Bio 183
Statistics in Research

A. Research designs

B. Cleaning up your data: getting rid of problems

C. Basic descriptive statistics

D. What test should you use?

What is “science” ?:

Science is a way of knowing….(anon.?)

Propositions/ideas/hypotheses that can’t be rejected are not falsifiable and are **not** “scientific”.

“Science is a way of trying not to fool yourself”

-Richard Feynman

Note: Not every project has a clear hypothesis

→ so is that project ‘science’?

It Depends:

→ Many projects are part of research **program.**

- some components may be purely **descriptive** and not test a hypothesis.

- that’s OK because hypotheses often need **background** info before they can be framed well
  (‘natural history’ – the field, lab (molecular sequences)

→ **Otherwise:** just isolated facts.

Philosophy of science: the scientific method

DONE

Philosophy → **How we understand the world**

Science → **A “method” for achieving this goal**

Develop theory

Formulate hypotheses - predictions

Test predictions

Design & Analysis → **How we test predictions of hypotheses**

Statistics → **Tool for quantitative tests**

→ A. Research designs (why do this before stats?)
    1. Basic principles
    2. Basic designs

B. Cleaning up your data: getting rid of problems

C. Basic descriptive statistics

D. What test should you use?

A. Research designs

1. Basic Principles:

a) Contrasts (manipulative or not)

→ explicit (between/among: groups, treatments, areas, etc.)

→ implicit (H_A vs a value of H_o that is known from another study or assumed)

b) Controls?

→ Why? What is a control?

→ Effects are due to the comparisons you study: "All else being equal…".

→ Do we always need controls? What are the controls if you compare 2 treatments (e.g. light, concentrations of X)?

More design concerns:

c) Removing bias in treatments, plots, observations

- → randomize: individuals, areas, location (in field, lab, greenhouse…)
- → always randomize??

d) Replicates

- → why?
  - variability
- → how many?
  - that depends!!
  - on average effect size, variation, what's affordable (money, time)

2. Basic, Simple Designs – to compare effects

- **Powerful designs** (& some examples)
  - compare imposed changes (manipulation)
  - compare unmanipulated areas, individuals or objects (‘natural experiment’) [But…]
  - before and after (same plot/area, same individual…)

Statistics in Research

A. Research designs

→ B. Cleaning your data: getting rid of problems
  1. Problem data points
  2. Normal or not? What to do?

C. Basic descriptive statistics

D. What test should you use?

B. Cleaning your data: common problems…..

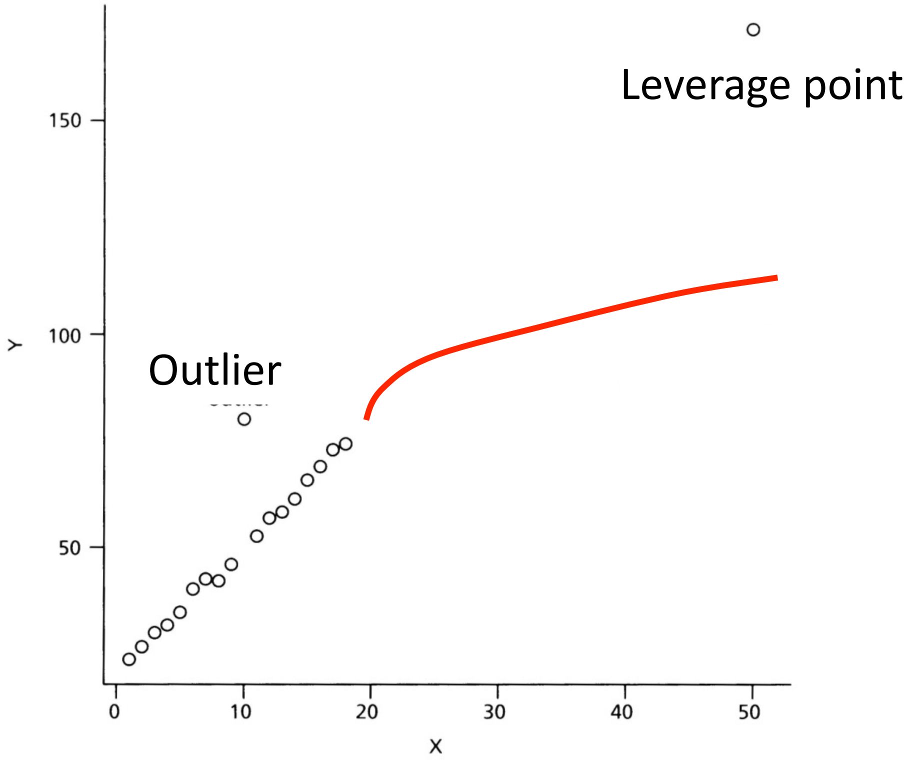
→ Data need to be examined <u>before</u> you do the statistics

→ Goal: To eliminate problems that lead to inappropriate analyses and/or interpretation.

Cleaning up your data:

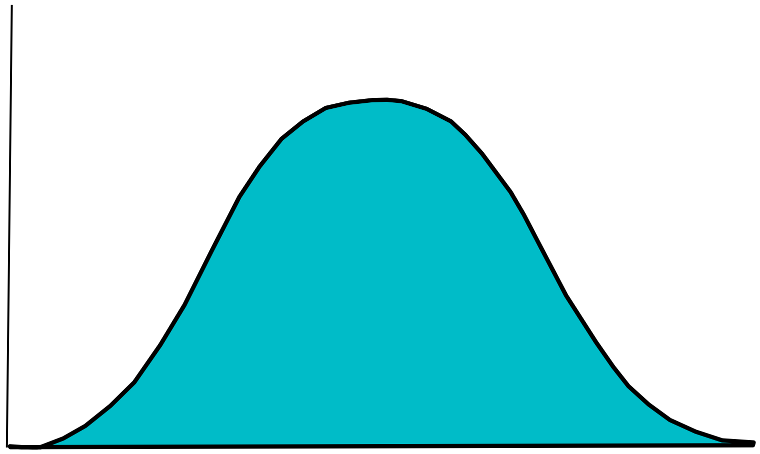
1. **Problem data points**:

- Any very odd points (and esp if you know why they are odd)?
- You can eliminate them sometimes!!
- This does NOT mean selecting the data points that best fits your hypotheses.



2. If your data are NOT normal

Normal = bell-shaped curve



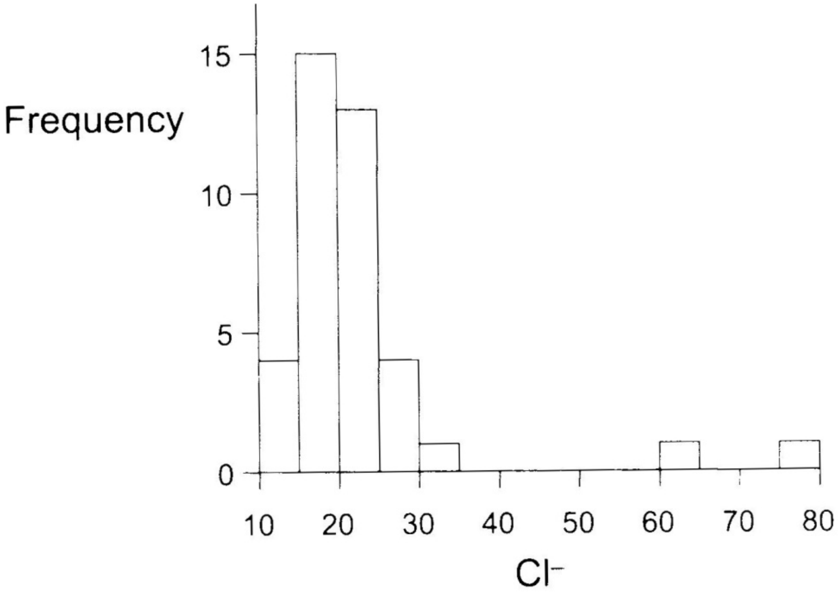
- probably don’t worry about this for your current projects

