

<p style="text-align: center;">Class 3, Chapter 2: Inferences using t- distributions</p> <p style="text-align: center;">2/4/09 W</p>	<p>Slide 1 Class 3, Chapter 2: Inferences using t-distributions</p> <p>NOTES:</p>
<p style="text-align: center;">HW 3 for Mon 2/9/09 9:50</p> <p style="text-align: center;">Submit as Myname-HW3.doc (or *.rtf)</p> <ul style="list-style-type: none"> ● Finish Chapter 2 and start on Chapter 3 “A closer look at assumptions” <ul style="list-style-type: none"> ▸ Read Sterne & Smith (2001) “Sifting the evidence” [Discusses p values & significance testing] ● Conceptual exercises, Chapter 2 <ul style="list-style-type: none"> ▸ 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) <ul style="list-style-type: none"> ▸ 2.21Bumpus’s data: weights of Bumpus’s birds 	<p>Slide 2 HW 3 for Mon 2/9/09 9:50</p> <p>NOTES:</p>
<p style="text-align: center;">HW 4 due Thus 2/12/09 11 am</p> <p style="text-align: center;">Submit as Myname-HW4.doc (or *.rtf)</p> <ul style="list-style-type: none"> ● Finish Ch 3 for Weds’ class <ul style="list-style-type: none"> ▸ Chapter 3: A closer look at assumptions ▸ Read <ul style="list-style-type: none"> ■ 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 <ul style="list-style-type: none"> ▸ 3.28 Pollen removal 	<p>Slide 3 HW 4 due Thus 2/12/09 11 am</p> <p>NOTES:</p>

Fisher's major contribution to statistics: randomization

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



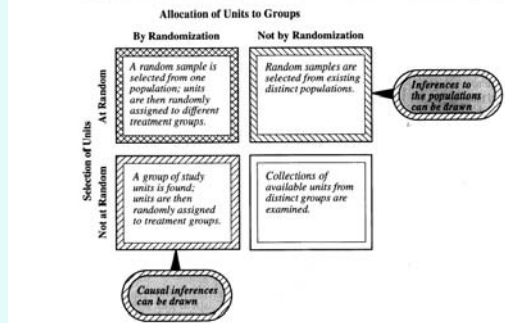
"The modern solution was first propounded by R. A. Fisher. We have already seen throughout this work that Fisher's contributions to statistical theory were remarkable and far-ranging. Nevertheless, it is probably no exaggeration to say that his advocacy of **randomization** in experimental design was 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:

Display 1.5 Statistical inferences permitted by study designs



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:

<p style="text-align: center;">Randomization</p> <p style="text-align: center;">From Kendall & Stuart's 'Advanced Theory of Statistics'</p> <ul style="list-style-type: none"> • 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. <p style="text-align: right;"><i>EEOS611</i></p>	<p style="text-align: center;">Slide 7 Randomization</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">Kendall & Stuart on Experiments</p> <p style="text-align: center;">Three classes of variables</p> <ul style="list-style-type: none"> • In any experiment the factors influencing the dependent variable are, explicitly or implicitly, divided by the experimenter into three classes: <ul style="list-style-type: none"> ▪ 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. <p style="text-align: right;"><i>EEOS611</i></p>	<p style="text-align: center;">Slide 8 Kendall & Stuart on Experiments</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">What makes a good experimenter?</p> <p style="text-align: center;">Kendall & Stuart (1977)</p> <p>"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."</p>	<p style="text-align: center;">Slide 9 What makes a good experimenter?</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

<p style="text-align: center;">Experimental design should include:</p> <p style="text-align: center;">Hurlbert (1984), posted on Blackboard/Vista4</p> <ul style="list-style-type: none"> • 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.' <p style="text-align: right;">EEOS611</p>	<p style="text-align: center;">Slide 10 Experimental design should include:</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">Randomized Experiments vs. Observational Studies</p> <ul style="list-style-type: none"> • 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) <p style="text-align: right;">EEOS611</p>	<p style="text-align: center;">Slide 11 Randomized Experiments vs. Observational Studies</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">Sample surveys vs. experiments</p> <p style="text-align: center;">Kendall & Stuart's "The Advanced theory of statistics" (1977)</p> <ul style="list-style-type: none"> • 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. <p style="text-align: right;">EEOS611</p>	<p style="text-align: center;">Slide 12 Sample surveys vs. experiments</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

