Class 3, Chapter 2: Inferences using t-distributions

2/4/09 W

Slide 1 Class 3, Chapter 2: Inferences using t-distributions

NOTES:

HW 3 for Mon 2/9/09 9:50

Submit as Myname-HW3.doc (or *.rtf)

- Finish Chapter 2 and start on Chapter 3 “A closer look at assumptions”
  - Read Sterne & Smith (2001) “Sifting the evidence” [Discusses p values & significance testing]
- **Conceptual exercises, Chapter 2**
  - Post ≥1 message & ≥1 reply to a message on the Blackboard Vista 4 discussion section.
- **Chapter 2 computation problems** (SPSS sdta on Blackboard Vista 4)
  - 2.21Bumpus’s data: weights of Bumpus’s birds

Slide 2 HW 3 for Mon 2/9/09 9:50

NOTES:

HW 4 due Thus 2/12/09 11 am

Submit as Myname-HW4.doc (or *.rtf)

- Finish Ch 3 for Weds’ class
  - Chapter 3: A closer look at assumptions
  - Read
    - Hayek & Buzas (1997, on sampling)
    - Hurlbert (1984) on Pseudoreplication
    - **Post one comment and one reply to issues raised in Hayek & Buzas or Hurlbert (1984)**
- Chapter 3 problem due Thus
  - **3.28 Pollen removal**

Slide 3 HW 4 due Thus 2/12/09 11 am

NOTES:

Fisher's major contribution to statistics: randomization

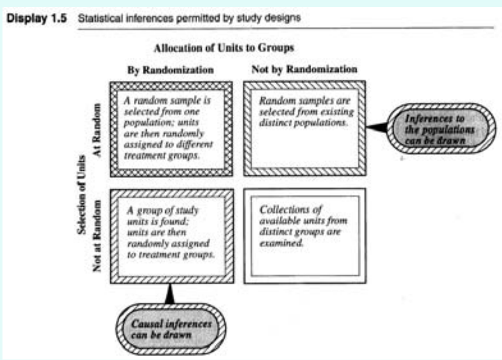
http://bmj.com/cgi/content/full/322/7280/0

"The modern solution was first propounded by R. A. Fisher. We have already seen throughout this work that Fisher's contributions to statistical theory were remarkable and far-ranging. Nevertheless, it is probably no exaggeration to say that his advocacy of randomization in experimental design was the most important and the most influential of his many achievements in statistics."
Kendall & Stuart 1977

EEOS611

Slide 4 Fisher's major contribution to statistics: randomization

NOTES:



Slide 5

NOTES:

Statistical inferences and chance mechanisms

- An **inference** is a conclusion that patterns in the data are present in some broader context
- A **statistical inference** is an inference justified by a probability model linking the data to the broader context

EEOS611

Slide 6 Statistical inferences and chance mechanisms

NOTES:

Randomization

From Kendall & Stuart's 'Advanced Theory of Statistics'

- **The principle of randomization is simply stated: Whenever experimental units are allocated to factor-combinations in an experiment, this should be done by a random process using equal probabilities.**
- Even if the relationship of the dependent variable with some unsuspected causal factor is not recognized until after the experiment, the validity of the inferences will not be impaired, provided that the factor's influence was "**randomized out"** of the experiment.

EEOS611

Slide 7 Randomization

NOTES:

Kendall & Stuart on Experiments

Three classes of variables

- In any experiment the factors influencing the dependent variable are, explicitly or implicitly, divided by the experimenter into three classes:
  - Those incorporated into the structure of the experiment
  - Those "randomized out" of the experiment
  - Those neither incorporated nor randomized out
- Classes 1 & 2 require positive action, affecting the layout of the experiment, or the randomization procedure employed. A factor may find its way into class (3) by simply being overlooked.

EEOS611

Slide 8 Kendall & Stuart on Experiments

NOTES:

What makes a good experimenter?

Kendall & Stuart (1977)

"A substantial part of the skill of the experimenter lies in his choice of factors to be randomized out of the experiment. If he is careful, he will randomize out all the factors which are suspected of being causally important but which are not actually part of the experimental procedure. But every experimenter necessarily neglects some conceivably causal factors; if this were not so, the randomization procedure required would be impossibly complicated. Thus the choice of what factors to be randomized out is essentially a matter of judgement."

Slide 9 What makes a good experimenter?

NOTES:

Experimental design should include:

Hurlbert (1984), posted on Blackboard/Vista4

- The nature of the experimental units to be employed
- The number and kinds of treatments and the properties of the responses that will be measured.
- Specification of how the treatments will be assigned to the available experimental units (replicates)
- The physical arrangement of the experimental units, (and often) the temporal sequence in which treatments are applied to and measurements made on the different experimental units.'

EEOS611

Slide 10 Experimental design should include:

NOTES:

Randomized Experiments vs. Observational Studies

- Randomized experiment: a chance mechanism used to assign subject to groups
- Observational study: group status beyond the control of the investigator
- **"Statistical inferences of cause-and-effect relationships can be drawn from randomized experiments, but not from observational studies"**
- "A **confounding variable** is related both to group membership and to the outcome. Its presence makes it hard to establish the outcome as being a direct consequence of group membership." (Male experience)

EEOS611

Slide 11 Randomized Experiments vs. Observational Studies

NOTES:

Sample surveys *vs.* experiments

Kendall & Stuart's "The Advanced theory of statistics" (1977)

- The distinction between the design of experiments and the design of sample surveys is fairly clear-cut, and may be expressed by saying that
  - **In surveys we make observations on a sample taken from a finite population of individuals, whereas in experiments we make observations which are in principle generated by a hypothetical infinite population, in exactly the same way that the tosses of a coin are.**
  - Of course, we may sometimes experiment on the members of a sample resulting from a survey, or even make a sample survey of the results of an (extensive) experiment, but the essential distinction between the two fields should be clear.

