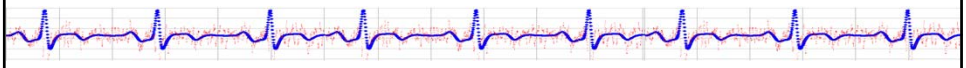
Empirical Research Methods in Information Science

IS 4800 / CS6350



Lecture 13

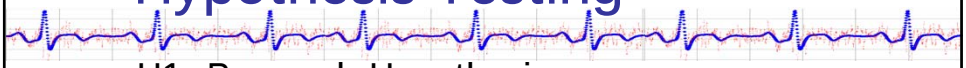
Between Subjects Experimental Designs

The Normal Curve

The Single-Observation Test

1

Review: Basic Process of Hypothesis Testing



- H1: Research Hypothesis:
  - Population 1 is different than Population 2
- H0: Null Hypothesis:
  - No difference between Pop 1 and Pop 2
- State test criteria (α, tails)
- Compute *p(observed difference|H0)*
  - 'p' = probability observed difference is due to random variation
- If p< *threshold* then reject H0 => accept H1
  - p typically set to 0.05 for most work
  - p is called the "level of significance"

6

Review

- Kinds of measures?
  - Nominal
  - Ordinal
  - Scale
  - Ratio

Nominal, Ordinal: Categorical

Scale, Ratio: Numeric

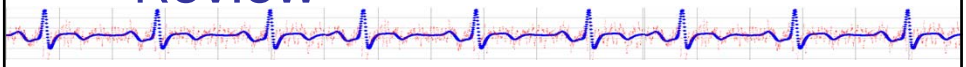
7

Review

- Kinds of analyses for Descriptive studies?
  - Descriptive
  - Chi-square goodness of fit [categorical]
  - *For demonstrations, too*

9

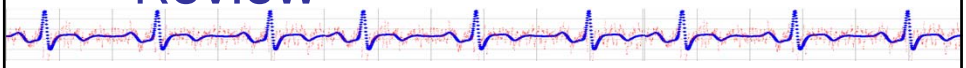
Review



- Kinds of analyses for Correlational studies?
  - Descriptive
  - Chi-square test for independence [categorical/categorical]
  - Correlation [numeric/numeric]
    - Point-biserial [numeric/categorical]
    - Spearman rank-order [ordinal/ordinal], [ordinal,numeric], or [numeric,numeric and not linear]

10

Review



- Kinds of analyses for Experimental studies?
  - Descriptive
  - Correlation (large number of IV values - *parametric*) [numeric/numeric] (atypical)
  - Chi-square test for independence [categorical/categorical]
  - t-test! [categorical/numeric]

11

t-test for independent means

- Tests association between binomial IV and numeric DV.
- Examples:
  - WizziWord vs. Word => wpm
  - Small vs. Large Monitors => wpd
  - Wait time sign vs. none => satisfaction

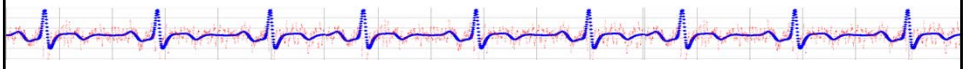
13

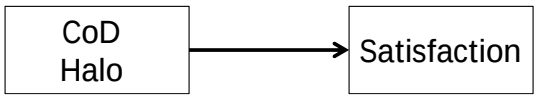
Understanding numeric measures

- Sources of variance
  - IV
  - Other uncontrolled factors (“error variance”)
- If (many) independent, random variables with the same distribution are added, the result is approximately a normal curve
  - The Central Limit Theorem

14

Example: Call of Duty vs. Halo



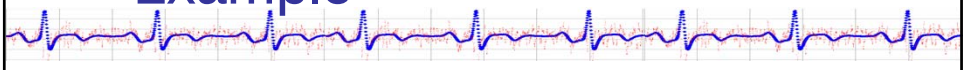


- What variables might affect Satisfaction?
- Typically, one subject's Satisfaction score =

  TrueSatisfaction + var1 + var2 + var3 + ...
- A sum of random variables.

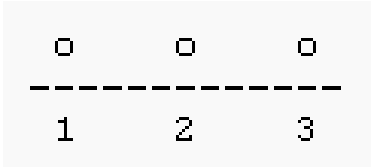
15

Why be normal? Example



- Suppose random variable X has distribution

$$X = \begin{cases} 1 & \text{with probability } 1/3, \\ 2 & \text{with probability } 1/3, \\ 3 & \text{with probability } 1/3. \end{cases}$$



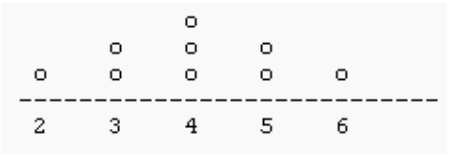
From wikipedia

16

Why be normal? Example

- Now, consider the distribution of X+X

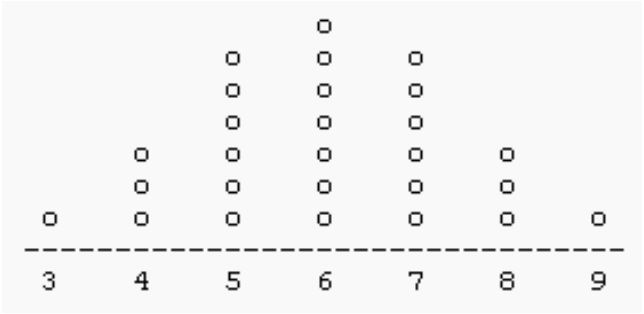
$$\left\{\begin{array}{rcl} 1+1 &=& 2 \\ 1+2 &=& 3 \\ 1+3 &=& 4 \\ 2+1 &=& 3 \\ 2+2 &=& 4 \\ 2+3 &=& 5 \\ 3+1 &=& 4 \\ 3+2 &=& 5 \\ 3+3 &=& 6 \end{array}\right\} = \left\{\begin{array}{ll} 2 & \text{with probability } 1/9 \\ 3 & \text{with probability } 2/9 \\ 4 & \text{with probability } 3/9 \\ 5 & \text{with probability } 2/9 \\ 6 & \text{with probability } 1/9 \end{array}\right\}$$



17

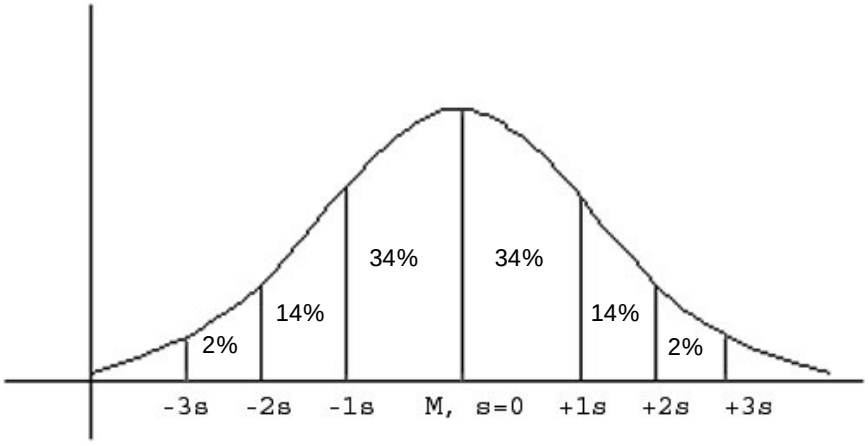
Why be normal? Example

- Now, consider the distribution of X+X+X



18

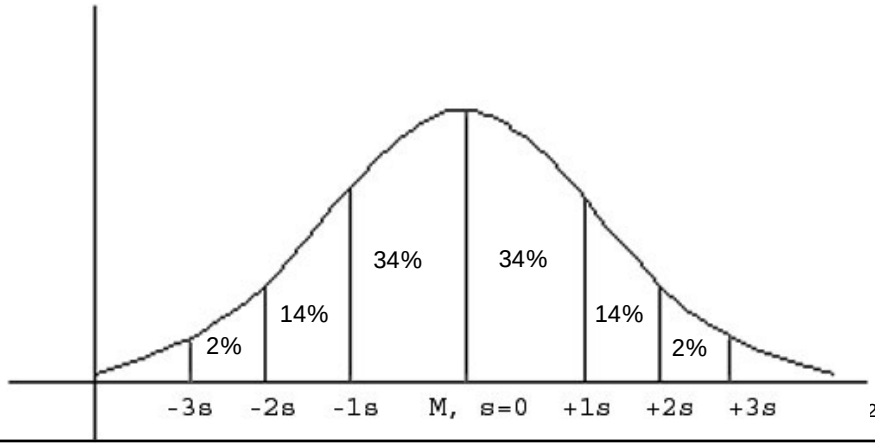
The Normal Curve



19

Estimating probabilities using the normal curve

- p(x>sample>y)



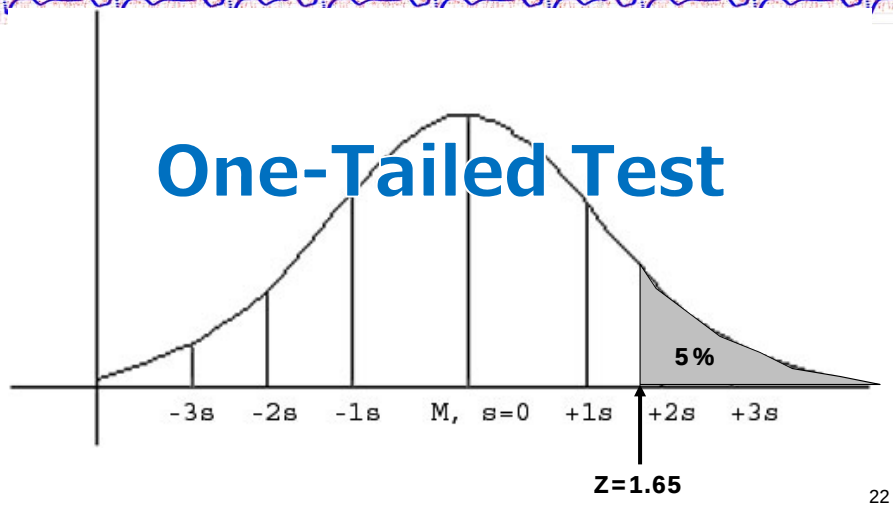
20

Example

- 3 die, each with 1,2,3 equiprobable
- M=6, StdDev=1.4
- What is the probability of
  - Rolling a number greater than 6?
  - Rolling a number greater than 7?
  - Rolling a number greater than 8?
  - Rolling a number greater than 9?

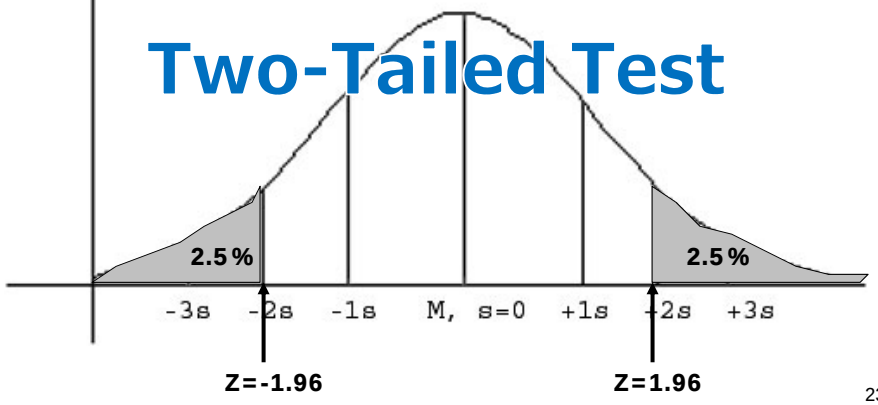
21

The most important parts of the normal curve (for testing)



22

The most important parts of the normal curve (for testing)



23

Hypothesis testing – one tailed

- Hypothesis: sample (of 1) will be significantly greater than known <u>population</u>
  - Population completely known (not an estimate)
- Example – WizziWord experiment:
  - H1: μ WizziWord > μ Word
  - Test criteria: $\alpha = 0.05$, one-tailed
  - Population (Word users): μ Word =150, σ=25
  - What level of performance do we need to see before we can accept H1?

24

Hypothesis testing – two tailed

- Hypothesis: sample (of 1) will be significantly different from known population distribution
- Example – WizziWord experiment:
  - H1: μ WizziWord $\neq \mu$ Word
  - $\alpha = 0.05$ (two-tailed)
  - Population (Word users): μ Word $=150$, $\sigma=25$
  - What level of performance do we need to see before we can accept H1?

25

Standard testing criteria for experiments

- $\alpha = 0.05$
- Two-tailed *why?*

26

Group Exercise

- For each problem, write
  1. What kind of study design is it?
  2. Two populations being compared
  3. Research & Null hypotheses in English
  4. Research & Null hypotheses in terms of Pop means
  5. Test criteria
  6. Test results
     - English
- Critique the study design

27

R
'norm' is the normal distribution

```
#by default, parameters are Z scores
dnorm     #probability density funct
dnorm(1.0) #height of normal curve at Z=1.0
dnorm(1.0,mean=1.2,sd=0.5)

pnorm     #cumulative density funct
pnorm(1.0)#area under curve for Z<=1.0

qnorm     #critical values (cutoffs)
qnorm(.95)#the Z score s.t. area below=.95
```

28

Don't try this at home

- You would never do a study this way.
- Why?
  - Can't control extraneous variables through randomization.
  - Usually don't know population statistics.
  - Can't generalize from an individual.

29

The meaning of not rejecting the null Hypothesis

- "The Power of a Nod and a Glance..."
- Study of differences between CONTent behaviors (e.g., acks) and EMOtional behaviors (e.g. confused expr) – both compared to ENVelope behaviors.
- Hypotheses:
  - ...
  - H3: there is no difference in speech overlaps between conditions CONT and EMO.
  - ...

30

Terminology Soup

- Parametric experimental design
  - IV is numeric & varied continuously
- Non-parametric experimental design
  - IV is categorical & varied in steps
- Parameteric statistic
  - Assumes numeric measure with normal distribution
- Non-parametric statistic
  - Makes no such assumptions

31

Between-Subjects Design

- Have two experimental conditions (treatments, levels, groups)
- Randomly assign subjects to conditions
- Measure numeric outcome in each group

32

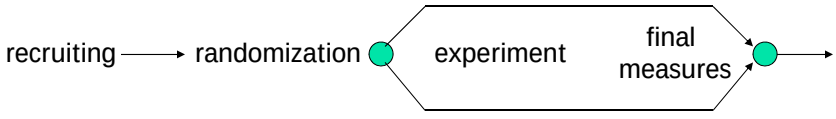
Between-Subjects Design

- Each group is a **sample** from a population
- Big question: are the populations the same (null hypothesis) or are they significantly different?

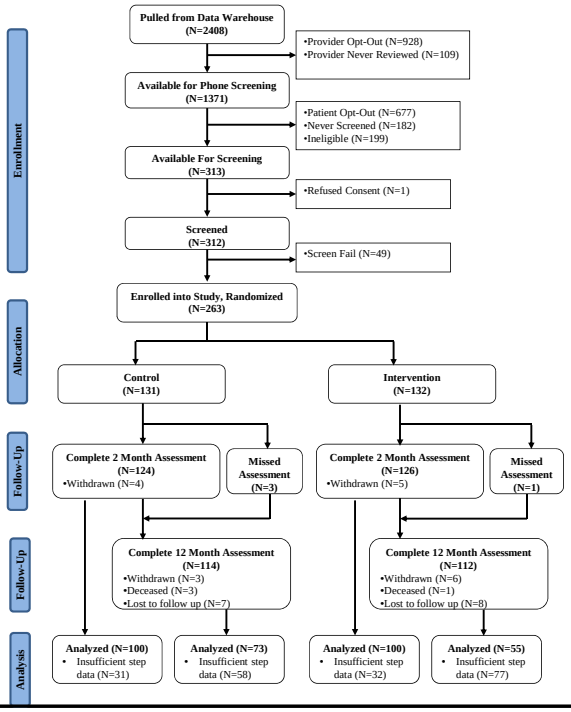
33

Sidebar: Randomization

- Crucial: method must not be applied subjectively
- Point in time at which randomization occurs is important



34



35

Intent-to-Treat

- You want to test a new support line ticket system.
- You randomize 20 support employees to use the new system, 20 to use the old one, then collect satisfaction and performance measures after one month.
- You discover that 6 of the employees using the new system stopped using it after a week.
- What do you do?

36

Intent-to-Treat

- Once a subject is randomized, every effort is made to include their outcome measures (DV) in the analysis
  - Even if they did not use the Intervention
  - Even if they went on vacation
  - Even if they died …
- Efficacy = IV/DV effect under ideal conditions (e.g., lab study) = "method effectiveness"
- Effectiveness (aka "use effectiveness") = IV/DV effect under real world conditions
- Intent-to-treat assesses "use effectiveness"

37

Sidebar: Randomization

- Simple randomization
  - Flip a coin
  - Random number generator
  - Table of random numbers
  - Partition numeric range into number of conditions
- Problems?

38

Sidebar: Randomization

- Blocked randomization
  - Avoids serious imbalances in assignments of subjects to conditions
  - Guarantees that imbalance will never be larger than a specified amount
  - Example: want to ensure that every 4 subjects we have an equal number assigned to each of 2 conditions => “block size of 4”
  - Method: write all permutations of N conditions taken B at a time (for B = block size)
    - Example: 1122, 1212, 2112, 2121, 2211, 1221
  - At the start of each block, select one of the orderings at random
  - Should use block sizes > 2

39

Sidebar: Randomization

- Stratified randomization
  - First stratify Ss based on measured factors (prior to randomization) (e.g., gender)
  - Within each strata, randomize
    - Either simple or blocked

Strata	Sex	Condition assignment
1	M	ABBA BABA...
2	F	BABA BBAA...

40

Sidebar: Control groups

- A controlled experiment ("experimental design") generally compares the results obtained from an experimental sample against a control sample, which is identical to the experimental sample except for the one aspect whose effect is being tested.
- You must carefully select your control group in order to demonstrate that only the IV of interest is changing between groups.
- The control group must also comprise a reasonable comparison.

41

Sidebar: Control groups

- Standard-of-care control (new vs. old)
- Non-intervention control
- "A vs. B" design (shootout)
- "A vs. A+B" design (e.g., S-O-C vs. S-O-C+intervention)

- Problem: the "intervention" may cause more than just the desired effect.
  - Example: giving more attention to intervention Ss in educational intervention
- Some solutions:
  - Attention control
  - Placebo control
  - Wait list control (also addresses measurement issues)

42

Sidebar: Control groups Related concepts

- Blind test – S does not know group
- Double blind test – neither S nor experimenter know

- Manipulation check
  - Test performed just to see if your manipulation is working. Necessary if immediate effect of manipulation is not obvious.
  - "Positive control" test for intervention effect
  - "Negative control" test for lack of intervention effect
  - Example:
    - Student Center Sign: ask students if they saw & read the sign

43

Homework

- Read B&A Ch 14 to 442
- Read Aron Ch 5, Ch 7 to 247 & Ch 8
  - Remedial stats:
    - Distributions of means
    - t distribution
    - Single sample t-test
    - t-test for independent means

44