<p style="text-align: center;">Do observational studies have value?</p> <ul style="list-style-type: none"> • 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. <p style="text-align: right;"><i>EEOS611</i></p>	<p>Slide 13 Do observational studies have value?</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">Inferences to populations</p> <ul style="list-style-type: none"> • 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. <p style="text-align: right;"><i>EEOS611</i></p>	<p>Slide 14 Inferences to populations</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>
<p style="text-align: center;">Selecting a random sample</p> <p style="text-align: center;">The type of sampling can dictate the analysis used.</p> <ul style="list-style-type: none"> • Simple random sampling • Stratified random sampling • Multilevel sampling (e.g., Regions, Lakes, areas within lakes) • Systematic sampling <ul style="list-style-type: none"> ▸ Quadrat samples ▸ Line transect samples: see Hayek & Buzas (1996) • Random cluster sampling (selecting blocks or grids at random) <ul style="list-style-type: none"> ▸ Lakes: Can adjust the probability of different types of lakes being sampled • Variable probability sampling <ul style="list-style-type: none"> ▸ EMAP sampling of estuaries • Adaptive sampling <ul style="list-style-type: none"> ▸ Adaptive cluster sampling (Thompson 1990) ▸ Randomized 'play the winner strategies' (Wei 1988, Biometrika 75: 603-606) 	<p>Slide 15 Selecting a random sample</p> <hr/> <p>NOTES:</p> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/> <hr/>

Simple Random Sampling Stratified Random Sampling Systematic Sampling

Figure 1.2 Comparison of simple random sampling, stratified random sampling and systematic sampling for plots in a rectangular study region, with chosen plots indicated by *.

Manly In Press

Slide 16
NOTES:

EMAP sampling, regular grid

1918 samples taken over 4 years

Virginian Province Sampling Sites
1990 - 1992

EEOS611

Slide 17 EMAP sampling, regular grid
NOTES:

EMAP probability-based sampling

Entire area divided into hexagons, with 1 sample per hexagon

Figure 1
EMAP SAMPLING GRID, with adjustments to density and density

1 dot hexagon 3 dot hexagon 4 dot hexagon 7 dot hexagon

low density 3 dot 4 dot 7 dot

EEOS611

Slide 18 EMAP probability-based sampling
NOTES:

MA Bay sampling

Random locations, but many rocky stations deleted

OS611

Slide 19 MA Bay sampling

NOTES:

Boston Harbor sampling

8 stations sampled since 1991: not selected randomly

No statistical inferences possible in a strict sense.

EEOS611

Slide 20 Boston Harbor sampling

NOTES:

Adaptive cluster sampling

Thompson (1990), estimate density with lower variance

You can't analyze these samples as if they were taken randomly

Manly *In Press*

Figure 4.1 Adaptive cluster sampling with a region of $n = 1$ objects in a quadrant. (a) Initial selection of three quadrants; (b) quadrants fully sampled.

Slide 21 Adaptive cluster sampling

NOTES:

Sampling designs in clinical trials

Solution to the Arrowsmith problem

Biometrika (1988), 75, 3, pp. 603-6
Printed in Great Britain

Exact two-sample permutation tests based on the randomized play-the-winner rule

By L. J. WEI
Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109-2029, U.S.A.

SUMMARY

In comparing two treatments in a clinical trial, the randomized play-the-winner rule tends to assign more study subjects to the better treatment. It is applicable when patients have delayed

EEOS611

Slide 22 Sampling designs in clinical trials

NOTES:

Zelen's play the winner rule

If the treatment works, continue using it

To meet the ethical requirement, Zelen (1969) introduced the play-the-winner rule with dichotomous responses into clinical trials. This rule can be described as follows: a success on a particular treatment generates a future trial on the same treatment with a new patient; a failure on a treatment generates a future trial on the alternate treatment. The play-the-winner rule can be implemented by placing in an urn balls marked with A whenever a success is obtained with treatment A or a failure with treatment B. Similarly balls marked with B are placed in the urn whenever a success is obtained with treatment B or a failure with treatment A. When a new patient enters the trial, the treatment assignment is determined by drawing a ball randomly from the urn without replacement; if the urn is empty, then the assignment is determined by the tossing of a fair coin. If the time to observe the response of patient to treatment is longer than the time between successive patient entries, the urn is usually empty and the play-the-winner rule has little value. When the response of the *n*th patient to treatment is known before the (*n*+1)st patient enters the trial, the play-the-winner rule can be modified so that after each success we continue to use the same treatment and after each failure we switch to the other treatment. Zelen (1969)

EEOS611

Slide 23 Zelen's play the winner rule

NOTES:

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

3. SIGNIFICANCE TESTS IN THE EXTRACORPORAL MEMBRANE OXYGENATION STUDY

Recently the rpw (1, 1) design was utilized in an interesting prospective controlled randomized study of the use of extracorporeal membrane oxygenation to treat newborns with respiratory failure (Bartlett et al., 1985; Cornell, Landenberger & Bartlett 1986). The control treatment was the conventional therapy and historically had probability of death of at least 0.8. The responses, either death or lung recovery, from the patients could be obtained within a few days after treatment. This seemed to be an ideal situation to use an adaptive design in allocating patients to treatment groups. For this study, the rpw (1, 1) assigned the first baby to the new treatment and the infant survived. However, the second baby, who was assigned to conventional therapy, died. Then, in part by chance and in part because of this failure and the early success of the new procedure, the next ten babies were all assigned to the new treatment and all survived. The trial was then terminated with the conclusion that the surgical procedure was superior to conventional treatment, using some information from the historical controls. This study has stimulated interesting discussion on the adaptive designs used in the trial among medical investigators and biostatisticians (Paneth & Wallenstein, 1985; Ware & Epstein, 1985).

12 babies: 1) New treatment (NT): Survived; 2) Conventional treatment-Died; 3-12) NT-S

$p=0.051$

Table 1. The exact permutational distribution $\text{pr}(S_{12} \geq s)$ for the ECMO study

	<i>s</i> = 6	<i>s</i> = 7	<i>s</i> = 8	<i>s</i> = 9	<i>s</i> = 10	<i>s</i> = 11
RPW (1, 1)	0.5	0.396	0.296	0.203	0.12	0.051
Complete randomization	0.5	0.275	0.114	0.033	0.006	0.001

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

NOTES:

Statistical inference
& Neyman-Pearson Hypothesis testing

Slide 25 Statistical inference

NOTES:

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

Slide 26 A probability model for randomized experiments

NOTES:

Display 1.5 Statistical inferences permitted by study designs

		Allocation of Units to Groups		
		By Randomization	Not by Randomization	
Selection of Units	At Random	A random sample is selected from one population; units are then randomly assigned to different treatment groups.	Random samples are selected from existing distinct populations.	Inferences to the populations can be drawn
	Not at Random	A group of study units is found; units are then randomly assigned to treatment groups.	Collections of available units from distinct groups are examined.	
		Causal inferences can be drawn		

Slide 27

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
 - $\delta \neq 0$ 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

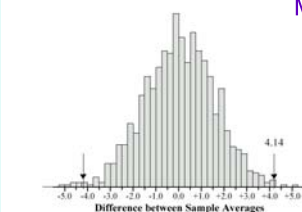
NOTES:

Randomization distribution

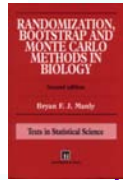
Can be done w/ Matlab & R, not SPSS

Display 1.8 p. 13

A histogram of differences between group averages, from 1,000 randomizations of the creativity study data



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 $4*1/(\text{desired precision of the p value})$
 - 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

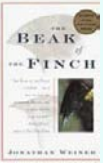
Slide 33 Sleuth Chapter 2

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

<p>Perished Average: .7279 SD: .0235 n: 24</p>	<pre> 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 </pre>	<p>Survived Average: .7380 SD: .0198 n: 35</p>
---	--	---

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 selection. 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 (cm³) 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	Average: 0.199
-2	Sample SD: 0.238
-1	n: 15
0	
1	23479
2	0139
3	
4	0
5	09
6	7
7	

Legend: | 6 | 7 represents 0.67 cm³

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 cm³ smaller for those with schizophrenia (about 11% smaller). A 95% confidence interval for the difference is from 0.07 to 0.33 cm³.

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.ucda.edu/history/people/>

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

NOTES:

Confidence Intervals: Egon Pearson's major contribution

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

Interpreting the size of a p-value

Is there evidence of a difference?

Don't use the Neyman Pearson decision rule approach 'significant' vs 'Non significant'

EEUS011

Slide 40 Confidence Intervals: Egon Pearson's major contribution

NOTES:

Confidence Intervals: Egon Pearson's major contribution

Estimates and confidence intervals for γ , the deflection of light around the sun, from 20 experiments

See Deb Mayo's book 'Error & the Growth of Knowledge'

EEOS611

Slide 41 Confidence Intervals: Egon Pearson's major contribution

NOTES:

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

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

NOTES:

Display 2.3
The sampling distribution of the sample average

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:

Display 2.4
The relationship between the population distribution and the sampling distribution of the average in random sampling

1 CENTER
The sampling distribution is centered on the population mean.

2 SPREAD
Sample averages are closer to the mean than single values: the sampling distribution has $SD(\bar{Y}) = \frac{\sigma}{\sqrt{n}}$

3 SHAPE
The shape of the sampling distribution will be more nearly normal than the shape of the population distribution.

EEOS611

Slide 44

NOTES:

Expressing uncertainty

$X \pm \delta$, >4 choices for δ {del}

- \pm 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
- \pm 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
- \pm half the 95% confidence interval
 - Assumes an underlying model for the data
 - For asymmetric 95% CIs in the statistics natural scale, provide the upper and lower 95% CI (for transformed data and results of Monte Carlo simulations)
- Analytical precision of the instrument or technique
 - E.g., the ChI a method has a certain analytical precision, This is rarely acceptable.
 - Hurlbert's $E(S_{(m)})$, 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, then 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

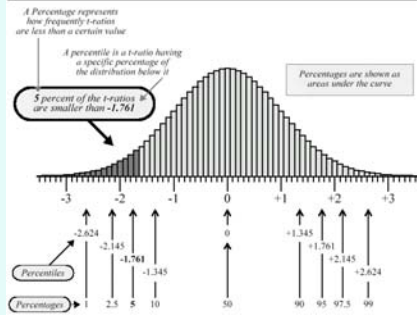
EEOS611

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

NOTES:

Display 2.5

Student's t-distribution on 14 degrees of freedom



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).

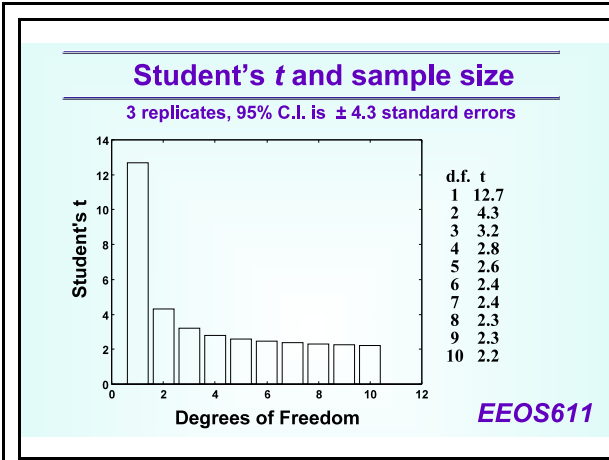
In practical terms, the number of degrees of freedom associated with a statistic is equal to the number of its independent components, i.e. the total number of components used in the calculation minus the number of parameters one had to estimate from the data before computing the statistic. For example, the number of degrees of freedom associated with a variance is the number of observations minus one (noted $v = n - 1$): 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:

95% CI for the mean difference

Problem 2.6 (page 37) Schizophrenia study

1. Compute differences. Obtain their average, \bar{Y} , and standard deviation, s .
 Y_i is the volume for the unaffected twin minus the volume for the schizophrenic twin ($i = 1, \dots, 15$).
 Sample average: $\bar{Y} = 199 \text{ cm}^3$
 Sample standard deviation: $s = 238 \text{ cm}^3$, 14 d.f.
2. Compute $SE(\bar{Y}) = s/\sqrt{n}$ and $d.f. = n - 1$.
 $SE(\bar{Y}) = 238/\sqrt{15} = 6615 \text{ cm}^3$
3. Paired *t* test for the hypotheses that the population mean difference is zero.
 Compute the *t*-statistic for this hypothesis: $(\bar{Y} - 0)/SE(\bar{Y})$.
 $t\text{-statistic} = 199/6615 = 3.236$
 Find the *p*-value (two-sided here) as the proportion of values in a t_{n-1} distribution as far or farther from zero than the observed *t*-statistic.
 two-sided *p*-value = .006 (from computer or Table A.2 and interpolation)
4. 95% confidence interval for population mean difference.
 Find the 97.5th percentile from the t -distribution or $n - 1$ degrees of freedom.
 $t_{(d.f., .975)} = 2.145$ (from Table A.2).
 95% confidence interval = $\bar{Y} \pm t_{(d.f., .975)} \times SE(\bar{Y})$
 $199 \text{ cm}^3 \pm 2.145 \times 6615 \text{ cm}^3$
 0.667 cm^3 to 0.333 cm^3

Slide 50 95% CI for the mean difference

NOTES:

A *t*-ratio for two-sample inference

Slide 51 A *t*-ratio for two-sample inference

NOTES:

Sampling distribution of the difference of averages

Display 2.7 (38): Application of the Central Limit Theorem

Facts about the sampling distribution of the difference of averages from two independent random samples (from statistical theory)

POPULATION DISTRIBUTIONS

SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN AVERAGES

1 **CENTER**: The sampling distribution is centered on the difference between population means.

2 **SPREAD**: The sampling distribution has $SE(\bar{Y}_2 - \bar{Y}_1) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

3 **SHAPE**: The shape of the sampling distribution will be approximately normal if the sample sizes are large relative to the shape of the population distributions.

Slide 52 Sampling distribution of the difference of averages

NOTES:

Pooled standard deviation

& standard error for the difference

This estimate assumes equal variances (Sleuth p 39)

$$s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1 + n_2 - 2)}}, \quad d.f. = n_1 + n_2 - 2.$$

$$SE(\bar{Y}_2 - \bar{Y}_1) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

EEOS611

Slide 53 Pooled standard deviation

NOTES:

SE of difference

Calculation of the pooled estimate of SD and the standard error for the difference between two sample averages; Bumpus' data

1 **SUMMARY STATISTICS**

Group	n	Average (in.)	Sample SD (in.)
1: Died	24	.7292	.02354
2: Survived	35	.7300	.01984

2 **THE POOLED SD**

$$s_p = \sqrt{\frac{(24-1)(.02354)^2 + (35-1)(.01984)^2}{(24 + 35 - 2)}}$$

$$= \sqrt{\frac{.026128}{57}}$$

These are the degrees of freedom associated with the pooled SD

$$= \sqrt{.0004584}$$

This is the pooled variance

Answer $\Rightarrow s_p = 0.02141$ inches

3 **THE STANDARD ERROR**

$$SE(\bar{Y}_2 - \bar{Y}_1) = 0.02141 \sqrt{\frac{1}{24} + \frac{1}{35}}$$

$$= 0.00567 \text{ inches} \leftarrow \text{Answer}$$

euth
lay 2.8
40) **EEOS611**

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
 degrees of freedom = $24 + 35 - 2 = 57$
 $t_{57}(.975) = 2.002$
 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

From Display 2.8
from tables of the t-distribution with 57 degrees of freedom

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:

Display 2.10

Was the difference consistent with chance?

The t-test for the hypothesis that the mean humerus lengths of sparrows that died is the same as the mean for sparrows that survived

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

$\bar{Y}_2 - \bar{Y}_1 = .73800 - .72792 = 0.01008$ From Display 2.8
 $SE(\bar{Y}_2 - \bar{Y}_1) = 0.00567$ inches
 degrees of freedom = $24 + 35 - 2 = 57$

$t\text{-statistic} = \frac{0.01008 - 0.0}{0.00567} = 1.778$

$P = .960$ from tables of the t-distribution with 57 degrees of freedom: 1.778 = $t_{57}(.960)$

1-sided p-value = .040 or 2-sided p-value = $2(.040) = .080$

Slide 58 Display 2.10

NOTES:

Randomization distribution

Can be done with Matlab & R, not SPSS
Display 2.11, page 46

- Creativity data
 - ▶ Randomly shuffle (500 times) the membership in intrinsic and extrinsic groups
 - ▶ Calculate the t-ratio for each random shuffle
 - ▶ Order the value of the t ratios from smallest to largest
 - ▶ For a 1-sided test, calculate how many the t ratios were larger (or smaller) than the observed t ratio, add 1, and divide the number of randomizations
 - ▶ For a 2-sided test, find the number c ratios whose absolute value exceed observed t ratios, add 1, and divide number of randomizations

Slide 59 Randomization distribution

NOTES:

Randomization doesn't solve problems with unequal variance

- Randomization is often superior to the t-distribution for 2-sample problems
- Randomization does not remedy violations of the assumptions of the t test.
 - ▶ The most common problem with Student's t test is the so-called Fisher-Behrens problem, testing the difference in the average if the distributions have different variances
 - ▶ This is an open question
 - ▶ Neither nonparametric approaches (see Chapter 4) nor randomization provide a clear solution

EEOS611

Slide 60 Randomization doesn't solve problems with unequal variance

NOTES:
