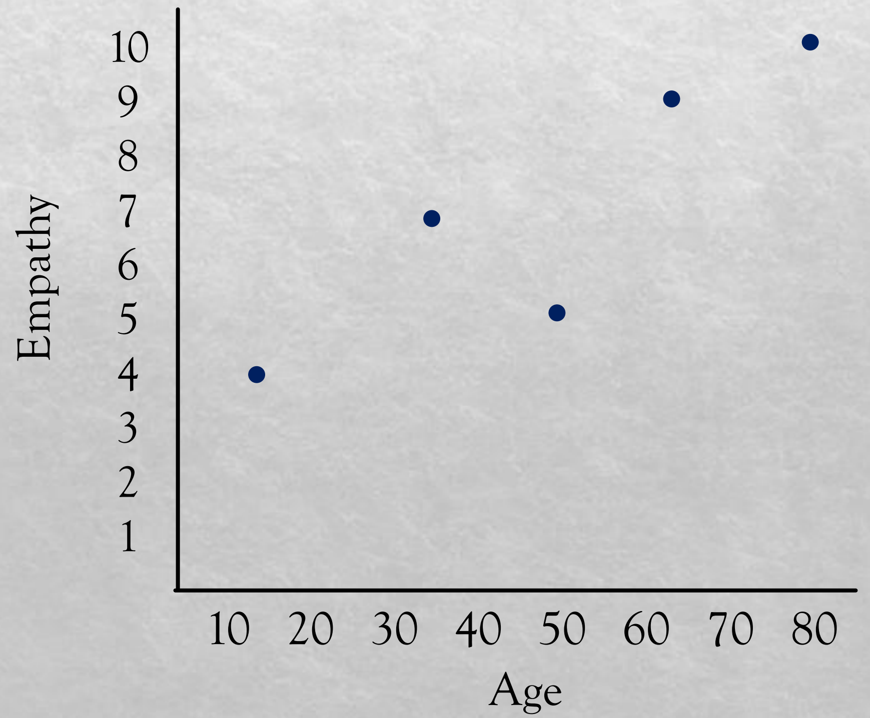
Examining Relationships Between Continuous Variables

PSY 612

Visualizing Relationships

- The simplest way of gauging whether a relationship exists between two continuous variables is by graphing the nature of their relationship.
  - For example, is there a relationship between how old people are and how much empathy they feel toward others?

Participant	Age	Empathy
1	50	5
2	15	4
3	35	7
4	80	10
5	65	9



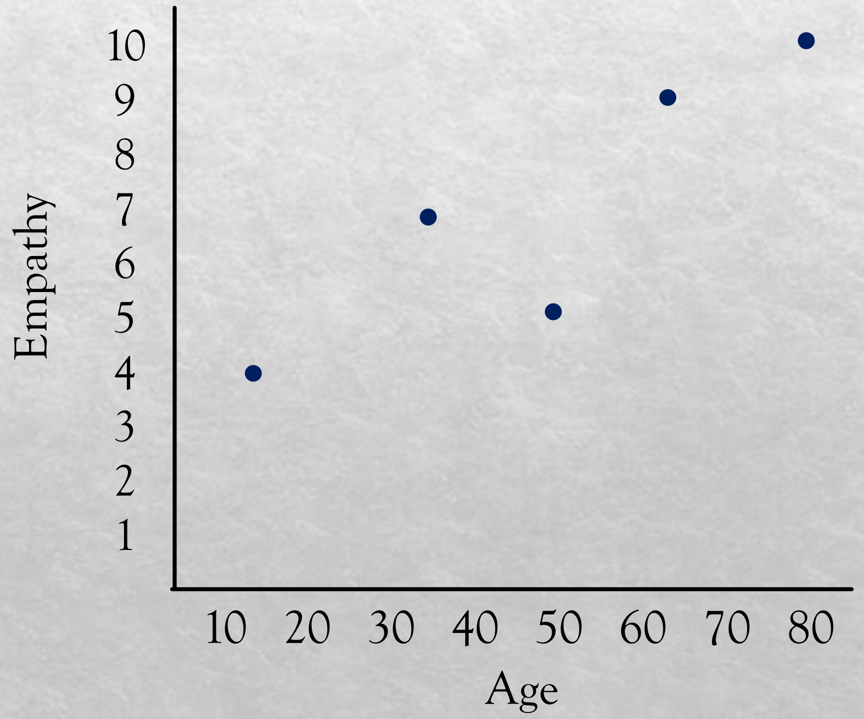
Quantifying Relationships

- **Covariance**
  - Is how people deviate from the mean of age (X) related to how people deviate from the mean of empathy (Y)?
  - Do the two variables ***covary?***