EEOS611

Slide 12 Sample surveys vs. experiments

NOTES:

Do observational studies have value?

- Establishing causation not always the goal of the study
- Establishing causation can be done in other ways.
- Analysis of observational data may lend evidence toward causal theories and suggest the direction for further research.

EEOS611

Slide 13 Do observational studies have value?

NOTES:

Inferences to populations

- Inferences to populations can be drawn from random sampling studies, but not otherwise
  - **Simple random sampling** (SRS): A simple random sample of size n from a population is a subset of the population consisting of *n* members selected in such a way that every subset of size n is afforded the same chance of being selected.
- Random sampling ensures that all subpopulations are represented in the sample in roughly the same mix as in the overall population.
- Statistical inference procedures incorporate measures of uncertainty that describe that chance.

EEOS611

Slide 14 Inferences to populations

NOTES:

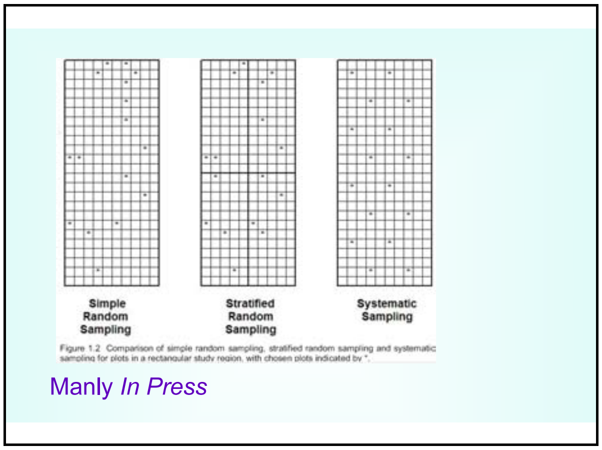
Selecting a random sample

The type of sampling can dictate the analysis used.

- Simple random sampling
- Stratified random sampling
- Multilevel sampling (e.g., Regions, Lakes, areas within lakes)
- Systematic sampling
  - Quadrat samples
  - Line transect samples: see Hayek & Buzas (1996)
- Random cluster sampling (selecting blocks or grids at random)
  - Lakes: Can adjust the probability of different types of lakes being sampled
- Variable probability sampling
  - EMAP sampling of estuaries
- Adaptive sampling
  - Adaptive cluster sampling (Thompson 1990)
  - Randomized 'play the winner strategies' (Wei 1988, Biometrika 75: 603-606

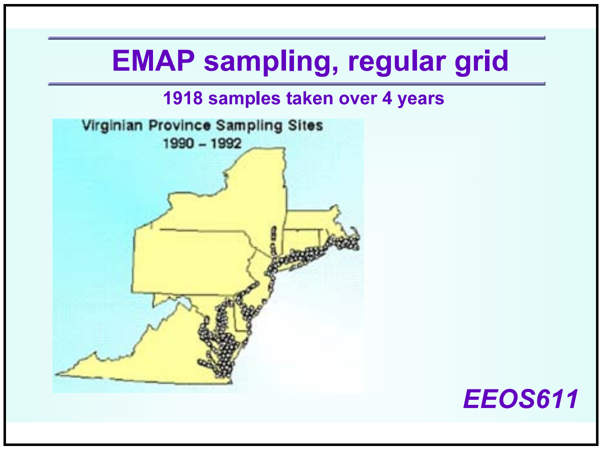
Slide 15 Selecting a random sample

NOTES:



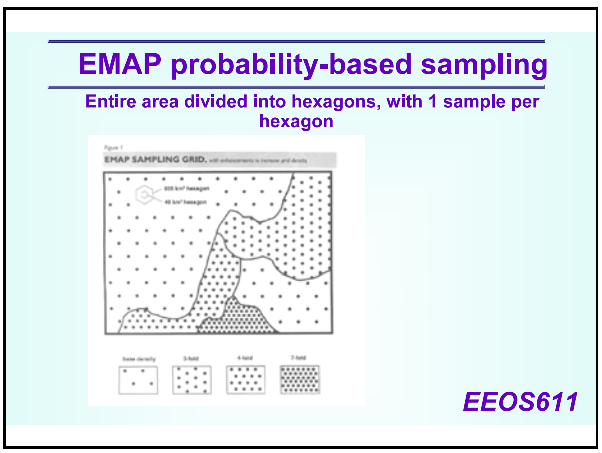
Slide 16

NOTES:



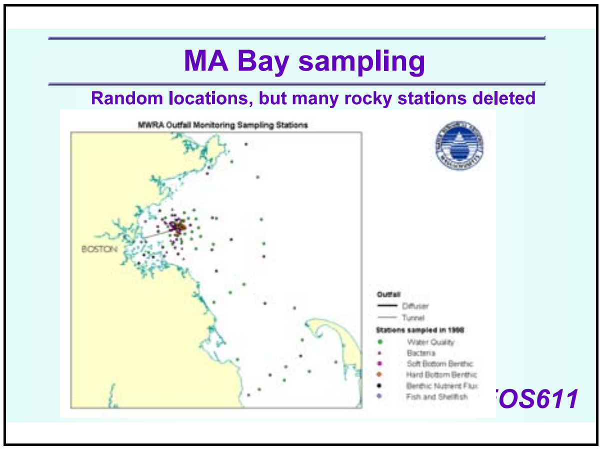
Slide 17 EMAP sampling, regular grid

NOTES:



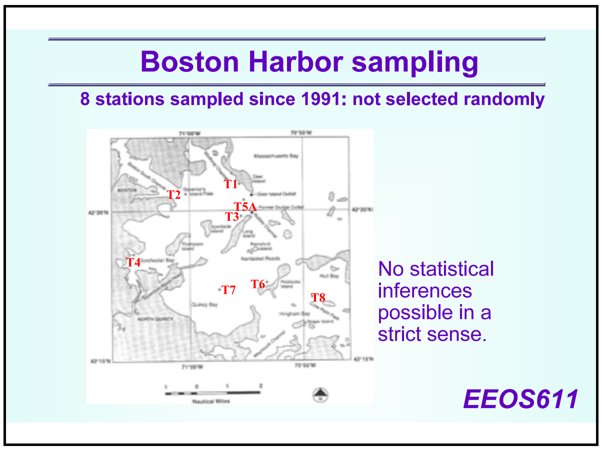
Slide 18 EMAP probability-based sampling

NOTES:



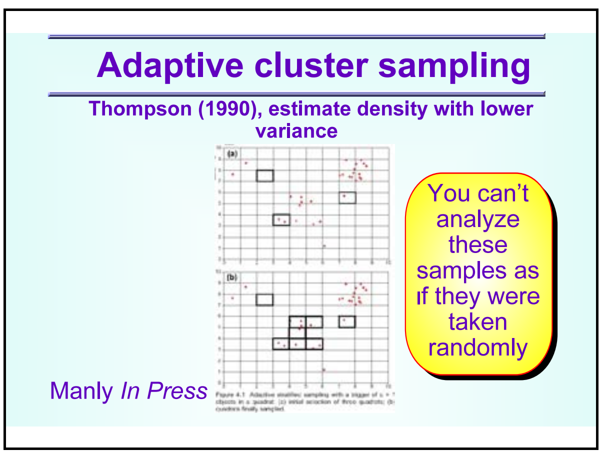
Slide 19 MA Bay sampling

NOTES:



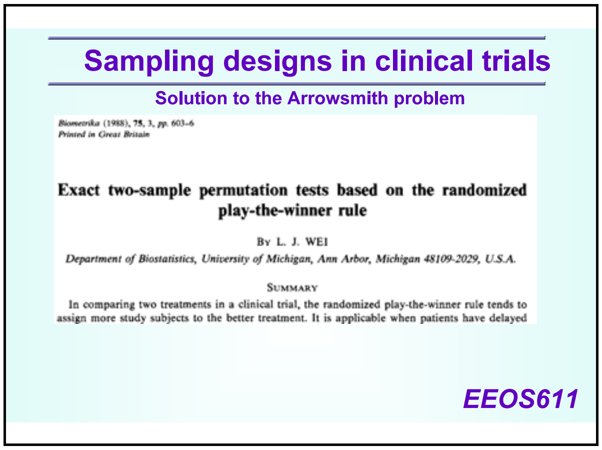
Slide 20 Boston Harbor sampling

NOTES:



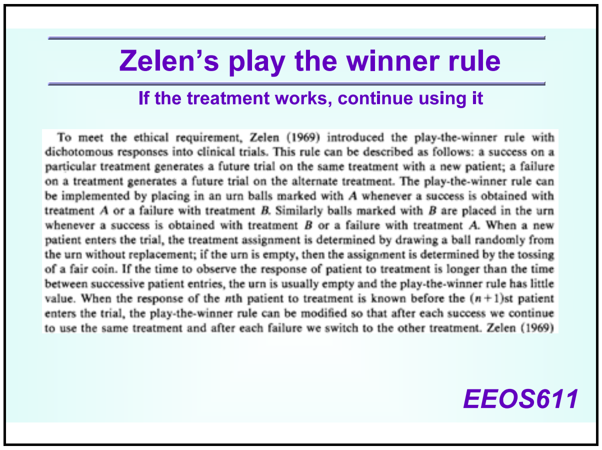
Slide 21 Adaptive cluster sampling

NOTES:



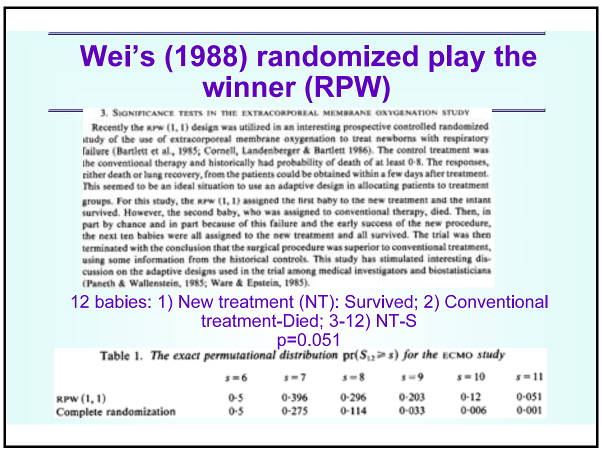
Slide 22 Sampling designs in clinical trials

