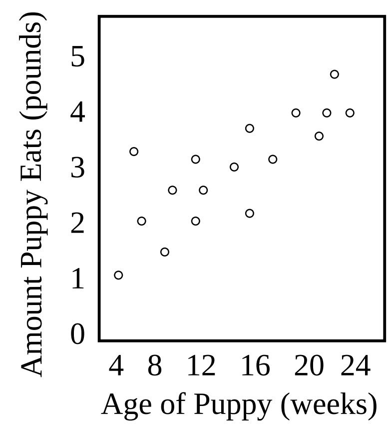
Pearson's r

- Scatterplots for 2 Quantitative Variables
- Research and Null Hypotheses for r
- Casual Interpretation of Correlation Results
  (& why/why not)
- Computational stuff for hand calculations

Displaying the data for a correlation:
With two quantitative variables we can display the bivariate relationship using a "scatterplot"

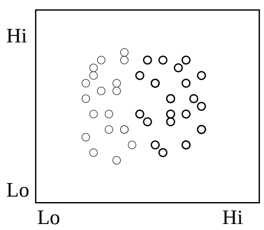
Puppy	Age (x)	Eats (y)
Sam	8	2
Ding	20	4
Ralf	12	2
Pit	4	1
Seff	24	4
…	..	..
Toby	16	3



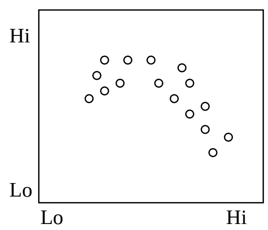
When examining a scatterplot, we look for three things...

- relationship
  - no relationship
  - linear
  - non-linear
- direction (if linear)
  - positive
  - negative
- strength
  - strong
  - moderate
  - weak

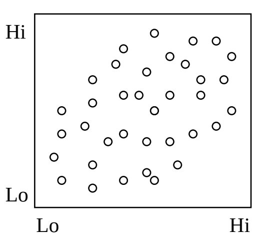
No relationship



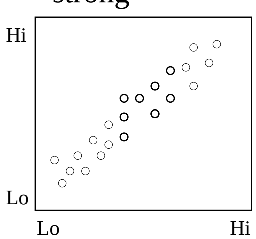
nonlinear, strong



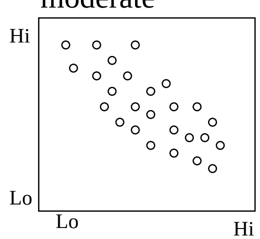
linear, positive, weak



linear, positive, strong

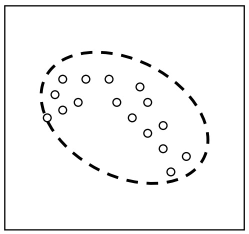


linear, negative, moderate

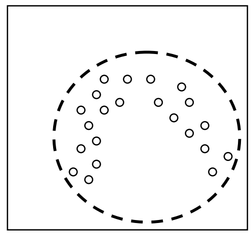


The Pearson's correlation (r) summarizes the direction and strength of the linear relationship shown in the scatterplot

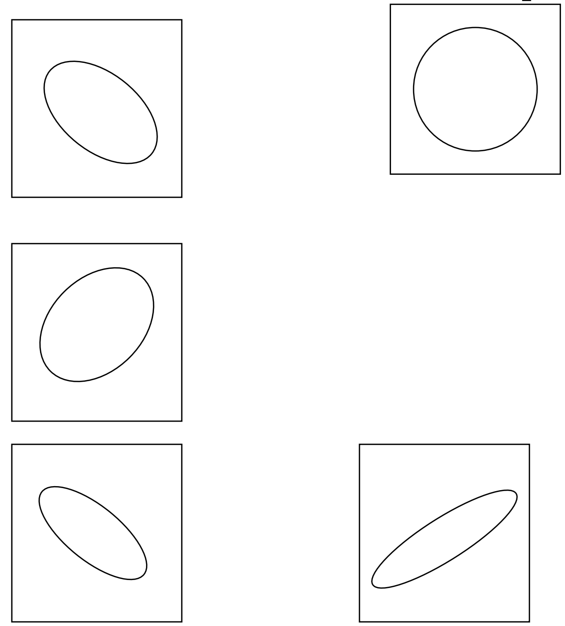
- r has a range from -1.00 to 1.00
  - 1.00 a perfect positive linear relationship
  - 0.00 no linear relationship at all
  - -1.00 a perfect negative linear relationship
- r assumes that the relationship is linear
  - if the relationship is not linear, then the r-value is an underestimate of the strength of the relationship at best and meaningless at worst



For a non-linear relationship, r will be based on a "rounded out" envelope -- leading to a misrepresentative r



Match the r values and the "envelopes" below



0.00

.30

-.40

-.70

.85

Stating Hypotheses with r ...