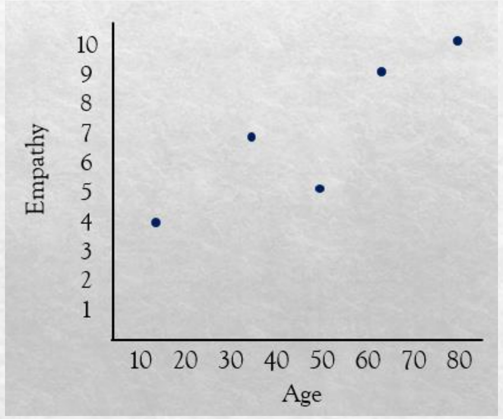
- *Sample:* $cov_{XY} = \dfrac{\Sigma(x-\overline{x})(y-\overline{y})}{n-1} = \dfrac{SP}{n-1}$

- *Population:* $cov_{XY} = \dfrac{\Sigma(x-\overline{x})(y-\overline{y})}{N} = \dfrac{SP}{N}$

- *Where n (or N) is the number of* **pairs** of scores



Calculating Covariance



Participant	Age	Empathy	Age - M_{Age}	Empathy - M_{Emp}	(Age – M_{Age})*(Empathy – M_{Emp})
1	50	5	1	-2	1*-2 = -2
2	15	4	-34	-3	-34*-3 = 102
3	35	7	-14	0	-14*0 = 0
4	80	10	31	3	31*3 = 93
5	65	9	16	2	16*2 = 32
	M_{Age} = 49	M_{Emp} = 7			SP = 225

$$cov_{XY} = \frac{\Sigma(x-\overline{x})(y-\overline{y})}{n-1} = \frac{SP}{n-1} = \frac{225}{5-1} = 56.25$$

Question: What are the units of this measure of covariance?

Interpreting Covariance

- Covariance is expressed in the units of the original variables
  - How can you gauge the strength of the relationship in this way? What is a high, or a low, covariance in years-empathy units?
  - Difficult to compare covariances across studies because not in standardized units

- Correlations are **standardized covariances** and thus are easier to compare across studies and to make judgments about the strength of a relationship.
  - Covariances are still important to know because they are used as the basis for more advanced analyses (especially covariance matrices)

Correlations

A **correlation** is a quantitative measure of the linear relationship between two continuous variables.

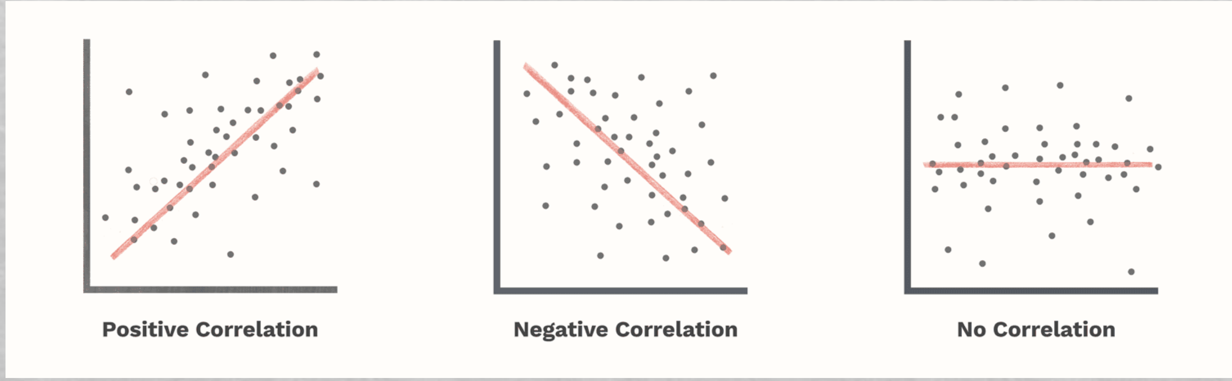
The formula for Pearson's correlation coefficient (*r*)

$$r = \frac{\Sigma(\mathrm{X} - M_x)(\mathrm{Y} - M_Y)}{\sqrt{\Sigma(\mathrm{X} - M_x)^2 \Sigma(\mathrm{Y} - M_Y)^2}} = \frac{SP}{\sqrt{(SS_{\mathrm{X}})(SS_{\mathrm{Y}})}} = \frac{\text{degree to which X and Y covary}}{\text{degree to which X and Y vary separately}}$$