NOTES:



Slide 23 Zelen's play the winner rule

NOTES:



Slide 24 Wei's (1988) randomized play the winner (RPW)

NOTES:

Slide 25 Statistical inference

Statistical inference

& Neyman-Pearson Hypothesis testing

NOTES:

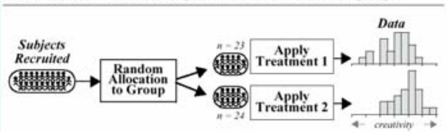
Slide 26 A probability model for randomized experiments

A probability model for randomized experiments

- The creativity study is an example
- An additive model: $Y^*=Y+\delta$

Display 1.6 p. 10

Illustration of a randomized experiment with two treatment groups

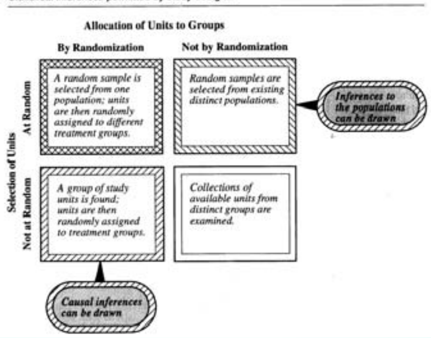


EECS611

NOTES:

Slide 27

Display 1.5 Statistical inferences permitted by study designs



NOTES:

Null & alternate hypotheses

Page 10

- "Is there a treatment effect?" must be translated into a model that can be tested statistically $Y^*=Y+\delta$, where δ is the treatment effect
- Create a test statistic
  - Assume a creativity parameter δ
    - $\delta=0$ is the null hypothesis
    - $\delta\neq0$ is the alternate hypothesis
  - Randomization distribution of the test statistic
  - The p-value of the test, derived from the randomization assumption

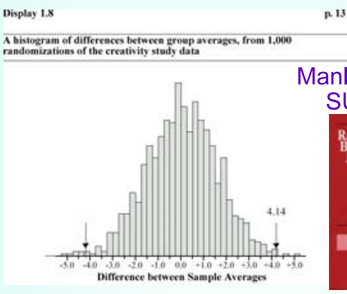
EEOS611

Slide 28 Null & alternate hypotheses

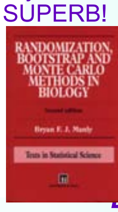
NOTES:

Randomization distribution

Can be done with Matlab & R, not SPSS



Manly's book is SUPERB!



EEOS611

Slide 29 Randomization distribution

NOTES:

Computing p values using randomization & Monte Carlo trials

- All possible permutations: not feasible for many studies
- Set the number of Monte Carlo simulations at about 4*1/(desired precision of the p value)
  - See: How many Monte Carlo Simulations Should You Run? See Gallagher's HO13-MCTRIALS.pdf
- Or, approximate the randomization distribution with a normal or *t* distribution

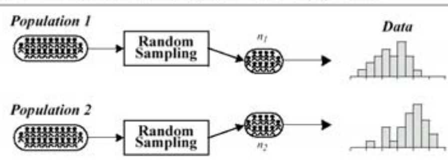
Slide 30 Computing p values using randomization & Monte Carlo trials

NOTES:

Measuring uncertainty in observational studies

Display 1.9 p. 14

Illustration of a random sampling study with two populations



EEOS611

Slide 31 Measuring uncertainty in observational studies

NOTES:

Related issues

- Relative frequency histograms
- Stem and leaf diagrams: poor in SPSS
- Box plots, box-and-whisker plot
- Standard statistical terminology
  - A parameter, a feature of a probability model. Parameters indicated by Greek letters.
  - Statistic: any quantity that can be calculated from the observed data.
    - Mean in statistical sleuth is over the entire population: it is a parameter
    - Standard deviation
  - Experimental units: the things to which treatments are applied

EEOS611

Slide 32 Related issues

NOTES:

Sleuth Chapter 2

Inference using t-distributions

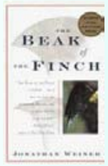
Slide 33 Sleuth Chapter 2

NOTES:

Weiner's account of Bumpus data

1994. The beak of the finch: a story of evolution in our time. Alfred A. Knopf, New York.

- **English sparrows had been introduced in New York's Central Park in 1851. An eccentric bird lover wanted to import every one of the birds in Shakespeare's plays to the United States. "So the birds were lying in the snow that morning in part because Shakespeare had written, '*There is a special providence in the fall of a sparrow.*'**
- **Last day of January 1898, huge storm, large number of English sparrows lay dead**



EEOS611

Slide 34 Weiner's account of Bumpus data

NOTES:

Bumpus sparrow data

Stem-and-leaf plot

Display 2.1 Humerus lengths (inches) of adult male house sparrows, 24 that perished and 35 that survived in a winter storm

Perished
Average: .7279
SD: .0235
n: 24

Survived
Average: .7380
SD: .0198
n: 35

Perished	Stem	Survived
9	65	
	66	
	67	
9	68	7
	69	
932	70	39
3	71	5
96600	72	13368889
988761	73	0033569
543	74	111139
422	75	12256
5	76	679
	77	0
	78	0

Legend: | 68 | 7 represents 0.687 inch.

