

Experiment Design for Computer Sciences

Week 04 – Inferential Statistics – Course Notes

Claus Aranha

`http://conclave.cs.tsukuba.ac.jp/`

Computer Science Department

2016-05-13 (updated May 13, 2016)

Course Notes/Warnings

Team Project – Don't procrastinate on your team project!

- I have so far only received 1 team e-mail;
- We are almost at the halfway point of the course!

Send me an e-mail (caranha@cs.tsukuba.ac.jp) with the names, student IDs, and mail addresses of your team members.

If you send me the topic of your project, I would be very happy too.

Review of Last Class

- Concepts of Statistics
 - Goats, and why the world sometimes does not follow our intuition
 - Difference between **sample** and **population**
 - The **Central Limit Theorem**
- DAE Chapter 03 – Point Estimators
 - Estimating values for the **population parameters** from the **sample parameters**;
 - **Bias** and **variance** of estimators;
- DAE Chapter 04 – Interval Estimators
 - **Confidence Interval** for a parameter
 - How to interpret the Confidence Interval;

Review of Last Class – Examples

In Manaba, there was a “Chapter04.R” file with example code to calculate Confidence Intervals in R (and plot them).

Please take a look at the code and apply the calculations to your own data. (If we have time, we will review the code later today).

This week, on EDCS

- DAE Chapter 05 – Inferential Statistics:

Where we take our concepts of statistical intervals, and use them to extrapolate information about the population;

Notes on the Lecture Notes 1

- Amos Nathan Tversky: Active Cognitive Psychologist on the field of human cognitive bias and handling of Risk. Humans are more irrational than we would feel comfortable with.
- Descriptive Statistics vs inference statistics;
 - Descriptive: Lets us characterize a population;
 - Inference: Lets us make decisions about a population;
Requires a **degree of certainty** about characteristics of a population;
 - When you are making a claim about research, in many cases you are not making a *description*, you are making a decision to claim something about the data. The decision is:
Is your claim *safe*?
Is your claim *supported by the data*?
How likely is your claim to verify/refute the data?

Notes on the Lecture Notes 1

- Hypothesis definition – Examples
- Karl Popper – Philosopher on the nature of truth. According to him, truth can only be obtained from observations. **BUT** the way that we obtain truth from observations must be well regulated. (Falsification, Black Swan Fallacy)
- Discussion of how to formulate scientific hypothesis;
- Peas Example – page 9: If mean of the sample is “much bigger” or “much smaller” than 500, then we refute the Null Hypothesis. But **How Much** is much bigger or much smaller? This is the key question – anyone has a good answer?

Suggested Reading

- **The Elements of Style(Willian Strunk Jr) – Recommended by Marco Antonio**
`http://www.amazon.com/
Elements-Style-William-Strunk-Jr/dp/
1557427283`
- **How to do Good Research and get it Published (Eamonn Keogh) – Recommended by Mateus Cruz**
`http://www.cs.ucr.edu/~eamonn/Keogh_
SIGKDD09_tutorial.pdf`