

# **Student Outreach Seminar**

September 18, 2007

Dong-Yun Kim

Department of Statistics

Virginia Tech

Good research starts by...

asking interesting, solid **questions:**

## Question #1:

After pouring in multi-million dollars, a pharmaceutical company developed a new drug and wonders: is it any good?

## Question #2:

Why do I get the feeling my brand-new MP3 players keep breaking down right after the warranty expires?

### Question #3:

People say green fields are turning to deserts in Africa. Is this true? Since when?

## Question #4:

They say it's in the genes. How do they know?

As statisticians, we approach those questions with:

- Appropriate statistical tools on the one hand;
- and plenty of common sense on the other.

## Topic #1: Sequential Test

- After surviving numerous pre-tests, the drug is put to the test using human subjects (Phase III clinical trial).
- Quite often human subjects join in different times;
- Large scale clinical trials takes long time to complete, so
- Complete data are sometimes years away.
- But we want the test result fast.

### **The question is: how?**

- Instead of waiting for the complete data, why don't we start testing now, using a portion of data first?
  - As more data become available, do the testing again and again and again...
  - Until we reach a conclusion?

## Topic #2: Change Point Analysis

- Your MP3 player may have a change point in life time distribution;
- After a certain time point, the failure rate increases dramatically.
- Statisticians use change point analysis to detect the change point and estimate it.
- Test of change point poses an interesting “dilemma”: under the null hypothesis (of no change point), the change point parameter disappears!
- One way to overcome this is using Profile Likelihood (PL) and compute the sup distribution of PL-process.

### Topic #3: Spatial and Temporal Analysis

- Amount of ground "greenness" in remote areas can be measured from satellite captured images.
- From these images, composite vegetation indices are computed.
- These indices are studied for any sign of change in vegetative production in the study area.
- The picture shows vegetation changes in a part of East Africa.
- Green squares represent vegetation increases in the past 20 years;
- Pink and red squares represent vegetation decreases in the same period.

## Topic #4: Statistical Genetics

- Which gene “causes” asthma? Are there obesity genes?

Two of the research areas in statistical genetics are:

- Association studies between particular genes and gene variations (SNP: Single Nucleotid Polymorphism), and phenotypes (eye color, height, obesity) and diseases;

- Testing and estimation for a QTL (Quantitative Trait Loci) in a long sequence of genetic markers.

- These areas usually require substantial computer simulation, with permutation test as a popular tool.

- Asymptotic theory can give those 'puters a break: why not just look up the critical value from a table, instead of spending days and weeks on the computer?

## Closing Remarks

- Learn as much as you can while you are a student (and keep doing it afterward), and
- Find one (or three) research topics you are really excited about.

Another question\*: What has been your favorite research discussion for the last few weeks?

\* Nobel Laureate Francis Crick (of double helix fame) chose to study molecular biology in this way.