16.400/453J

Human Factors Engineering

# Design of Experiments I



Massachusetts Institute of Technology

# Human Factors Experiments



- Why do a human factors experiment?
  - To find out whether a hypothesis about a question "is true"
  - To explore the relationship between variables
  - To develop and validate model to predict performance
  - Concept validation
  - Improve product design
- When not to do a human factors experiment
  - Question can be resolved by analysis or based on existing data
  - There are no critical consequences
  - Deeper understanding is not required

# Research Methods



- Quantitative
  - With or without humans
    - Natural phenomenon
    - Physical Experiments
    - Mathematical modeling
    - Optimization
  - With Humans
    - Performance models
    - Surveys
    - Experiments
- Qualitative (w/humans)
  - Observation
    - e.g., observe pilots flying
  - Case studies
    - e.g., NASA ASRS reports
  - Usability testing
    - e.g., Electronic Flight Bag
    - Can be quantitative
  - (Open-response) surveys
  - Focus groups
  - Interviews

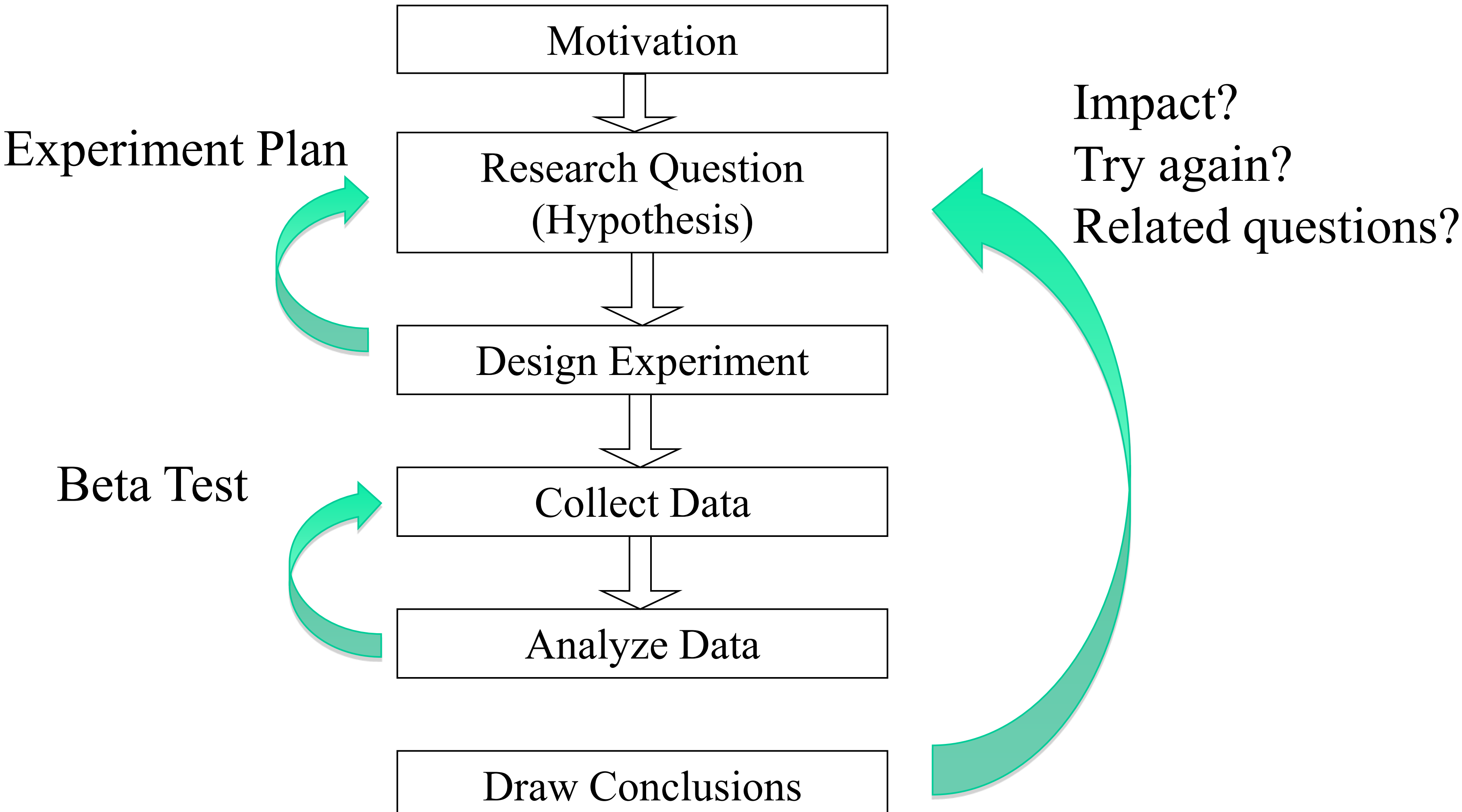
# The Basics



- Understanding the relationship between objectives (research question) and variables is critical for quantitative research
  - Clearly map your goals to your test
  - Field vs. laboratory research
    - Tradeoffs between realism vs. control, generalizability
- Planning in advance is a must
  - Includes how data will be analyzed
- The importance of statistics

# The Experimental Design Process





# DOE Terminology I



- Independent variables vs. Dependent Variables
  - What you are manipulating vs. What you are measuring
- Measuring a variable (discrete vs. continuous)
  - Nominal/Categorical (e.g., label, multiple choice answer)
  - Ordinal (e.g., military rank, self-report rating)
  - Interval (e.g., temperature, date)
  - Ratio scale (e.g., length, time)
- Descriptive Statistics vs. Inferential Statistics
  - Describing your data vs. drawing inferences

# Types of Independent Variables



- Control condition
  - Baseline is not necessarily “no treatment” (e.g., placebo)
- Levels of a variable
  - 2 levels, can use simple “t-test” for statistical inference
    - e.g., 2 levels of “Experience” (novice, expert)
  - 3 or more levels, more complicated tests
    - e.g., 3 levels of “Air Traffic Density” (low, medium, high)
    - ANOVA, paired comparisons, etc.
    - Next lecture & other courses
- Within-subjects and Between-subjects
  - e.g., Air Traffic Density vs. Experience

# Types of Dependent Variables



- Performance-based, e.g.,
  - Reaction time (< 1 sec) or Response time (> 1 sec)
  - Accuracy or errors
- Subjective, e.g.,
  - Preference
  - Free response
- Psychophysiologic response, e.g.,
  - Pulse rate, blood pressure
- Meta-metrics (inferred), e.g.,
  - Workload, Situation Awareness

# DOE Terminology II



- Computer Programs
  - Excel, SAS, SPSS, MatLab, R
  - Plan your data recording format for the software
- Samples vs. populations
  - Avoid sampling bias

# Exercise: Design of Stove Top Control



- Motivation?
- Research Question?
  - Independent variables?
    - Within/between?
    - Continuous or discrete?
  - Dependent variables?
    - Subjective, objective?
- User task/instructions?
  - What does the subject see? What does the subject do?
  - Any particular emphasis to motivate the subject?
  - How long/hard is this task?
- Data analysis?
- Example conclusion that could be drawn?

# Descriptive Statistics



- Measures of central tendency
  - Mean, median, mode, (range)

  "Subject age ranged from 20 to 70 years, with a mean age of 32."

  "Pilots had a median experience of 9775 flight hours."

  "Most of the pilots held Air Transport Ratings (100), but some held only Instrument Ratings (30), and a few held only Visual Flight Ratings (6)."

- Measures of "spread"
  - Variance, standard deviation

  "Pilots had a mean experience of 9775 flight hours with a standard deviation of 550 hours."

# Measures of Central Tendency



- A fancy way to say average
- Roman letters represent *statistics* (samples)
- Greek letters represent *parameter* (populations)

Mean $\bar{X} = \frac{\sum X}{n}$ $\mu = \frac{\sum X}{N}$

Median $\tilde{X}$

Halfway point in data array

Median of 1, 3, 4, 2, 3, 5, 1?

What about 1, 3, 4, 2, 5, 1? 2.5

- Don't forget about skew!

# Measures of Central Tendency, cont.



## Mode: Value that occurs most often

The only measure of central tendency for nominal/categorical data (e.g,. response to a multiple choice question)

*How many pets do you own?*

Sample responses 0, 1, 2, 1, 2, 2, 3. Mode = 2

Sample responses 1, 3, 4, 2, 5, 6. Mode = X 0?

Sample responses 1, 3, 0, 2, 3, 5, 1 1,3 - Bimodal

Midrange = rough estimate = $\dfrac{X_{\min} + X_{\max}}{2}$

# Measures of Variance



- Variance = average of the squares of the distance of each value from the mean
  - If individual data points are near the mean, then variance is small.
  - Standard deviation is square root of the variance

$$\sigma^2 = \frac{\sum (X-\mu)^2}{N} \qquad \sigma = \sqrt{\sigma^2}$$

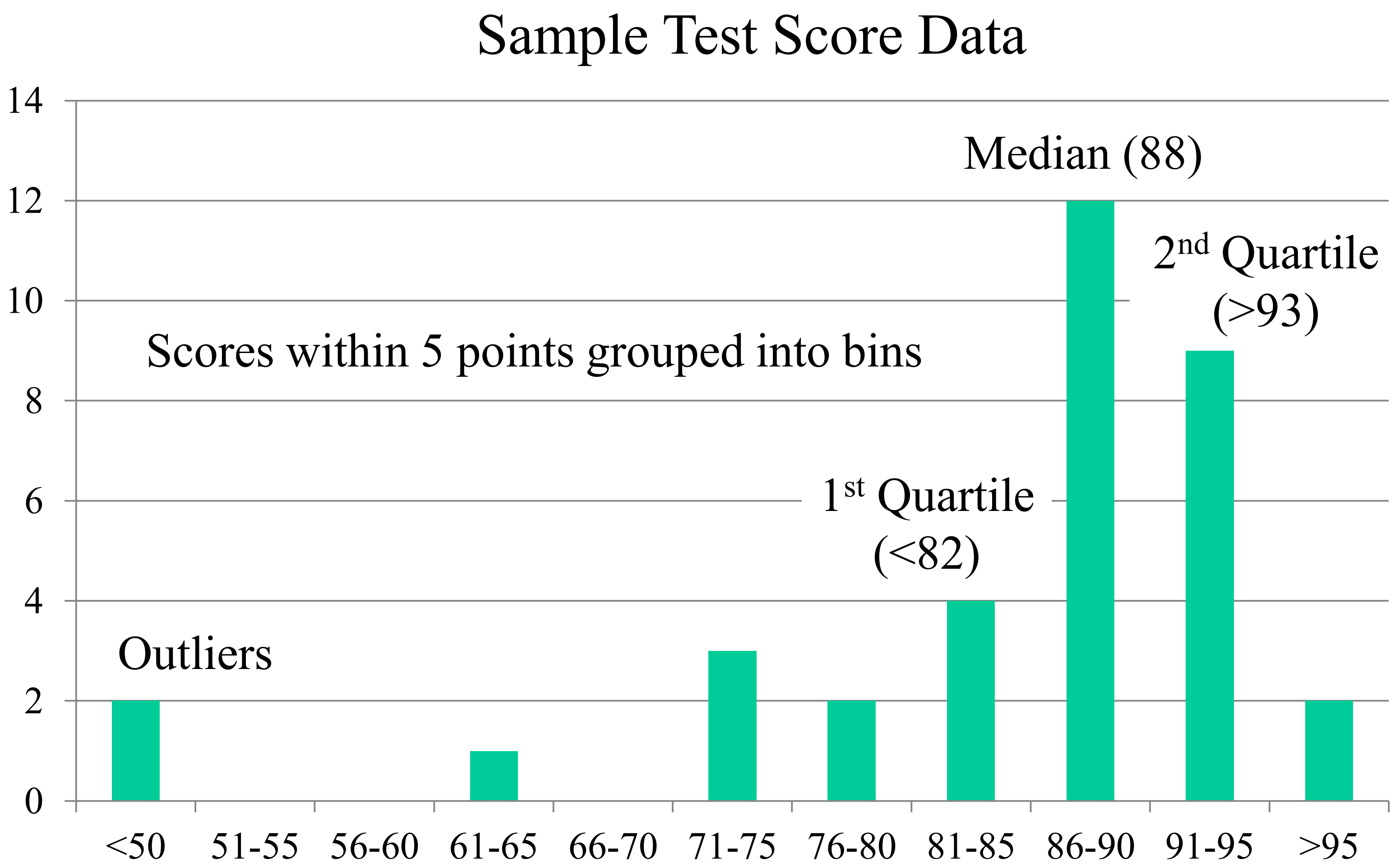
Population vs. sample

$$s^2 = \frac{\sum (X-\bar{X})^2}{n-1} = \frac{\sum X^2 - \left(\sum X\right)^2 / n}{n-1}$$

Unbiased estimate

# Visualizing the Data Set - Histogram

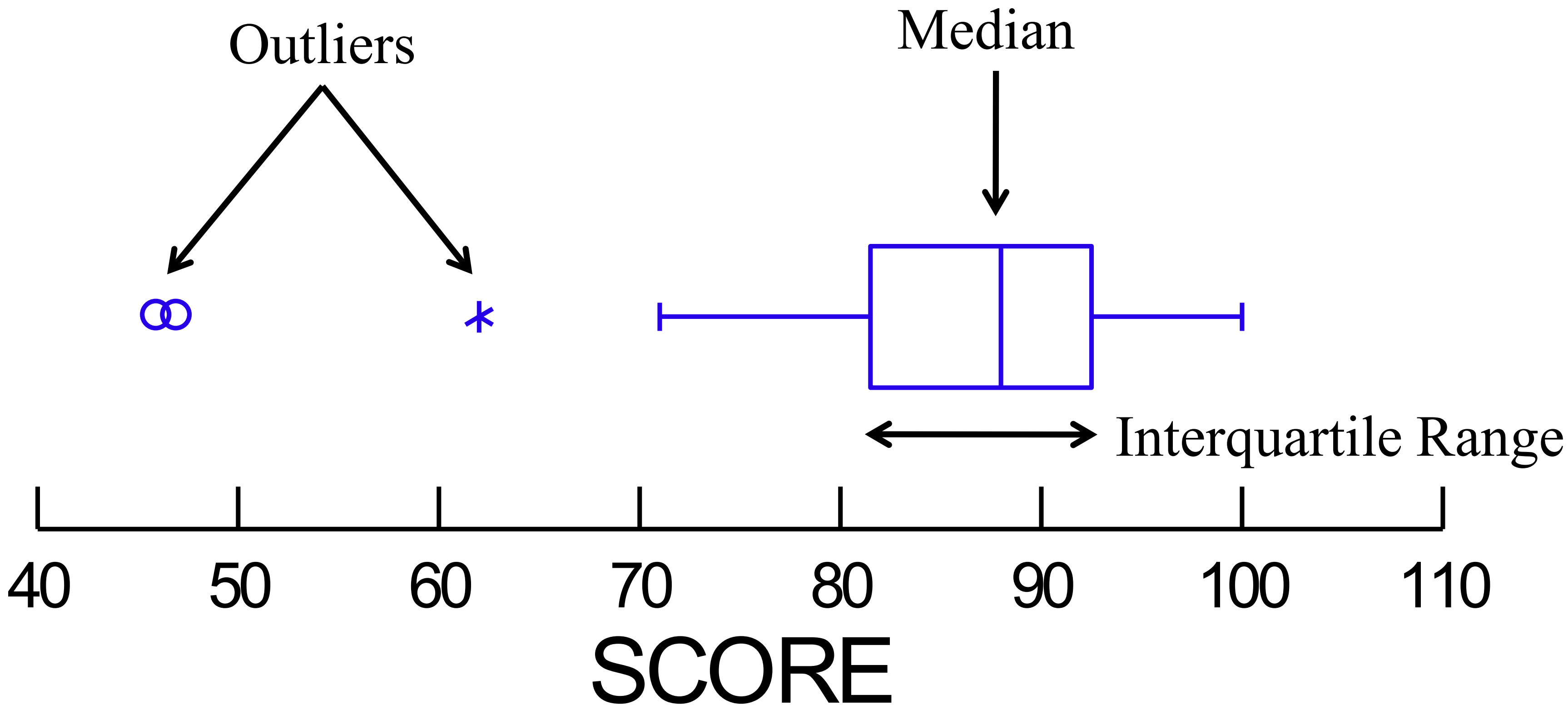





What is plotted on the y-axis? Number of cases (frequency)

# Visualizing the Data Set - Box plot





# A Simple Experiment



- Motivation
  - To illustrate experiment design and data analysis
- Specific research question
  - Are men taller than women on average?

  $H_0$: $\mu_m \leq \mu_f$ vs. $H_a$: $\mu_m > \mu_f$
- Independent Variable
  - Between subjects, male/female
- Dependent Variable
  - Height in inches (or cm)
- Number of subjects?
- Distributions? Sample means? Inferences?

# Example Experiment: Aeronautical Charting Standards



- Motivation
  - Need to develop industry standards/recommendations for electronic chart symbols, lines, and linear patterns
  - What line patterns on charts should be standardized, and what specific patterns should be recommended?
- Specific research questions
  - What line patterns are used regularly? (by type of operation, pilot experience etc.)
  - What line patterns are well recognized? (by pilot experience)
    - First cut based on subject matter expert input

# Example: Standards for Lines on Charts



- Independent Variables
  - Within-subjects: Specific line patterns of interest (7)
  - Between-subjects: Pilot experience (type of chart used)
- Dependent Variables
  - Accuracy (correct/incorrect) (judged written response)
- Task

Line pattern (or ?): ____________________

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **Low Confidence** | | | **Medium Confidence** | | | **High Confidence** |

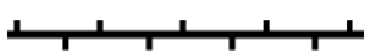
Controlled Airspace 

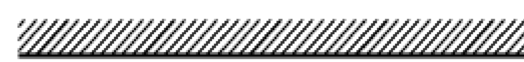
ARTCC 

Comm Boundary 

Time Zone 

International Boundary 

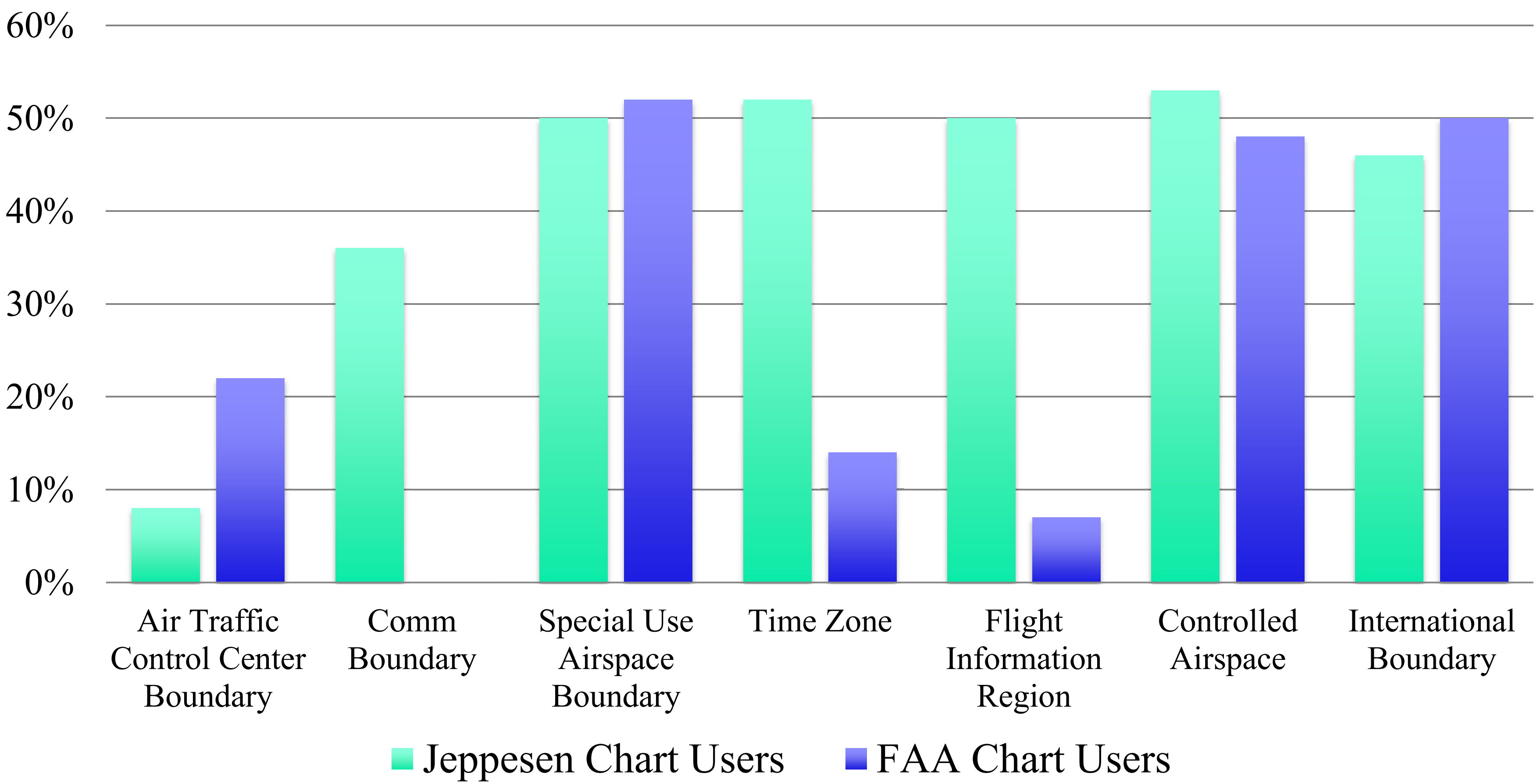
FIR 

Special Use Airspace Boundary 

# Descriptive Statistics - Example



**% of Pilots Who Correctly Identified Test Line Patterns by Type of Chart Used (Pilot Experience)**



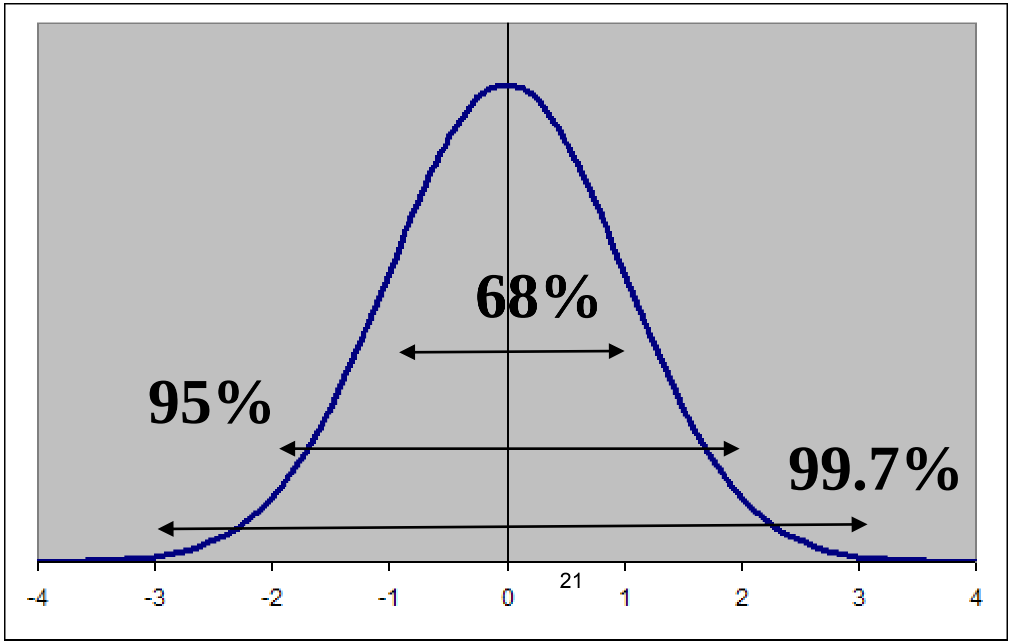



*See Chandra, 2009 DOT-VNTSC-FAA 09-03 for full results

# Inferential Statistics



- Continuous probability distribution
- Probability that some variable is <, >, or between 2 values
  - How do we determine what is a statistically significant finding?



# Standard Score



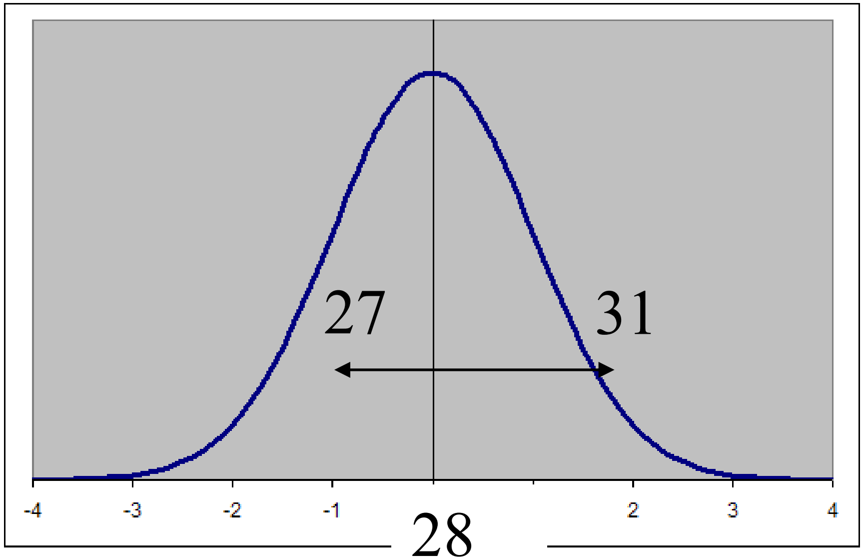
- Normal distribution $N(\mu,\sigma^2)$ vs. Standard Normal distribution N(0,1)
  - Comparing apples to oranges
  - Also known as z-score
  - http://www.statsoft.com/textbook/sttable.html
- The number of of standard deviations that a value falls above or below
- Test statistic

$$z = \frac{X - \overline{X}}{s} \qquad \frac{\text{observed value} - \text{expected value}}{\text{standard error}}$$

# Standard Score Example



- The average reaction time for a search task is 28 secs +/- 2 secs. What is the probability that someone randomly selected to perform the task will be 1) between 27 & 31 secs and 2) > 30.2 secs?
- Part 1 - Draw!




$$z_1 = \frac{27-28}{2} = -.5 \quad z_2 = \frac{31-28}{2} = 1.5$$

Area $z_1$ = .1915

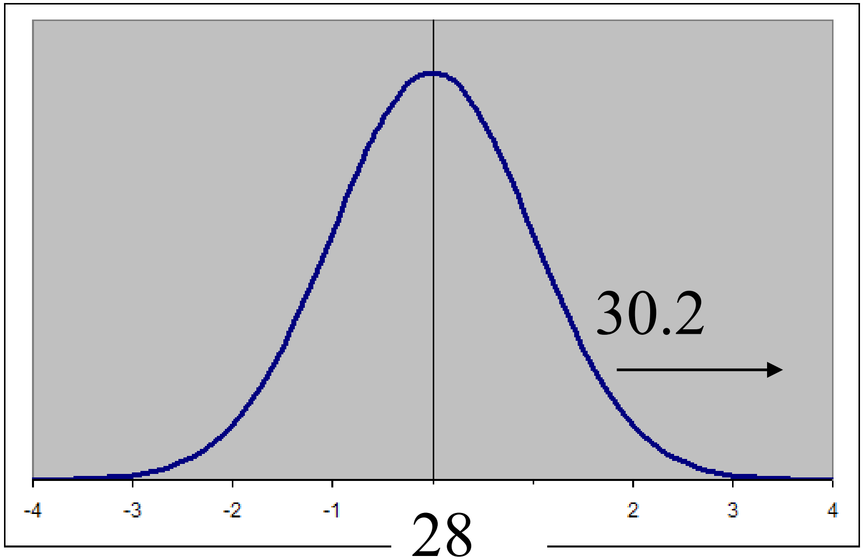
Area $z_2$ = .4332

$P[27<X<31] = .1915 + .4332 = .6247$

# Standard Score Example, Part II



- The average reaction time for a search task is 28 secs +/- 2 secs. What is the probability that someone randomly selected to perform the task will be 1) between 27 & 31 secs and 2) > 30.2 secs?
- Draw!




$$z_1 = \frac{30.2 - 28}{2} = 1.1$$

Area $z_1$ = .3643

P[X>30.2] = .5 -.3643 = .1357

# Confidence Intervals: Means

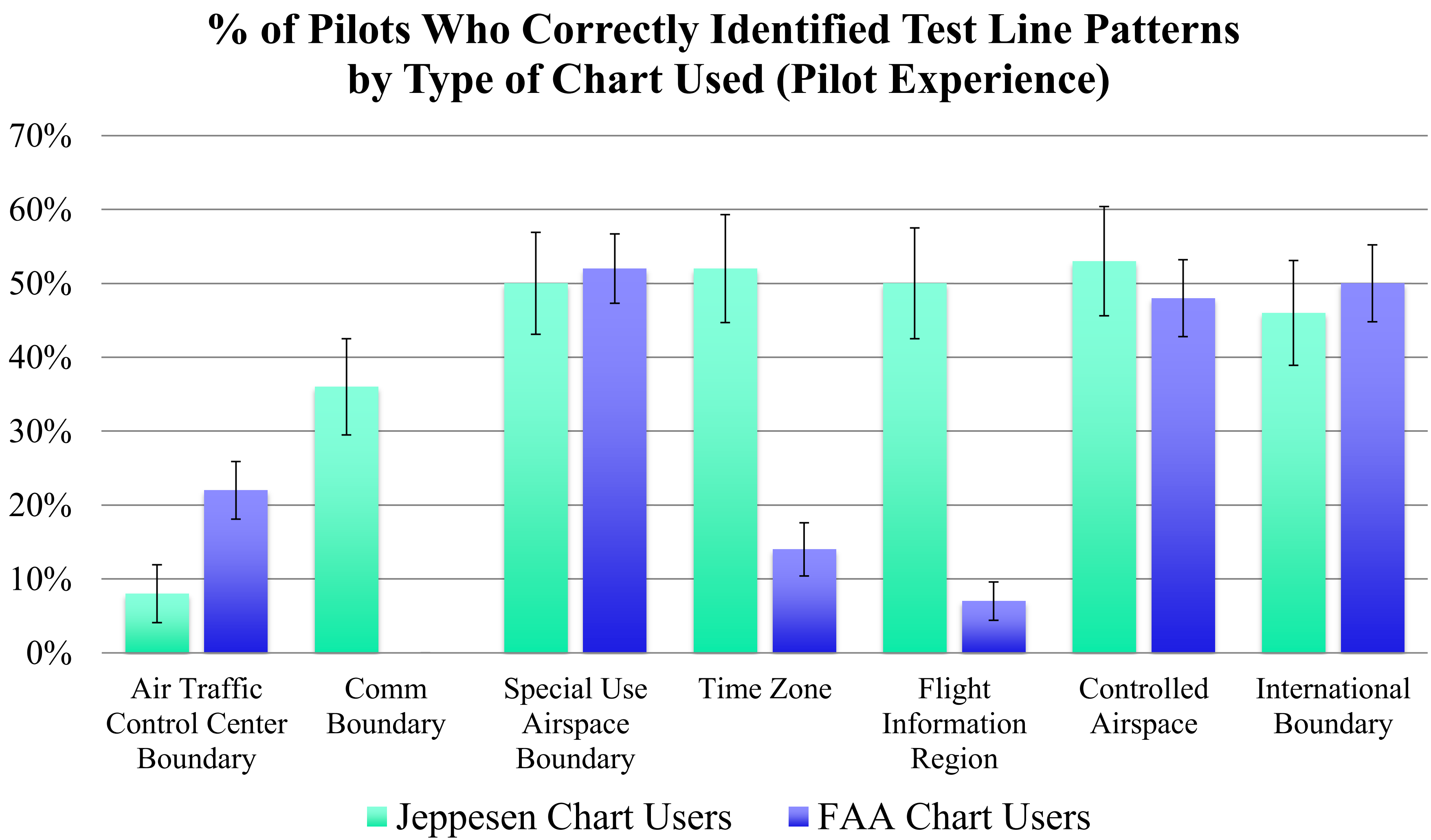


- Interval estimate
  - Range of values that estimates a parameter
- Confidence level – probability that estimate will contain the parameter
- Standard error - standard deviation of the sampling distribution of a statistic.

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

# Standard Error - Example








*See Chandra, 2009 DOT-VNTSC-FAA 09-03 for full results

# Confidence Interval Example



- Estimate the average age of a student population with 95% confidence:
  - SD is known to be 2 yrs (previous studies)
  - Mean of sample of 50 students is 23.2 yrs

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$22.6 < \mu < 23.8$$

*But what is the catch?*

# Small Samples: Student's t Distribution

- Use Standard normal when:
  - σ known, normal distribution
  - σ unknown, $n \geq 30$
  - If these conditions are not met, use t distribution (aka "Student's t")
- For t distribution:
  - variance > 1
  - A family of curves based on degrees of freedom approaching standard normal as sample size increases
- DOF = Number of values that are free to vary after a sample statistic has been computed
  - Which curve to use - http://www-stat.stanford.edu/~naras/jsm/TDensity/TDensity.html

# Critical t-value Examples



For tables online:

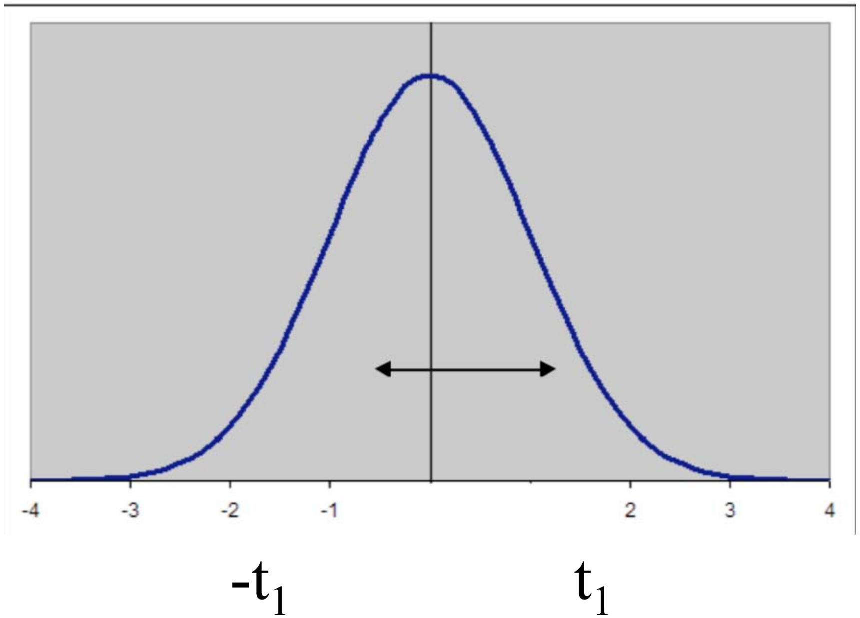
http://www.statsoft.com/textbook/distribution-tables/#t

Given DOF = 9, find $t_1$ given

a) $P[X>t_1] = .05$
1.83 (one-tailed)

b) $P[X<-t_1] + P[X>t_1]= .01$
3.25 (two-tailed)

c) $P[X<-t_1]=.01$
2.82 (one tailed)




For more practice and information:
http://simon.cs.vt.edu/SoSci/converted/T-Dist/activity.html

# Confidence Intervals, revisited.



- Sample size still must be approximately normal
- t tables: http://www.statsoft.com/textbook/sttable.html

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

DOF = n-1

# Confidence Interval Example Revisited



- Estimate the average age of a student population with 95% confidence:
  - SD is known to be 2 yrs
  - Mean of sample of 8 students is 23.2 yrs

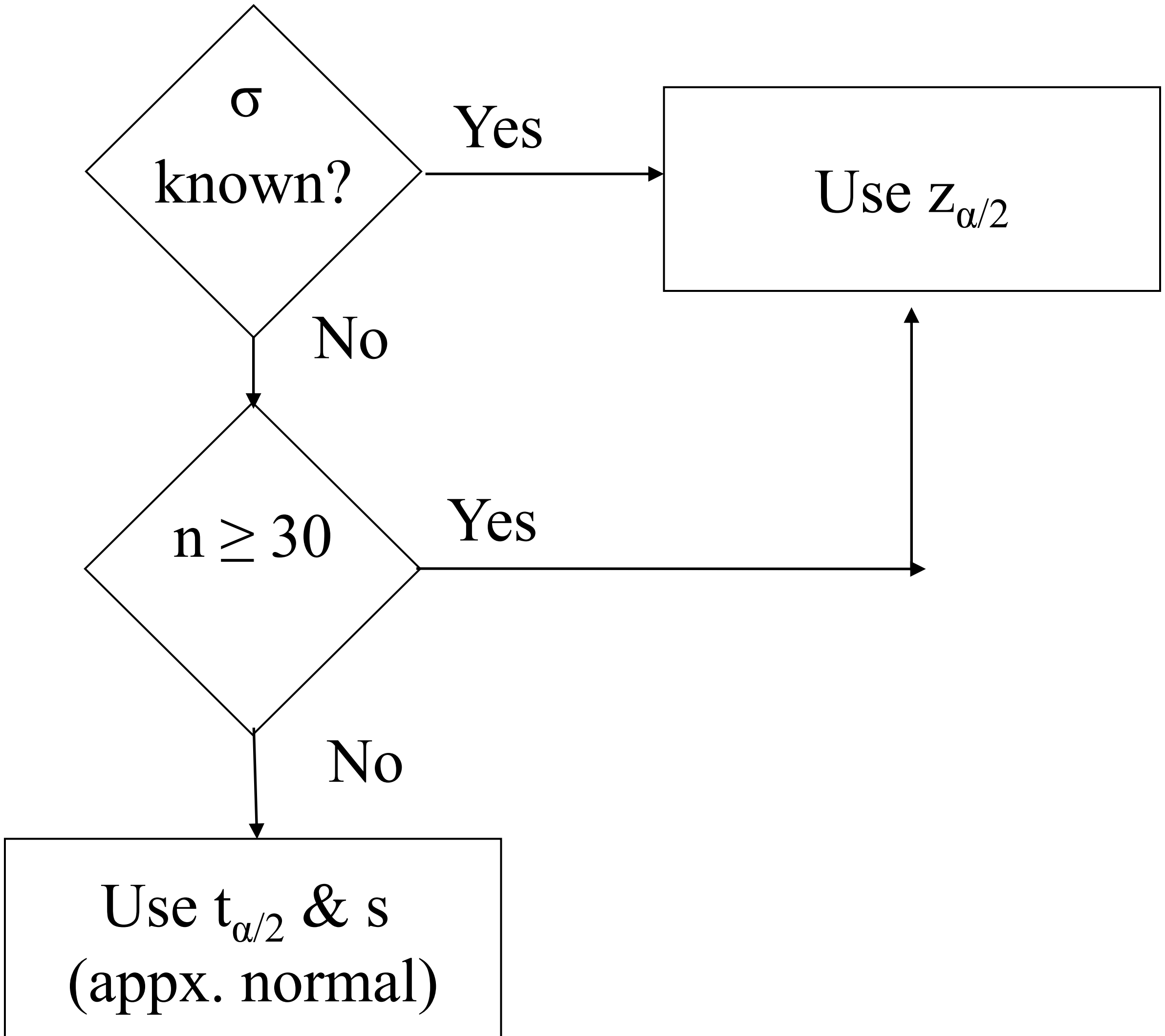
$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$22.6 < \mu < 23.8 \text{ (N=50)}$$

$$21.5 < \mu < 24.9 \text{ (N=8)}$$

# t vs. z?




σ known?
Yes
Use $z_{\alpha/2}$
No
$n \geq 30$
Yes
No
Use $t_{\alpha/2}$ & s (appx. normal)

# Questions?



- Next lecture on hypothesis testing and more advanced statistical tests
- Pset due Sept $27^{th}$

MIT OpenCourseWare
http://ocw.mit.edu

16.400 / 16.453 Human Factors Engineering
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.