Every RH must specify ...
- the variables
- the direction of the expected linear relationship
- the population of interest
- Generic form ...

There is a no/a positive/a negative **linear** relationship between X and Y in the population represented by the sample.

Every H0: must specify ...
- the variables
- that no linear relationship is expected
- the population of interest
- Generic form ...

There is a no **linear** relationship between X and Y in the population represented by the sample.

For each of the following show the envelope for the H0: and the RH:

People who have more depression before therapy will be those who have more depression after therapy. →

Depression after

H0: - - - -

RH: ——

Depression before

Errors

H0: - - - -

RH: ——

Study

← Those who study more have fewer errors on the spelling test

Performance isn't related to practice. →

Practice

H0: - - - -

RH: ——

Performance

For each of the following show the envelope for the H0: and the RH:

People who score better on the pretest will be those who tend to score worse on the posttest →

Post-test

H0: - - - -

RH: ——

Pretest

Depression

H0: - - - -

RH: ——

Sessions

← You can't predict depression from the number of therapy sessions

I predict that larger turtles will eat more crickets. →

Crickets

H0: - - - -

RH: ——

Size

What "retaining H0:" and "Rejecting H0:" means...

- When you retain H0: you're concluding…
  - The linear relationship between these variables in the sample ***is not*** strong enough to allow me to conclude there is a linear relationship between them in the population represented by the sample.
- When you reject H0: you're concluding…
  - The linear relationship between these variables in the sample ***is*** strong enough to allow me to conclude there is a linear relationship between them in the population represented by the sample.

Deciding whether to retain or reject H0: when using r ...

When computing statistics by hand

- compute an "obtained" or "computed" r value
- look up a "critical r value"
- compare the absolute value of the obtained r to the critical value
  - if |r-obtained| < r-critical Retain H0:
  - if |r-obtained| > r-critical Reject H0:

When using the computer

- compute an "obtained" or "computed" r value
- compute the associated p-value ("sig")
- examine the p-value to make the decision
  - if p > .05 Retain H0:
  - if p < .05 Reject H0:

Practice with Pearson's Correlation (r)

The RH: was that younger adolescents would be more polite.

A sample of 84 adolescents were asked their age and to complete the Politeness Quotient Questionnaire

Retain or Reject H0: ???

Support for RH: ???

obtained r = -.453 critical r= .254

Again...

The RH: was that older professors would receive lower student course evaluations.

A sample of 124 Introductory Psyc students from 12 different sections completed the Student Evaluation. Profs' ages were obtained (with permission) from their files.

Retain or Reject H0: ???

Support for RH: ???

obtained r = -.352 p = .431

Statistical decisions & errors with correlation ...

Statistical Decision	In the Population: - r	$r = 0$	+ r
- r ($p < .05$)	Correct H0: Rejection & Pattern	Type I "False Alarm"	Type III "Mis-specification"
$r = 0$ ($p > .05$)	Type II "Miss"	Correct H0: Retention	Type II "Miss"
+ r ($p < .05$)	Type III "Mis-specification"	Type I "False Alarm"	Correct H0: Rejection & Pattern

Remember that "in the population" is "in the majority of the literature" in practice!!

About causal interpretation of correlation results ...

We can only give a causal interpretation of the results if the data were collected using a true experiment

- random assignment of subjects to conditions of the "causal variable" (IV)
  -- gives initial equivalence.
- manipulation of the "causal variable" (IV) by the experimenter
  -- gives temporal precedence
- control of procedural variables
  -- gives ongoing eq.

Most applications of Pearson's r involve quantitative variables that are subject variables -- measured from participants

In other words -- a Natural Groups Design -- with ...

- no random assignment -- no initial equivalence
- no manipulation of "causal variable" (IV) -- no temporal precendence
- no procedural control -- no ongoing equivalence

Under these conditions causal interpretation of the results is not appropriate !!

A bit about computational notation for r …

As before, sort the datafrom the study into two columns – one for each variable (X & Y).

Make a column of squared values for each variable (X^2 & Y^2)

- sum each column -- making a ΣX, ΣX^2, ΣY, ΣY^2

Make a column that's the product of each participants scores

- sum the products to get ΣXY

Practice		Performance		
X	X^2	Y	Y^2	XY
3	9	5	25	15
5	25	8	64	40
4	16	6	36	24
12	50	19	125	79
ΣX	ΣX^2	ΣY	ΣY^2	ΣXY

A bit about computational notation for r, continued …

The computations for r are slightly different –but all the various calculations will use combinations of these five terms – be sure you are using the correct one !

ΣX ΣX^2 ΣY ΣY^2 ΣXY

Other symbols you'll need to know are…

- N = total number of participants