The denominator makes *r* unitless and stay within the range of -1 to +1

Interpreting Correlations

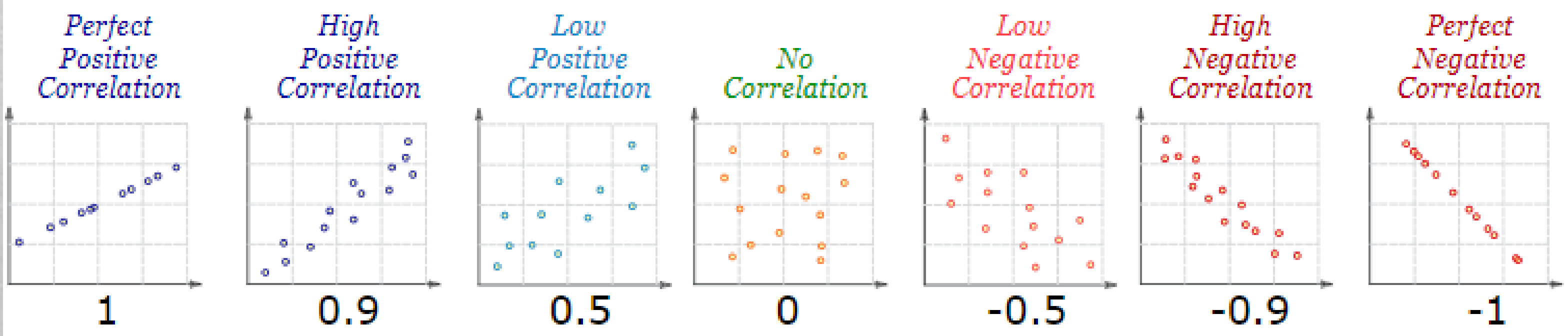
- **Direction:** The sign of the correlation indicates the direction of the relationship
  - Positive (+): as values tend to increase (or decrease) on the X variable, they also tend to increase (or decrease) on the Y variable
  - Negative (-): as values tend to increase on the X variable, they tend to decrease on the Y variable
  - None: There is no systematic relationship between the two variables



Interpreting Correlations

- **Strength**: the numerical value of the correlation indicates the strength of the linear relationship between the variables

- The closer the points are to lying perfectly on the pattern of a straight line, the stronger the relationship
  - The strength of the relationship increases as the correlation approaches ±1
  - The strength decreases as the correlation approaches 0



Calculating Correlations

Question: How would you describe the direction and strength of this correlation?

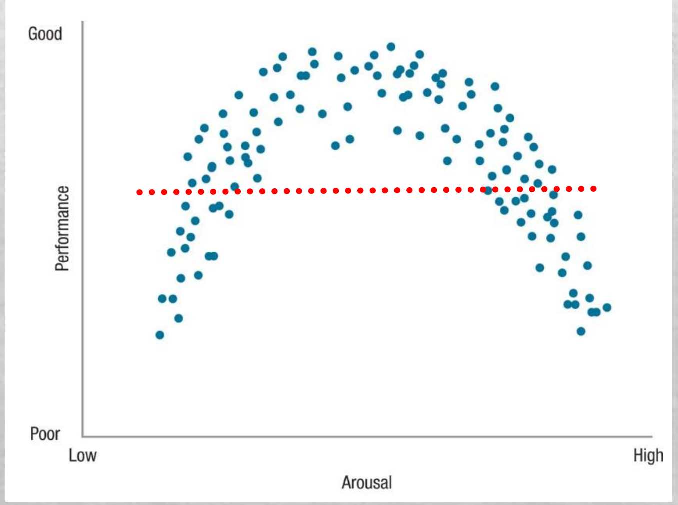
$$r = \frac{\Sigma(X - M_x)(Y - M_Y)}{\sqrt{\Sigma(X - M_x)^2 \Sigma(Y - M_Y)^2}} = \frac{SP}{\sqrt{(SS_X)(SS_Y)}} = \frac{225}{\sqrt{(2570)(26)}} = 0.87$$

