MA 677 - Spring 2019, Final project - version 29april

April 29, 2019

Instructions

The purpose of this project is for you to engage with the material we have discussed in the second part of MA677. Some of the questions are primarily computational, but in all cases, you should explain your approach and your conclusions.

During this semester, the publication of the supplemental issue of the American Statistician dealinig specifically with issues of statistical inference has provided resource and a challenge for students in this course.

As we discussed in class,There will be one more question dealing with decision theory added to this project.

Statistics and the Law

In 1977, the U.S. Congress passed the Community Reinvestment Act which had provisions that were intented to motivate financial institutions to meet the needs of communities where they did business. enactment of the Equal Credit Oppportunity Act (Equal Credit Opportunity Act) and compiled under the Home Mortgage Disclosure Act of 1975 (HMDA)

provided to tools for minority groups to gain equal access to credit

In order to establish in reality the equality of opportunity these laws provided on paper, legal action in the courts was necessary and evidence was required.

Once of the most effective organizations acting as a champion of equal rights was ACORN (Association of Community Organizations for Reform Now). In what has become famous testimony before the hous committee on Banking, Financie and Urban Affairs in 1992, ACORN made a statstical arguement that the difference between the rates of mortgage application refusals of wihte applicants and minority applicatns constitued evidence of discrimination. Your job is to use ACORN's data and create the arguements that (1) the data are sufficient evidence of discrimination to warrant corrective action and (2) the data are not sufficient.

Column labels

Column Label	Definition
BANK	name of bank
MIN	refusal rate for minority applicants
WHITE	refusal rate for white applicants
HIMIN	ref usal rate for high income minority applicant
HIWHITE	refusal rate for high income white applicants

Loan data

BANK	MIN	WHITE	HIMIN	HIWHITE
HARRIS TRUST	20.90	3.7	21.4	2.2
NCNB TEXAS	23.23	5.5	8.0	8.0
CRESTAR	23.10	6.7	11.3	3.6
MERCANTILE	30.40	9.0	17.3	5.5
1ST NB COMMERCE	42.70	13.9	38.0	7.6
TEXAS COMMERCE	62.20	20.6	33.3	10.3
COMERICA	39.50	13.4	33.6	9.4
FIRST OF AMERICA	38.40	13.2	29.5	7.3
BOATMANS NATL	26.20	9.3	21.7	7.4
FIRST COMM_L	55.90	21.0	39.1	15.8
PROVIDENT NAT_L	49.70	20.1	36.6	15.3
WORTHEN	44.60	19.1	28.6	10.1
HIBERNIA NAT_L	36.40	16.0	32.9	9.2
SOVRON	32.00	16.0	21.0	13.0
BELL FEDERAL	10.60	5.6	5.8	4.2
SEC PAC AZ	34.30	18.4	24.2	14.1
CORE STATES	42.30	23.3	38.3	15.0
CITIBANK AZ	26.50	15.6	27.3	16.1
MF_ERS HANOVER	51.50	32.4	41.3	25.1
CHEMICAL	47.20	29.7	41.1	26.8

Comparing Suppliers

Acme Student Products sources ornithopters from high schools where students make orithopters as projects in a kinetics sculptor class. Not all of the ornithopers fly. Not all of them look good enough. Acme sells them all after evaluating them as shown in this table:

Rating	Flies	Looks Good	Price
Flying Art	Yes	Yes	$75
Display Art	No	Yes	$45
Dead Bird	NO	No	$10

Acme is currently working with three high school:

a) Area 51 Regional High

b) BDV American Borstal

c) Giffen Prep

	Dead Bird	Display Art	Flying Art
Area 51	12	23	89
BDV	8	12	62
Giffen	21	30	119

Revenue aside, which of the three schools produces the higher quality ornithopters, or are do they all produce about the same quality?

How deadly are sharks?

If you have spent any time in the ocean enjoying activities such as swimming, surfing, sailing, or fishing, you may have seen a shark or two. It might have made you nervous. Of course, a little knowledge is helpful. Hammerhead sharks, for example, rarely attack humans (but are killed in great numbers by ignorant people).

In the past year, an interesting shark attack dataset has been available on Kaggle. The data clearly show that surfing is an ocean sport that accounts for a large percentage of shark attacks on humans. Personally, I have always believed that the sharks in Australia were, on average, a more vicious lot than the sharks in the United States. Now, that you have the data, please help me sort out how U.S. sharks compare with Australian sharks. Explain your analysis in terms that are simple but technically correct, make sure to include an analysis of statistical power.

Power analysis

The R package pwr is an implementation of the methods in Jacob Cohen's *statistical power analysis for the behavioral sciences* (google the title).

In class we discussed, d, the effect size for t tests. Now read the opening pages of Chapter 6 in Cohen's book (pp 179-185) where he explains the use of the arcsine transformation which is implemented in the pwr function ES.h. The problem ES.h addesses is illustrated by Cohen in is example on Page 180 where he notes that in testing the parameter of a binomial distribution, the power to detect the difference between hypothetical parameters .65 and .45 is .48 while the power to detect the difference between hypothetical parameters .25 and .05 is .82, even though the difference between both pairs of values is .20.

Explain the use of the arcsine transformation. How does it work? Why does it work?

Estimators

Use the Method of Moments and MLE to find estimators as described in these three cases.

Exponential

$X_1, \cdots, X_n$ are independent draws from an exponential distribution, $exp(\lambda)$. Find the MLE of λ.

A new distribution

$$f(x) = \begin{cases} (1-\theta) + 2\theta x & \text{for} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the method of moments estimator for θ.

Find the MLE for θ.

Rain in Southern Illinois

In the early 1960's, Floyd Huff and Stanley Changnon of the Illinois State Water Survey at the University of Illinois conducted a study to determine the natural variability of rainfall during summer storms in southern Illinois. They used two raingage networks to collect data during four summers in 1960 through 1964.

In the <illinois storms.zip> zipfile you will find rainfall data consisting of the average rainfall for each summer storm in the years under studey. I have also put a copy of the Changnon/Huff article. The data in the file are from Table 1 in the article.

Your job is to explore the distribution of the rainfall data. We have done this in a variety of ways this semester. You may find that the fitdistrplus package is helpful, but you are not required to use it.

As you explore the data consider what they mean. Are the four years similar? Where some years wetter? If some years were wetter, was it because there were more storms? Or, was it because storms produced more rain?

In their article that Changnon and Huff concluded that the gamma distribution was a good fit for their data. What other distributions might they have considered? Do you agree with Changnon and Huff? Why? Why not?

Using the gamma distribution as your model, produce estimates of the parameters using both the method of moments and maximum likelihood. Use the bootstrap to estimate the variance of the estimates. Compare the estimates which estimates would you present? Why?

Analysis of decision theory article

Refer to:

Charles F. Manski (2019) Treatment Choice with Trial data: Statistical Decision Theory Should Supplant Hypothesis Testing, The American Statistician, 73:sup1, 296-304.

Derive equations (10a), (10b), (10c) in Section 3.2.2.

Use R to reproduce the calculations in Table 1 which is explained in 3.2.3. Describe what you have done and what it means in the context the the treatment decision used as an illustration in the Manski article.