2.3 Basics of experimental research

Components of an experiment:
- Treatments – Different techniques, devices, or procedures to test
- Units - i.e. human subjects in HCI with specific characteristics
- Assignment method – Way in which units are assigned different treatments

Eg. QWERTY vs DVORAK
   Treatment = type of keyboard
   Units = participants
   - Need participants with NO keyboard experience???
   Assignment method = Toss of a coin (random)

Randomization
Randomization is important. Nobody should be able to predict which subjects get assigned which treatments.
   - Effectively controls HIDDEN factors

Randomize other factors: i.e. if you have 15 tasks, randomize the order of the tasks
   - Helps control learning curve issues, or some tasks easier than others.

2.4 Significance Tests
Why? Consider:

1. Mike’s height is 6’2”. Mary’s height is 5’8”. So Mike is taller than Mary.
2. The average height of three males is 5’5”. The average height of three females is 5’10”. So females are taller than males.

1 is ok because we are comparing a known quantity and drawing a conclusion about that quantity.
What’s wrong with 2?
   - Common Sense?
   - I can find 3 other males and 3 other females easily → counter example is easy
   - There are only 3 individuals in each group. The sizes of the comparison groups are too small
   - The individuals in the male group and the female group are not representative of the general population.

These last 2 have more statistical roots.

When you cannot test a whole group, you test a subset: this is called sampling.
2.4.2 Type I and Type II errors:
- Type I – False Positive or alpha-error (greek alpha)
  Mistake of rejecting the null hypothesis when it is true and should not be rejected
- Type II – “false negative” or beta-error (greek beta)
  o Mistake of not rejecting the null hypothesis when it is false and should be rejected.

Eg.
H0: The defendant is innocent
H1: The defendant is guilty

Chart of Reality vs. Jury Decision

Type I errors are generally considered to be worse than type II errors.

Eg. Touch screen ATM NO difference vs difference.
- Type I error causes unneeded investments
- Type II error maybe is just a lost opportunity

Eg. Drug trial.
- Bad to put people on a drug with no positive effects

2.4.3 Controlling type I and Type II

Probability of making Type I error
“significance level” or p value
Statistical power of a test 1 – beta
Probability of successfully rejecting a null hypothesis when it is false and should be rejected.

Type I errors are worse than Type II

Significance test
P < 0.05 (5%) is generally accepted as significant

Type II: Large sample size can reduce type II errors

2.5 Limits on experimental research

Many limits due to:
  Difficulty in constructing a well-defined and testable hypothesis
  Strict control factors – out of the control of testing when using human subjects
    Experience, abilities, etc...
  Lab based observations