:

- *reject* the null hypothesis
- there *is* a statistically significant positive correlation

Pearson's product moment (r) - conclusions

As with many statistical tests, determine significance based on:

- *significance level*
- *sample size*
- *critical value*

If the test statistic is <u>greater than</u> the critical value:

- *reject* the null hypothesis
- there *is* a statistically significant positive correlation

If the test statistic is <u>less than</u> the critical value:

- *fail to reject* the null hypothesis
- there is no evidence of a statistically significant positive correlation

Pearson's product moment (r) - mayfly conclusions

- *significance level* (typically 5%)
- *sample size* (df = number of pairs - 2 = 8 - 2 = 6)
- *critical value* (0.707)
- *test stasistic* ($r = 0.84$)
- the *test stasistic* is ?? than the *critical value*

Degrees of freedom	Significance 5%	Significance 1%
1	0.997	1
2	0.95	0.99
3	0.878	0.959
4	0.811	0.917
5	0.754	0.874
6	0.707	0.834
7	0.666	0.798
8	0.632	0.765
9	0.602	0.735
10	0.576	0.708
12	0.532	0.661
14	0.497	0.623
16	0.468	0.59
18	0.444	0.561
20	0.423	0.537
22	0.404	0.515
24	0.388	0.496
26	0.374	0.478
28	0.361	0.463
[illegible]	[illegible]	0.449

Pearson's product moment (r) - mayfly conclusions

So, $r > 0.707$:

- reject the null hypothesis
- accept the alternative hypothesis

There *is* a statistically significant positive correlation between stream flow and mayfly abundance!

Mayfly conclusions

Spearman's rank correlation coefficient:

- $r_s > 0.738$:
- reject the null hypothesis
- accept the alternative hypothesis

Pearson's correlation coefficient:

- $r > 0.707$:
- reject the null hypothesis
- accept the alternative hypothesis

Conclusion:

- same regardless of assumptions!
- there *is* a statistically significant positive correlation between stream flow and mayfly abundance!

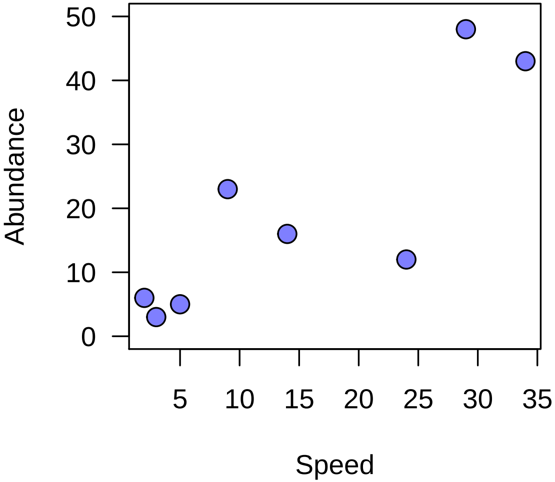
Correlations in practice

- we *could* do it by hand (UGH!)
- we *prefer* to use built in functions
  - in `Excel`
  - in `R`

Example: the Mayfly data

```
head(mayfly)
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
```

Spearman: $r_s = 0.81$
Pearson: $r = 0.84$



Spearman's rank test in Excel

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

In Excel

- an automated routine doe snot exist
- have to do it by hand
- :(

Pearson's product moment in `Excel`

Pearson's correlation coefficient:

- $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$
- `CORREL(response, predictor)`

Slope of the best fit line:

- $m = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2}$
- `SLOPE(response, predictor)`

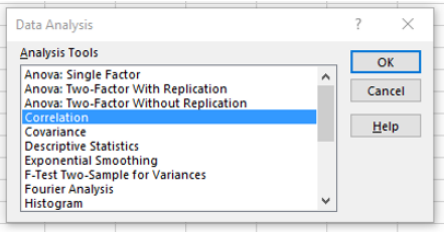
Intercept of the best fit line:

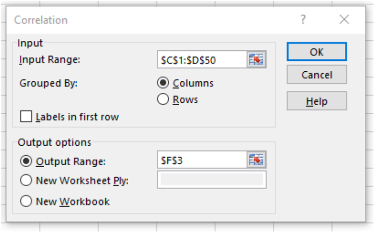
- $c = \bar{y} - m\bar{x}$
- `INTERCEPT(response, predictor)`

Pearson's product moment in `Excel`

Pearson's correlation coefficient:

- $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$
- Analysis toolpack - `Correlation`





Spearman's rank test in R

```
#make the mayfly 'data.frame' by hand:
mayfly <- data.frame(Speed = c(2,3,5,9,14,24,29,34),
                     Abundance = c(6,3,5,23,16,12,48,43))
```

Spearman's rank test in R

```
#make the mayfly 'data.frame' by hand:
mayfly <- data.frame(Speed = c(2,3,5,9,14,24,29,34),
                     Abundance = c(6,3,5,23,16,12,48,43))
mayfly
  Speed Abundance
1     2         6
2     3         3
3     5         5
4     9        23
5    14        16
6    24        12
7    29        48
8    34        43
```

Spearman's rank test in R

$$r_s = 1 - \frac{6 \times \sum D^2}{n(n^2 - 1)}$$

- Spearman correlation coefficient (r_s)

```
cor(var1, var2, method='spearman')
```

- Spearman correlation coefficient *and* significance test

```
cor.test(var1, var2, method='spearman')
```

Spearman's rank test in R

- Spearman correlation coefficient (r_s)

```
cor(mayfly$Speed, mayfly$Abundance, method='spearman')
[1] 0.8095238
```

Spearman's rank test in R

- Spearman correlation coefficient (r_s)

```
cor(mayfly$Speed, mayfly$Abundance, method='spearman')
[1] 0.8095238
```

- Spearman correlation coefficient *and* significance test

```
cor.test(mayfly$Speed, mayfly$Abundance, method='spearman')

    Spearman's rank correlation rho

data:  mayfly$Speed and mayfly$Abundance
S = 16, p-value = 0.02178
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8095238
```

Pearson's product moment in R

- $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$
- Pearson's correlation coefficient (r)

```
cor(var1, var2, method='pearson')
```

- Pearson's correlation coefficient *and* significance test

```
cor.test(var1, var2, method='pearson')
```

Pearson's product moment in R

- Pearson's correlation coefficient (r)

```r
cor(mayfly$Speed, mayfly$Abundance, method='pearson')
```

Pearson's product moment in R

- Pearson's correlation coefficient (r)

```
cor(mayfly$Speed, mayfly$Abundance, method='pearson')
[1] 0.8441408
```

Pearson's product moment in R

- Pearson's correlation coefficient (r)

```
cor(mayfly$Speed, mayfly$Abundance, method='pearson')
[1] 0.8441408
```

- Pearson's correlation coefficient *and* significance test

```
cor.test(mayfly$Speed, mayfly$Abundance, method='pearson')
```

Pearson's product moment in R

- Pearson's correlation coefficient (r)

```
cor(mayfly$Speed, mayfly$Abundance, method='pearson')
[1] 0.8441408
```

- Pearson's correlation coefficient *and* significance test

```
cor.test(mayfly$Speed, mayfly$Abundance, method='pearson')

    Pearson's product-moment correlation

data:  mayfly$Speed and mayfly$Abundance
t = 3.8568, df = 6, p-value = 0.008393
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3442317 0.9711386
sample estimates:
      cor
0.8441408
```

Linear equation in `R`

To calculate the slope and the intercept of the best fit line:

- use a *linear model*
- in `R` use the function `lm(respnse ~ predictor)`

Linear equation in R

To calculate the slope and the intercept of the best fit line:

- use a *linear model*
- in R use the function `lm(respnse ~ predictor)`

```
lm(mayfly$Abundance ~ mayfly$Speed)

Call:
lm(formula = mayfly$Abundance ~ mayfly$Speed)

Coefficients:
 (Intercept)  mayfly$Speed
       1.867         1.176
```