But for your information:

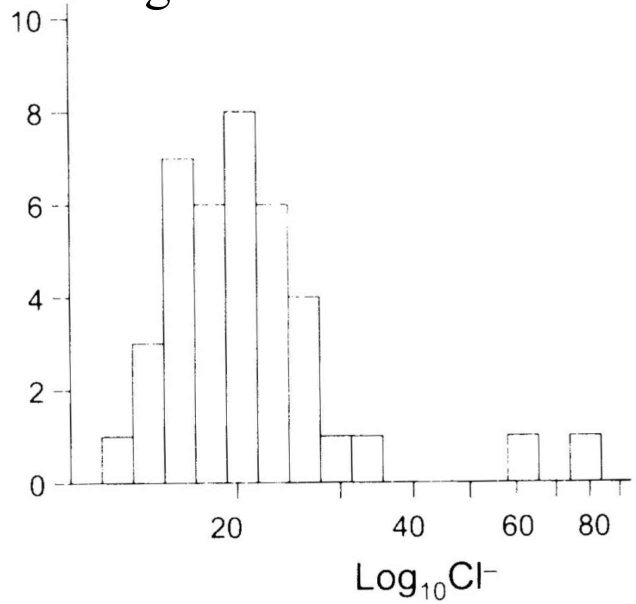
i) Transformations:

Some basic ones to try for continuous arithmetic data: log, √, 1/x, x^2

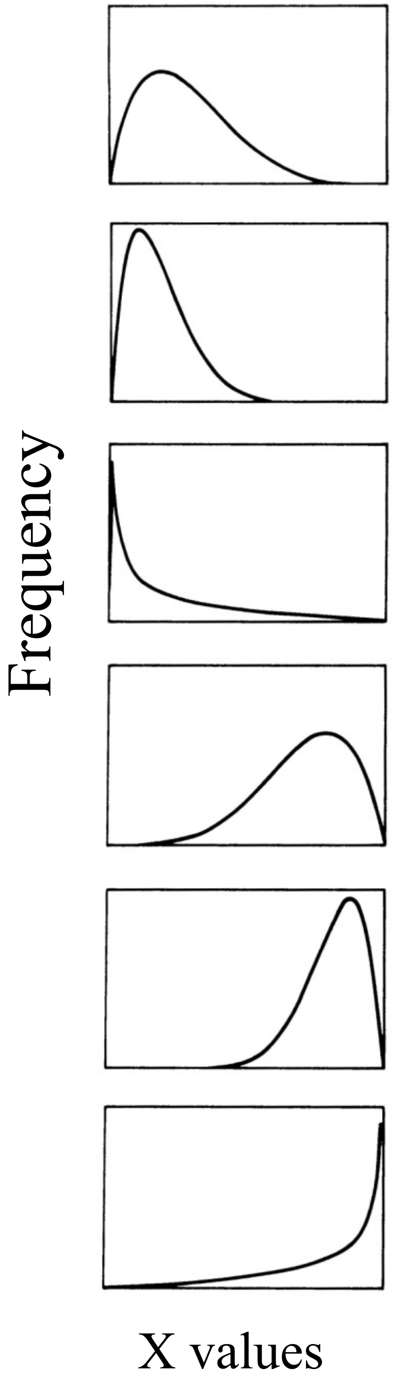
Raw data



Log transform: → more normal



TRANSFORMATION



Square Root NEWX = SQRT(X)

Log NEWX = LOG10(X)
with 0: NEWX = LOG10(X+C)

Inverse NEWX = 1/X
with 0: NEWX = 1/(X+C)

Reflect & Square Root
NEWX = SQRT(K-X)

Reflect & log NEWX = LOG10(K-X)

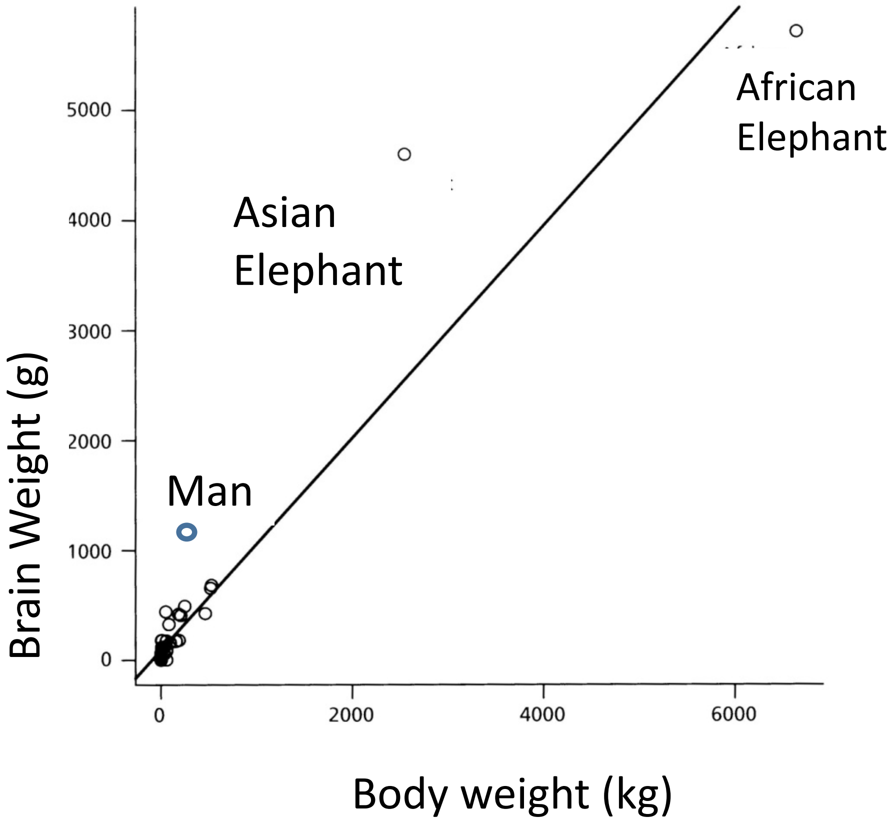
Reflect & Inverse
NEWX = 1/(K-X)

C = constant added to each score so that the smallest score is 1 (for base 10)

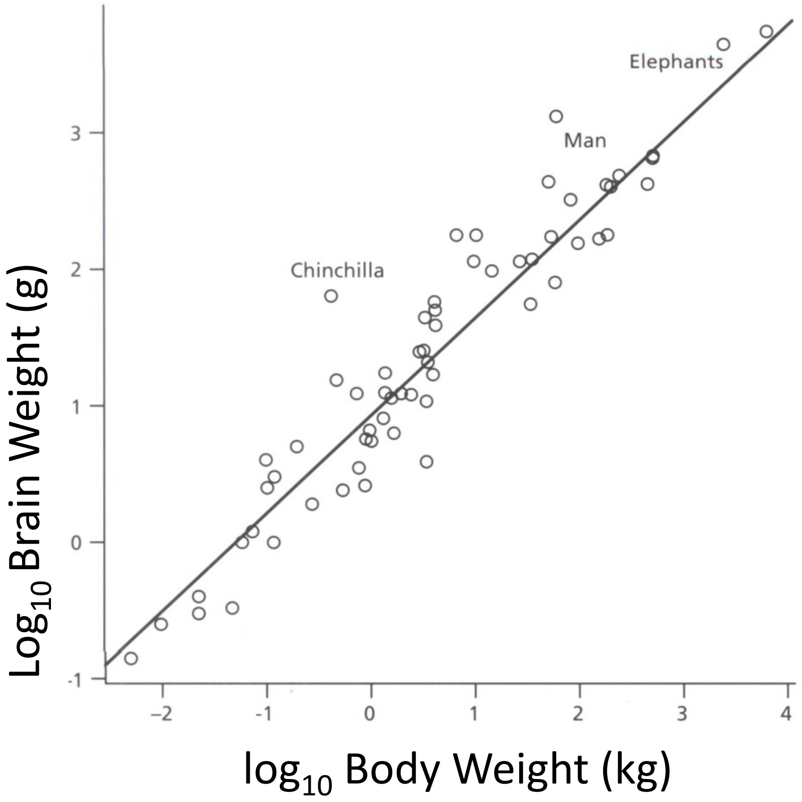
K = constant from which each score is subtracted so that the smallest score is 1; usually = largest score + 1.

