16.400/453J

Human Factors Engineering

Design of Experiments II

Massachusetts Institute of Technology

Review

- Experiment Design and Descriptive Statistics
  - Research question, independent and dependent variables, histograms, box plots, etc.
- Inferential Statistics (in Bluman Chapt 6 & 7)
  - Parameter estimates for a population given *sample* mean ($\overline{X}$) & *sample* variance (s^2)
  - Distributions
    - Normal, N(μ, σ), vs. standard normal, N(0, 1)
    - Student's t (degrees of freedom)
    - Binomial (in text & problem set)
  - Probabilities of occurrence
  - Confidence Intervals
  - Sample sizes

Formulas

	Population	Sample, (Estimators for the population)
Mean	μ	$\overline{X}$
Variance	$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$	$s^2 = \frac{\sum (X-\overline{X})^2}{n-1}$
Standard Deviation	σ	s

More Formulas

	Population of Sample Means (Central Limit Theorem)	Estimators for Population of Sample Means
Mean	μ	$\bar{X}$
Variance	σ^2 / N	$s^2 / (n-1)$
Standard Deviation	$\sigma / \sqrt{N}$ aka "Standard Error of the Mean"	$s / \sqrt{n-1}$

To convert N(μ, σ^2) to N(0, 1): $z = \dfrac{X - \mu}{\sigma}$ or $z = \dfrac{X - \bar{X}}{s}$

When N < 30 (small sample), use *t* statistics instead of *z*.

Other Topics in Bluman

- Chapter 8
  - Hypothesis Testing
  - Proportions (z test statistic)
  - Variances (Chi-square statistic)
- Chapter 9
  - Two-sample tests for means, variance, and proportion
    - Large samples, small samples
    - Dependent and independent means

Blue items will be covered in this lecture…

Hypothesis Testing

- Null hypothesis (H_o):
  the independent variable has no effect
- Alternative hypothesis (H_a):
  any hypothesis that differs from the null
- Significance (p-value)
  - How likely is it that we can reject the null hypothesis?

Hypothesis Testing & Errors

Outcome	True in the World: H_o	True in the World: H_a
Reject H_o	Type I Error α	Correct Decision $(1-\beta)$ (power)
Accept H_o	Correct Decision $(1-\alpha)$	Type II Error β

Alpha & Beta

http://www.intuitor.com/statistics/T1T2Errors.html
http://www.intuitor.com/statistics/CurveApplet.html

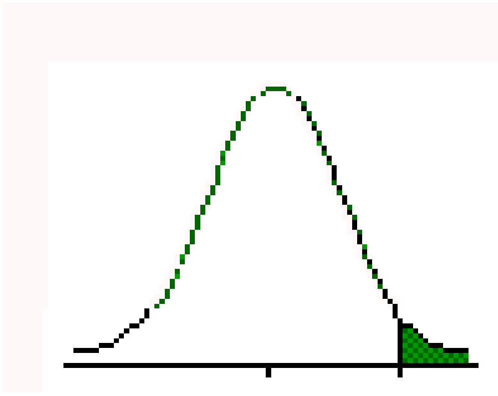
Hypothesis & Tails

	One-Tailed Left	One-Tailed Right	Two-Tailed
H_o	$\mu \geq k$	$\mu \leq k$	$\mu = k$
H_a	$\mu < k$	$\mu > k$	$\mu \neq k$
Example	If people try your diet, they will lose weight.	If people try your exercise routine, their muscle mass will increase.	If you test students more often, their grades will change.(maybe up, maybe down)

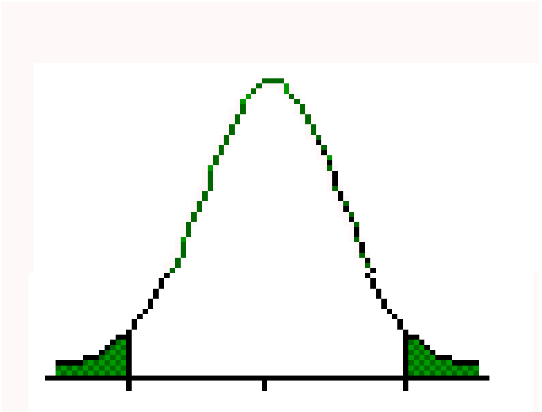
- Your hypothesis will guide your selection
- Two-tailed test tells you if values are or are not different; you have no *a priori* expectations for the direction of the difference.

One-Tailed vs. Two-Tailed Tests

H_o: $\mu \leq 0$ versus H_a: $\mu > 0$



H_o: $\mu = 0$ versus H_a: $\mu \neq 0$



How to Use Confidence Intervals

- Confidence Level = 1- α
  - α, Type I error, rejecting a hypothesis when it is true
- Commonly used confidence intervals and critical values for the *standard normal* distribution
  - 90% $z_{\alpha/2} = 1.65$
  - 95% $z_{\alpha/2} = 1.96$
  - 99% $z_{\alpha/2} = 2.58$
- Whether a hypothesis "is true" translates to *how likely or unlikely it is that the data were obtained due to chance*
  - Adjusting the size of the confidence interval adjusts the likelihood that the data could have occurred by chance.

Original Procedure

- State hypothesis and identify claim
- Find critical value
- Compute test statistic
- Decide whether or not to accept or reject
- Summarize results
- Applies to *z* & *t* tests for means

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

$$\text{Test statistic} = \frac{\text{observed value} - \text{expected value}}{\text{standard error}}$$

$$t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Hypothesis Testing Example

- A researcher claims the average salary of assistant professors > $42,000 – is this true?

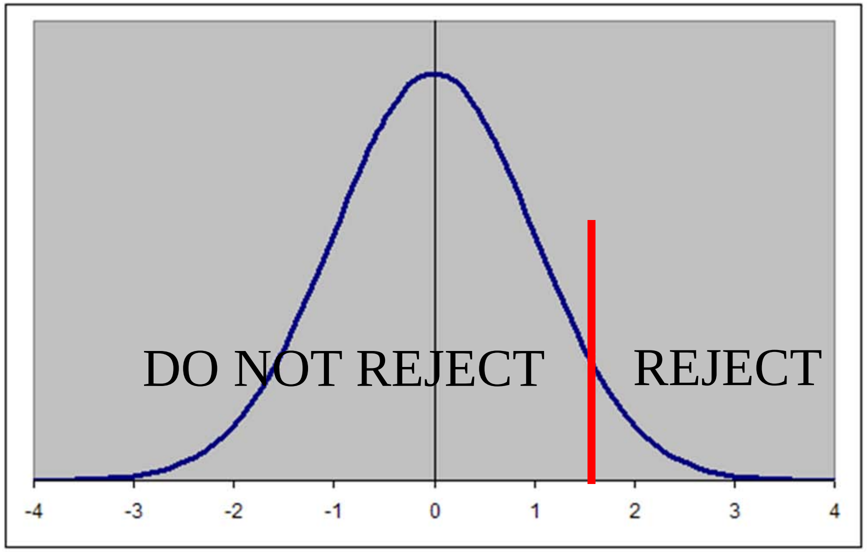
Sample of 30 has a mean salary of $43,260, σ = $5,230 ($\alpha$ = .05)

H_o: $\mu \leq$ $42,000

H_a: $\mu >$ $42,000

$$z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

- Critical value?
  - Right tailed test, z = 1.65
  - Test value = 1.32
- Can't reject the null
  - What if sample < 30?



A More Useful Procedure

- State hypothesis and identify claim
- ~~Find critical value~~ Compute test statistic
- ~~Compute test statistic~~ Find *p*-value
- Decide whether or not to accept or reject
- Summarize results

- You get a little more information from this approach…

Hypothesis Testing Example Again

- A researcher claims the average salary of assistant professors > \$42,000 – is this true

Sample of 30 has a mean salary of \$43,260, σ = \$5,230 (α = .05)

H_o: $\mu \leq$ \$42,000

H_a: $\mu >$ \$42,000

$$z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

Test Statistic = 1.32

p-value of 1.32 = .0934 (0.5-0.4066)

  - Using standard normal table

- Also cannot reject (but marginal!)

Hypothesis Testing: t Tests for a Mean

- Use when σ is unknown and/or n < 30

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

DOF = n-1

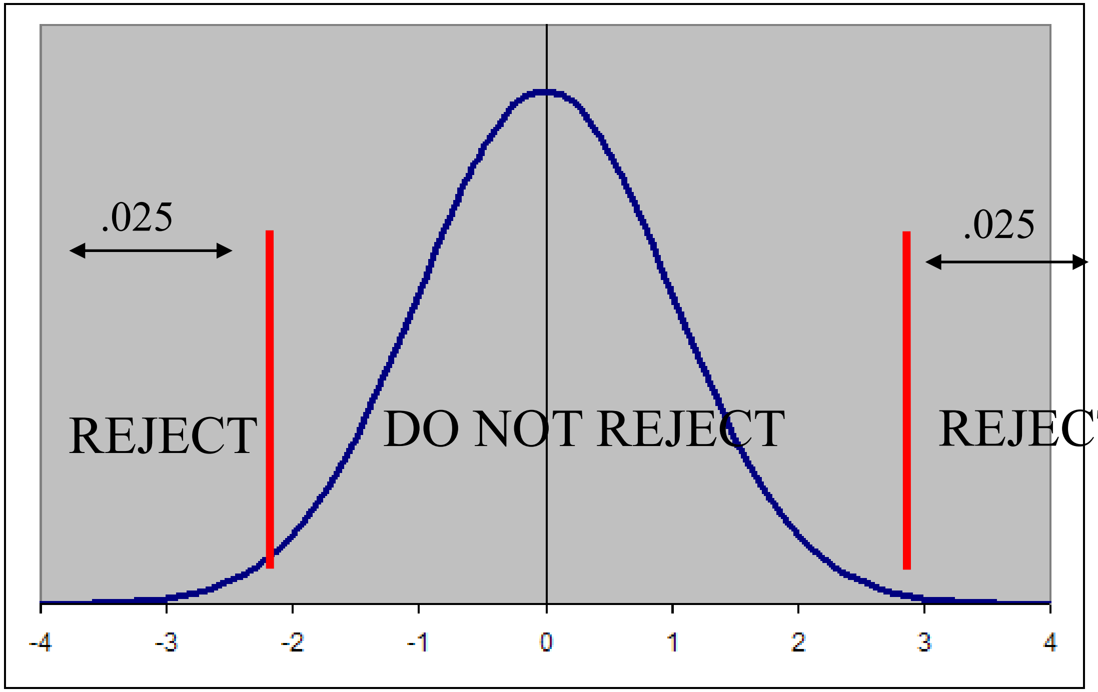
- Example: Job placement director claims average nurse starting salary is $24K, is this true?

  Sample of 10 nurses has a mean of $23,450, SD = $400 (α = .05)

- H_o vs. H_a?
  - •Critical value = +/-2.26 ….why?
  - •Test statistic = - 4.35
  - •Reject - draw a picture!
    - 2 tailed appropriate?

Two Tailed *t*-Test

Sample of 10 nurses, mean of \$23,450, SD = \$400 (α = .05)



- Critical value = +/-2.26 ….why?
- Test statistic = - 4.35

Two Sample Tests for Large Populations

- Assumptions:
  - Independent samples, between subjects
  - Population normally distributed
  - SD known or sample size > 30

H_o: $\mu_1 = \mu_2$ and H_a: $\mu_1 \neq \mu_2$ *can be restated as*
H_o: $\mu_1 - \mu_2 = 0$ and H_a: $\mu_1 - \mu_2 \neq 0$

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}}$$

- Either one or two tailed

Pooled variance = sum of the variances for each population

Two Sample for Small Populations

- *t* tests
- Assumptions:
  - Independent samples
  - Population normally distributed
  - Sample size < 30
- Unequal vs. equal variances of population
  - To determine whether variances are different, use F test
  - If variances are equal, use pooled variance. (Otherwise, out of scope)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

DOF = smaller of (n_1-1) or (n_2-1)

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\right)}\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

DOF = $n_1 + n_2 - 2$

Pooled variance = sum of the variances for each population

If sample sizes are very different, use a weighted average.

t-tests for Matched (Dependent) Samples

- Matched samples (within subjects)
  - Learning, medical trials, etc.
  - ~ normally distributed data

$$t = \frac{\bar{D} - \mu_D}{s_D / \sqrt{n}} \quad \bar{D} = \frac{\sum D}{n} \quad s_D = \sqrt{\frac{\sum D^2 - \frac{\left(\sum D\right)^2}{n}}{n-1}}$$

$D = X_1 - X_2$, DOF = n-1, μ_D is part of the hypothesis

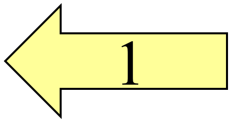
Comparing Variances

- F-test is one option
  - Independent samples, normally distributed population
  - Ratio of two Chi-square distributions
  - Other, more complicated, but better options exist
- DOF for numerator: n_1-1,
- DOF for denominator: n_2-1
- s_1 is larger of 2 variances
- http://www.statsoft.com/textbook/sttable.html
- If DOF not listed, use lower (to be conservative)

$$F = \frac{s_1^2}{s_2^2}$$

Comparing Variances - Example

- Hypothesis: SD for exam grade for males is larger for males than females – is this true (α = 0.01)?
  - Males: n = 16, s = 4.2
  - Females: n = 18 s = 2.3



$$H_0 : \sigma_1^2 \leq \sigma_2^2 \quad H_a : \sigma_1^2 > \sigma_2^2$$

DOF(n) = 15, DOF(d) = 17, table critical value = 3.31

$$F = \frac{s_1^2}{s_2^2} = \frac{4.2^2}{2.3^2}$$

= 3.33, so reject the null

Other Tests

- Linear regression
  - Correlations
- Analysis of variances (ANOVA)
  - Testing the differences between two or more independent means (or groups) on one dependent measure (either a single or multiple independent variables).
  - Uses the F test to test the ratio of variances
  - Most flexible tests (for mixed models, repeated measures etc.)

Correlations Example

Correlations among the pilot characteristics and experience from Chandra (2009) DOT-VNTSC-FAA-09-03.

	VFR/IFR Pilots (143 VFR/130 IFR)	Private VFR (177)	FAA Chart Experience (177)	Jeppesen Experience (123)	Flight Length	International (58)	Air Transport (76)	Flight Hours
VFR/IFR Pilots	1	0.74	0.56	-0.46	-0.46	-0.55	-0.65	-0.73
Private VFR	0.74	1	0.60	-0.44	-0.52	-0.56	-0.72	-0.68
FAA Chart Experience	0.56	0.60	1	-0.55	-0.45	-0.54	-0.67	-0.63
Jeppesen Experience	-0.46	-0.44	-0.55	1	0.34	0.35	0.43	0.53
Flight Length	-0.46	-0.52	-0.45	0.34	1	0.59	0.57	0.57
International	-0.55	-0.56	-0.54	0.35	0.59	1	0.60	0.64
Air Transport	-0.65	-0.72	-0.67	0.43	0.57	0.60	1	0.71
Flight Hours	-0.73	-0.68	-0.63	0.53	0.57	0.64	0.71	1

All values are significant at $p < 0.01$. Strong positive correlations appear in the top left and bottom right. Strong negative correlations appear in the bottom left and top right.

Sample sizes are given in parentheses in top row.

VFR = 1, IFR = 0

Flight length has 4 categories ranging from short to long.

ANOVA

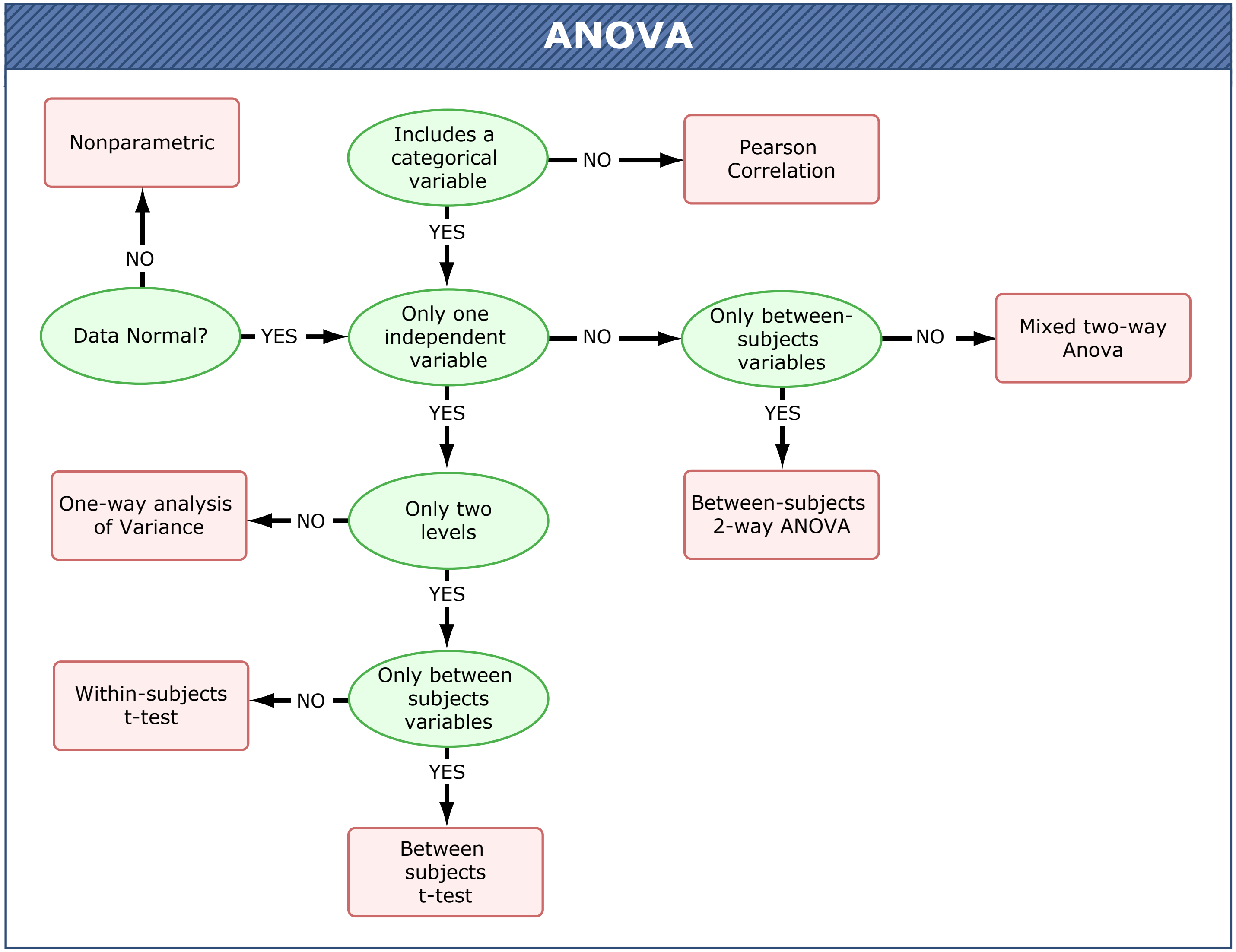
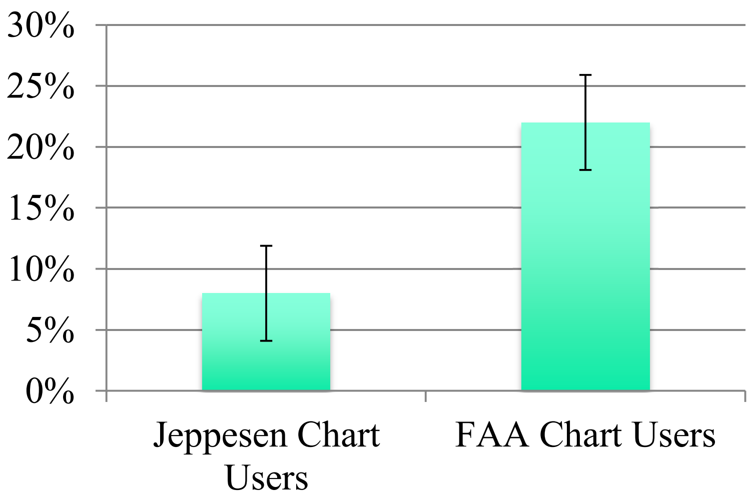


Image by MIT OpenCourseWare.

Aeronautical Charting Example Continued

Accuracy in Identifying Air Traffic Control Center Boundary



	Jeppesen Chart Users	FAA Chart Users
N	50	117
Mean	0.08	0.22
Std Dev	0.274	0.418
Std Error	0.039	0.039
Variance	0.075	0.174

Pooled variance = 0.145

Z score, assuming equal sample size = 2.56, $p < 0.01$

Z score corrected for unequal sample sizes = 2.18, $p < 0.05$

ANOVA	**Output From SPSS**	**Sum of Squares**	df	Mean Square	F	Sig.
ARTCC	Between Groups	0.709	1	0.709	4.891	**0.028**
	Within Groups	23.902	165	0.145		
		Sum of Squares	df	Mean Square	F	Sig.

Practical Questions

- What confidence level (α) should I use?
  - Number of simultaneous tests performed
  - Statistical significance vs. operational significance
  - Relative differences between effects when many comparisons are made
  - Ability to explain the effect aside from the statistics
- What test should I use when…?
- How many subjects should I test?
  - Face validity, resource considerations, power calculations

In an imperfect world…

- Complex designs
  - Repeated measures
  - Mixed models
    - Combination of within & between subjects
  - Lots of "trials" required
  - Lots of subjects required
- Unplanned complexities
  - Missing data
  - Unequal cell sizes
  - Experiment confounds
- Additional course work
  - Mathematical and/or practical perspectives

16.400 / 16.453 Human Factors Engineering
Fall 2011