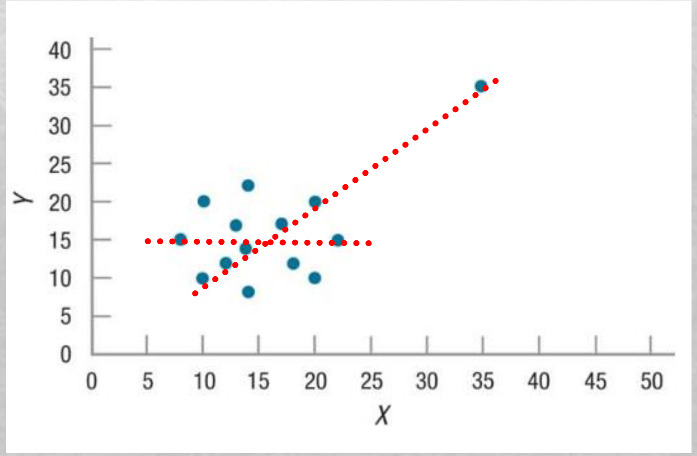
Age	Empathy	Age - M_{Age}	Empathy - M_{Emp}	(Age – M_{Age}) * (Empathy – M_{Emp})	(Age – M_{age})^2	(Empathy – M_{Emp})^2
50	5	1	-2	-2	1	4
15	4	-34	-3	102	1156	9
35	7	-14	0	0	196	0
80	10	31	3	93	961	9
65	9	16	2	32	256	4
M_{Age} = 49	M_{Emp} = 7			SP = 225	SS_X = 2570	SS_Y = 26

Potential issues with correlations

1. **Nonlinearity**: When there is a nonlinear relationship, using Pearson's correlation will give misleading results

2. **Outliers**: outliers in the data can severely affect the measure of the correlation between two variables

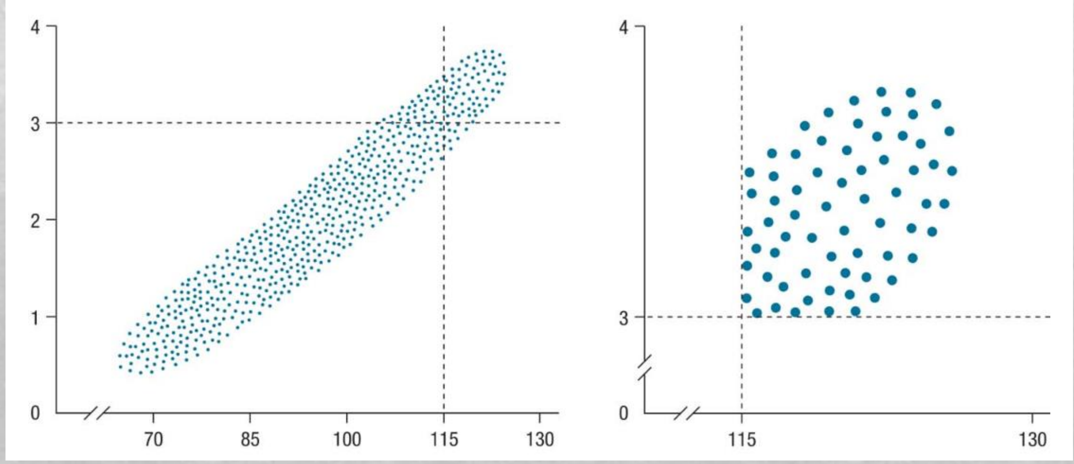




Potential issues with correlations

3. **Restriction of range**: when the range of the variable on the x-axis or y-axis is restricted, you may not see a relationship that exists across a wider range of values on the X or Y variable.

Test Score



Average Minutes Spent Studying per Exam

Testing the Significance of a Correlation

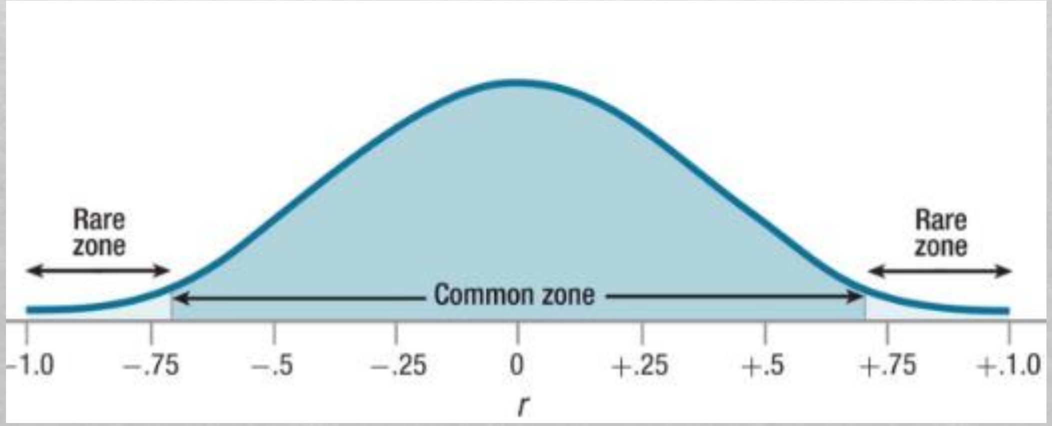
- Step 1: Visualize the relationship between the two variables using a scatterplot
  - Pearson's correlation *assumes* that the nature of the relationship between the variables is linear
  - Other assumptions of the test include: 1) The X and Y variables are normally distributed, 2) participants are independent, 3) the variability in scores on Y is approximately the same across the range of X

- Step 2: State the null hypothesis
  - H_0: $\rho = 0$
  - H_1: $\rho \neq 0$

Testing the Significance of a Correlation

- Step 3: Specify alpha (your willingness to make a Type I error) *prior to* inspecting the data

- Step 4: Construct a sampling distribution representing the results one would expect to obtain *if the null hypothesis is true*



Testing the Significance of a Correlation

- Step 5: Calculate the correlation between your two variables

$$r = \frac{\Sigma(\mathrm{X}-M_x)(\mathrm{Y}-M_Y)}{\sqrt{\Sigma(\mathrm{X}-M_x)^2\Sigma(\mathrm{Y}-M_Y)^2}} = \frac{SP}{\sqrt{(SS_{\mathrm{X}})(SS_{\mathrm{Y}})}} = \frac{225}{\sqrt{(2570)(26)}} = 0.87$$

- Step 6: Determine whether the correlation is significant or non-significant by either
  - Comparing the correlation to an *r*-critical value based on $n - 2$ degrees of freedom
    - n = number of pairs of scores
  - Or using R to obtain a *p*-value

Testing the Significance of a Correlation

- Step 7: Calculate the effect size, r^2
  - The proportion of variability in Y that is related to X

- For our example:
  - $r^2 = (0.87)^2 = .7569$
  - Approximately 76% of the variation in empathy scores is associated with empathy's relationship to age

Conventions for r^2
Small = 0.02
Medium = 0.13
Large = 0.26

Covariance & Correlations in R

- Covariance in R

```
age <- c(50,15,35,80,65)
empathy <- c(5,4,7,10,9)

cov(age,empathy)
```

```
> cov(age,empathy)
[1] 56.25
```

If you have more than two continuous variables, you can produce a **covariance (or correlation) matrix** (you'll discuss this in lab this week!)

Covariance & Correlations in R

- Correlation in R

```
age <- c(50,15,35,80,65)
empathy <- c(5,4,7,10,9)

cor(age,empathy)
```

```
> cor(age,empathy)
[1] 0.8704208
```

Question: What does the confidence interval mean?

- Testing the significance of a correlation

```
corr.test(age,empathy)
```

```
Call:corr.test(x = age, y = empathy)
Correlation matrix
[1] 0.87
Sample Size
[1] 5
These are the unadjusted probability values.
  The probability values  adjusted for multiple tests are in the p.adj object.
[1] 0.05
```

```
corr_age_emp <- corr.test(age,empathy)
corr_age_emp$p
```

```
> corr_age_emp$p
[1] 0.05489211
```

```
> corr_age_emp$ci
              lower         r     upper          p
NA-NA -0.05104602 0.8704208 0.9913709 0.05489211
```

```
> cor(age,empathy)^2
[1] 0.7576324
```

APA-Style Reporting

- Although the correlation between age and empathy was strong and positive, it was non-significant, $r(3) = 0.87$, $p = .055$, 95%CI[-0.05, 0.99], $r^2 = 0.76$.