Logs: can use any base, but base 10 has bigger effect than natural logs.

Common graph in physiology



Better comparison on log-log plot

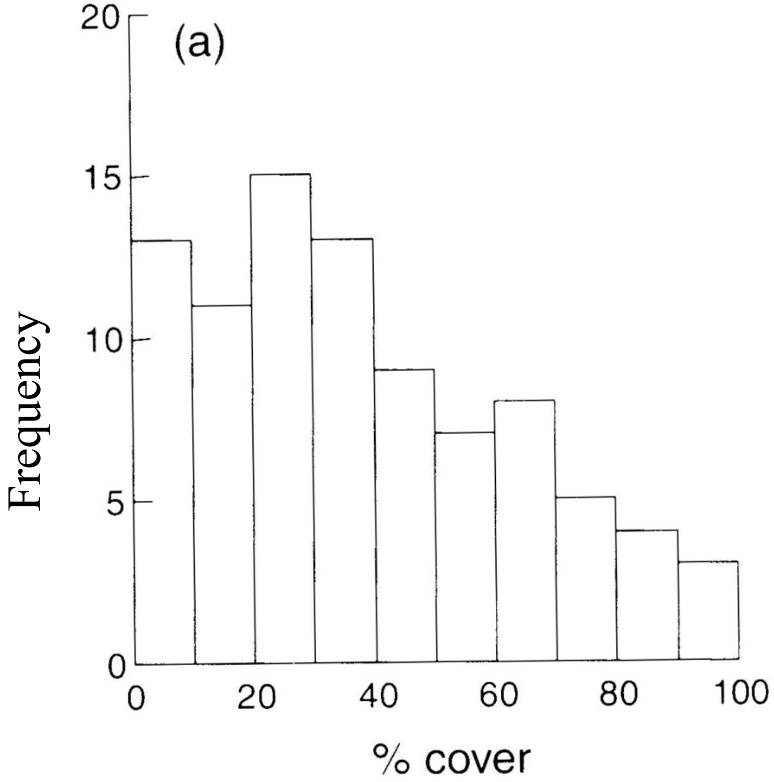


ii. Proportions – are constrained between 0 and 1:

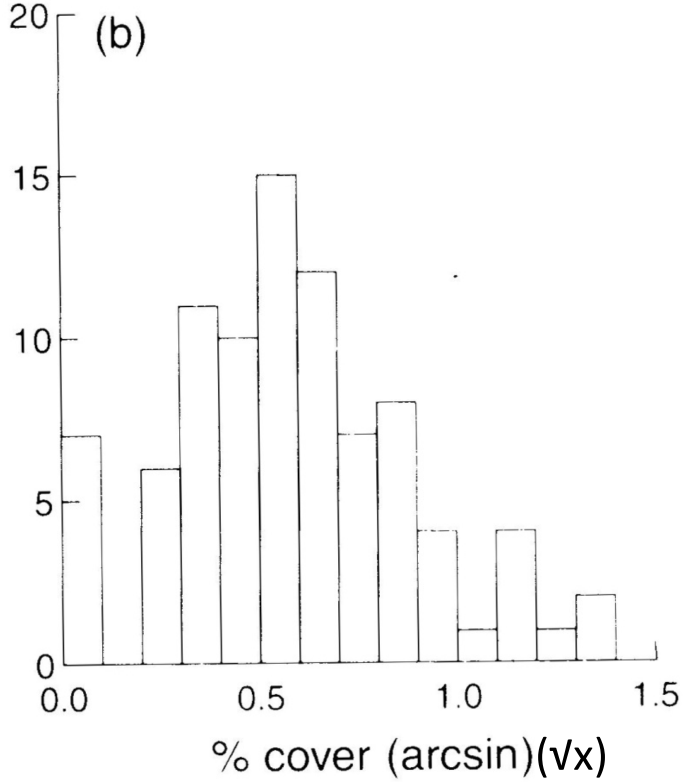
→ transform with: arcsin(√x)

(NOTE: this is very contested)

Not a normal distribution



A more normal distribution.



NOTE: If you can’t get your data to look normal, even with a transform

→ **Non-parametric statistics** –

- median, mode,
- estimates of spread
- tests of hypotheses

A. Research designs

B. Cleaning up your data: getting rid of problems

→ C. Basic descriptive statistics: patterns

D. What test should you use?

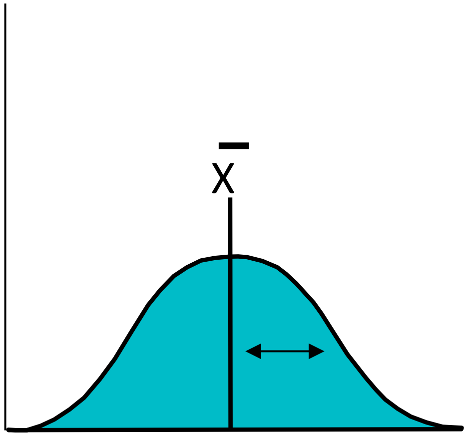
Different project designs → different options

Type of Data	**Basic Analyses**	**Relationships**
1. One dataset → one time and/or place	Descriptive stats = "Data mining" (mean, var) Graphs	---
2. One dataset → compare to Ho	"	One sample tests
3. Several datasets → compare over time, or place If experiment → compare treatments	Descriptive stats by treatment, levels, etc. Graphs	Compare distributions (mean, var) of diff treatments, variables… Paired comparisons Multiple comparisons Compare groups

1. Basic analyses: Describe your data = Descriptive stats

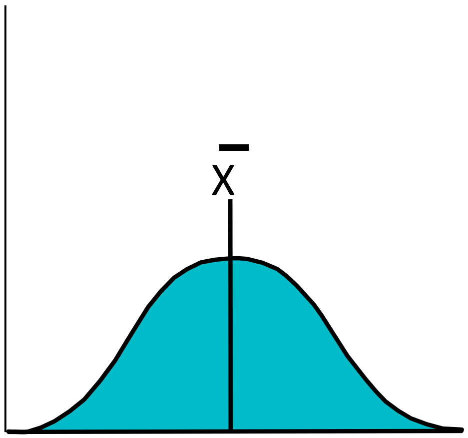
→ Not test an hypothesis
- central tendency: mean, median
- variation

Assume: a normal “bell”– shaped distribution
→“parametric”



If not normal
→ “nonparametric”

Parametric descriptions:



Estimate means

$$\bar{x} = \frac{\Sigma x_i}{n}$$

Estimate the error (variation) associated with the mean:

$$S = \frac{\Sigma (X_i - \bar{x})^2}{n} = \text{Variance} = \sigma^2$$

Standard Deviation (sd) = $\sqrt{S} = \sigma$

Standard Error of the Mean (SEM) = sd / $\sqrt{n}$

What is the difference between sd and SEM??

A. Research designs

B. Cleaning up your data: getting rid of problems

C. Basic descriptive statistics

→ D. What test should you use?
  1. Hypothesis testing
  2. Basic tests
  3. Binomial key to some widely used tests

1. Basic concepts of hypothesis testing

Steps:

1) Assume the **null** hypothesis is true

if we don't detect a difference, then null H_o is true.

2) **Test → Compare measurements**

generally test the **mean difference** between two sample distributions (e.g. the numbers from the experiment or set of observations) or against an "expected value"; summarized as a **test statistic** (t, F, …..).

3) **Determine the probability** that the difference between the sample distributions is due to chance alone, rather than the hypothesized effect (translate test statistic to **calculated p-value**)

4) **Compare calculated p-value with a critical p-value**: (now done for you in stats packages)

assign significance (i.e., whether accept or reject null hypothesis)

Need: probabilities

What is the probability of rejecting/accepting an hypothesis?

α = probability of rejecting Ho when it is true.

The level of uncertainty that you are willing to accept that you reject Ho when it is true.

→ Type-I error.

By convention $\alpha = 0.05$ (but not always).

α = critical p-value

Example

H_A= More mussels settle inside adult distribution
H_o= No difference in mussel settlement inside vs outside adult distribution

Calculated p-value	Critical p-value (α)	Decision
0.50	0.05	Do not reject H_o
0.051	0.05	Do not reject H_o**
0.050	0.05	Reject H_o, not H_A
0.049	0.05	Reject H_o, not H_A

α = probability the 2 samples are the same (i.e., H_o is true) – when they're not

**This is hard to deal with. Can say it is suggestive. Some say .05 is too stringent in some cases. But depends on *power* of the test….because

Goals of testing:

1. Reject Ho when it is false and accept Ho when it is true

2. Avoiding errors:

α = probability of rejecting Ho when it is true = Type-I error.

ß = probability of accepting Ho when it is false = Type-II error.

	Decision	
Truth	Accept H_O	Reject H_O
H_O true	no error (1-α)	Type I error (α)
H_O false	Type II error (β)	no error (1-β)

precautionary principle

Power

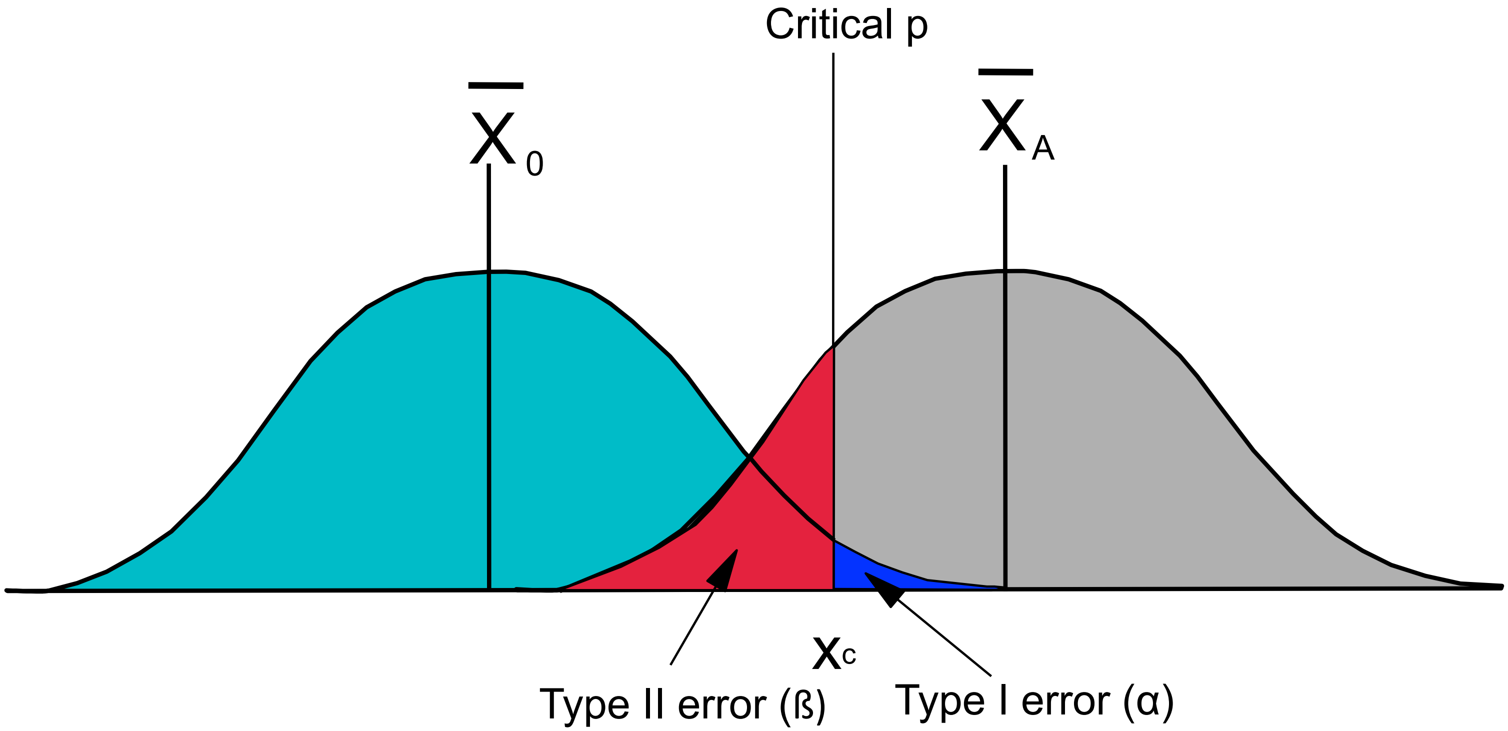
Power of Test = probability (1-ß) of rejecting Ho when it is false and not committing a Type II error.

The more powerful the test, the more likely you are to correctly conclude than an effect exists when it really does.

How to control Type II error

This will maximize statistical power to detect real impacts

	Decision	***Decision***
Truth	**Accept H_0**	**Reject H_0**
$H_O$true	no error (1-α)	Type I error (α)
$H_O$false	Type II error (ß)	no error (1- ß)



Type I and Type II Errors are ‘controlled’ by:

Nature:

- Intrinsic variation
- Differences between means

Scientist:

- Inadvertent measurement/sampling errors
- Vary critical P-values**

Know → α and ß are inversely related for a given N

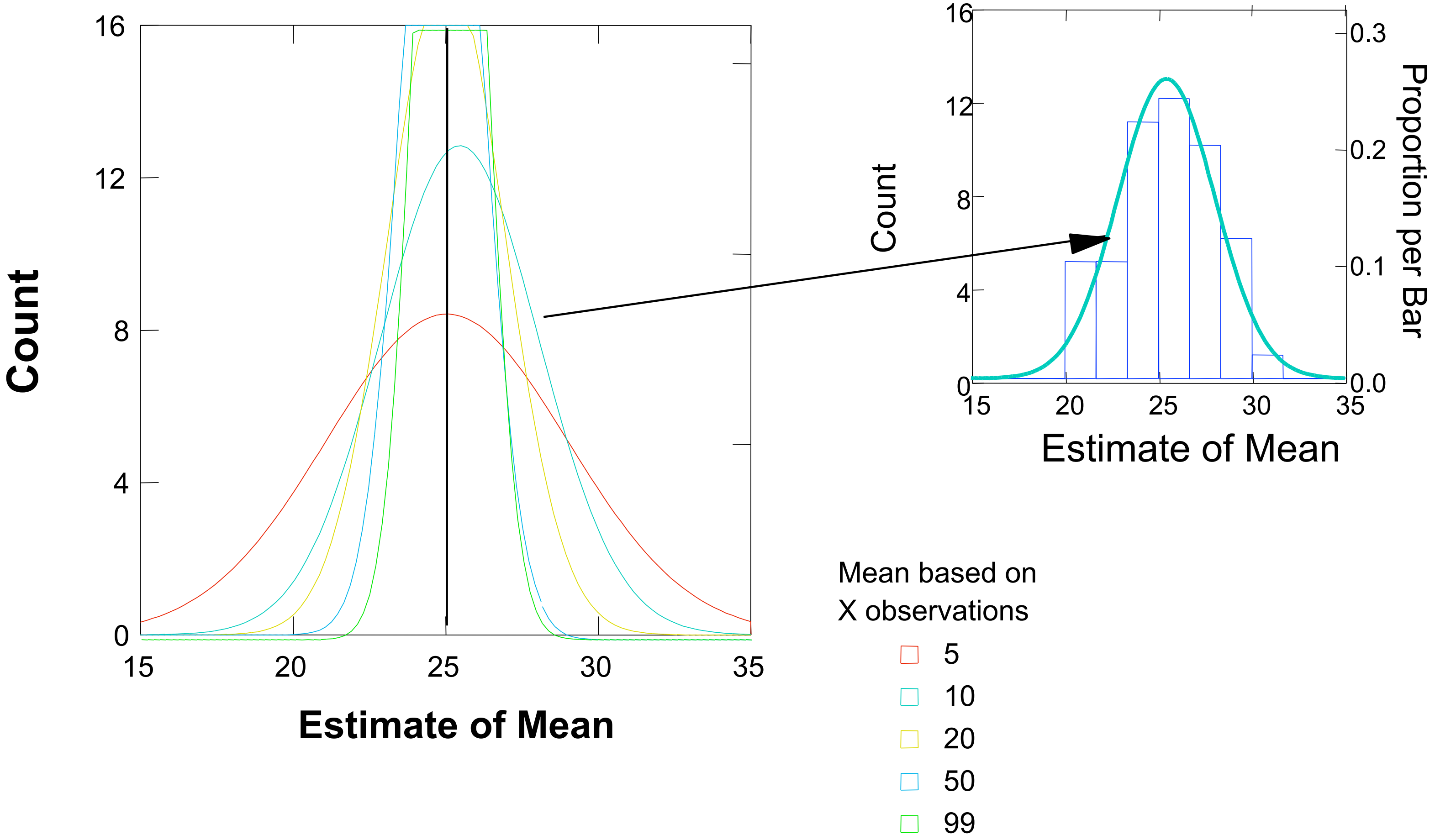
Problem: how to minimize both α and ß?

- Increase replication** → variation around the mean declines as number of observations increase

Reflect design considerations

Effect of number of observations on estimate of Mean

Frequency distributions of sample means

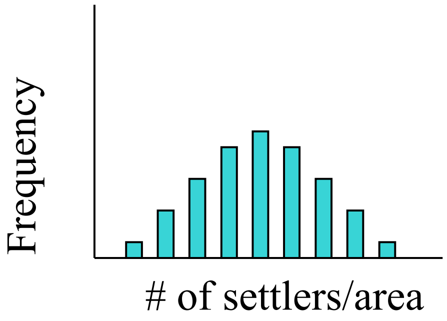


The estimated mean approaches true mean with larger sample size!
("Central Limit Theorem")

Notice how variation around estimated mean increases with lower sample size!

2. Basic Stats Testing:
a) Only one sample:

Ex 1. Settlement distribution of mussels in the intertidal.
Is this significantly different from…??



A one-sample test:

$$\frac{\bar{X}_{\text{sample}} - \text{comparison mean}}{\text{SEM}}$$

Comparison mean can be any reasonable value. Commonly:
a. 0
b. Previous sample mean

This formula generates a ‘test statistic’: t =
Find $P_{(t)}$ =
Compare to $P_{(\text{critical})}$

IMPT NOTE: some stats packages assume the comparative mean = 0. Be careful if that isn’t what you want.

b) > 1 sample:

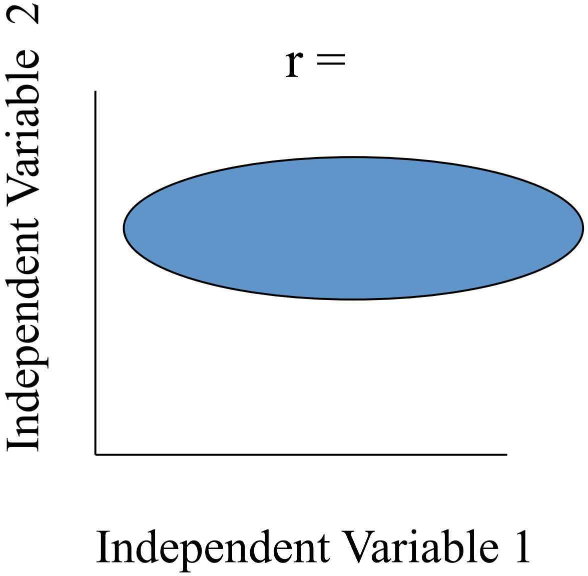
→ Associations/Relationships among the variables?

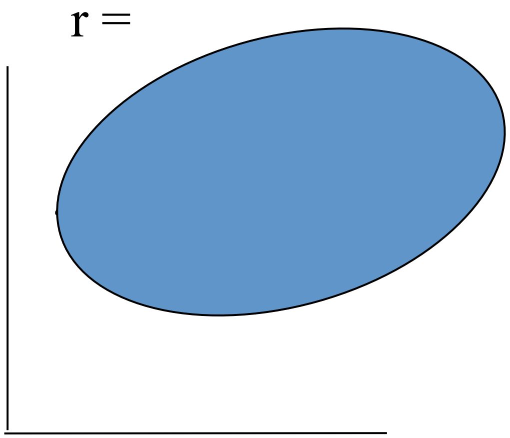
i. <u>Correlations between two variables – how much association??</u>

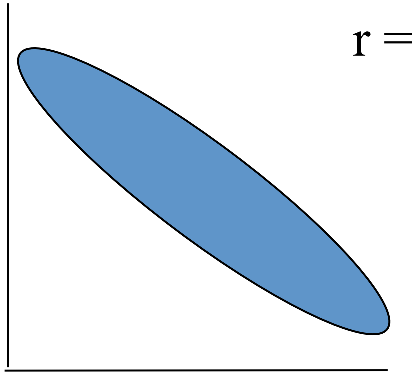
→ Correlation is just a pattern and does not indicate cause and effect

- If both variables are normally distributed -- standard correlation (Pearson's)
- If either or both are NOT normal
  → rank the data (x and y separately) and then…
  → do a correlation on ranks (Spearman's or Kendall's).

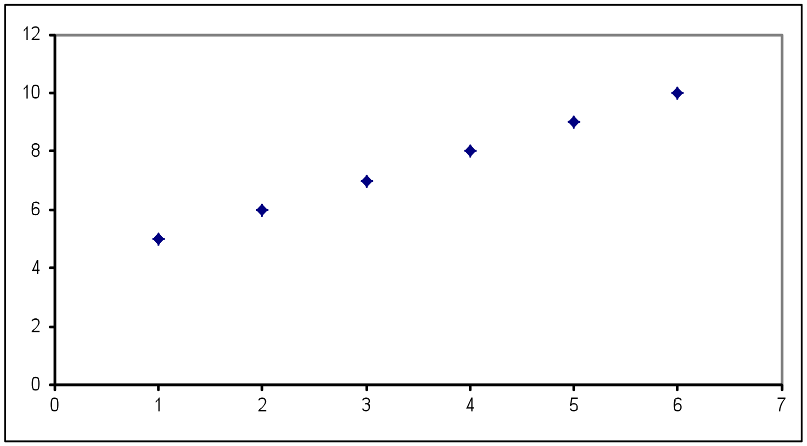
→ Correlation coefficient = r
(what is the range of possible values?)

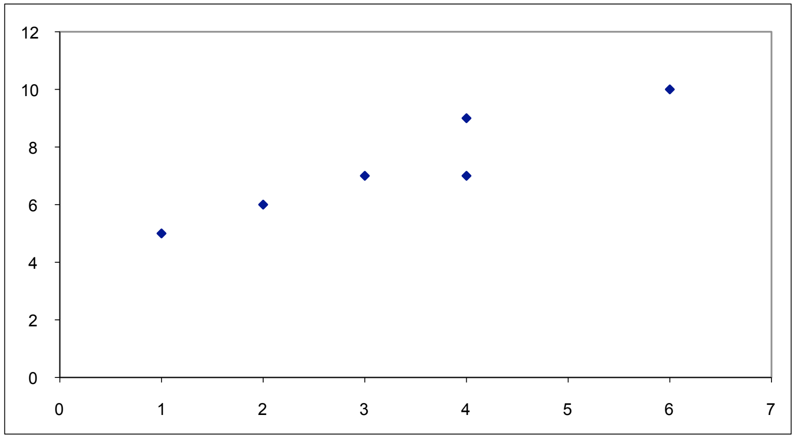






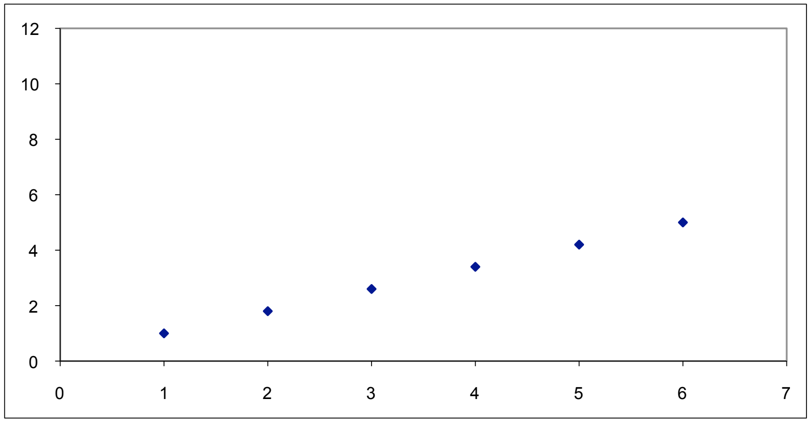
Test for significance:
t-test, mostly Ho = 0
N


0
2
4
6
8
10
12
0
1
2
3
4
5
6
7


0
2
4
6
8
10
12
0
1
2
3
4
5
6
7

r = 1.0

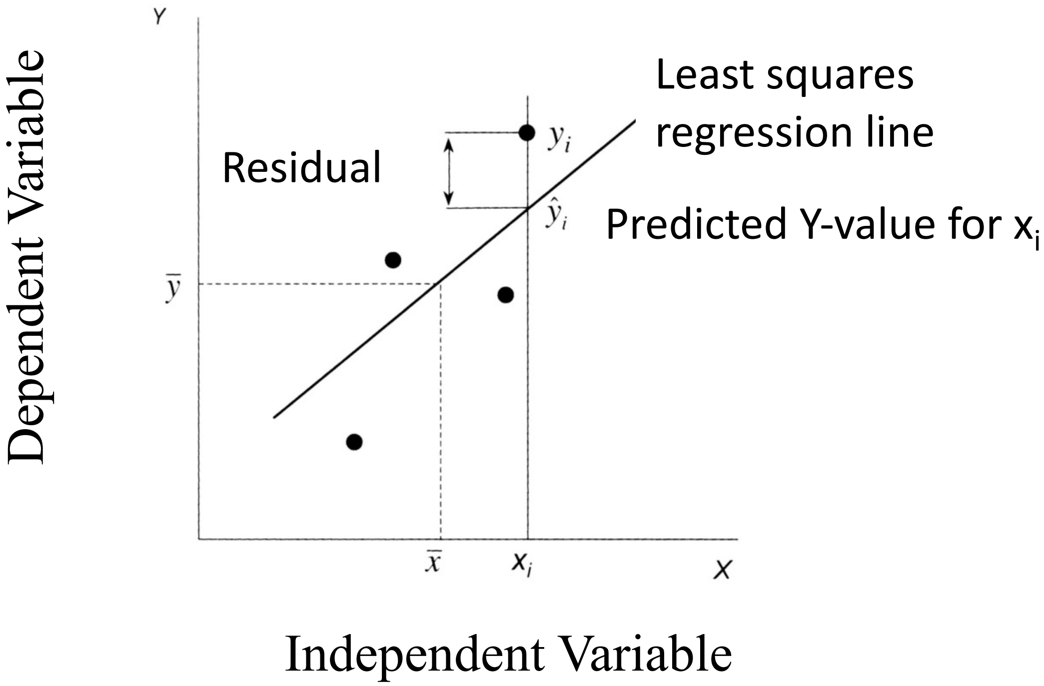
r = 0.94


0
2
4
6
8
10
12
0
1
2
3
4
5
6
7

r = 1.0

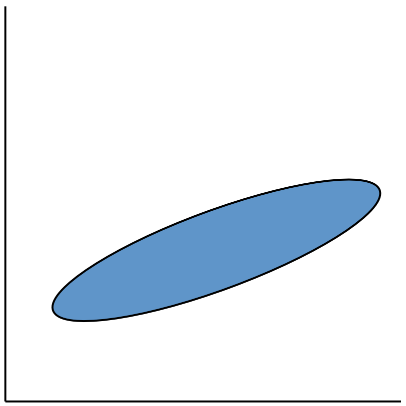
ii. Regression between two variables: predicting Y from X

X = independent variable: **not** usually normal (often “uniform” distribution)
Y = dependent variable: normal



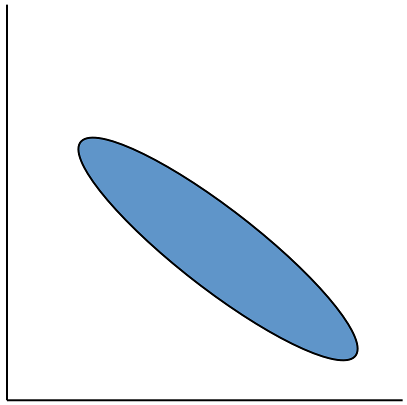
What statistics?

- relationship (slope): t-statistic (or F-statistic)
- goodness of fit: r^2



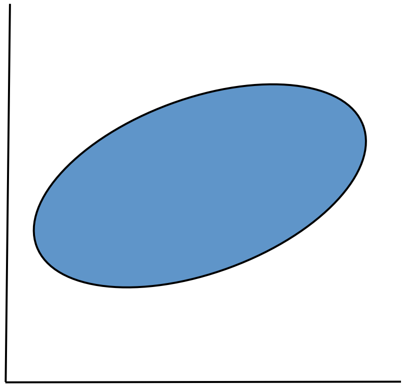
$t = a$

$r^2 =$



$t =$

$r^2 =$



$t =$? Compared to above?

→ Same as above

$r^2 =$ (why?)

iii. Some basic tests of differences between categories/treatments

Example 2: two samples

General/working hypothesis – larval settlement determines adult distribution

Specific hypothesis – more mussel larvae settle in areas inside the adult distribution than in areas outside it

Observation 1:

	Number outside	Number inside
	0	10
	0	15
	0	18
	0	12
	0	13
Mean	0	13

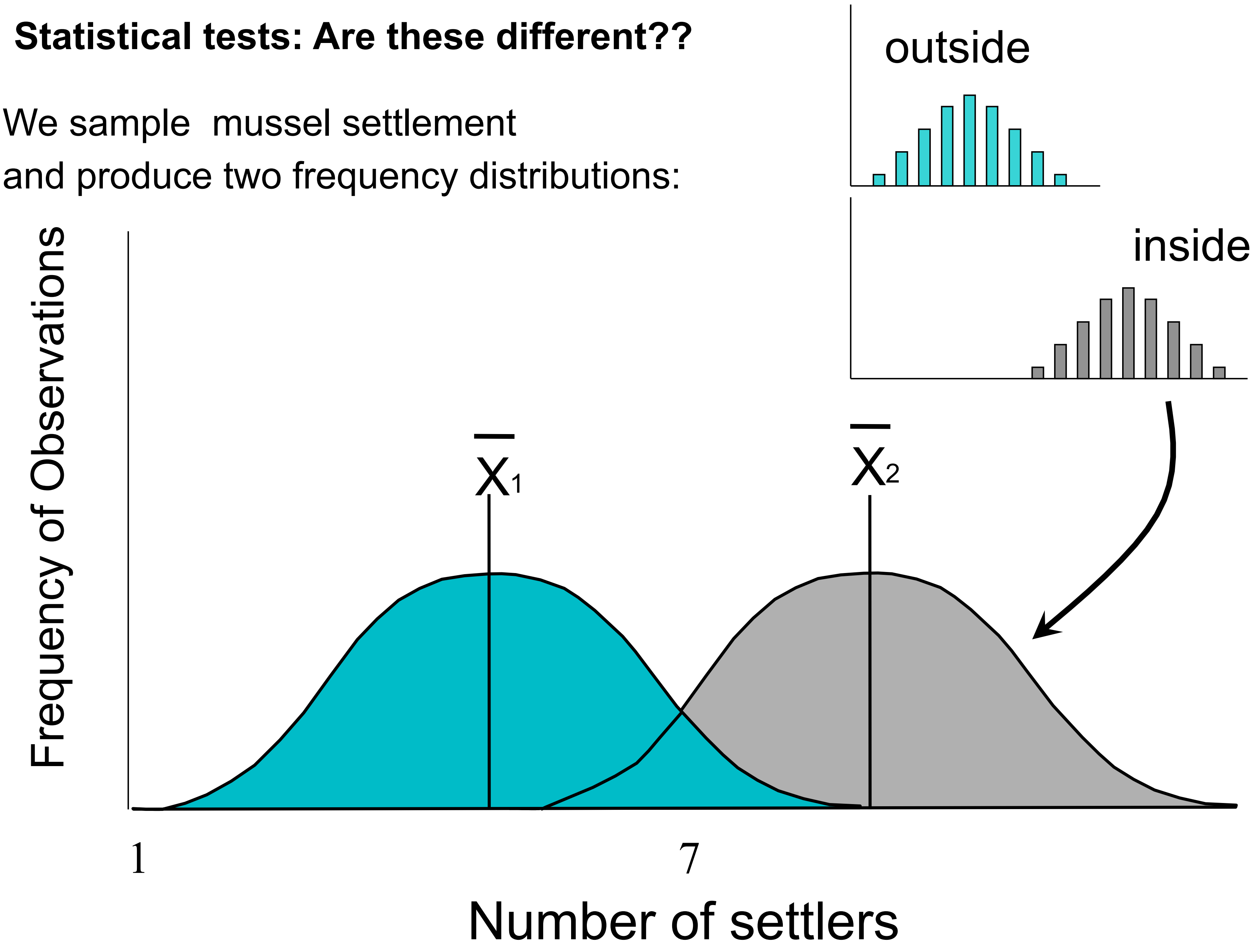
Observation 2:

	Number outside	Number inside
	3	10
	5	7
	2	9
	8	12
	7	8
Mean	5	9.2

- What counts as a difference? How different do samples ‘need’ to be to be considered statistically different?
- Are in/out for Obs 1 different?
- For Obs 2?

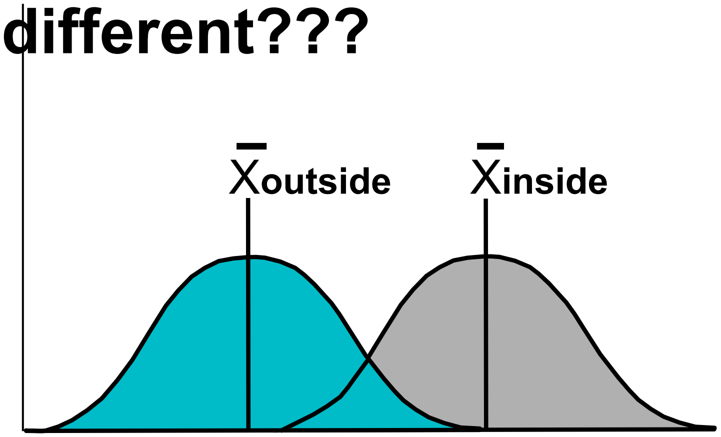
Statistical tests: Are these different??

We sample mussel settlement
and produce two frequency distributions:



Comparing two samples: Are they different???

How do you compare distributions?



a) Are both distributions normal? Or at least the same distribution?

b) Compare means, taking into account the confidence you have in your estimate of the means (variation).

Calculate a standardized difference between the means of the two distributions (again, taking into account their error):

$$\frac{\bar{X}_{inside} - \bar{X}_{outside}}{\sqrt{2\,s^2 / n}} = \frac{\bar{X}_{inside} - \bar{X}_{outside}}{\mathbf{SEM}}$$

This generates a 'test statistic' (e.g., t, F).

Then ask…"is the value of the statistic due to random chance or to a real difference"??

Probability that the estimated difference is not due to random chance

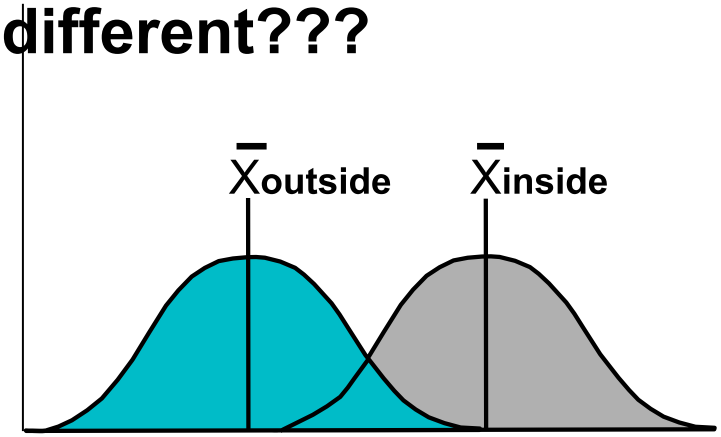
1. Calculate the **probably (P) of getting the test value** (table or stats package)
2. Compare to **critical P-value**

Comparing two samples: Are they different???

Outside	Inside
3	10
5	7
2	9
8	12
7	8

2-sample t-test

$$\frac{\bar{X}_{inside} - \bar{X}_{outside}}{\mathbf{SEM}}$$



Means = 5.5 9

t-value	-2.049	
$P(t < t_c)$ One-tail	0.043	H_A = sample 1 < sample 2 *OR*
t_c One-tail	1.943	H_A = sample 1 > sample 2
$P(t < t_c)$ Two-tail	0.086	H_A = sample 1 is different from sample 2
t_c Two-tail	2.447	

Not all data sets are based on measurements.

→ Data may be frequencies in discrete categories of a sample.

Example: Eye color in genetic crosses (flies)
H_o: no difference in frequency among groups
H_A: significant variation

Color	Frequency in sample*
Red	30
Brown	50
Yellow	20

* Score 300 flies

→ Reject or accept H_o?
How test?

Color	Obs Freq.	Obs Numbers	Expected (Ho)
Red	30	90	
Brown	50	150	
Yellow	20	60	

= 300 counted

??
100 flies

Color	Obs Freq.	Obs Numbers	Expected (Ho)	O-E	$\frac{(O-E)^2}{E}$
Red	30	90	100	10	1
Brown	50	150	100	50	25
Yellow	20	60	100	40	16

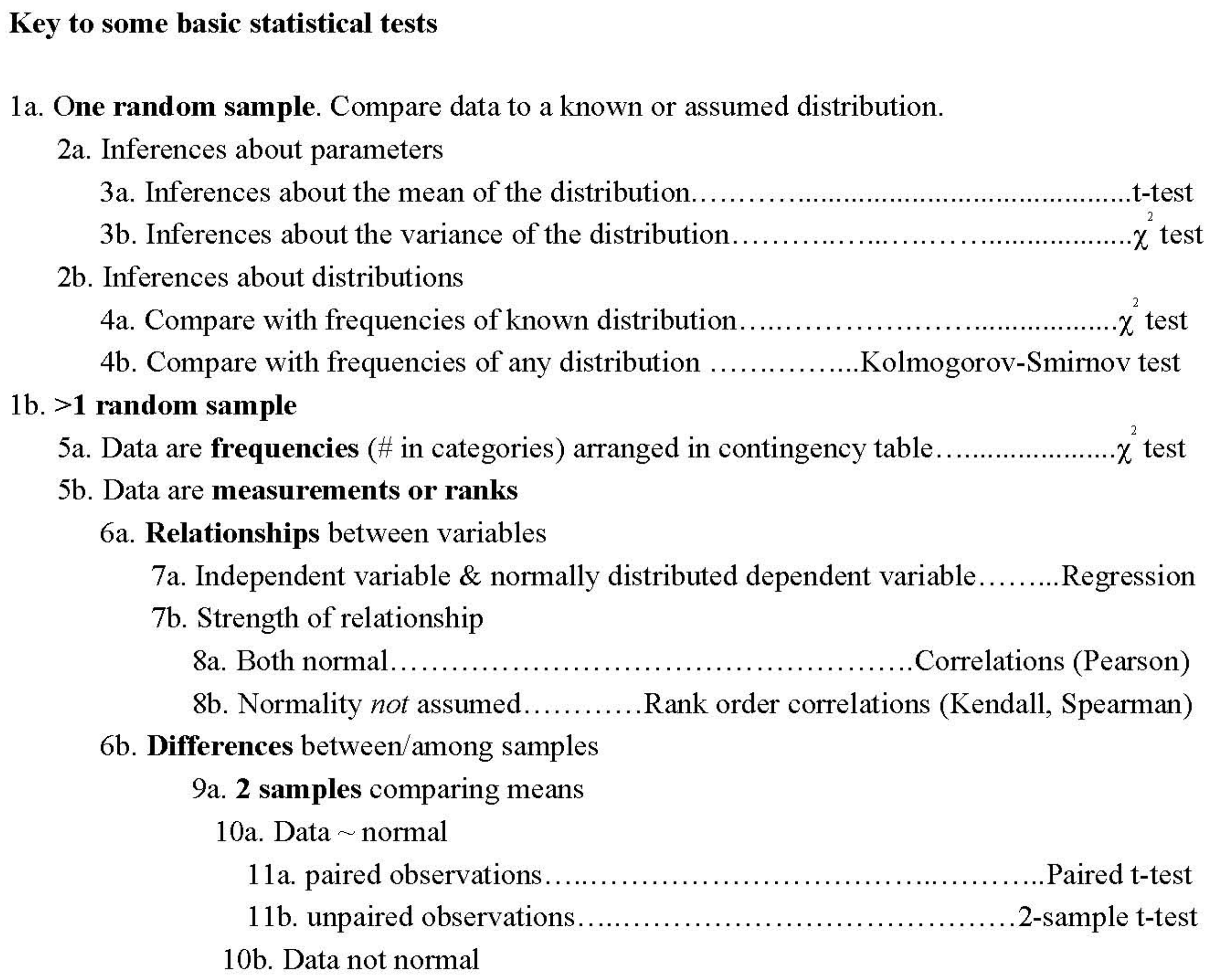
χ^2 test: $\Sigma \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

$\chi 2 = 42$

Compare P_{test} and $P_{critical}$

$P_{test} < 0.0001$ → reject H_o

Now it gets a bit more complicated…..



Key to some basic statistical tests

- 1a. **One random sample**. Compare data to a known or assumed distribution.
  - 2a. Inferences about parameters
    - 3a. Inferences about the mean of the distribution………...t-test
    - 3b. Inferences about the variance of the distribution………..…..…..……...................χ^2 test
  - 2b. Inferences about distributions
    - 4a. Compare with frequencies of known distribution…..…………………..................χ^2 test
    - 4b. Compare with frequencies of any distribution ……..……...Kolmogorov-Smirnov test
- 1b. **>1 random sample**
  - 5a. Data are **frequencies** (# in categories) arranged in contingency table…....................χ^2 test
  - 5b. Data are **measurements or ranks**
    - 6a. **Relationships** between variables
      - 7a. Independent variable & normally distributed dependent variable……...Regression
      - 7b. Strength of relationship
        - 8a. Both normal……………………………………………Correlations (Pearson)
        - 8b. Normality *not* assumed…………Rank order correlations (Kendall, Spearman)
    - 6b. **Differences** between/among samples
      - 9a. **2 samples** comparing means
        - 10a. Data ~ normal
          - 11a. paired observations…..…………………………..………..Paired t-test
          - 11b. unpaired observations…..…………………………………2-sample t-test
        - 10b. Data not normal

6a. **Relationships** between variables

7a. Independent variable & normally distributed dependent variable……...Regression

7b. Strength of relationship

8a. Both normal…………………………………………….Correlations (Pearson)

8b. Normality *not* assumed…………Rank order correlations (Kendall, Spearman)

6b. **Differences** between/among samples

9a. **2 samples** comparing means

10a. Data ~ normal

11a. paired observations…..…………………………..………..Paired t-test

11b. unpaired observations…..…………………………………2-sample t-test

10b. Data not normal

12a. paired observations

13a. sign of difference important……………………………..…..Sign test

13b. size of difference important…………..……Wilcoxon Rank Sum test

12b. unpaired observations……..…………………...Wilcoxon Rank Sum test

9b. **>2 samples**

13a. Normal

14a. 1-factor…………………………………………………1-way ANOVA

14b. >1 factor…………………………………………multi-way ANOVA

13b. Not normal

15a. 1-factor………………………………..…………Kruskal-Wallis test

15b. complete randomized blocked design………..Friedman's test (~χ^2 test)

Adapted From: http://www.oakleafsystems.net/StatKey/Keyx.html