Linear equation in R

To calculate the slope and the intercept of the best fit line:

- use a *linear model*
- in R use the function `lm(respnse ~ predictor)`

```
summary(lm(mayfly$Abundance ~ mayfly$Speed))

Call:
lm(formula = mayfly$Abundance ~ mayfly$Speed)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.080  -2.481  -0.580   3.975  12.042 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)   
(Intercept)    1.8667     5.7912   0.322  0.75813   
mayfly$Speed   1.1756     0.3048   3.857  0.00839 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.05 on 6 degrees of freedom
Multiple R-squared:  0.7126,	Adjusted R-squared:  0.6647 
F-statistic: 14.87 on 1 and 6 DF,  p-value: 0.008393
```

Its all related!

```
cor.test(mayfly$Speed, mayfly$Abundance, method='pearson')

    Pearson's product-moment correlation

data:  mayfly$Speed and mayfly$Abundance
t = 3.8568, df = 6, p-value = 0.008393
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.3442317 0.9711386
sample estimates:
      cor
0.8441408
```

```
summary(lm(mayfly$Abundance ~ mayfly$Speed))

Call:
lm(formula = mayfly$Abundance ~ mayfly$Speed)

Residuals:
    Min      1Q  Median      3Q     Max
-18.080  -2.481  -0.580   3.975  12.042

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.8667     5.7912   0.322  0.75813
mayfly$Speed   1.1756     0.3048   3.857  0.00839 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.05 on 6 degrees of freedom
Multiple R-squared:  0.7126,    Adjusted R-squared:  0.6647
F-statistic: 14.87 on 1 and 6 DF,  p-value: 0.008393
```

Its all related!

```
cor.test(mayfly$Speed, mayfly$Abundance, method='spearman')

    Spearman's rank correlation rho

data:  mayfly$Speed and mayfly$Abundance
S = 16, p-value = 0.02178
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8095238
```

```
summary(lm(mayfly$Abundance ~ mayfly$Speed))

Call:
lm(formula = mayfly$Abundance ~ mayfly$Speed)

Residuals:
    Min      1Q  Median      3Q     Max
-18.080  -2.481  -0.580   3.975  12.042

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.8667     5.7912   0.322  0.75813
mayfly$Speed   1.1756     0.3048   3.857  0.00839 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.05 on 6 degrees of freedom
Multiple R-squared:  0.7126,    Adjusted R-squared:  0.6647
F-statistic: 14.87 on 1 and 6 DF,  p-value: 0.008393
```

Group exercise: salamander length correlations

Correlations in `R`:

1. Read in the data
2. Make a plot of 'total length' in response to 'SVL' length
3. Decide whether to use the Pearson or Spearman test

Group exercise: salamander length correlations

Correlations in `R`:

- Read in the data
- Make a plot of ‘total length’ in response to ‘SVL’ length
- Decide whether to use the Pearson or Spearman test
- Conduct the analysis:
  - what is the correlation coefficient?
  - what is the intercept of the relationship?
  - what is the slope of the relationship?
  - is the relationship significant?

Group exercise: salamander length correlations

Correlations in `Excel`:

- Create an Excel worksheet that show the same values you got in `R`:
  - the correlation coefficient
  - the intercept of the relationship
  - the slope of the relationship
  - the *p*-value

Group exercise: salamander length correlations

Correlations in `Excel`:

- Create an Excel worksheet that show the same values you got in `R`:
  - the correlation coefficient
  - the intercept of the relationship
  - the slope of the relationship
  - the *p*-value (see below)
- **Submit the workbook**

To compute the *p*-value, you need to use `T.DIST.2T()` in Excel. This needs t, *test statistic*, which you can calculate that using r (the correlation coefficient):

$$t = \frac{r \times \sqrt{n-2}}{\sqrt{1-r^2}}$$