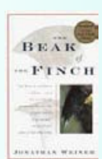
Slide 35 Bumpus sparrow data

NOTES:

Bumpus's sparrow data

From Weiner (1994,p. 227-228) "The Beak of the Finch"

- **In the early 1970s, Peter Grant reanalyzed Bumpus's data, "He concluded that Bumpus had actually seen not one but two kinds of natural slection. For the female sparrows the storm was stabilizing. The event killed the largest and the smallest but preserved the mean, just as Bumpus had said. In the males, however, the pressure of the storm was directional, pushing the birds toward smaller size. The reanalysis of Bumpus's classic data helped inspire the Grants' first trip to the Galapagos."**



Slide 36 Bumpus's sparrow data

NOTES:

Anatomical abnormalities & schizophrenia

Case 2.2: 15 pairs of twins, paired *t* test

Display 2.2

Differences in volumes (cm^3) of left hippocampus in fifteen sets of monozygotic twins where one twin is affected by schizophrenia

Pair #	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

Differences

Stem	Leaves
-2	
-1	9
-0	
0	23479
1	0139
2	3
3	
4	0
5	09
6	7
7	

Average: 0.199
Sample SD: 0.238
n: 15

Legend: | 6 | 7 represents 0.67 cm^3

Slide 37 Anatomical abnormalities & schizophrenia

NOTES:

Case 2.2 Statistical Summary

Sleuth, p. 31

There is substantial evidence that the mean difference in the left hippocampus volumes between schizophrenic individuals and their nonschizophrenic twins is nonzero (two-sided *p-value = 0.006, from a paired t test). It is estimated that the the mean volume is 0.20 cm^3 smaller for thoose with schizophrenia (about 11% smaller). A 95% confidence interval for the difference is from 0.07 to 0.33 cm^3.*

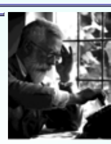
Slide 38 Case 2.2 Statistical Summary

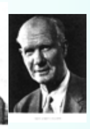
NOTES:

Statistical Summary includes elements of Fisher, Neyman-Pearson & Deming

- **Fisher**
  - Randomization & causation
  - P values
- **Neyman-Pearson**
  - Critical values: significant vs. Non-significant
  - 95% confidence Intervals
- **A. E. Deming effect sizes**

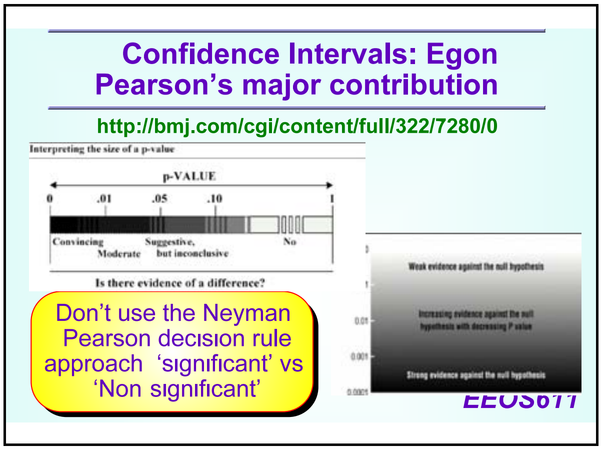
http://www.stat.ucla.edu/history/people/





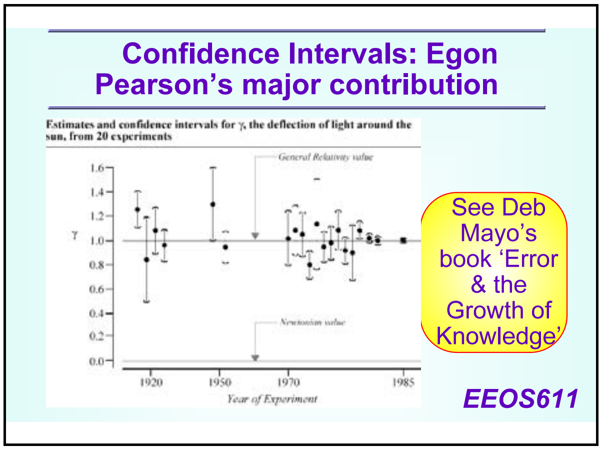

Slide 39 Statistical Summary includes elements of Fisher, Neyman-Pearson & Deming

NOTES:



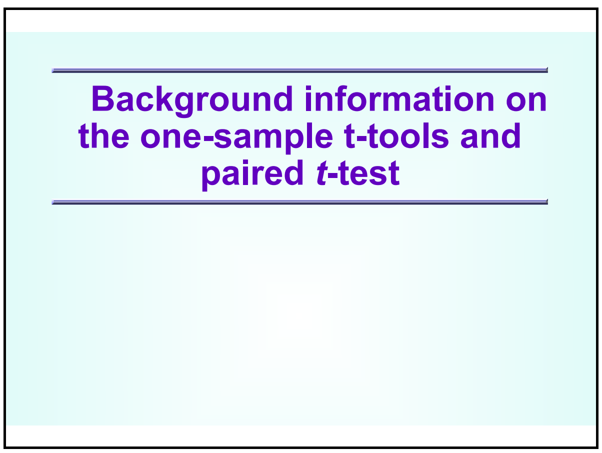
Slide 40 Confidence Intervals: Egon Pearson's major contribution

NOTES:



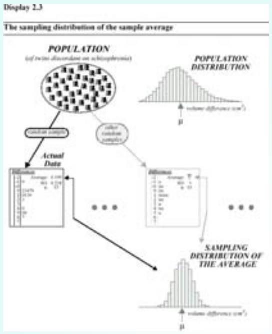
Slide 41 Confidence Intervals: Egon Pearson's major contribution

