# Statistics - Challenges and Rewards

## A point to ponder

Speaker: Nabendu Pal

Department of Mathematics
University of Louisiana at Lafayette,
Lafayette, Louisiana, 70504, USA

- **Q. Why Statistics?**

**A. (a) It is one of the most promising scientific areas with a huge potential.**

**(b) It acts as a catalyst in most of the scientific studies/research.**

**(c) As a result, it has a tremendous job prospect.**

- **Looking back -**

**Many useful results in Statistics which we take for granted were derived by the mathematicians in early 18th century. Initially, the research was pioneered by the mathematicians, but then slowly statistics emerged as a separate scientific branch though it uses all the tools that math can offer.**

- **How many Crows in the Kingdom?**

**One day Emperor Akbar (emperor of India, mid 16th century) and Birbal (Akbar's chief minister) were taking a walk in the palace gardens. It was a nice summer morning and there were plenty of crows happily playing around the pond. While watching the crows, a question came into Akbar's head. He wondered how many crows were there in his kingdom.**

**Since Birbal was accompanying him, he asked Birbal this question. After a moment's thought, Birbal replied,**

**"There are ninety-five thousand four hundred and sixty-three (95,463) crows in the Kingdom".**

**Amazed by his quick response, Akbar tried to test him again, "What if there are more crows than you answered?"**

**Without hesitating Birbal replied, "If there are more crows than my answer, then some crows are visiting from other neighboring kingdoms".**

**"And what if there are less crows", Akbar asked.**

**"Then some crows from our kingdom have gone on holidays to other places".**

- **Most of the early results of statistical theory fall under the 'Theory of Probability' which was mainly motivated by gambling.**

- **But more application oriented results were derived from the early 20th century.**

## (A) Various Branches of Statistics (application-wise):

## (a) Design of Experiments:

- Originally started with agricultural studies to see if two or more fertilizers are equally effective or not.
- Essentially deals with allocating resources, and how we should collect data to optimize the info thus gathered.
- This concept is now widely used in Medical studies / Clinical Trials, Sociological studies, Psychometric tests, etc.

# (b) Multivariate Analysis:

- Deals with multidimensional measurements.
- A patient's height, weight, blood pressure, BMI, Cholesterol level, etc. are recorded simultaneously.
- What are the connections among all these measurements? Do we need all these measurements?
- Widely used in psychometric studies,
- Studying personality traits ("e-harmony.com", dating services)

# (c) Time Series and Forecasting:

- **Gather past data, see the trend, and then predict the future.**
- **A lot of this is used in projecting 'global warming' in near and distant future.**
- **Also used in studying the performance of company stocks (IBM, Microsoft, Kraft, Intel, etc.) predict how a portfolio would perform**
- **Used by the investment managers, bankers, brokers, futures traders, and pure speculators.**
- **A lot of economists use this in economic research**
- **Fusion of Statistics & Economics = ECONOMETRICS**

# (d) Decision Theory and Bayesian Analysis:

- Risk analysis, how to make a decision, what is an optimal decision?
- This has a lot of applications in Finance and Insurance industry.
- This has its roots in mathematical 'Game Theory'
- In Economics this is called the 'Utility Theory'.
- Recall the famous movie 'A Beautiful Mind' made a couple of years back. It was about the life of mathematician / statistician / economist Prof. John Nash who won Nobel Prize in 1994 for his work (how to attain optimal equilibrium, known as 'Nash Equilibrium') in Decision Theory.
- After 9-11, researchers are using various decision theoretic techniques to assess 'Bio-terrorism' and / or in Disaster Management. (Think of hurricane Katrina.)
- (Department of Homeland Security is particularly interested in it.)

# (e) Reliability and Life Testing:

- This is about application of statistics in engineering problems.
- How long a car battery lasts? What is the average life of a car battery?
- How do we set the warranty time (36 months or 48 months) for car batteries?
- A satellite is sent to the orbit. What is the chance that it will last more than five years?
- System engineering - Parallel system, Series system - how do they work?
- How to build a more reliable system?
- In oil industry, we want to know the relation between the pipe thickness and its probability of having any damage / rupture, etc.
- The accuracy of a missile is measured by CEP (Circular Error Probable) which is estimated by firing a number of test missiles. Smaller CEP indicates a more reliable missile.
- For example, The Joint Direct Attack Munitions (JDAM), a guided air to surface weapon developed jointly by the USAF and DoN has a CEP of 11 yards.

# (f) Statistical Quality Control:

- **How to monitor products in a manufacturing process meeting certain quality requirements? Is the production process in control? or out of control?**
- **The idea is to market better products.**

# There are other areas too, like –

- Queuing Theory
- Pattern Recognition
- Sample Survey

**From a theoretical side one can look at various areas of statistics as:**

- **Parametric Inference**
- **Non-parametric Inference**
- **Statistical Computations**
- **Stochastic Processes**

**Latest trends in Statistical studies identify two areas as hot areas:**

**1. Biostatistics**

**All possible applications of Statistics in biological sciences which may include :**

- **environmental studies (air/water contamination)**
- **ecological studies (birth/death rates of species)**
- **Spread of a disease (SARS, AIDS)**
- **Also covers genetics, human genome project.**

- **A lot of medical professionals are getting involved in it. Pharmaceutical companies are pouring in a huge amount of resources.**

**Q. Reason?**

**A. More and more complex diseases, Pollution, Ever increasing aging population, etc.**

# 2. Data Mining

- **Deals with huge datasets.**
- **A lot of Computer Science people are involved in it and are studying algorithms to handle massive datasets.**
- **How to find trends and outliers in massive datasets.**
- **Due to internet, and online searches and/or transactions, we have now massive datasets on the users' habits, tastes, likes and dislikes, or other profile info.**

**How to find trends?**

- **'Amazon.com' or 'ebay.com' now use the transaction history to increase their business.**
- **After 9-11, the Data Mining techniques are getting fresh attentions to study individuals' travel patterns, online search patterns, telephone call patterns, etc.**
- **Computer Scientists are interested because such large data handling requires efficient algorithms and machine learning.**
- **A related area: Chaos Theory**

# (B) Job Prospect

- **(a) Academic：Requires graduate degrees - mostly doctoral**
- **(b) Non-academic：Industry – requires MS degree most of the time. But pharmaceutical companies or other companies with large R&D do require doctoral degree.**

- **Government – all degrees (BS, MS, PhD) can come handy**
  - **(i) US Fish & Wildlife**
  - **(ii) FDA (Food & Drug Administration)**
  - **(iii) Census Bureau**
  - **(iv) EPA (Environmental Protection Agency)**
  - **(v) BLS (Bureau of Labor Statistics)**
  - **(vi) NSA (National Security Agency)**
  - **(vii) NIH (National Institute of Health)**
  - **(viii) CDC (Center for Disease Control)**
  - **(ix) SSA (Social Security Administration)**
  - **(x) CPSC (Consumer product & Safety Commission)**
  - **(xi) DNR (Department of Natural Resources)**
  - **(xii) DoJ (Department of Justice)**

## (C) Professional Organizations

### (a) The American Statistical Association

(i) has chapters in every state

(ii) has committees looking after various interest / focus groups

- Minorities in Statistics,
- Women in Statistics,
- Business Statistics,
- Statistics in Law,
- Environmental Statistics
- Geostatistics
- Biostatistics

**(b) Institute of Mathematical Statistics**

**(c) Biometric Society**

**(d) International Statistical Institute**

**(e) Various other organizations based in other countries**

- **Top Ten Reasons to Become a Statistician：**
  - **10. Deviation is considered normal.**
  - **9. We feel complete and sufficient.**
  - **8. We are mean lovers.**
  - **7. Statisticians do it discretely and continuously.**
  - **6. We are right 95 percent of the time.**
  - **5. We can safely comment on someone's posterior distribution.**
  - **4. We may not be normal but we are transformable.**
  - **3. We never have to say we are certain.**
  - **2. We are honestly significantly different.**
  - **1. No one wants our jobs.**