Sampling and Bias Review Quiz

Mallick Hossain

University of Pennsylvania

Course Survey

Thanks to everyone that filled out the midterm evaluation!

Course Survey Results

1. Strengths
    - Enthusiastic and high-energy teaching style
    - Comfortable and collaborative atmosphere
    - Welcomes questions (“His priority is to have us learn the material”)
2. Suggested Improvements
    - More examples and applications, especially sample exam questions as we go along would be great
    - Discuss R in class, learning it separately is difficult without clear links to the course material
    - Take up an R assignment before the R project is due to get more practice of what is expected
    - Make homework due more often or have some form of weekly evaluation

Class Logistics

1. Have you gathered the data for your R project?
    - The correct answer is “Yes”
2. Have you compiled summary statistics of your data?
    - The correct answer is “Yes,” but I will also accept “Once I figure out why R keeps giving me an error message.”
    - This week’s R assignment will get you started on your R project. I will take it up in two weeks.
3. How should I submit my R project?
    - It should be typed it either RMarkdown or TEX. Let me know if this will pose a difficulty (but only after you’ve completed the R Tutorial on RMarkdown).
    - To encourage you to practice with either of these programs, I will give you 2 bonus points for submitting homework typed up in RMarkdown or TEX. (For reference, that’s about an extra 15-20% on your homework grade)

Midterm Questions

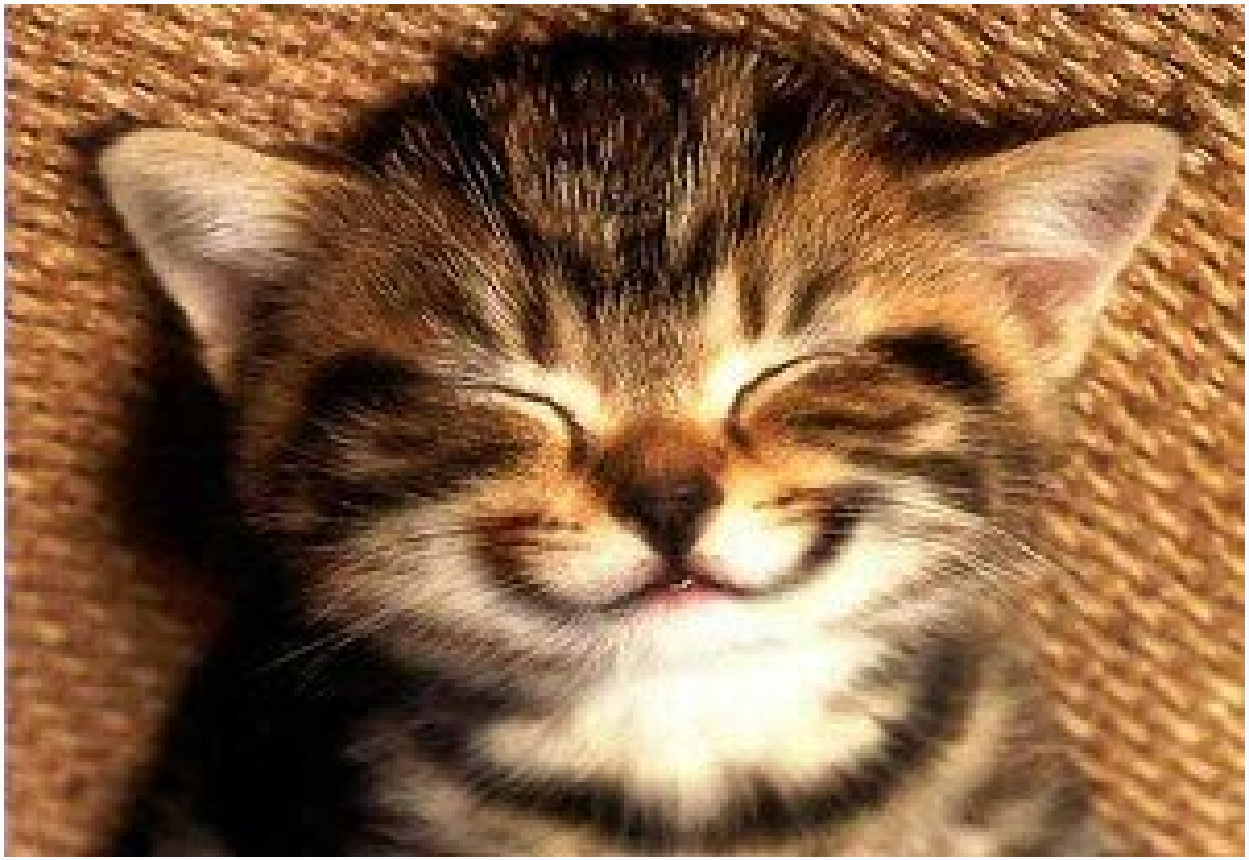
1. When is the next midterm?
    - Next week

Class Logistics

Midterm Questions

1. Can you hold extra office hours since the midterm is next week?
    - Sure! Come by my office between 11-1 on Friday if you have questions!
2. Post on Piazza!

Class Logistics



Homework Questions

1. How do you get the CDF from the PDF?
   - Given a PDF $f(x)$, the CDF can be found by computing $\int_{-\infty}^{x_0} f(x)dx$
2. What does "what is the distribution" mean?
   - This means you should completely specify the random variable that describes whatever I am asking about. Example, if X_1 and X_2 are independent standard normal RVs, what is the distribution of $X_1 + X_2$?
   - $X_1 + X_2 \sim N(0, 2)$
3. What are degrees of freedom?
   - Roughly, it is the number of values that are allowed to vary when computing a statistic. For example, if you are computing the deviations from the mean, you have $n - 1$ degrees of freedom since they have to sum to 0.

Real Questions

1. In real life, are surveys carried out with or without replacement? Why?
    - Surveys are carried out without replacement (i.e. we do not call the same person twice). While this does affect the probability of each person getting chosen (since the unsampled population size changes), given a big population and a small relative sample size, the effect is negligible.
2. What is the difference between an estimator and an estimate?
    - An estimator is a procedure for estimating some parameter. An estimate is the actual result of the procedure given the data. Think of an estimator as a function (f) while an estimate is the function evaluated at a point ($f(x)$).

Real Questions

3. What does “unbiased” mean?
    - Mathematically, this means that the bias is 0, where the bias of an estimator $\hat{\theta}$ is defined as $E[\hat{\theta}] - \theta_0$. In other words, the expectation of the estimator is the true parameter.
4. What is efficiency?
    - Efficiency can only be estimated relative to another estimator. Given two estimators $\hat{\theta}_1, \hat{\theta}_2$, we say that $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$
5. What is mean-squared error?
    - $MSE(\hat{\theta}) = Bias(\hat{\theta})^2 + Var(\hat{\theta})$

Real Questions

6. What does consistency mean?
    - An estimator is consistent if $\lim_{n \to \infty} MSE(\hat{\theta}_n) = 0$
7. Ready for the next lecture?

Real Questions

Sponsored by Home Depot

Not really