NOTES:



Slide 42 Background information on the one-sample t-tools and paired t-test

NOTES:

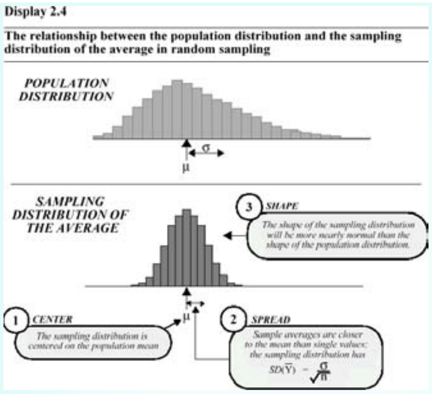


No matter what the underlying distribution, the sampling distribution of the sample averages will be 'more nearly normal' than the underlying distribution. This is a result of the **Central Limit Theorem**.

http://mathworld.wolfram.com/CentralLimitTheorem.html

Slide 43

NOTES:



EEOS611

Slide 44

NOTES:

Expressing uncertainty

X ± δ, >4 choices for δ {del}

- ± standard deviation
  - Advantage: expressed in the same units as the parameter being estimated
  - Not a good choice, some statisticians argue that it is inappropriate for a sample statistic
- ± standard error, also called the standard deviation of the average or sd of the mean or the standard error of the mean
  - $SE(\bar{x}) = s/\sqrt{n}$, d.f.=(n-1)
  - Advantage: statistically appropriate
  - Needs the sample size for interpretation
- ± half the 95% confidence interval
  - Assumes an underlying model for the data
  - For asymmentric 95% CI's in the statistics natural scale, provide the upper and lower 95% CI (for transformed data and results of Monte Carlo simuations
- Analytical precision of the instrument or technique
  - E.g., the Chl *a* method has a certain analytical precision. This is rarely acceptable.
  - Hurlbert's $E(S_{100})$, a measure of species richness, has an analytical precision based on the sampling properties of the hypergeometric distribution.
  - Polling data has an analytical precision based on var=p*(1-p), but this rarely expresses the other sources of variability in a poll

Slide 45 Expressing uncertainty

NOTES:

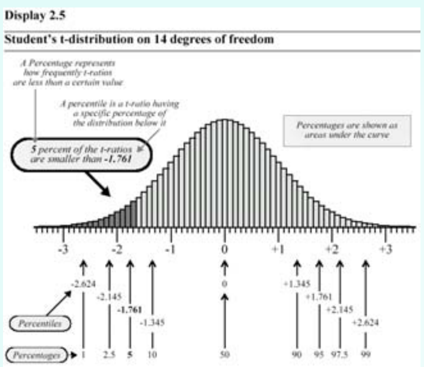
The Z-ratio & t-ratio based on a sample average

- Z-ratio = (Estimate - Parameter)/ SD(Estimate)
  - If the sampling distribution is normal, than the sampling distribution of Z is standard normal
  - Mean zero and standard deviation of 1
  - Z distribution provided in Appendix A.1
  - t-ratio = (Estimate - Parameter)/SE (Estimate)
  - If $\bar{x}$ is the average in a random sample of size n from a normally distributed population, the sampling distribution of *t* is described by the Student's *t* distribution on *n-1* degrees of freedom

EEOS611

Slide 46 The Z-ratio & t-ratio based on a sample average

NOTES:



Slide 47

NOTES:

Degrees of freedom **Box 1.2**

Statistical tests of significance often call upon the concept of degrees of freedom. A formal definition is the following: "The degrees of freedom of a model for expected values of random variables is the excess of the number of variables [observations] over the number of parameters in the model" (Kotz & Johnson, 1982).

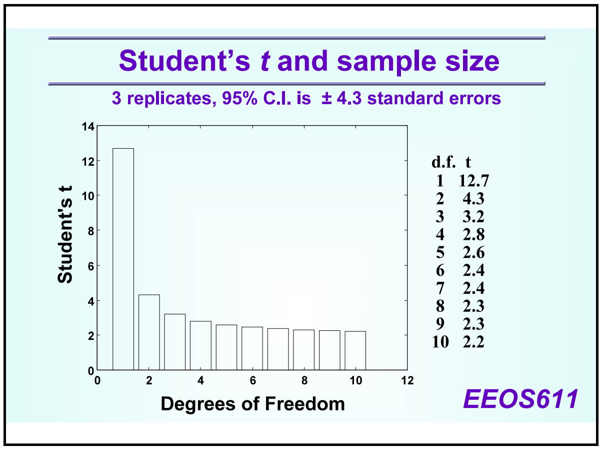
In practical terms, the number of degrees of freedom associated with a statistic is equal to the number of its independent components, i.e. the total number of components used in the calculation minus the number of parameters one had to estimate from the data before computing the statistic. For example, the number of degrees of freedom associated with a variance is the number of observations minus one (noted $\nu = n - 1$): n components $(x_i - \bar{x})$ are used in the calculation, but one degree of freedom is lost because the mean of the statistical population is estimated from the sample data; this is a prerequisite before estimating the variance.

There is a different *t* distribution for each number of degrees of freedom. The same is true for the *F* and χ^2 families of distributions, for example. So, the number of degrees of freedom determines which statistical distribution, in these families (*t*, *F*, or χ^2), should be used as the reference for a given test of significance. Degrees of freedom are discussed again in Chapter 6 with respect to the analysis of contingency tables.

Legendre & Legendre (1989) Numerical Ecology 2nd Ed.

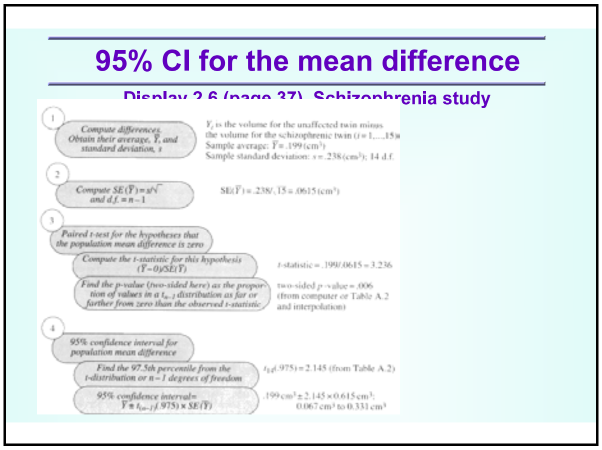
Slide 48

NOTES:



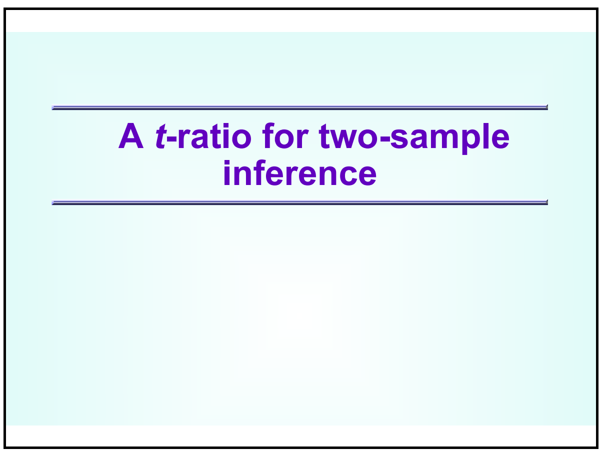
Slide 49 Student's t and sample size

NOTES:



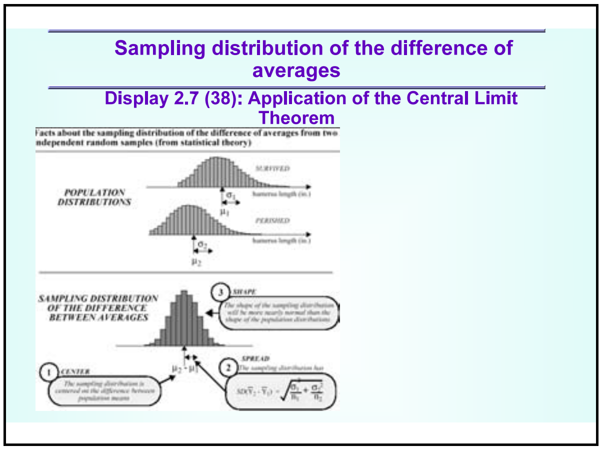
Slide 50 95% CI for the mean difference

NOTES:



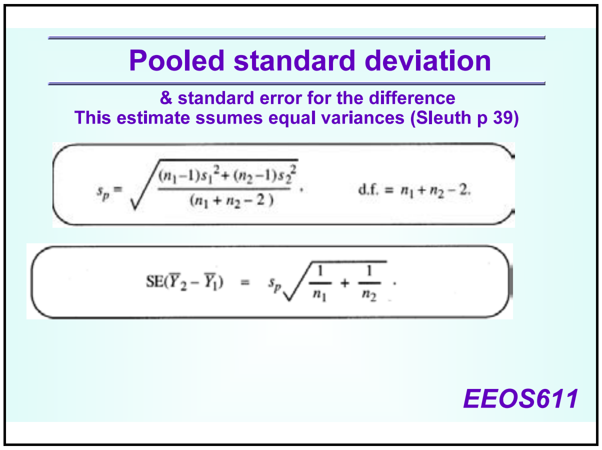
Slide 51 A t-ratio for two-sample inference

NOTES:



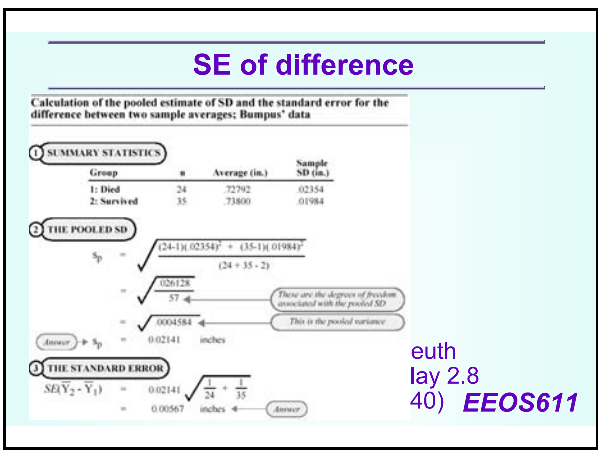
Slide 52 Sampling distribution of the difference of averages

NOTES:



Slide 53 Pooled standard deviation

NOTES:



Slide 54 SE of difference

NOTES:

95% Confidence Limits

For the difference between means

100(1−α)% Confidence Limits for the Difference Between Means

$$(\bar{Y}_2 - \bar{Y}_1) \pm t_{df}(1-\alpha/2)SE(\bar{Y}_2 - \bar{Y}_1).$$

"A 95% confidence interval will contain the parameter if the t-ratio from the observed data happens to be one of those in the middle 95% of the sampling distribution. Since 95% of all possible pairs of samples lead to such t-ratios, **it is safe to say that the procedure of constructing a 95% CI is successful in 95% of its applications."**

It is incorrect to say that there is a 95% probability that the true parameter is within the 95% CI. **That** probability is either 0 or 1. Bayesians have a different interpretation of p values.

