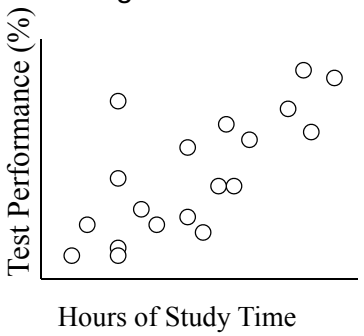
Pearson's X^2

- Correlation vs. X^2 (which, when & why)
- Qualitative/Categorical and Quantitative Variables
- Contingency Tables for 2 Categorical Variables
- Research and Null Hypotheses for X^2
- Causal Interpretation for X^2 Results
- Computational stuff for hand calculations

Pearson's r Vs. X^2

- Pearson's Correlation (r)
  - 2 quantitative variables
  - **LINEAR** relationship
  - range = -1 to +1



- Pearson's Chi Square (X^2)
  - 2 qualitative variables
  - **PATTERN** of relationship
  - range = 0 to + infinity

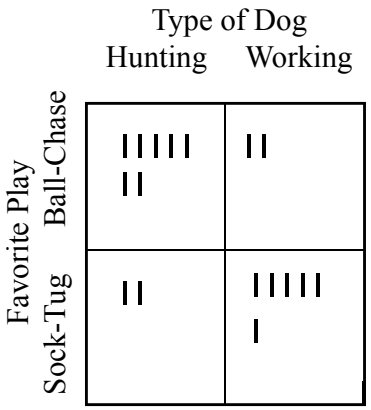
Food Preference	Turtle Type: Painted	Turtle Type: Snapper
crickets	5	15
"duck weed"	19	1

Practice -- would you use r or X^2 for each of the following bivariate analyses?
Hint: Start by determining if each variable is qual or quant !

- GPA & GRE — r
- Age & Shoe Size — r
- Preferred Pet Type & Preferred Toy Type — X^2
- Leg Length & Hair Length — r
- Age and Preferred Type of Pet — ANOVA -- psyche!
- Gender & Preferred Type of Car — X^2
- Grade (%) & Hrs. Study — r

With two qualitative variables we can display the bivariate relationship using a “contingency table”

Puppy	Type (col)	Play (row)
Sam	work	tug
Ding	hunt	chase
Ralf	hunt	tug
Pit	work	tug
Seff	hunt	chase
…	..	..
Toby	hunt	chase



When examining a contingency table, we look for two things...

- whether or not there is a pattern
- if so, which row tends to “go with” which column?

Pattern: A&1 B&2

Rows	A	B
2	15	34
1	36	15

no pattern

Rows	A	B
2	25	24
1	26	25

Pattern: A&2 B&1

Rows	A	B
2	35	14
1	16	35

Describe each of the following ...

Chips \ Crackers	Boys	Girls
Crackers	12	44
Chips	30	16

boys prefer chips & girls prefer crackers

Chips \ Crackers	Boys	Girls
Crackers	17	14
Chips	13	16

no pattern

Chips \ Crackers	Boys	Girls
Crackers	42	14
Chips	10	36

boys prefer crackers & girls prefer chips

Chips \ Crackers	Boys	Girls
Crackers	32	44
Chips	30	16

girls prefer crackers & boys have no preference

The Pearson's Chi-square (X^2) summarizes the relationship shown in the contingency table

- X^2 has a range from 0 to ∞ (infinity)
  - 0.00 absolutely no pattern of relationship
  - "smaller" X^2 -- weaker pattern of relationship
  - "larger" X^2 - stronger pattern of relationship
- However...
  - The relationship between the size of X^2 and strength of the relationship is more complex than for r (with linear relationships)
    - you will seldom see X^2 used to express the strength of the bivariate relationship

Stating Hypotheses with X^2 ...

Every RH must specify ...

- the variables
- the specific pattern of the expected relationship
- the population of interest
- Generic form ...

There is a pattern of relationship between X & Y, such that . . . . . in the population represented by the sample.

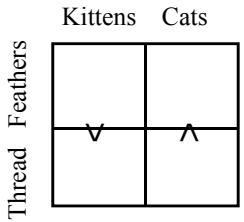
Every H0: must specify ...

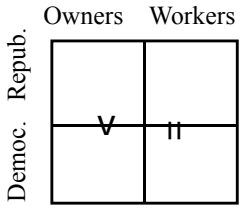
- the variables
- that no pattern of relationship is expected
- the population of interest
- Generic form ...

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

For each of the following use <, > & = to portray the RH:

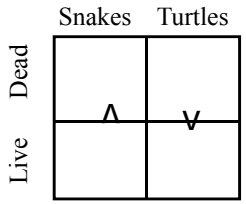
Kittens prefer feathers, whereas cats prefer thread





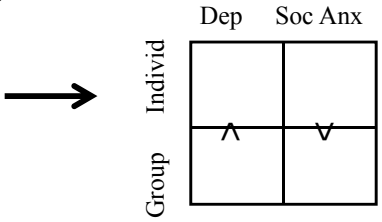
Owners tend to be Republican, while Workers show no preference

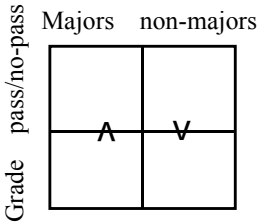
Snakes prefer live crickets and turtles prefer dead crickets



For each of the following use <, > & = to portray the RH:

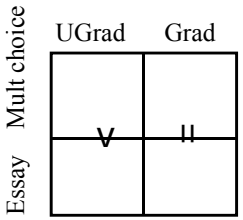
Depressed patients prefer group Tx while those with social anxiety prefer individual Tx





Majors tend to take the course for a a grade, while non-majors tend to take it pass/no-pass

Undergraduate students prefer multiple choice tests, while graduate students don't care



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

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

Deciding whether to retain or reject H0: when using X^2

When computing statistics by hand

- compute an "obtained" or "computed" X^2 value
- look up a "critical X^2 value"
- compare the two
  - if X^2 -obtained < X^2 -critical Retain H0:
  - if X^2 -obtained > X^2 -critical Reject H0:

When using the computer

- compute an "obtained" or "computed" X^2 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:

Statistical decisions & errors with X^2 ...

Statistical Decision	In the Population: that specific pattern	no pattern	any other pattern
that specific pattern ($p < .05$)	Correct Decision	Type I "False Alarm"	Type III "Mis-specification"
no pattern ($p > .05$)	Type II "Miss"	Correct Decision	Type II "Miss"
any other pattern ($p < .05$)	Type III "Mis-specification"	Type I "False Alarm"	Correct Decision

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

Testing X^2 RH: -- different "kinds" of RH: & it matters!!!

"Proportion" type RH:

RH: A greater proportion of those who do the "on web" exam preparation than of those who do the "on paper" version will pass the exam.

"Implied Proportion" Type of RH:

RH: Those who do the "on web" exam preparation will do better than those who do the "on paper" version.

"Pattern" type RH:

RH: More of those who do the "on web" exam preparation assignment will pass the exam, whereas more of those who do the "on paper" version fill fail the exam.

Testing X^2 RH: -- different "kinds" of RH: & it matters!!!

"Proportion" type RH:

RH: A greater proportion of girls than of boys will prefer crackers.

	Boys	Girls
Crackers	12	44
Chips	30	16

X^2=19.93, p<.001

Both RH:s supported !!

Girls 44/60 = .73

Boys 12/42 = .29

Girls 44 > 16 & Boys 12 < 3

"Pattern" type RH:

RH: More girls will prefer crackers and more boys will prefer chips.

	Boys	Girls
Crackers	32	44
Chips	30	16

X^2=6.12, p=.013

Only "Proportion" RH supported !!

Girls 44/60 = .73

Boys 32/62 = .52

Girls 44 > 16 But Boys 32 >

Testing X^2 RH: -- one to watch out for…

Sometime, instead of …

RH: A greater proportion of those do the "on web" exam preparation than of those who do the "on paper" version will pass the exam.

You'll get… → This is not a good way to express a X^2 RH: !!!!

RH: More of those who do the "on web" exam preparation assignment will perform better on the exam than those who do the "on paper" version.

You have to be careful about these kinds of "frequency" RH:!!!

X^2 works in terms of proportions, not frequencies! And, because you might have more of one group than another, this can cause confusion and problems…

Testing X^2 RH: -- one to watch out for…

Instead of …

RH: A greater proportion of girls than of boys will prefer crackers.

You'll get… → This is not a good way to express a X^2 RH: !!!!

RH: More girls than boys will prefer crackers.

	Boys	Girls
Crackers	20	20
Chips	40	10

X^2=9.00, p=.003

The number of boys & girls is same 20 = 20 …

But X^2 tests for differential proportion of that category not for differential number of that category…

Girls 20/30 = .66 > .33 = 20/40 Boys

About causal interpretation of X^2 ...

Applications of Pearson's X^2 are a mixture of the three designs you know

- True Experiment
- Non-Experiments

But only those data from a True Exp can be given a causal interpretation …

- 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.

You must be sure that the design used in the study provides the necessary evidence to support a causal interpretation of the results !!

Practice with Statistical and Causal Interpretation of X^2 Results

RH: Those who do the "on web" exam preparation assignment will perform better on the exam than those who do the "on paper" version.

	Paper	Web
Pass	28	39
Fail	22	11

X^2 obtained = 5.47, critical X^2 = 3.84

Retain or Reject H0: ??? Reject!

Support for RH: ??? Yep ! 39/50 > 28/50

Design: Before taking the test, students were asked whether they had chosen to complete the "on Web" or the "on paper" version of the exam prep. The test was graded pass/fail.

Type of Design ??? Natural Groups Design

Causal Interpretation? Nope!

What ***CAN*** we say from these data ??? There's an association between type of prep and test performance.

RH: More of those who do the "on web" exam preparation assignment will pass the exam and more of those who do the "on paper" version will fail.

	Paper	Web
Pass	21	27
Fail	23	24

X^2 obtained = .26, p = .612

Retain or Reject H0: ??? Retain!

Support for RH: ??? Nope !

Design: Students in the morning laboratory section were randomly assigned to complete the "on Web" version of the exam prep, while those in the afternoon section completed the "on paper" version. Student's were "monitored" to assure the completed the correct version. The test was graded pass/fail.

Type of Design ??? Quasi Experiment

Causal Interpretation? Nope!

What ***CAN*** we say from these data ??? There's no association between type of prep and test performance.

RH: More of those who do the "on web" exam preparation assignment will pass the exam and more of those who do the "on paper" version will fail.

	Paper	Web
Pass	21	37
Fail	23	14

X^2 obtained = 6.12, p = .013

Retain or Reject H0: ??? Reject!

Support for RH: ??? **Partial: 37 > 14, but 23 = 21**

Design: One-half of the students in the T-Th AM lecture section were randomly assigned to complete the "on Web" version of the exam prep, while the other half of that section completed the "on paper" version. Students were "monitored" to assure the completed the correct version. The test was graded pass/fail. Only data from students in the T-TH AM class were included in the analysis.

Type of Design ??? True Experiment

Causal Interpretation? Yep!

What ***CAN*** we say from these data ??? That type of prep influences test performance.

About calculations for hand computations…

Many find the X^2 hand computations to be easier than the others !!!

This may be do to the near absence of $\sum$ notation !

Just be sure that you get the same “total” when calculated as the sum or row totals and the sum of column totals !!!

	Paper	Web	Total
Pass	21	27	48
Fail	23	24	47
Total	44	51	95

$95 = 48 + 47 = 44 + 51$