Slide 55 95% Confidence Limits

NOTES:

CI for difference of means

Sleuth 2e Display 2.9 (41)

Construction of a 95% confidence interval for the difference between the mean humerus lengths of sparrows that died and that survived

Group	n	Average (in.)	SD (in.)
1: Died	24	.72792	.02354
2: Survived	35	.73800	.01984

$\bar{Y}_2 - \bar{Y}_1$ = .73800 - .72792 = 0.01008

$SE(\bar{Y}_2 - \bar{Y}_1)$ = 0.00567 inches

From Display 2.8

degrees of freedom = 24 + 35 - 2 = 57

$t_{57}(.975)$ = 2.002

from tables of the t-distribution with 57 degrees of freedom

Half-width = (2.002)(0.00567) = 0.01136

Lower 95% confidence limit = 0.01008 - 0.01136 = -0.00128 inches

Upper 95% confidence limit = 0.01008 + 0.01136 = 0.02144 inches

Slide 56 CI for difference of means

NOTES:

Testing a hypotheses about the difference of means

$$t\text{-statistic} = \frac{(\bar{Y}_2 - \bar{Y}_1) - [\text{Hypothesized value for } (\mu_2 - \mu_1)]}{SE(\bar{Y}_2 - \bar{Y}_1)}.$$

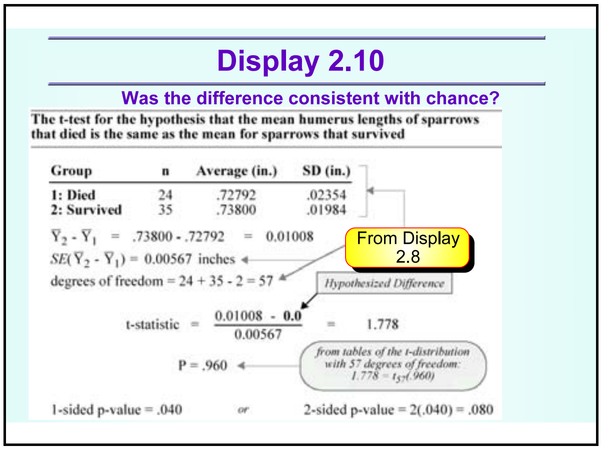
"The p-value for a *t*-test is a probability of obtaining a t-ratio as extreme or more extreme than the t-statistic in its evidence against the null hypothesis, if the null hypothesis is correct." (Sleuth 2nd ed. p. 42). [Bayesians do not use this interpretation.]

A large p value means that the study is not capable of excluding the null hypothesis as a possible explanation ... **It is wrong to conclude that the null hypothesis is true.**

EEOS611

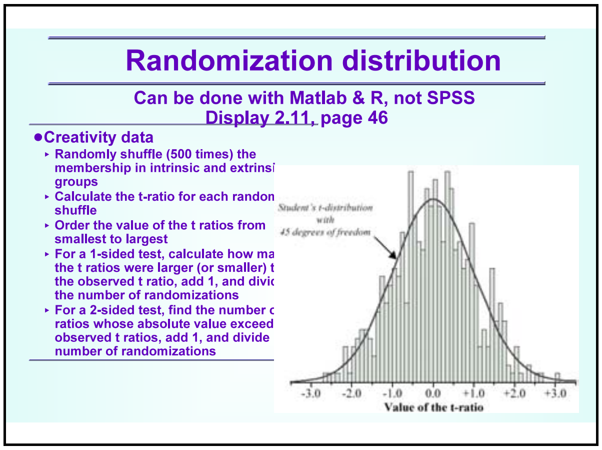
Slide 57 Testing a hypotheses about the difference of means

NOTES:



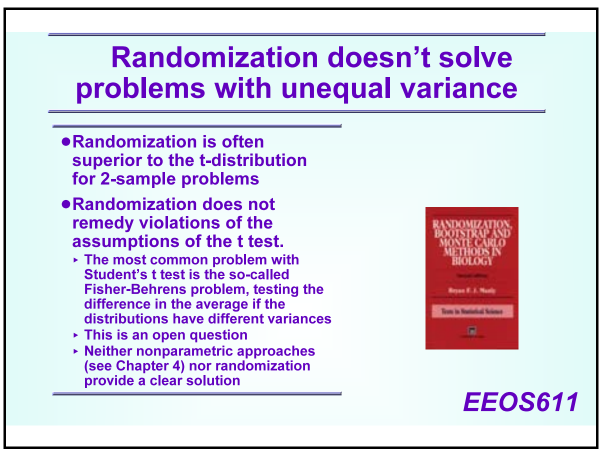
Slide 58 Display 2.10

NOTES:



Slide 59 Randomization distribution

NOTES:



Slide 60 Randomization doesn't solve problems with unequal variance

NOTES: