

Mid-Term Examination (30 points total)

1. (10 points) In an Unemployment Insurance experiment, a random sample of size N was drawn from the population of newly unemployed persons. Each member of the sample was offered a "wage-subsidy" treatment: If the unemployed person should obtain employment within 15 weeks of filing for unemployment insurance, the employer would receive a wage subsidy. Among the unemployed persons offered the wage-subsidy treatment, 70 percent accepted the offer and 30 percent chose not to accept the offer. Among the subjects who accepted the offer, 40 percent obtained employment within 15 weeks. Among the subjects who did not accept the offer, 50 percent obtained employment within 15 weeks.

Let $t = 1$ denote the wage subsidy treatment and $t = 0$ denote no wage subsidy. Let $y(t) = 1$ if, under treatment t , an unemployed person would obtain employment within 15 weeks of filing for unemployment insurance; $y(t) = 0$ otherwise. Let $P[y(t) = 1]$ denote the probability that a randomly drawn unemployed person would obtain employment within 15 weeks if he were to receive treatment t . Let $z = 1$ if an experimental subject accepts the offer of the wage subsidy and let $z = 0$ otherwise. Suppose that $N = \infty$.

A. Suppose that a researcher has the empirical evidence from the experiment but has no other information. What can the researcher infer about the *relative treatment effect*, defined by the ratio $P[y(1) = 1]/P[y(0) = 1]$?

B. Suppose the researcher assumes that all unemployed persons want to obtain employment as soon as possible and, moreover, that all unemployed persons know their values of $[y(0), y(1)]$. Using the empirical evidence and these assumptions, now what can the researcher infer about $P[y(1) = 1]/P[y(0) = 1]$?

2. (10 points) In the setting described in question 1, define $R \equiv P[y(1) = 1 | z = 1]/P[y(0) = 1 | z = 0]$ and assume that $0 < R < 1$. Let $R_N \equiv P_N [y(1) = 1 | z = 1]/P_N [y(0) = 1 | z = 0]$ denote the sample ratio of employment success of treatment 1 relative to treatment 0. (Thus, $R_N = .40/.50$ in the available random sample.)

A. Suppose that $N = 100$. What is $\text{Var}(R_N)$?

B. What is $\lim_{N \rightarrow \infty} \text{Prob}(|R_N - R| > .00001)$?

C. What is $\lim_{N \rightarrow \infty} \text{Prob}(R_N = R)$?

3. (10 points) Let (y, x, ε) be a trivariate random variable. Assume that $y = x^\varepsilon$ and that ε is statistically independent of x . Assume that ε is distributed Bernoulli with parameter p , where $0 < p < 1$. Assume that x has a continuous distribution with finite mean and support $(0, \infty)$.

A. What is the best predictor of y given x under square loss?

B. What is the best predictor of y given x under absolute loss?

C. What is the best predictor of x given y under square loss?

Final Examination (50 points total)

Each year many students ride bicycles from Andersen Hall to the athletic center. Consider the population of such bicycle trips. Let y denote the speed of a trip in kilometers per hour (kph) and let x denote the number of centimeters of snow cover on the ground when the trip is made. A random sample of N trips made during a given year is drawn and $(y_i, x_i, i = 1, \dots, N)$ are observed. The questions below concern the problem of predicting trip speed.

Let $f(x) \equiv \alpha + \beta x$ denote the population best linear predictor of y given x under square loss. Let $f_N(x) \equiv \alpha_N + \beta_N x$ be its sample analog. Suppose that in the sample drawn, $N = 100$, $\alpha_N = 30$, and $\beta_N = -1$.

1. (10 points) Two students report different estimates of the sampling distribution of the parameter estimate β_N . Student A reports the asymptotic normal approximation to the sampling distribution of β_N and student B reports the ideal bootstrap approximation to the sampling distribution.

a. One student reports that the sampling distribution has mean -1 and the other reports that the mean of the sampling distribution does not exist. Which student reports which finding? Explain.

b. One student reports that the sampling distribution places positive probability on the event $\{\beta_N = -1\}$ and the other reports that this event has zero probability. Which student reports which finding? Explain.

2. (15 points) A student points out that $f_N(x) < 0$ if $\beta_N < 0$ and $x > -\alpha_N/\beta_N$. The student thinks it nonsensical to predict a negative trip speed and so suggests an alternative sample predictor of y . This alternative is $g_N(x) \equiv \max[0, f_N(x)]$.

a. What is the asymptotic behavior of $g_N(x)$ at each $x \geq 0$? Explain.

b. What is the asymptotic behavior of $\sqrt{N}[g_N(x) - f(x)]$ at each $x \geq 0$? Explain.

3. (10 points) Two students want to learn $P(y \geq 20 \mid x = 5)$, the probability that trip speed is at least 20 kph in trips made when the snow cover is 5 centimeters. Student A uses the cell mean estimate

$$\frac{1}{N_5} \sum_{i \in N(5)} 1[y_i \geq 20],$$

where $N(5)$ is the subset of observations with $x = 5$ and N_5 is the number of such observations. Student B uses a uniform kernel estimate with bandwidth 2. Discuss the considerations that you think are most relevant in comparing the two estimates.

4. (15 points) You are about to make the trip from Andersen Hall to the athletic center on a day with 10 cm of snow cover. You can either walk or ride your bicycle. You know that all students walk at 4 kph when the snow cover is 10 cm. You also know that $P(y > 4 \mid x = 10) = 1$; that is, all bicycle trips made when the snow cover is 10 cm have a trip speed higher than 4 kph. Should you conclude that bicycling is faster than walking on this trip? Explain.

Midterm Examination (30 points total)

A survey firm conducting an election poll can contact voters by any of three modes: mail, telephone, or home interview. Let z denote the mode that the firm uses to contact a voter. Suppose that the firm contacts 1200 voters; 400 by mail ($z = 0$), 400 by telephone ($z = 1$), and 400 by home interview. These voters are asked if they want the Democratic party to regain control of Congress in the year 2000 election. The possible responses are no ($y = 0$), indifferent ($y = \frac{1}{2}$), and yes ($y = 1$). Suppose that all voters have a value of y , but some of them choose not to respond to the survey. Here are the data obtained:

Response to Survey Question	Contact Mode		
	Mail ($z = 0$)	Telephone ($z = 1$)	Home Interview ($z = 2$)
$y = 0$	100	100	100
$y = \frac{1}{2}$	100	100	100
$y = 1$	100	100	100
no response	100	100	100

When answering the questions below, consider these 1200 voters to be the population of interest, not a sample drawn from a larger population.

1. (12 points) You are asked to predict y conditional on the event ($z = 1$). Given the available data, what can you deduce about

- the best predictor of y conditional on ($z = 1$), under absolute loss?
- the best predictor of y conditional on ($z = 1$), under square loss?
- the best linear predictor of y given z under square loss, evaluated at ($z = 1$)?

2. (6 points) Given the available data, what can you deduce about $P(z = 1 | y = 1)$?

(Hint: The sub-population with $y = 1$ may contain persons who do not respond to the survey.)

3. (12 points) Let $y(t)$ denote the response that a voter would give if he or she were contacted by mode t , for $t = 0, 1, 2$. Given the available data, what can you deduce about the average treatment effect $E[y(1)] - E[y(0)]$ if

- the data constitute the only information available.
- voters who do not respond to the survey have the same distribution of $[y(0), y(1), y(2)]$ as those who do respond.

Final Examination (50 points total)

As in Problem Set 5, consider the population of bicycle trips from Andersen Hall to the athletic center. As earlier, let y denote the speed of a trip in kilometers per hour (kph) and let x denote the number of centimeters of snow cover on the ground when the trip is made. A random sample of N trips is drawn and $(y_i, x_i, i = 1, \dots, N)$ are observed. The questions below concern the problem of predicting trip speed.

Let $f(x) \equiv \beta e^{-x}$ denote the *best exponential predictor* of y given x under square loss. That is, β solves the problem $\min_{\beta} E(y - \beta e^{-x})^2$. Let β_N be the sample analog of β . Assume that $P(x = 0) > 0$ and $P(x > 0) > 0$.

1. (15 points) Derive the limiting distribution of $\sqrt{N}(\beta_N - \beta)$ as $N \rightarrow \infty$. Show the steps in your proof. State your finding as explicitly as possible.

2. (20 points) Two students want to estimate $E(y | x = 0)$, the expected trip speed when there is no snow cover. One student recommends evaluation of the sample best exponential predictor $\beta_N e^{-x}$ at $x = 0$, yielding β_N as the estimate. The other student recommends estimating $E(y | x = 0)$ by its sample analog; that is, by the cell mean

$$\gamma_N \equiv \{\sum_i y_i \cdot 1[x_i = 0]\} / \{\sum_i 1[x_i = 0]\}.$$

A. Suppose that the heuristic “asymptotic” distributions of β_N and γ_N (as defined in Goldberger) correctly give the sampling distributions of these two estimates. Under this assumption, what is the mean square error of each estimate? State your findings as explicitly as possible.

B. Discuss the considerations that you think are most relevant in comparing the estimates β_N and γ_N .

3. (5 points) Suppose that $N = 2$ and that the observed values of (y, x) are $(16, 0)$ and $(10, 1)$. Compute the joint bootstrap sampling distribution of (β_N, γ_N) .

4. (10 points) Suppose that the fastest possible bicycle speed on trips from Andersen Hall to the Athletic Center is 30 kph. You examine the observations $(y_i, x_i, i = 1, \dots, N)$ and find that $y_i < 30$ for $i = 1, \dots, N - 1$, but you also find that $(y_N, x_N) = (35, 0)$. Hence you conclude that observation N must be in error. Present alternative approaches to inference on $E(y | x = 0)$ in this setting. State the assumptions under which each approach is valid.

Mid-Term Examination: Monday October 30

(30 points total)

Consider the population of nations in the world at a specified date. For each nation, let

$y = 1$ if GDP per capita is above \$5000; $= 0$ otherwise.

$w = a$ if the nation is a dictatorship; $= b$ if a monarchy; $= c$ if a democracy

$v =$ population size.

Let $P(y, w, v)$ denote the distribution of (y, w, v) across nations.

1. Suppose you know the conditional distribution $P(y|w, v)$ for all support values of (w, v) . However, you do not know the distribution $P(w, v)$. Consider the problem of predicting y given only knowledge of w , not v . What can you deduce about the following quantities? Be as specific as possible.

A (4 points). $P(y = 1 | w = c)$

B (4 points). $M(y | w = c)$

2. For $t \in [a, b, c]$, let $y(t)$ denote the value that the outcome y would take if a nation were to have political system t . Consider the problem of choosing t to solve the maximization problem

$$\max_{t \in [a, b, c]} P[y(t) = 1].$$

Suppose you know the distribution $P(y, w, v)$. What can you conclude about the optimal value of t in the following informational situations:

A (4 points). You have no other information.

B (4 points). You know that $P[y(t) = 1 | v] = P[y(t) = 1]$, for all values of (t, v) .

C (4 points). You know that $P[y(t) = 1 | w, v] = P[y(t) = 1 | v]$, for all values of (t, w, v) .

3. Suppose the world contains uncountably many nations. A random sample of N nations is drawn, with realizations (y_i, w_i, v_i) , $i = 1, \dots, N$. However, some of the sample realizations are lost due to an error in processing the data. Let V_N denote the sample average of v using only the sample data that are not lost. As $N \rightarrow \infty$, describe the limiting behavior of V_N if the data processing error has each of the following forms:

A (3 points). Every third sample realization is lost.

B (3 points). Sample realization i is lost if and only if $w_i = a$.

C (4 points). Sample realization i is lost if and only if $v_i < v_{i-1}$.

Final Examination: Wednesday December 6
(50 points total)

As this examination is written, the certified Florida presidential vote totals are as follows:

Bush = 2,912,790 Gore = 2,912,253 other candidates = 137,898.

Thus, Bush leads Gore in the certified totals by 537 votes. The certified totals do not include 14,000 disputed ballots in Miami-Dade and Palm Beach Counties. Each disputed ballot showed no vote for president when counted by machine. The Florida Supreme Court must decide whether to now count these ballots by hand.

Suppose that the $N = 14,000$ disputed ballots will be hand counted and added to the certified vote totals. For each disputed ballot $i = 1, \dots, N$, let $h(i) = 1$ if the hand count reveals that ballot i is a vote for Gore, let $h(i) = -1$ if it reveals that the ballot is a vote for Bush, and let $h(i) = 0$ otherwise.

The questions below concern the problem of predicting the outcome of a hand count before it is undertaken. When answering these questions, assume that the N disputed ballots are a random sample from an infinite population of potential disputed ballots. Assume that a hand count of the entire population would reveal this distribution of outcomes:

$$\Pr(h = 1) = g \quad \Pr(h = -1) = b \quad \Pr(h = 0) = 1 - g - b, \quad \text{where } 0 < g, 0 < b, \text{ and } g + b \leq 1.$$

Let $\theta(g, b)$ denote the probability that a hand count of the disputed ballots will make the total Florida vote for Gore exceed that for Bush.

1. (10 points) Describe a Monte Carlo procedure to estimate $\theta(g, b)$ at specified values of (g, b) . How would you measure the precision of the Monte Carlo estimate?
2. (10 points) Use Chebychev's inequality to place a bound on $\theta(g, b)$. Be explicit.
3. (10 points) Use asymptotic distribution theory, including the informal "asymptotic distribution" discussed in Goldberger, to approximate $\theta(g, b)$.
4. (10 points) Suppose that there is not enough time to hand count all $N = 14,000$ ballots before the December 12 electoral college deadline. There is only enough time to count $J = 1,000$ ballots. Given this, suppose that the court were to hand count ballots $i = 1, \dots, J$ and to multiply the result by 14. Let $\phi(g, b)$ denote the probability that this partial hand count approach will make the total Florida vote for Gore exceed that for Bush. Use asymptotic distribution theory to evaluate the conditions under which Gore should prefer the partial hand count approach to a full count of the 14,000 disputed ballots. Be explicit.
5. (10 points) On December 4, Judge N. Sanders Saul of Leon County Circuit Court ruled that he would not order a hand count of the disputed ballots. His decision including the following statement:
 "In this case, there is no credible statistical evidence and no other competent substantive evidence to establish by a preponderance a reasonable probability that the results of the statewide election in the state of Florida would be different from the result which has been certified by the State Elections Canvassing Commission." (reported in the New York Times, December 5, 2000, p. A18).
 In this statement, Judge Saul uses the term "reasonable probability." How would you interpret the judge's use of this term if the judge is a frequentist statistician? If he is a Bayesian?

Midterm Examination (Thursday November 1)
30 points total

A public health researcher wants to learn the fraction of adult Americans who have received a dose of the smallpox vaccine. Let

$y = 1$ if a person received the vaccine; $= 0$ otherwise.

The researcher wants to learn $P(y = 1)$, the probability that an adult American received the vaccine.

The researcher draws a very large random sample of adult Americans and asks the sample members if they have received the vaccine. All persons drawn into the sample agree to be interviewed and to answer the question truthfully. However, some persons do not recall if they have received the vaccine and so do not respond. Thus, let

$z = 1$ if a person recalls whether he or she received the vaccine; $= 0$ otherwise.

The researcher observes y for persons with $z = 1$, but does not observe y for persons with $z = 0$. Suppose that the researcher also learns the sex ($x = \text{male}$ or female) of each sample member.

Let $P(y, z, x)$ denote the population distribution and suppose that the sample evidence reveals the following:

$$\begin{array}{ll} P(y = 1, z = 1, x = \text{male}) = 0.3 & P(y = 1, z = 1, x = \text{female}) = 0.28 \\ P(y = 0, z = 1, x = \text{male}) = 0.1 & P(y = 0, z = 1, x = \text{female}) = 0.16 \\ P(z = 0, x = \text{male}) = 0.1 & P(z = 0, x = \text{female}) = 0.06 \end{array}$$

(Note: Interpret the sample frequencies as if they are population probabilities).

1. (4 points) Using the sample evidence alone, what can the researcher conclude about $P(y = 1)$?
2. (4 points) Assume that the fraction of males who received the vaccine equals the fraction of females who received the vaccine. That is,

$$P(y = 1 | x = \text{male}) = P(y = 1 | x = \text{female}).$$
 Now what can the researcher conclude about $P(y = 1)$?
3. (7 points) Assume that among persons of each sex, those who recall whether they received the vaccine are more likely to have received it than are those who do not recall. That is,

$$\begin{array}{l} P(y = 1 | x = \text{male}, z = 1) \geq P(y = 1 | x = \text{male}, z = 0) \\ P(y = 1 | x = \text{female}, z = 1) \geq P(y = 1 | x = \text{female}, z = 0). \end{array}$$
 Now what can the researcher conclude about $P(y = 1)$?
4. (7 points) Assume that persons who received the vaccine are more likely to recall than are those who did not receive the vaccine. That is, suppose that

$$P(z = 1 | y = 1) \geq P(z = 1 | y = 0).$$
 Now what can the researcher conclude about $P(y = 1)$? (Note: This assumption does not condition on x)
5. (8 points) Assume that there exists a blood test which reveals whether or not a person has received the vaccine. Consider the sub-population of persons for whom $z = 0$. Suppose that the researcher administers this blood test to a small random sample of N persons drawn from this sub-population. He finds that N_1 of these persons did receive the vaccine, while $N - N_1$ did not.
 - (a) Use the original sample evidence and this new evidence to form a point estimate of $P(y = 1)$, and prove that this estimate is consistent as $N \rightarrow \infty$.
 - (b) Consider the mean square error of the point estimate as a function of N . How large must N be to guarantee that the mean square error is less than 0.01?

Final Examination (Monday December 10)
50 points total

Background

Please read the attached article “Hints of an Alzheimer’s Aid in Anti-Inflammatory Drugs,” New York Times, November 22, 2001. This article, which reports on a recent study in the New England Journal of Medicine, states that

$M = 10,275$ persons 55 or older were contacted to participate in the study

$N = 6,989$ of these persons actually participated in the study.

$M - N = 3,286$ persons were contacted but did not participate.

Assume that all M persons contacted are eligible for the study; thus the $M - N$ persons who do not participate are eligible persons who are not willing to participate. (This is a simplifying assumption; in fact, some of the $M - N$ were not eligible for the study.) Assume that the M persons contacted are a random sample drawn from a population of potential study participants.

Let t denote a possible treatment, with

$t = 0$ if the person never uses anti-inflammatory drugs during the study period

$= 1$ if the person uses anti-inflammatory drugs for less than two years

$= 2$ if the person uses anti-inflammatory drugs for at least two years.

Let z denote the treatment that a person actually receives. Thus $z = 0, 1, \text{ or } 2$.

Let $y(t)$ denote the outcome if the person were to receive treatment t , with

$y(t) = 1$ if the person develops dementia during the study period

$= 0$ otherwise.

Let $y \equiv y(z)$ denote the outcome that a person actually experiences.

Let w denote willingness to participate in the study, with

$w = 1$ if a person is willing to participate

$= 0$ otherwise.

Let $N(y, z)$ denote the number of study participants who receive treatment z and experience outcome y . Information provided in the newspaper article, combined with some auxiliary computations, reveals that

$$N(0, 0) = 2343$$

$$N(0, 1) = 4123$$

$$N(0, 2) = 230$$

$$N(1, 0) = 210$$

$$N(1, 1) = 80$$

$$N(1, 2) = 3.$$

Let $P[y(0), y(1), y(2), z, w]$ denote the joint population distribution of potential outcomes $[y(0), y(1), y(2)]$, received treatments z , and willingness-to-participate w .

For each value of t , let $P(y = 1 | z = t, w = 1)$ denote the probability of dementia among persons who receive treatment t and are willing to participate in the study. Let $P_N(y = 1 | z = t, w = 1)$ be the sample estimate of this probability. Thus $P_N(y = 1 | z = t, w = 1) = N(1, t) / [N(0, t) + N(1, t)]$.

Questions

1. (15 points) To compare treatments 0 and 2, the researchers who performed this study focused attention on the *observable relative risk* statistic, defined to be this ratio of outcome probabilities:

$$\theta \equiv P(y = 1 | z = 2, w = 1) / P(y = 1 | z = 0, w = 1).$$

To estimate θ , they used the sample analog

$$\hat{\theta}_N \equiv P_N(y = 1 | z = 2, w = 1) / P_N(y = 1 | z = 0, w = 1).$$

What is the limiting distribution of $\sqrt{N}(\hat{\theta}_N - \theta)$? Show your derivation clearly.

2. (10 points) Consider the exact sampling distribution of $\hat{\theta}_N$. Is each statement below true or false. Explain.

A. (5 points) “The mean of $\hat{\theta}_N$ is completely undefined.”

B. (5 points) “The median of $\hat{\theta}_N$ is nearly well-defined.”

3. (15 points) A critic of the study complains that the observable relative risk statistic $\hat{\theta}$ does not appropriately compare treatments 0 and 2. Instead, this critic recommends that attention should focus on the *counterfactual relative risk* statistic, defined to be this ratio of outcome probabilities:

$$\beta \equiv P[y(2) = 1] / P[y(0) = 1].$$

Assume no prior information about the distribution $P[y(0), y(1), y(2), z, w]$. Use the sample data to estimate the identification region for β . Show your derivation clearly and give the numerical result.

4. (10 points) Continue to assume that the objective is to learn β . Suppose that you are the researcher conducting the study and that you have research funds that can be spent in one of two ways, being:

Option A: You can double the size of the random sample from $M = 10,275$ to $2M = 20,550$.

Option B: You can provide participation incentives to the $M - N = 3,286$ persons who originally did not participate in the study, and thereby persuade half of these non-participants (i.e., 1,643 persons) to join the study.

Discuss how you would choose between these options. Explain your reasoning.

New York Times, November 22, 2001, page A18

Hints of an Alzheimer's Aid in Anti-Inflammatory Drugs

By GINA KOLATA

Middle-age and elderly people who took anti-inflammatory drugs like ibuprofen or naproxen for at least two years were apparently protected from Alzheimer's disease, according to a new study by scientists in the Netherlands. Their likelihood of getting Alzheimer's dementia was one-sixth that of people who did not take the drugs.

The study, published today in *The New England Journal of Medicine*, offers hope for preventing Alzheimer's but falls short of being definitive, experts said. They cautioned that the findings did not mean that people should dose themselves with anti-inflammatories, which can have serious side effects, to prevent Alzheimer's.

The researchers, led by Dr. Bruno H. C. Stricker of the Erasmus Medical Center in Rotterdam, invited every person 55 or older in a suburb of Rotterdam, Ommoord, to participate in their study. Of 10,275 who were asked, 6,989 agreed and were eligible. None had Alzheimer's. But by the end of the study, 293 had developed it.

Some participants took anti-inflammatories and some did not, by their own choice and not the researchers' design. Most who took the drugs used them for arthritis.

In the group that took no anti-inflammatory drugs, 210 out of 2,553 developed dementia. But 3 out of 233 who had taken the drugs for at least two years developed it. The dose did not appear to be important, the researchers found.

The effect was specific for Alzheimer's. The drugs did not help other diseases like ministrokes, which can cause memory loss and confusion.

The sole drugs that seemed to affect the risk of Alzheimer's were anti-inflammatories. Aspirin, unlike other anti-inflammatory drugs, did not appear to have a protective effect, possibly, the researchers said, because participants were taking very low doses.

Dr. John Breitner, an Alzheimer's disease researcher at the Johns Hopkins Bloomberg School of Public Health, said the new study was impressive. Unlike many others, it followed healthy people for a long time, an average of 6.8 years, to see who came down with Alzheimer's disease. The investigators also had impeccable records of the patients' use of medicines.

"It's a fantastic study, and they showed a very powerful effect," said Dr. Breitner, who is directing a study of anti-inflammatory drugs in people at risk of becoming ill with Alzheimer's.

A potential flaw in the Dutch study is that the scientists did not randomly assign people to take active drugs or dummy pills for comparison purposes. For that reason, they cannot be certain it was the drugs and not some other characteristic of the patients that made the difference.

Because the drugs have side effects like serious, sometimes fatal, stomach bleeding, medical experts advise healthy people to await the results of randomized trials now under way before taking anti-inflammatory drugs other than aspirin in the hope of preventing Alzheimer's.

A strength of the study is that the researchers had highly reliable information about what drugs the participants took, because for most of the study period nonsteroidal anti-inflammatory drugs were available in the Netherlands only by prescription, and the Dutch keep excellent records on the use of prescription drugs.

By contrast, in earlier studies, researchers had no good way of knowing what drugs people had taken. Some studies with patients suggested that anti-inflammatories prevented Alzheimer's, whereas others did not.

Studies in the laboratory and in animals had also suggested that anti-inflammatory drugs prevented Alzheimer's disease or delayed its onset.

Alzheimer's disease experts say that clinical trials that may confirm or refute the Dutch study will be completed soon.

In one study supported by the National Institute on Aging, people who already have Alzheimer's disease are taking either dummy pills or anti-inflammatory drugs to see whether the drugs can slow the progression of the disease. The researchers will have an answer by the middle of 2002, said Dr. Neil Buckholtz, chief of the Dementias of Aging branch at the institute.

Another study, also supported by the institute, involves healthy people 70 or older with family histories of Alzheimer's. Those subjects, too, are taking anti-inflammatories or dummy pills. That study will take at least five more years, Dr. Buckholtz said.

For now, said Dr. Steven DeKosky, who directs the Alzheimer's Disease Research Center at the University of Pittsburgh, the Dutch study is encouraging.

"This is solid circumstantial evidence," Dr. DeKosky said. "It is probably the strongest evidence so far that anti-inflammatory drugs can quiet things down perhaps enough to delay the onset of the disorder.

"Now it's up to the clinical trials to show us whether or not they are an effective prevention or an effective treatment strategy."

Midterm Examination (Tuesday November 5)
30 points total

Today – November 5, 2002 – is election day in Illinois. A researcher would like to learn the frequencies with which men and women vote in the election. Thus, consider the population of Illinois residents who are registered to vote and let

$x = 1$ if a registered voter is male,

$= 0$ if a registered voter is female.

$y = 1$ if a registered voter does vote in today's election.

$= 0$ if a registered voter does not vote in today's election.

The researcher would like to learn $P(y = 1 | x = 1)$ and $P(y = 1 | x = 0)$. The researcher has no prior information about the population distribution $P(y, x)$.

1. (10 points) Suppose that the Secretary of State of Illinois maintains a complete record of who is registered, their sex, and whether they vote. After the election, the Secretary of State draws two separate random samples of registered voters, one being a sample of $N(1)$ males and the other a sample of $N(0)$ females, and gives the researcher the sample realizations of (y, x) ; here $N(1)$ and $N(0)$ are predetermined numbers. Suppose that the researcher uses the empirical voting frequencies $P_{N(1)}(y = 1 | x = 1)$ and $P_{N(0)}(y = 1 | x = 0)$ to estimate $P(y = 1 | x = 1)$ and $P(y = 1 | x = 0)$. In particular, he uses $P_{N(1)}(y = 1 | x = 1) - P_{N(0)}(y = 1 | x = 0)$ to estimate $P(y = 1 | x = 1) - P(y = 1 | x = 0)$.

(a) What is the mean square error (m. s. e.) of this estimate? Be explicit.

(b) Suppose that the sum $N = N(0) + N(1)$ of the two sample sizes is fixed, but the researcher can request any allocation of N between males and females. What value of $[N(0), N(1)]$ achieves the minimum m. s. e.?

(c) Can the researcher use the result in (b) to choose $[N(0), N(1)]$? If so, how? If not, why not?

2. (10 points) Suppose that a political polling firm maintains a complete record of registered voters. After the election, the polling firm draws a random sample of registered voters and asks the sampled persons their sex (x) and whether or not they voted in the election (y). All persons reveal their sex but some do not respond to the voting question. Let $z = 1$ if a person reveals y and $z = 0$ otherwise. Assume that when persons do answer questions, they always do so truthfully. Suppose that the polling firm continues to draw persons at random until it obtains $N(1)$ males who do respond to the voting question; here $N(1)$ is a predetermined number. The firm gives the researcher the sample realizations of (y, x) for these $N(1)$ persons.

(a) Suppose that the researcher uses the reported sample voting frequency $P_{N(1)}(y = 1 | x = 1, z = 1)$ to estimate $P(y = 1 | x = 1)$. What is the m. s. e. of this estimate? Be explicit, writing the m. s. e. as a function of sample size and of those features of $P(y, z | x)$ which are and are not point-identified by the sampling process.

(b) Assume that the researcher knows the response probability $P(z = 1 | x = 1)$. Using this information and that revealed by the sampling process, what is the identification region for the m. s. e.?

3. (10 points) Suppose that a political polling firm places observers at all voting locations and records the sex of all persons who vote. Thus, the polling firm learns $P(x = 1 | y = 1)$ and $P(x = 0 | y = 1)$. Suppose that the researcher wants to learn the ratio $P(y = 1 | x = 1) / P(y = 1 | x = 0)$. What does knowledge of $P(x | y = 1)$ reveal about the ratio $P(y = 1 | x = 1) / P(y = 1 | x = 0)$? Be explicit.

Final Examination (Monday December 9)
50 points total

A large body of empirical research in experimental economics has analyzed the behavior of players in the *ultimatum game*. The ultimatum game is a two-player game of proposal and response. One player (the proposer) is given an endowment of T dollars and is asked to offer part of this money to the other player (the responder). The proposer decides how many dollars to offer the responder; thus, the feasible offers are the integers $[0, 1, 2, \dots, T]$. The responder then decides whether to accept or reject the offer. If the responder accepts an offer of t dollars, the responder receives t dollars and the proposer keeps the remaining $T - t$ dollars. If the responder rejects the offer, both players receive nothing.

Suppose that N pairs of players are drawn at random from a population, with the first player in each pair designated the proposer and the second designated the responder. Each pair plays the ultimatum game in an anonymous manner; that is, players do not see who each other are. Let $x \in [0, 1, 2, \dots, T]$ denote the offer made by a proposer. Let $y = 1$ if the responder accepts the offer and $y = 0$ if he rejects it. Then the data obtained from the N plays of the games are $[(y_i, x_i), i = 1, \dots, N]$.

1. (15 points) The experimenter wants to estimate the conditional response probability $P(y = 1 | x = 0)$; that is, the probability that the responder accepts an offer of zero dollars. To do this, she uses the empirical conditional response probability

$$P_N(y = 1 | x = 0) \equiv (\sum_{i=1, \dots, N} y_i \cdot 1[x_i = 0]) / (\sum_{i=1, \dots, N} 1[x_i = 0]),$$

provided that this quantity is well-defined. If $P_N(y = 1 | x = 0)$ is not well-defined, she uses the overall empirical response probability $P_N(y = 1) \equiv (\sum_{i=1, \dots, N} y_i) / N$. Thus, the estimate of $P(y = 1 | x = 0)$ is

$$\theta_N \equiv \begin{cases} P_N(y = 1 | x = 0) & \text{if } \sum_{i=1, \dots, N} 1[x_i = 0] > 0, \\ P_N(y = 1) & \text{if } \sum_{i=1, \dots, N} 1[x_i = 0] = 0. \end{cases}$$

Characterize fully the behavior as $N \rightarrow \infty$ of $\sqrt{N}[\theta_N - P(y = 1 | x = 0)]$. [Hint: The asymptotic behavior of this quantity depends on the specifics of the population distribution $P(y, x)$.]

2. (15 points) The experimenter is considering conducting an additional M plays of the game, with new randomly drawn players. A budget of B dollars is available to pay players when offers are accepted. The cost of M plays is the random variable $C_M \equiv T(\sum_{j=1, \dots, M} y_j)$. The experimenter is concerned that the cost of the M plays may exceed the budget. Use the informal "asymptotic distribution" of a sample mean to obtain an estimate of $\text{Prob}(C_M > B)$. Be explicit.

3. (15 points) A game theorist wants to learn the distribution $P(x)$ of offers. The theorist believes that players are entirely motivated by self-interest and that this is common knowledge. Given this, he deduces that:

- (a) when $x = 0$, responders are indifferent between accepting and rejecting the offer,
- (b) when $x > 0$, responders always accept the offer.

Going further, he deduces that proposers never offer more than $x = 1$. Therefore, the game theorist asserts that $P(x > 1) = 0$. Hence, $P(x = 0) + P(x = 1) = 1$.

The above reasoning leaves unspecified the magnitude of $P(x = 0)$. [Note: It suffices to consider $P(x = 0)$ because the theorist believes that $P(x = 1) = 1 - P(x = 0)$.] The theorist places a subjective prior probability distribution ϕ_0 on $P(x = 0)$. Given the data $(x_i, i = 1, \dots, N)$, what is the theorist's posterior subjective distribution on $P(x = 0)$? Be explicit.

4. (5 points) A researcher asserts the following two-part hypothesis:

- (a) every responder who accepts the offer he actually receives would also accept any larger offer.
- (b) every responder who rejects the offer he actually receives would also reject any smaller offer.

Using the data $[(y_i, x_i), i = 1, \dots, N]$, is this hypothesis testable? If yes, explain how. If not, explain why not.

Midterm Examination (40 points total)

The Illinois State Police monitors driving on the interstate highways that pass through Illinois. A lawsuit has been filed asserting that the State Police discriminates against out-of-state drivers when officers stop vehicles for speeding on the part of the Northwest Tollway (I-90) near the Wisconsin border. The plaintiff asserts that discrimination against out-of-state drivers exists if

$$(1) \quad \alpha \equiv P(y = 1 | z = 0) - P(y = 1 | z = 1) > 0.$$

These probabilities pertain to the population of vehicles that cross the Illinois-Wisconsin border on I-90. Here $y = 1$ if the Illinois State Police stopped a vehicle for speeding and $y = 0$ otherwise. The variable $z = 1$ if the vehicle had an Illinois license plate and $z = 0$ otherwise. As evidence of discrimination, the plaintiff in the lawsuit has provided this empirical evidence:

$$P(y = 1 | z = 0) = 0.02, \quad P(y = 1 | z = 1) = 0.01.$$

In defense, the State Police has asserted that inequality (1) does not imply discrimination. The State Police asserts that discrimination exists if

$$(2) \quad \beta \equiv P(y = 1 | s > 65, z = 0) - P(y = 1 | s > 65, z = 1) > 0,$$

where s is vehicle speed and where 65 miles per hour is the posted speed limit.

Questions 1-3 ask what one can learn about β when empirical knowledge of $P(y | z)$ is combined with certain assumptions. Thus, suppose that $P(y=1 | z=0) = 0.02$ and $P(y=1 | z=1) = 0.01$ when answering these questions.

1. (8 points) Assume that the police are more likely to stop vehicles that exceed the speed limit than ones that do not exceed the speed limit, whether the vehicle belongs to a resident or not. Thus, assume that

$$(3) \quad P(y = 1 | s > 65, z = j) > P(y = 1 | s \leq 65, z = j), \quad j = 0, 1.$$

Given this assumption, what can one learn about β ?

2. (8 points) Assume that the police never stop vehicles driven at or below the speed limit. Also, more than half of all vehicles exceed the speed limit, whether the vehicle belongs to a resident or not. Thus, assume that

$$(4a) \quad P(y = 1 | s \leq 65, z = j) = 0, \quad j = 0, 1;$$

$$(4b) \quad P(s > 65 | z = j) > 0.5, \quad j = 0, 1.$$

Given these assumptions, what can one learn about β ?

3. (8 points) Assume that the police follow a strict rule when deciding whether to stop a vehicle, this being to stop the vehicle if and only if its speed exceeds 80 miles per hour. Thus, assume that

(5) $y = 1 \Leftrightarrow s > 80$.

Given this assumption, what can one learn about β ?

4. (8 points) Now suppose that $P(y|z = 0)$ and $P(y|z = 1)$ are not known. However, someone has been able to learn the conditional probabilities $P(y = 1 | s, z)$ for all values of (s, z) . Given knowledge of $P(y = 1 | s, z)$, what can one learn about β ?

5. (8 points) Suppose that a researcher is able to learn the distributions $P(s|z = 0)$ and $P(s|z = 1)$ of vehicle speeds of non-residents and residents. The researcher reports these distributional statistics:

$$E(s|z = 0) = 70, \quad E(s^2|z = 0) = 4925.$$

$$E(s|z = 1) = 75, \quad E(s^2|z = 1) = 5629.$$

Given this information, what can one learn about $P(s > 80 | z = 0)$ and $P(s > 80 | z = 1)$?

Final Examination (40 points total)

This “percent-chance” subjective-expectations query has been posed recently to persons sampled in the Michigan Survey of Consumers:

The next question is about investing in the stock market. Please think about the type of mutual fund known as a diversified stock fund. This type of mutual fund holds stock in many different companies engaged in a wide variety of business activities. Suppose that tomorrow someone were to invest one thousand dollars in such a mutual fund. Please think about how much money this investment would be worth one year from now. What do you think is the percent chance that this one thousand dollar investment will increase in value in the year ahead, so that it is worth more than one thousand dollars one year from now?

The results are as follows:

persons responding	sample average	sample std. dev.	sample quantiles			persons not responding
			25	50	75	
3257	42.0	28.6	20	50	60	286

The first column in the table gives the number of sampled persons who responded to the query, the middle columns give various sample statistics for the respondents, and the last column gives the number of nonrespondents. When answering the questions below, assume that all persons are capable of responding to the query, although some choose not to do so. Assume that the N persons sampled in the Survey of Consumers are a random sample of the population of interest. Let P(y, z) denote the population distribution of (y, z), where y is the percent-chance subjective expectation and z is a binary variable indicating whether a person responds to the query (z = 1) or not (z = 0).

1. (10 points) A researcher wants to estimate the population mean E(y). To do this, she uses the average response of sample members who responded to the query; that is,

$$\theta_N \equiv (\sum_{i=1, \dots, N} y_i \cdot 1[z_i = 1]) / (\sum_{i=1, \dots, N} 1[z_i = 1]).$$

Characterize fully the behavior of $\sqrt{N}[\theta_N - E(y)]$ as $N \rightarrow \infty$.

2. (10 points) A researcher wants to learn $E(y|z = 1)/[100 - E(y|z = 1)]$. Propose a consistent estimate and use the data in the table to give its informal “asymptotic distribution.”

3. (10 points) Let $w \equiv y/(100 - y)$ be the odds that a person places on the mutual fund increasing in value. A researcher wants to estimate $E(w|z = 1)$ and $Med(w|z = 1)$. Can these two quantities be estimated consistently using the data shown in the table? If a quantity can be consistently estimated using the data in the table, give the estimate. If a quantity cannot be consistently estimated, explain why not.

4. (10 points) Consider a person who responds $y = 50$ to the survey query. Let v denote the value of the mutual fund one year from now, where “now” means the date when the survey is administered to the person. Suppose that, at the time of the survey, the person places a subjective distribution on v. What does the response $y = 50$ reveal about the person’s subjective mean and median for v?

Midterm Examination (40 points total, Thursday October 28)

1. (6 points) County election officials want to make a point prediction of the number of persons in a tiny village who will vote in the election next week. The village has two eligible voters, denoted $j = 1$ and 2 . Let $y_j = 1$ if person j will vote and $y_j = 0$ otherwise. The election officials know that the voting probabilities for the two voters are as follows: $P(y_1 = 1) = P(y_2 = 1) = 0.5$ and $P(y_1, y_2) = P(y_1)P(y_2)$.

- What is the best predictor of $y_1 + y_2$ under square loss?
- What is the best predictor of $y_1 + y_2$ under absolute loss?

2. (6 points) Consider question 1 again, except that now $P(y_1, y_2) \neq P(y_1)P(y_2)$. Instead, the election officials know that $P(y_1 = y_2) = 1$.

- What is the best predictor of $y_1 + y_2$ under square loss?
- What is the best predictor of $y_1 + y_2$ under absolute loss?

3. (7 points) Many polling organizations use a *likely-voter* model to predict which eligible voters will actually vote in the election. These pollsters report voting intentions only for those classified as *likely voters*. Recently, a pollster has reported these findings for the national population of eligible voters:

40 percent of eligible voters are likely to vote and 60 percent are unlikely to vote. Within the group of *likely voters*, 45 percent prefer Kerry, 45 percent prefer Bush, and 10 percent are undecided.

Accept these findings as correct; thus, ignore sampling error and the possibility of voting for candidates other than Kerry and Bush. Assume that persons who say they prefer one candidate to the other will not change their minds before the election. Assume that all *likely voters* will actually vote in the election. Suppose that nothing is known about the actual voting behavior of *unlikely voters*. In this setting, what can you conclude about the fraction of actual voters who will vote for Kerry?

4. (7 points) Consider question 3 again, except that now information is also available about the voting behavior of *unlikely voters*. In particular,

25 percent of the *unlikely voters* will actually vote. Within the group of *unlikely voters*, 40 percent prefer Kerry, 40 percent prefer Bush, and 20 percent are undecided.

Now what can you conclude about the fraction of actual voters who will vote for Kerry?

5. (7 points) After each presidential election in the United States, the National Election Survey (NES) asks Americans citizens about their voting behavior. Let

$y = 1$ if a person actually voted in the election; $y = 0$ otherwise.

$s = 1$ if a person tells the NES interviewer that he or she did vote; $s = 0$ otherwise.

t = election year to which the survey pertains.

$P_t(y, s)$ = joint distribution of (y, s) in the American population in election year t .

Suppose that, in 1996 and 2000, the NES asked all American citizens about their voting and that all such persons responded to the survey. Suppose that 58 percent of citizens reported $s = 1$ in 1996 and that 60 percent reported $s = 1$ in 2000. Thus, the data reveal that $P_{1996}(s = 1) = 0.58$ and $P_{2000}(s = 1) = 0.60$.

It is widely thought that NES respondents tend to overstate their true voting behavior. Call this

Assumption 1: $P_t(y = 1) < P_t(s = 1)$, all t .

It is sometimes thought that the survey provides accurate information about trends. Call this

Assumption 2: $P_t(y = 1) - P_{t-4}(y = 1) = P_t(s = 1) - P_{t-4}(s = 1)$, all t .

Let Assumptions 1 and 2 hold. What do the NES data for 1996 and 2000 reveal about $P_{2000}(y = 1)$?

6. (7 points) Consider question 5 again, but without making Assumptions 1 and 2. Instead, you know that the actual rate of voting in 2000 was $P_{2000}(y = 1) = 0.50$. What do this fact and the NES data reveal about $P_{2000}(y = 1 | s = 0)$ and $P_{2000}(y = 1 | s = 1)$?

Final Examination (40 points total, Thursday December 2)

Background: In late 2003, the pharmaceutical firm Amgen began a clinical trial to evaluate GDNF, a new treatment for advanced Parkinson's Disease. A total of N patients with advanced Parkinson's Disease agreed to participate in the trial. The *treatment group* consisted of n subjects who received GDNF; these n persons were randomly drawn from the N patients who agreed to participate in the trial. The *control group* consisted of the remaining $N - n$ subjects, who received a placebo. The trial was double blind; that is, the subjects and the physicians treating them did not know which subjects were in the treatment group and which were in the control group.

Amgen halted the trial after six months. The stated reason was that the firm had found no statistically significant difference in the mean outcomes of the treatment and control groups.

Assumptions and Notation: Let J denote the population of all persons with advanced Parkinson's Disease. Suppose that some members of J are willing to participate in the clinical trial and others are not, with $z_j = 1$ if person j is willing to participate and $z_j = 0$ if not. Assume that the N patients who actually did participate in the trial are a random sample from the sub-population with $z = 1$.

Define these treatments for advanced Parkinson's Disease:

$t = 0$: The standard treatment administered to persons in population J .

$t = 1$: Receipt of GDNF, with full knowledge of the patient and his physician

$t = 2$: Receipt of GDNF under the double-blind conditions of the clinical trial

$t = 3$: Receipt of the placebo, with full knowledge of the patient and his physician

$t = 4$: Receipt of the placebo under the double-blind conditions of the clinical trial

Let $y(t)$ denote the outcome of interest under treatment t . This outcome is binary, with $y(t) = 1$ denoting a success and $y(t) = 0$ a failure of the treatment. Let \bar{y}_2 and \bar{y}_4 denote the observed average outcomes in the treatment and control groups, respectively.

1. (10 points) Amgen used the trial outcomes to test the null hypothesis $H_0: E[y(2)|z = 1] = E[y(4)|z = 1]$. Suppose that Amgen used a one-tailed test of the form: Reject H_0 if $\bar{y}_2 - \bar{y}_4 > k$, where k is a chosen constant. Is it possible to choose k so as to make the probability of a Type 1 error equal 0.05? If so, show how. If not, explain why not and propose a reasonable method for choosing k which ensures that the probability of a Type 1 error is no greater than 0.05. (Note: Do not use asymptotic arguments when answering this question.)

2. (10 points) Suppose that subjects in the trial are randomly assigned to the treatment group with probability q and to the control group with corresponding probability $1 - q$, where $0 < q < 1$. What is the informal "asymptotic distribution" of $\bar{y}_2 - \bar{y}_4$?

3. (10 points) A Bayesian statistician wishes to use the Amgen data to form a posterior distribution for $E[y(2)|z = 1]$. He requests Amgen to provide the values of $(n, N - n, \bar{y}_2, \bar{y}_4)$. However, the firm only gives him the values of (n, \bar{y}_2) , reasoning that the values of $(N - n, \bar{y}_4)$ carry no information about $E[y(2)|z = 1]$. Considering this matter from the Bayesian perspective, in what circumstances is Amgen correct/incorrect that $(N - n, \bar{y}_4)$ is uninformative about $E[y(2)|z = 1]$?

4. (10 points) A frustrated physician complains that the Amgen trial data are not very helpful to him in choosing treatments. His patients are randomly drawn from J and he assigns treatments with full knowledge of the patient and physician; hence, the physician wants to know $E[y(1)]$, not $E[y(2)|z = 1]$. Suppose that

(a) The Amgen sample sizes n and $N - n$ are large enough that the trial reveals $E[y(2)|z = 1]$ and $E[y(4)|z = 1]$ with essentially no sampling error.

(b) The physician knows the fraction $P(z = 1)$ of persons who are willing to participate in the trial.

(c) The physician knows that $y_j(1) \geq y_j(2)$ for each person $j \in J$.

Given this information, what can the physician conclude about $E[y(1)]$?

Midterm Examination
October 25, 2006: 40 points total

1. This question concerns survey research performed by a sports sociologist on the baseball perspectives of American married couples as the 2006 World Series between the Detroit Tigers and St. Louis Cardinals is about to begin. The researcher draws a random sample of couples and inquires about their preference for the winning team. However, the survey has some nonresponse.

Let the population of married couples be described by the probability distribution $P(z_t, y_t, t = 0, 1)$. Here $t = 0$ denotes the wife and $t = 1$ denotes the husband.

$y_t = 1$ if person t prefers that the Tigers win the 2006 World Series, $= 0$ otherwise.

$z_t = 1$ if person t reports his or her preference on the survey, $= 0$ if t does not respond to the question.

The survey findings are given in the table below.

wife response	husband response		
	$z_1 = 1, y_1 = 0$	$z_1 = 1, y_1 = 1$	$z_1 = 0$
$z_0 = 1, y_0 = 0$	0.20	0.05	0.05
$z_0 = 1, y_0 = 1$	0.05	0.30	0.05
$z_0 = 0$	0.10	0.10	0.10

Assume that persons who respond to the question always do so truthfully. Assume that the response frequencies in the table are population probabilities, not just sample estimates of those probabilities. When answering the questions below, use the empirical evidence alone.

- A. (6 points) What can the researcher infer about $P(y_0 = y_1)$, the probability that a (wife, husband) couple share the same preference?
- B. (6 points) What can the researcher infer about $E(y_1 - y_0)$, the mean difference in the preference of husbands and wives?
- C. (6 points) What can the researcher infer about $M(y_1 - y_0)$, the median difference in the preference of husbands and wives?
- D. (6 points) What can the researcher infer about $M(y_1) - M(y_0)$, the difference between the median preference of husbands and wives?

2. Let $P(y, z, x)$ be the population distribution of (y, z, x) . All three variables are binary, taking the value 0 or 1. All population realizations of (z, x) are observed. Realizations of y are observed when $z = 1$.

Several assumptions are given below. Take each in turn and determine

- (i) whether the assumption is refutable
- (ii) the identification region for $P(y = 1)$ under the assumption.

In each case explain your answer.

A. (4 points) Assume that $P(y|x = 1, z = 1)$ stochastically dominates $P(y|x = 0, z = 1)$.

B. (8 points) Assume that $P(y|x = 1)$ stochastically dominates $P(y|x = 0)$.

C. (4 points) Assume that the *linear probability model* $P(y = 1 | x) = \alpha + \beta x$ holds, where (α, β) are parameters.

Final Examination
November 29, 2006: 40 points total

1. IPD Section 11.3 considered choice between a status quo treatment $t = a$ and an innovation $t = b$ using data from a randomized experiment with partial compliance. The outcome is binary, $\alpha \equiv P[y(a) = 1]$ and $\beta \equiv P[y(b) = 1]$. The objective is to choose a treatment allocation $\delta \in [0, 1]$ that maximizes the population rate of treatment success.

Suppose that two experiments are performed. In both, subjects assigned to treatment a must comply but those assigned to treatment b can choose not to comply. The experiments are identical except that, in the second one, subjects assigned to treatment b are paid \$100 if they comply with the assigned treatment. The analogous subjects in the first experiment do not receive this payment. The findings are as follows:

Experiment 1: $\alpha = 0.35$ $P(y = 1 | \zeta = b, z = b) = 0.38$ $P(z = b | \zeta = b) = 0.68$.

Experiment 2: $\alpha = 0.35$ $P(y = 1 | \zeta = b, z = b) = 0.30$ $P(z = b | \zeta = b) = 0.80$.

Assume that offering \$100 for compliance has no effect on the population distribution of treatment response $P[y(\cdot)]$. Consider a planner who observes the findings of both experiments.

A. (6 points) What is (are) the Bayes treatment rule(s) for a planner who places a uniform subjective distribution on the identification region for β ?

B. (6 points) What is (are) the minimax-regret treatment rule(s)?

2. Suppose that each member of a population makes a binary choice, whether or not to purchase one unit of a specified product. Let $d(\cdot)$ denote the population demand function; thus, $d(t) \in [0, 1]$ is the fraction of the population who would make the purchase if price equals t . Let the realized price be p and the realized demand be $d \equiv d(p)$.

You observe the realized (p, d) , which satisfy $0 < p < 1$ and $0 < d < 1$. You know that $d(t) \in [0, 1]$ for all t and that $d(\cdot)$ is a weakly downward sloping function of t . You have no other information. Your objective is to maximize revenue. That is, you want to choose t to solve the problem $\max_{t \geq 0} t \cdot d(t)$. Answer these questions, giving your reasoning:

A. (4 points) What are the feasible states of nature?

B. (8 points) What prices are dominated?

C. (6 points) What is (are) the maximin price(s)?

3. (10 points) Let the set of feasible treatments be $T = [0, \infty)$. Let $Y = [0, 1]$ be the logical domain of the outcomes. A researcher poses a homogeneous linear-response model $y_j(t) = \beta t + \epsilon_j$. The researcher observes the realized distribution $P(y, z, v)$, where v is a covariate. He makes the covariance assumptions of IPD Section 7.7, applies “the” instrumental variable estimator, and reports that $\beta = \text{Cov}(v, y) / \text{Cov}(v, z)$.

A critic of this research comments as follows: “If the reported value of β is non-zero, some part of the researcher’s maintained assumptions must be wrong.” Is the critic correct or not? Explain.

Midterm Examination: Thursday November 1
(40 points total)

Let the grading of a pass-fail course be based on the student's total score on a set of examinations and problem sets. The range of feasible total scores is the real interval $[0, 100]$.

Let y = a student's total score. Let $w = 1$ if the student passes the course and $w = 0$ if he fails. Let $P(y, w)$ be the distribution of (y, w) in the population of students.

When answering all of the questions below, you know that grades are assigned as follows: $w = 0$ if $y \leq 60$ and $w = 1$ if $y > 60$. You also know that the fraction of the class who pass is $P(w = 1) = 0.8$.

1. (5 points) Given only the above information, what can you infer about $E(y)$?
2. (5 points) Given only the above information, what can you infer about $M(y)$?
3. (10 points) Suppose that $P(y)$ is known to be the uniform distribution on the interval $[\alpha, \beta]$, for some values of α and β . What is the joint identification region for the pair of parameters α and β ?
4. (10 points) Suppose that the student population is surveyed and asked to report their scores. Let $z = 1$ if a student reports his score on the survey and $z = 0$ otherwise. Assume that all reported scores are accurate. Suppose the survey reveals that $P(z = 1) = 0.5$ and that $P(y|z = 1)$ is the uniform distribution on the interval $[50, 90]$. What can you infer about $E(y)$?
5. (10 points) Let x denote gender, with $x = 1$ if male and $x = 0$ if female. You are told that the gender composition of the class is as follows: $P(x = 1) = 0.7$ and $P(x = 0) = 0.3$. What can you infer about the fractions of males and females who pass the course; that is, about $P(w = 1 | x = 1)$ and $P(w = 1 | x = 0)$?

End of Term Examination: Thursday December 6
(40 points total)

1. There are two treatments for patients diagnosed with a disease, $t = a$ and $t = b$. The patients in a study population have been treated, with $z = a$ for half of the patients and $z = b$ for the remaining half. A physician obtains data on these treatment decisions and observes partial data on the number of years, denoted y , that each patient lives after treatment. Here are the available data on the distribution of different values of y :

Years of Life after Treatment	Treatment		
	($z = a$)	($z = b$)	
$y = 0$	0.10	0.12	0.22
$y = 1$	0.25	0.30	0.55
$y = 2$ to 10	0.15	0.08	0.23
	0.50	0.50	1

When answering the questions below, assume that 10 years is the maximum length of life for patients diagnosed with the disease, under both treatments.

A. (10 points) Given the available data, what can the physician deduce about the average treatment effect $E[y(b)] - E[y(a)]$?

Suppose that the physician must treat a new population of patients with the same distribution of treatment response. The physician's objective is to maximize mean years of life after treatment.

B. (4 points) What is the maximin treatment rule?

C. (8 points) What is the minimax-regret treatment rule?

D. (8 points) Suppose that the physician declares himself to be a Bayesian and chooses to assign all new patients to treatment b . What can you conclude about his subjective beliefs regarding relevant unobserved quantities? Be specific.

2. (10 points) In their April 2006 NBER working paper "What Does Certification Tell Us About Teacher Effectiveness: Evidence from New York City?" T. Kane, J. Rockoff, and D. Staiger wrote this about their methodology (page 17):

"To generate estimates of teachers' effectiveness in raising student achievement, we estimate the following regression with student-level data:

$$A_{it} = \beta_g X_{it} + \gamma_g Y_{it} + \zeta_g Z_{it} + \delta W_{it} + \pi_{gt} + \varepsilon_{it}$$

where A_{it} represents the math or reading test score of student i in year t , X_{it} represents student characteristics, Y and Z are the mean characteristics of the students in student i 's classroom and school respectively in year t , W_{it} represents characteristics of the teacher to which the student is assigned in year t , and π is a fixed effect for the grade in which student i is enrolled and the year t in which we observe him/her."

The authors interpret parameter δ as measuring the effect of teacher characteristics on student outcomes. What assumptions about treatment selection and response must they make to justify this interpretation? Be specific.

Midterm Examination: Thursday October 30
(40 points total)

1. A bank must decide when to approve loan requests from a population of loan applicants. There are two possibilities, being to reject an application ($z = 0$) or approve it ($z = 1$). Assume that all loans have size 1 and the real interest rate is 0.05. When the bank rejects an application, it receives profit 0 for sure. When it approves a loan, it receives profit $y = 0.05$ if the borrower repays the loan and $y = -1$ if the borrower does not repay. The bank observes this distribution for the lending decisions made last year:

$$P(z = 1) = 0.8 \quad P(y = -1 | z = 1) = 0.02 \quad P(y = 0.05 | z = 1) = 0.98.$$

Suppose that the bank must make lending decisions this year to a new population of applicants. Assume that the marginal profit distribution $P(y)$ is the same as it was last year. A bank officer suggests eliminating the loan approval process and instead approving all loans.

a. (5 points) Using the empirical evidence alone, what can the bank conclude about the mean profit per loan it would earn if it were to approve all loans?

b. (5 points) Using the empirical evidence alone, what can the bank conclude about the median profit per loan it would earn if it were to approve all loans?

c. (5 points) Suppose the bank knows that $P(y|z = 1)$ stochastically dominates $P(y|z = 0)$. Now what can it conclude about the mean profit it would earn if it were to approve all loans.

d. (5 points) Continue to assume that $P(y|z = 1)$ stochastically dominates $P(y|z = 0)$. It was assumed throughout the questions above that the marginal profit distribution in the new population of loan applicants is $P(y)$. Someone objects that this assumption is not credible, but states that it would be reasonable to assume that the new profit distribution is stochastically dominated by $P(y)$. Given this assumption, what can the bank conclude about mean profit if it were to approve all loans?

2. Consider the population of home owners in Cook County who obtain mortgage loans. Each borrower either repays the loan in full or defaults, in which case he repays nothing. Let $z = 1$ if a borrower repays the loan and $z = 0$ if he defaults. Each borrower has a binary credit rating x , with $x = 1$ indicating a “prime” rating and $x = 0$ a “sub-prime” rating. Each borrower also has a binary indicator w of home location, with $w = 1$ if the home is in the city of Chicago and $w = 0$ if it is in the remainder of Cook County. Let $P(z, x, w)$ be the population distribution of (z, x, w) .

a. (8 points) A researcher observes that the overall repayment rate is $P(z = 1) = 0.8$ and the overall fraction of prime borrowers is $P(x = 1) = 0.9$. He maintains this assumption:

$$A1: P(z = 1 | x = 1) \geq P(z = 1 | x = 0).$$

Given the empirical evidence and assumption A1, what can the researcher conclude about $P(z = 1 | x = 1)$ and $P(z = 1 | x = 0)$?

b. (12 points) A researcher observes that

$$\begin{array}{ll} P(z = 1 | w = 1) = 0.70 & P(z = 1 | w = 0) = 0.90. \\ P(x = 1 | w = 1) = 0.85 & P(x = 1 | w = 0) = 0.95. \end{array}$$

He maintains this assumption:

$$A2: z \text{ is statistically independent of } w, \text{ conditional on } x.$$

Given the evidence and assumption A2, what can the researcher conclude about $P(z = 1 | w = 1, x = 1)$ and $P(z = 1 | w = 1, x = 0)$?

1. A bank must decide when to approve loan proposals from a population of potential borrowers. There are two possibilities, being to reject an application ($t = a$) or approve it ($t = b$). Assume that all loans have size 1 and the real interest rate is 0.04. When the bank rejects an application, it receives profit $y(a) = 0$ for sure. When it approves a loan, it receives profit $y(b) = 0.04$ if the borrower repays the loan and $y(b) = -1$ if the borrower does not repay.

Suppose that the bank observes the frequency distribution of the profit outcomes of the lending decisions made last year. However, the bank does not know why some applications were approved and others rejected. In the table below, $z = a$ if an application was rejected and $z = b$ if it was approved.

Profit Outcome	Loan Decision		
	($z = a$)	($z = b$)	
$y = -1$	0	0.01	0.01
$y = 0$	0.50	0	0.50
$y = 0.04$	0	0.49	0.49
	0.50	0.50	1

Suppose that the bank must make lending decisions this year to a new population of borrowers whose composition is the same as last year. The bank has no covariate information to enable it to systematically differentiate among potential borrowers this year. The bank's objective is to maximize profit.

A. (4 points) What is the maximin lending rule?

B. (6 points) What is the minimax-regret lending rule?

C. (6 points) Suppose that the bank management is risk-neutral Bayesian and chooses to approve loans to all potential borrowers. What can you conclude about management's subjective beliefs regarding relevant unobserved quantities? Be specific.

D. (4 points) Suppose the bank learns that applications last year were approved randomly. Now what is the minimax-regret lending rule?

2. This question concerns study effort and performance on an examination. Suppose that a population of five students ($j = 1, 2, 3, 4, 5$) took the exam, which was graded on a 0–10 scale. Prior to the exam, each student could study 0, 1, or 2 days. Let $y_j(t)$ denote the score that student j would achieve if he or she were to study t days. Let $z_j \in \{0, 1, 2\}$ be the number of days that student j actually studied and let $y_j \equiv y_j(z_j)$ be the score that this student actually achieved. Let it be common knowledge that $y_j(0) = 0$ for all $j \in J$; thus, studying is a necessary condition to achieve a positive score. This information about actual study effort and exam performance is available:

$$(z_1, y_1) = (1, 0); \quad (z_2, y_2) = (1, 3); \quad (z_3, y_3) = (1, 6); \quad (z_4, y_4) = (2, 6); \quad (z_5, y_5) = (2, 8).$$

A. (10 points) A researcher assumes that studying enhances exam performance, but with decreasing marginal returns. Formally, the researcher assumes that the response function $y_j(\cdot)$ is weakly increasing and concave for each $j \in J$. What can he infer about the mean and median returns to studying a second day, defined as $E[y(2) - y(1)]$ and $M[y(2) - y(1)]$?

B. (10 points) Another researcher assumes that studying affects exam performance linearly. Formally, this researcher assumes that $y_j(t) = \beta_j t$, where the parameter β_j can vary across students. What can he infer about $E[y(2) - y(1)]$ and $M[y(2) - y(1)]$?

Midterm Examination: Thursday October 29 (40 points total)

Various surveys have posed probabilistic expectations questions to large random samples of populations. An example is this job-loss question in the Survey of Economic Expectations: "I would like you to think about your employment prospects over the next 12 months. What do you think is the percent chance that you will lose your job during the next 12 months?"

A response can take any percent-chance value in the range $[0, 100]$. A persistent empirical finding has been that a larger fraction of respondents state exactly "50 percent" than give other responses. This finding has led to much speculation about the appropriate interpretation of the data when a person states "50 percent."

Assume that all members of a population respond to a question of the above format, asking about some future event. Let r denote a person's response and let x be a real covariate with support X . Assume that all persons report (r, x) , so the population distribution $P(r, x)$ is known.

Let y denote the percent-chance that a person actually places on occurrence of the event under study. In general, the values of y and r may differ. However, assume that reported and actual subjective probabilities always coincide when a person reports a value other than 50; that is, $r \neq 50 \Rightarrow y = r$.

Suppose that the objective is to learn $E(y|x)$ at a specified covariate value on the support X . Answer the questions below. Be as explicit as possible.

1. (4 points) A researcher conjectures that persons who respond $r = 50$ have actual beliefs y in the interval $[25, 75]$. Given this assumption and the empirical evidence, what can the researcher conclude about $E(y|x)$?
2. (4 points) A researcher conjectures that, among persons with covariates x , a known fraction $q(x)$ of those who state $r = 50$ report their actual beliefs and the remaining $1 - q(x)$ do not. Given this assumption and the empirical evidence, what can the researcher conclude about $E(y|x)$?
3. (8 points) A researcher conjectures that $E(y|x)$ is a weakly increasing function of x . Given this assumption and the empirical evidence, what can the researcher conclude about $E(y|x)$? Characterize the identifying power of the assumption, relative to the case with no assumptions.
4. (8 points) A researcher conjectures that $E(y|x)$ is a weakly increasing, linear function of x . Given this assumption and the empirical evidence, what can the researcher conclude about $E(y|x)$?
5. (8 points) Which of the assumptions posed in questions 1 through 4 are refutable, and which are non-refutable? Explain.
6. (8 points) Suppose that you know $E(r|x)$ and $P(r = 50|x)$, but you do not know other features of $P(r, x)$. Given only this empirical evidence and making none of the assumptions posed in questions 1 through 4, what can you conclude about $E(y|x)$?

Final Examination: Thursday December 3 (40 points total)

Research on treatment response has mainly assumed that a person's outcome varies only with his own treatment, not with those of other members of the population. However, personal outcomes often vary with the treatment of others. Vaccination is a leading example. Vaccination of an individual may not only protect this person from illness but may reduce the rate of illness among unvaccinated persons.

Suppose that a planner must choose whom to vaccinate against the H1N1 virus in a large population of observationally identical persons. Assume that vaccination always prevents a vaccinated person from becoming ill. Let $p(t) \in [0, 1]$ be the fraction of unvaccinated persons who become ill when the vaccination rate is t . Then the fraction of the population who become ill is $p(t)(1 - t)$.

Suppose the planner wants to minimize a social cost function with two additive components. These are the harm caused by illness and the cost of vaccination. Let a denote the monetized mean harm caused by illness and let c denote the mean cost per vaccination. The social cost of vaccination rate t is

$$S(t) = ap(t)(1 - t) + ct.$$

The first term on the right-hand side gives the aggregate cost of illness, and the second gives the aggregate cost of vaccination.

Suppose that the planner wants to solve the problem $\min_{t \in [0, 1]} S(t)$. The planner knows the values of (a, c) , with $0 < c < a$. However, he has only partial knowledge of the function $p(\cdot)$.

1. (8 points) Suppose the planner knows that $p(\cdot)$ is one of two functions. One is $p(t) = 1$, all $t \in [0, 1]$. The other is $p(t) = 1 - t$. Using this information, what vaccination rates are weakly dominated? Be explicit.

2. Maintaining the same knowledge as in question 1,

A. (6 points) what vaccination rate minimizes subjective expected social cost when the planner places subjective probability 0.5 on each feasible function $p(\cdot)$? Be explicit.

B. (6 points) what is the minimax vaccination rate? Be explicit.

3. Suppose the planner knows that $p(\cdot)$ has the form $p(t) = \rho$, all $t \in [0, 1]$, he but does not know the value of the constant ρ . Using this information,

A. (6 points) what vaccination rates are weakly dominated? Be explicit.

B. (6 points) what is the minimax-regret vaccination rate? Be explicit.

4. (8 points) Suppose the planner observes the realized vaccination rate and illness rate in a study population. Let r denote the observed vaccination rate and let q denote the realized illness rate within the group who are unvaccinated. The planner makes these assumptions:

Assumption 1: The study and the treatment populations have the same function $p(\cdot)$.

Assumption 2: $p(t)$ is weakly decreasing in t .

Using this information, what can the planner conclude about social cost $S(t)$, for each $t \in [0, 1]$? Be explicit.

Midterm Examination: Thursday October 28 (40 points total)

1. Suppose that a product is offered at price p per unit. Suppose that each person j in population J chooses between purchase of zero or one unit of the product. Let

x_j be a binary covariate for person j , taking the value zero or one.

e_j be the resource endowment available to person j . It is the case that $e_j > 0$.

v_j be the valuation person j places on the product. That is, v_j is the highest price the person would be willing to pay for one unit. Suppose that $0 \leq v_j \leq e_j$.

$z_j = 1$ if person j chooses to purchase one unit at price p , and $z_j = 0$ if j does not purchase the product.

Let $P(v, x, e, z)$ denote the population distribution of the variables (v, x, e, z) . Suppose that one observes (x_j, e_j, z_j) for each person $j \in J$. However, one does not observe v_j .

A. (6 points) What can one conclude about $E(v)$?

B. (7 points) What can one conclude about $M(v)$?

C. (7 points) Suppose a researcher assumes that $E(v|x=1) = E(v|x=0)$. What can this researcher conclude about $E(v)$?

D. (6 points) Consider the group of persons with endowment level k . The observed fraction of this group who purchase the product is $P(z=1|e=k)$. Consider a hypothetical situation in which each person in this group has endowment $m > k$ rather than k . If asked to predict the purchase behavior of the group in the hypothetical case, how would you reply? Explain your reasoning.

2. This fall the Northwestern Wildcats football team will play an away game against the University of Wisconsin. Consider the population of NU undergraduate students. Let $P(y, x)$ be the population distribution of (y, x) . For each member of this population, let

x = student year: 1 = freshman, 2 = sophomore, 3 = junior, 4 = senior.

y = subjective probability the student places on Northwestern winning the game; thus, $y \in [0, 1]$.

Suppose that the Daily Northwestern surveys the entire population, asking each student to predict who will win the game. Rather than ask students for their subjective probabilities y , the survey asks each student to provide a point prediction, with $w = 1$ for NU and $w = 0$ for Wisconsin. Assume that all students answer the question, giving the response $w = 1$ if $y > 1/2$ and $w = 0$ if $y < 1/2$. Suppose the survey result is $P(w=1) = 1/5$.

A. (7 points) Suppose that all students bet on the winner of the game, as follows. A dollar bet on NU pays 10 dollars if NU wins and 0 if it loses. A dollar bet on Wisconsin pays 2 dollars if Wisconsin wins and 0 if it loses. Suppose that each student maximizes his or her expected payoff on a dollar bet. Knowing that $P(w=1) = 1/5$, what can you conclude about the fraction of students who bet on NU?

B. (7 points) Conditioning the survey results by student year, suppose that

$$P(w=1|x=1) = 1/2, \quad P(w=1|x=2) = P(w=1|x=3) = P(w=1|x=4) = 1/10.$$

A researcher assumes that, within the group of students in year x , the subjective probability y has a uniform distribution on the interval $[0, a + bx]$, where (a, b) are parameters. Knowing the values of $P(w=1|x)$, what can you conclude about the researcher's assumption? Explain your reasoning.

Final Examination: Thursday December 2 (40 points total)

1. Let J be a population, each member j of which receives one of the two treatments a or b . Let $y_j(t)$ be a bounded outcome under treatment t , with $y_j(t)$ taking values in the unit interval $[0, 1]$. Let $P(y, z)$ be the observed population distribution of realized outcomes and treatments.

Suppose it is found that $y \perp z$; that is, y is statistically independent of z . In the absence of other information, some researchers make assertions about the identification region $H\{E[y(b)] - E[y(a)]\}$ for the average treatment effect. Evaluate these assertions:

Assertion A (5 points): $y \perp z \Rightarrow H\{E[y(b)] - E[y(a)]\} = \{0\}$.

Assertion B (5 points): $y \perp z \Rightarrow 0 \in H\{E[y(b)] - E[y(a)]\}$.

Assertion C (5 points): $y \perp z \Rightarrow 0$ is the center of $H\{E[y(b)] - E[y(a)]\}$.

In each case, explain your answer.

2. Let J be a population, each member j of which may receive treatment a or b . Let $y(t)$ be a binary outcome under treatment t . Thus, $y_j(t)$ takes the value 0 or 1. Let $P(y, z)$ be the observed population distribution of realized outcomes and treatments. Suppose it is found that

$$P(y = 1 | z = a) = P(y = 1 | z = b) = P(z = a) = P(z = b) = 1/2.$$

A. (5 point) In the absence of other information, what can you conclude about the median treatment effect $M[y(b)] - M[y(a)]$? Be specific.

B. (5 points) Consider the assumption of monotone treatment response, $y_j(b) \geq y_j(a)$, all $j \in J$. Using this assumption, what can you conclude about the average treatment effect $E[y(b) - y(a)]$? Be specific.

3. (15 points) Let J be a population, each member j of which may receive a real-valued treatment t , taking values in the unit interval $[0, 1]$. Let $y_j(t)$ be a bounded outcome under treatment t , with $y_j(t)$ taking values in the unit interval $[0, 1]$. Let F denote the space of concave functions mapping $[0, 1] \rightarrow [0, 1]$. Let it be known that $y_j(t) = f(t)$ for some $f(\cdot) \in F$. Thus, it is known that all persons have the same concave treatment response function, but the specific form of this function is not known.

A planner can assign each agent any feasible action. Thus, the planner can choose any element of the Cartesian Product set $[0, 1]^{|J|}$. Let $(w_j, j \in J)$ be any treatment allocation. Suppose that the planner wants to choose an allocation to maximize the mean outcome. Thus, the planner wants to choose $(w_j, j \in J)$ to maximize $\int f(w_j) dP(j)$. Suppose that the planner considers two treatment allocations, being

Allocation I: a fractional allocation assigning $1/3$ of the members of the population to each of the three treatment values $w = 0$, $w = 1/2$, and $w = 1$.

Allocation II: the singleton allocation assigning all members of the population to treatment value $w = 1/2$.

What, if anything, can you say about which of these two allocations the planner should prefer? Explain.

Midterm Examination: Thursday October 27 (40 points total)

1. (10 points) Consider the population of economics Ph.D. students at Northwestern. For each student j , let $y_j = 1$ if j owns an iPhone and $y_j = 0$ otherwise. Let x_j be the age of student j , in years. Suppose that the table below gives the available empirical evidence on the joint distribution of iPhone ownership and ages of the economics Ph.D. students:

	<u>Age in Years</u>					
	24	25	26	27	28	
<u>iPhone ownership</u>						
$y = 1$		0.05	0.08	0.10	0.06	0.01
$y = 0$		0.05	0.08	0.06	0.06	0.01
y unknown	0.10	0.04	0.04	0.08	0.18	

Assume that iPhone ownership increases with age; thus, $P(y = 1|x)$ increases with x . Combining this assumption and the empirical evidence, what can you conclude about $P(y = 1)$, the overall fraction of iPhone ownership?

2. Let $T = \{a, b\}$ be a set of two treatments administered to members of study population J . For each $j \in J$, let the potential outcomes $y_j(a)$ and $y_j(b)$ be binary variables, taking the value 0 or 1. Let z_j denote the treatment received by j and let y_j denote his realized outcome. Suppose that you observe that $P(y = 1) = 0.8$ and $P(z = b) = 0.6$. Given this empirical evidence alone, what can you conclude about

(A). (6 points) $P[y_j(b) = 1]$?

(B). (6 points) $P[y_j(a) = y_j(b)]$?

3. Let J be a large population of (husband, wife) couples. Within couple j , let j_1 denote the husband and j_2 the wife. Let treatment $t = 1$ denote vaccination of a person against a disease and let $t = 0$ denote the absence of vaccination. For each $(s, t) \in \{0, 1\} \times \{0, 1\}$, let $[y_{j_1}(s, t), y_{j_2}(s, t)]$ denote the bivariate health outcome that couple j would experience if the husband were to receive treatment s and the wife were to receive t . Outcomes are binary, with $y_{jk}(s, t) = 1$ if person k in household j remains healthy and $y_{jk}(s, t) = 0$ if this person becomes ill with the disease. Let (z_{j_1}, z_{j_2}) denote the treatments actually received by couple j and let $(y_{j_1}, y_{j_2}) \equiv [y_{j_1}(z_{j_1}, z_{j_2}), y_{j_2}(z_{j_1}, z_{j_2})]$ denote their actual outcomes. Suppose that the population distribution of treatments and outcomes is as follows:

$$P(z_1 = 1, z_2 = 1) = P(z_1 = 0, z_2 = 0) = 0 \quad P(z_1 = 1, z_2 = 0) = P(z_1 = 0, z_2 = 1) = \frac{1}{2}$$

$$\begin{aligned} P(y_1 = 0, y_2 = 0 | z_1 = 1, z_2 = 0) &= 0 & P(y_1 = 0, y_2 = 1 | z_1 = 1, z_2 = 0) &= 0 \\ P(y_1 = 1, y_2 = 0 | z_1 = 1, z_2 = 0) &= 0.1 & P(y_1 = 1, y_2 = 1 | z_1 = 1, z_2 = 0) &= 0.9 \end{aligned}$$

$$\begin{aligned} P(y_1 = 0, y_2 = 0 | z_1 = 0, z_2 = 1) &= 0 & P(y_1 = 0, y_2 = 1 | z_1 = 0, z_2 = 1) &= 0.4 \\ P(y_1 = 1, y_2 = 0 | z_1 = 0, z_2 = 1) &= 0 & P(y_1 = 1, y_2 = 1 | z_1 = 0, z_2 = 1) &= 0.6. \end{aligned}$$

The questions below concern prediction of the health outcomes that would occur if no one in the population were vaccinated. When answering the questions, assume throughout that the outcomes of each couple depend only on the treatments that this couple receives, not on the treatments received by other couples. Each question uses additional assumptions that apply only within the context of the question posed.

A. (6 points) Assume that couples were randomly assigned to treatments, with the allocation shown above. Given this assumption and the data shown above, what can you deduce about $P[y_1(0, 0) = 1, y_2(0, 0) = 1]$?

B. (12 points) Assume that treatment response is individualistic. That is, the outcome experienced by a person depends only on the treatment that this person receives, not on the treatments received by his or her spouse or any other member of the population. Also assume that husbands and wives have statistically independent treatment response. That is, $P[y_1(s, t), y_2(s, t)] = P[y_1(s, t)] \cdot P[y_2(s, t)]$ for all (s, t) . Given these assumptions and the data, what can you deduce about $P[y_1(0, 0) = 1, y_2(0, 0) = 1]$?

Final Examination: Thursday December 1 (40 points total)

1. Consider a planner who must allocate a large population of observationally identical persons to two treatments, labeled a and b. The feasible allocations to (a, b) are $(1 - \delta, \delta)$, where $\delta \in [0, 1]$. Outcomes are binary, equal to 0 or 1. The planner's objective is to maximize welfare

$$W(\delta) = (1 - \delta) \cdot P[y(a) = 1] + \delta \cdot P[y(b) = 1].$$

The planner observes the distribution of realized treatments and outcomes in three study populations where treatments were previously assigned, called J_1 , J_2 , and J_3 . These are

Study Population J_1 : $P_1(y = 1 | z = a) = 1.0$, $P_1(y = 1 | z = b) = 0.5$, $P_1(z = a) = 0.5$.

Study Population J_2 : $P_2(y = 1 | z = a) = 0.5$, $P_2(y = 1 | z = b) = 0.8$, $P_2(z = a) = 0.5$.

Study Population J_3 : $P_3(y = 1 | z = a) = 0.0$, $P_3(y = 1 | z = b) = 0.2$, $P_3(z = a) = 0.5$.

The planner assumes that the probabilities of potential outcomes are the same in all of these populations. That is, $P[y(t) = 1] = P_1[y(t) = 1] = P_2[y(t) = 1] = P_3[y(t) = 1]$ for $t \in \{a, b\}$. Given this information and no other assumptions or data,

A. (6 points) What can the planner conclude about $P[y(a) = 1]$ and $P[y(b) = 1]$?

B. (3 points) What is the minimax-regret choice of δ ?

C. (3 points) What is the maximin choice of δ ?

D. (3 points) What choice of δ maximizes subjective expected welfare if the planner places a uniform distribution on the identification region for $\{P[y(a) = 1], P[y(b) = 1]\}$?

2. Consider a large population of observationally identical persons, each of whom must choose for himself between treatments a and b. Each person must choose a specific treatment—he cannot choose a randomized strategy. Outcomes are binary, equal to 0 or 1. Each person knows that $P[y(a) = 1] = 0.4$ and $P[y(b) = 1] \in [0.3, 0.6]$. Each person wants to choose the treatment with the higher probability of success. That is, each person wants to solve the problem $\max_{t \in \{a, b\}} P[y(t) = 1]$. Suppose that the population is composed of three types of persons:

Type 1 chooses a treatment that minimizes maximum regret.

Type 2 chooses a maximin treatment.

Type 3 places a uniform subjective distribution on the identification region for $P[y(b) = 1]$ and chooses a treatment that maximizes subjective expected welfare.

A. (8 points) What treatments do Types 1, 2, and 3 choose?

B. (5 points) Let (π_1, π_2, π_3) be the fractions of each type in the population. Suppose that $\frac{1}{2}$ of the population chooses treatment a and $\frac{1}{2}$ chooses b. What can you conclude about the values of (π_1, π_2, π_3) ?

C. (5 points) Consider a particular person j. Suppose that he wants to solve the problem $\max_{t \in \{a, b\}} y_j(t)$ rather than the problem $\max_{t \in \{a, b\}} P[y(t) = 1]$ posed earlier. What is the result if he applies the minimax-regret criterion to the revised problem?

3. (7 points) Consider a planner who must allocate a large population of observationally identical persons to two treatments, labeled a and b. Treatment outcomes are binary, taking the value 0 or 1. Let $P[y(a) = 1] = \alpha$ and $P[y(b) = 1] = \beta$. Suppose that the planner knows α , where $0 < \alpha < 1$. He has no a priori knowledge of β . However, he observes the outcome of a classical randomized experiment in which a random sample of size N are drawn and assigned to b. Let n denote the number of subjects for whom the realized outcome is $y = 1$. Let $\delta(\cdot): [0, 1, 2, \dots, N] \rightarrow [0, 1]$ denote a statistical treatment rule specifying a choice of treatment allocation as a function of n . The planner evaluates a rule by its frequentist expected welfare

$$W(\delta, P, N) = \alpha + (\beta - \alpha) \cdot E[\delta(n)].$$

To help him select a rule, suppose that the planner asks a consultant for advice. The consultant recommends that the planner choose a rule with two properties:

Monotonicity: $\delta(\cdot)$ should be a weakly increasing function of n .

Diversification: $\delta(\cdot)$ should satisfy $0 < \delta(n) < 1$ for all values of n .

Is the consultant's advice reasonable? Explain.

Midterm Examination: Thursday November 8 (40 points total)

Let J denote a population of children born in a certain country. Suppose that the parents of each newborn child submit an application for children's health insurance to a monopolist health insurance agency. The feasible decisions t are *denial of insurance* ($t = A$) and *granting of insurance* ($t = B$) to the newborn child.

Prior to making a decision, the insurance agency may order a physical examination that yields one of two results. The exam result R may be a designation of *high* ($R = H$) or *low* ($R = L$) health. The status quo policy is to grant insurance only if the agency orders the exam and the result is a designation of high health. Let $z_j = 1$ if the agency orders the exam for child j and $z_j = 0$ otherwise. Then j is granted health insurance if and only if $z_j = 1$ and $R_j = H$.

Let the outcome of interest be whether a child lives to at least age 18. Assume that receipt of the physical examination has no direct effect on life span. However, receipt of health insurance may affect life span. For $t \in \{A, B\}$, let $y_j(t) = 1$ if child j lives to at least age 18 and $y_j(t) = 0$ if the child dies before age 18. Let y_j denote the realized outcome for child j . Given the status quo policy, $y_j = y_j(B)$ if ($z_j = 1, R_j = H$) and $y_j = y_j(A)$ otherwise.

You observe (y_j, z_j) for each $j \in J$. You observe R_j if and only if $z_j = 1$. Hence, you observe the population distributions $P(y, z)$ and $P(y, R | z = 1)$. Suppose that these observed distributions are as follows:

$$P(y = 0, z = 0) = 0.1$$

$$P(y = 1, z = 0) = 0.1$$

$$P(y = 0, z = 1) = 0.1$$

$$P(y = 1, z = 1) = 0.7.$$

$$P(y = 0, R = L | z = 1) = 0.1$$

$$P(y = 1, R = L | z = 1) = 0.025$$

$$P(y = 0, R = H | z = 1) = 0.025$$

$$P(y = 1, R = H | z = 1) = 0.85.$$

1. (10 points) A proposed change to the status quo policy is to give all children the physical examination, leaving in place the decision to grant insurance only when the exam has result H. Given the empirical evidence alone, what can you conclude about the fraction of children who would reach age 18 under the proposed policy.
2. (10 points) Assume that treatment response is monotone, with receipt of insurance never decreasing life span; that is, $y_j(B) \geq y_j(A)$ for all $j \in J$. Reconsider the proposal specified in question 1, bringing to bear this assumption.
3. (10 points) Another proposal is to drop the status quo policy entirely. Instead, all children would receive health insurance. Given the empirical evidence alone, what can you conclude about the fraction of children who would reach age 18 under this proposal?
4. (10 points) As in question 2, assume that treatment response is monotone. In addition, assume that the examination result is a monotone instrumental variable. Specifically, children who are tested and have result H have life prospects that are at least as good as those with result L; that is, $P[y(t) = 1 | z = 1, R = H] \geq P[y(t) = 1 | z = 1, R = L]$ for each value of t . Now reconsider the proposal specified in question 3, bringing to bear these assumptions.

Final Examination: Thursday December 6 (40 points total)

1. A firm must choose the magnitude of a scalar input x into a production process. The input can take any value in the interval $[0, 100]$. Output is $f(x)$. The firm pays a dollar per unit of x and receives a dollar per unit of output. Thus, profit is $f(x) - x$.

The firm's objective is to maximize profit. However, the firm has only partial knowledge of its production function $f(\cdot)$. It observes that an identical firm has set $x = 40$ and has achieved output $f(40) = 200$, yielding profit $200 - 40 = 160$. The firm also knows that $f(\cdot)$ is a weakly increasing and weakly concave function such that $f(0) \geq 0$. However, it does not know more about $f(\cdot)$.

A. (8 points) Which if any values of x are weakly dominated? Explain your answer.

B. (8 points) Suppose that the firm uses the maximin criterion to choose x . State the criterion and solve for the maximin choice.

C. (4 points) An economist suggests that the firm should maximize expected profit, using a uniform subjective distribution on the space of feasible production functions. A probability theorist responds that this may not be technically feasible. Comment as best you can on why the probability theorist is concerned.

2. The recent revelations regarding doping by Lance Armstrong and other members of his cycling team have brought renewed attention to the decisions of athletes to use banned performance-enhancing drugs and how these decisions may be affected by efforts to detect and punish doping.

Let J denote the population of athletes who participate in a specified sport. Let $t \in [0, \infty)$ measure the potential intensity of sanctions; that is, efforts to detect and punish doping in this sport. Assume that each athlete $j \in J$ makes a binary decision $y_j(t)$ to engage or not engage in doping, knowing the intensity of the prevailing sanctions. Specifically, $y_j(t) = 1$ if athlete j chooses to dope when the intensity of sanctions is t and $y_j(t) = 0$ otherwise. Let $v_j \in (-\infty, \infty)$ measure the net value that j places on doping, taking into account its performance-enhancing effects, health consequences, and the athlete's views about the ethics of doping. Assume that athlete j chooses $y_j(t) = 1$ if $v_j - t > 0$ and $y_j(t) = 0$ otherwise. Thus, he chooses to dope if the net value of doping exceeds the costs imposed by sanctions.

Suppose that athletes compete in separate American and National leagues, denoted $x = A$ and $x = N$ for short. Each league has the same number of athletes, so $P(x = A) = P(x = N) = 1/2$. Each league currently has its own status quo sanctions policy, with league A setting $t = 1$ and league N setting $t = 3$. Suppose that the U.S. Anti-Doping Agency would like to learn the fraction of athletes who would choose to dope if a new sanctions policy would set $t = 2$ in both leagues. Thus, the agency would like to learn the choice probability $P[y(2) = 1]$.

A. (5 points) Suppose that random drug testing enables you to learn that the fractions of athletes in each league who engage in doping under its status quo policy are $P[y(1) = 1 | x = A] = 0.6$ and $P[y(3) = 1 | x = N] = 0.2$ respectively. Given this empirical evidence and the behavioral assumptions stated above, what can you conclude about $P[y(2) = 1]$?

B. (5 points) Consider again the scenario of question 1. Add the assumption that athletes in leagues A and B have the same distribution of doping preferences. Thus, assume that $P(v | x = A) = P(v | x = N)$. Now what can you conclude about $P[y(2) = 1]$?

C. (5 points) Consider again the scenario of question 2. Add the assumption that the distribution of v in both

leagues is uniform on some interval $[\alpha, \beta]$, where $[\alpha, \beta]$ are a priori unknown parameters. Now what can you conclude about $P[y(2) = 1]$?

D. (5 points) Considering the behavioral model posed in the second paragraph of the introduction, a researcher conjectures that an athlete's evaluation of doping may depend on the strength of the competition he faces, which may depend on the doping decisions of other athletes. To express this idea, the researcher modifies the behavioral model. He assumes that athlete j in league X chooses $y_j(t) = 1$ if $v_j + \gamma \cdot P[y(t) = 1 | x] - t > 0$ and $y_j(t) = 0$ otherwise, where γ is an a priori unknown parameter. The researcher points out that if $\gamma \neq 0$, then an *equilibrium* value of $P[y(t) = 1 | x]$ solves a certain equation. What equation must $P[y(t) = 1 | x]$ solve to be an equilibrium? For what values of γ is it possible that there may exist multiple equilibria; that is, multiple solutions to the equation?

Midterm Examination: Thursday October 31 (40 points total)

1. Each spring, the Current Population Survey (CPS) of the U.S. Census Bureau asks a large random sample of American households to report their income for the previous year. However, many sampled households do not report their incomes. Let y denote household income, measured in thousands of dollars. Let $z = 1$ if a household would report its income if sampled and $z = 0$ if not. Suppose that you want to use the CPS to learn the fraction $P(y \leq 50)$ of American households with income less than or equal to \$50,000.

Ignoring finite-sample imprecision, suppose that the CPS data reveal that $P(y \leq 50|z = 1) = 0.5$ and $P(z = 1) = 0.6$. When answering each question below, combine this empirical evidence with the assumption and evidence stated in the question.

A. (5 points) Assume that all households that do not report their income either have low or high income. Specifically, $z = 0 \Rightarrow y \leq 10$ or $y \geq 100$. What can you conclude about $P(y \leq 50)$?

B. (6 points) Assume that the CPS randomly divides sampled households into two groups, $g = 1$ and $g = 2$. It attempts to interview households in group 1 during the month of March and those in group 2 during April. Suppose that the group-specific findings are

$$\begin{aligned} P(y \leq 50|g = 1, z = 1) &= 0.4, & P(z = 1|g = 1) &= 0.6. \\ P(y \leq 50|g = 2, z = 1) &= 0.6, & P(z = 1|g = 2) &= 0.6. \end{aligned}$$

What can you conclude about $P(y \leq 50)$?

C. (6 points) Assume that households with income above \$50,000 are less likely to report than ones with income less than or equal to \$50,000. That is, $P(z = 1|y > 50) < P(z = 1|y \leq 50)$. What can you conclude about $P(y \leq 50)$?

D. (5 points) Assume that all households with income up to \$100,000 report their income and that no households with income above \$100,000 do so. What can you conclude about $P(y \leq 50)$?

2. Let J denote a population of persons who are ill with a life-threatening disease. Let $t = 0$ and $t = 1$ denote two mutually exclusive and exhaustive treatments. Let $y_j(t) = 1$ if person j would survive with treatment t and $y_j(t) = 0$ if the person would die. Let z_j denote the treatment that person j actually receives and let $y_j = y_j(z_j)$ be his realized outcome. Let observation of realized treatments and outcomes in the population yield the following empirical evidence:

$$P(z = 1) = 0.5 \qquad P(y = 1|z = 0) = 0.5 \qquad P(y = 1|z = 1) = 0.5.$$

A. (6 points) Assume that treatment response follows this linear-threshold model:

$$y_j(t) = 1 \text{ if } \beta_j t + u_j > 0, \quad = 0 \text{ otherwise,}$$

where (β_j, u_j) are unobserved attributes of person j . Combining this assumption and the empirical evidence, what can you conclude about $P[y(1) = 1]$, the survival probability if everyone were to receive treatment 1?

B. (6 points) Continue to assume the linear-threshold model. Also assume that $\beta_j > 0$ for all persons j . Combining these assumptions and the evidence, what can you conclude about $P[y(1) = 1]$?

C. (6 points) Continue to assume the linear-threshold model. Also assume that physicians observe β_j and choose treatments as follows: $z_j = 1$ if $\beta_j > 0$ and $z_j = 0$ otherwise. Combining these assumptions and the evidence, what can you conclude about $P[y(1) = 1]$?

Modern societies legislate numerous requirements for regulatory approval of private activities. Suppose that each member j of a population considers whether to apply for approval of some activity. Let $v_j > 0$ be the value that person j associates with the activity and let $c_j > 0$ be the cost to j of submitting an application. Suppose that the regulatory agency chooses a $\delta \in [0, 1]$ and uses a randomizing device to make an independent decision on each submitted application, approving person j 's application with probability δ and rejecting it with probability $1 - \delta$.

Suppose that persons know their own values of (v, c) and the agency's chosen value of δ when deciding whether to apply for approval, but they do not know whether their particular applications will be approved. Let $z_j(\delta) = 1$ if j chooses to submit an application when the approval probability is δ , and $z_j(\delta) = 0$ if he chooses not to apply. Assume that j chooses to apply if the expected benefit exceeds the cost. Thus, $z_j(\delta) = 1[\delta v_j > c_j]$. Let $P(v, c)$ denote the distribution of (v, c) values across the population.

1. Consider the situation of the agency prior to choosing a value of δ . Suppose that the agency knows how the population makes application decisions. The agency knows that $P(v > c) > 0$ but has no other knowledge of $P(v, c)$. Suppose that the agency wants to choose δ to maximize an additive social welfare function. Let the social cost of an application by person j equal its private cost c_j . Let the social value of person j undertaking the activity be $v_j + \alpha$, where α is known to the agency. Thus, the activity has an externality of known magnitude α .

Given the above, the net contribution to welfare derived from person j is $z_j(\delta) \cdot [\delta(v_j + \alpha) - c_j]$ if the agency chooses approval rate δ . Welfare function $W(\delta)$ is the mean welfare across the population, namely

$$W(\delta) = E\{z(\delta) \cdot [\delta(v + \alpha) - c]\}.$$

- A. (10 points) Suppose that $\alpha \geq 0$. Show that there exists a unique dominant value of δ . What is it?
 B. (5 points) Suppose that $\alpha < 0$. What is the maximin value of δ ?
 C. (10 points) Suppose that $\alpha < 0$. Replace the assumption $(c_j > 0, j \in C)$ with the assumption $(c_j = 0, j \in C)$. Also assume $P(0 < v \leq K) = 1$, where K is a known real number. What is the minimax-regret value of δ ?

2. Suppose you know how the population makes application decisions but you do not know $P(v, c)$. Suppose you observe that the agency sets $\delta = 1/4$ and that the fraction of persons who apply is $P[z(1/4) = 1] = 1/2$. You are asked to predict the fraction of persons who would apply if the agency were to approve $3/4$ of the applications rather than $1/4$.

- A. (5 points) What prediction can you make using your knowledge and the empirical evidence?
 B. (5 points) In addition to the knowledge and evidence specified above, suppose you are told that the quantity v/c has a uniform distribution on an interval $[0, K]$ for some real K . You are not told the value of K . Now what prediction can you make?
 C. (5 points) Suppose you are told instead that v/c has a uniform distribution on an interval $[4 - L, 4 + L]$ for some real L . You are not told the value of L . Now what prediction can you make?

3. Reconsider the application decision of person j . Suppose that j knows (v_j, c_j) , where $v_j > c_j$. However, he does not know the agency's chosen value of δ when deciding whether to apply for approval. The person only knows that his utility outcome will be zero if he does not apply, $v_j - c_j$ if he applies and the application is approved, and $-c_j$ if he applies and the application is rejected. Suppose that j wants to maximize utility. Let $z_j = 1$ if j chooses to submit an application and $z_j = 0$ if he chooses not to apply.

- A. (5 points) What is j 's maximin application choice?
 B. (5 points) What is j 's minimax-regret application choice?

Midterm Examination: Tuesday November 4 (40 points total)

Consider the population of students who enrolled in an economics Ph.D. program in 2004. The program had two mutually exclusive and exhaustive fields of study: microeconomics ($t = A$) and macroeconomics ($t = B$). Each student had to decide on entry which field to study. Let $z_j = A$ if student j decided to study micro and $z_j = B$ if he studied macro. Assume that treatment response is individualistic. Let $y_j(t) \in [0, 1]$ measure the degree of success of the career outcome of student j if he or she should study field t . Let $y_j \equiv y_j(z_j)$ be the success of j in his chosen field.

It has been hypothesized that success in micro and macro is associated with the composition of a student's genome. The hypothesis asserts that there exists a *macro gene*, possessed by some students but not others, such that students who possess the macro gene are innately able to understand macroeconomic theory while students who do not possess the gene must struggle to understand it. Existence of the macro gene has not been verified, but analysis of a student's DNA enables one to measure whether a student has a binary genetic marker x , taking the value 0 or 1. It has been conjectured that the presence of this observable marker may be statistically associated with presence of the unobserved macro gene.

Suppose that, on entry to the program in 2004, each student was required to submit DNA to the DGS. The DGS used the DNA to measure the value of x for each student and kept the results secret. Ten years later, in 2014, the DGS measured the career success y of each student. Thus, in 2014, the DGS was able to learn the joint distribution $P(y, z, x)$ of student (career success, field of study, genetic marker). Some of the findings were as follows:

$$E(y|x = 0, z = A) = 1/2, \quad E(y|x = 1, z = A) = 1/2, \quad E(y|x = 0, z = B) = 2/5, \quad E(y|x = 1, z = B) = 3/5.$$

$$P(z = B|x = 0) = 1/4, \quad P(z = B|x = 1) = 3/4, \quad P(x = 1) = 1/2.$$

1. (10 points) Given this evidence, the DGS asks you to predict what mean career outcomes would have been under an alternative policy in which students did not choose their own fields of study. Instead, students with $x = 1$ would have been required to study macro and those with $x = 0$ would have been required to study micro. Let

$$\gamma \equiv E[y(A)|x = 0]P(x = 0) + E[y(B)|x = 1]P(x = 1)$$

denote the mean career outcome that would have occurred under the alternative policy. What can you conclude about the value of γ in the absence of assumptions restricting treatment response and selection?

2. (10 points) The DGS asks you to assume that x is informative about treatment response in the sense that $x_j = 0 \Rightarrow y_j(A) \geq y_j(B)$ and $x_j = 1 \Rightarrow y_j(B) \geq y_j(A)$ for all j . Now what can you conclude about γ ?

3. (10 points) The DGS asks you to assume that x is an instrumental variable in the sense that $E[y(t)|x = 0] = E[y(t)|x = 1]$ for $t = A$ and $t = B$. Now what can you conclude about the value of γ ?

4. (5 points) The DGS asks you to assume that student field choices in 2004 were statistically independent of $[y(\cdot), x]$. Now what can you conclude about the value of γ ?

5. (5 points) The DGS asks you to predict what mean career outcomes would have been under a different policy, in which students are told their values of x upon entry to the program and then choose their own fields of study with knowledge of x . Let δ denote the mean career outcome under this policy. What can you conclude about the value of δ in the absence of assumptions restricting treatment response and selection?

A researcher studies the voting decisions of a population J of citizens in a recent election. Let $y_j = 1$ if person j voted in the election and $y_j = 0$ if he did not vote.

The researcher assumes that person j chose to vote if and only if the utility of voting exceeded zero. He assumes that the utility of voting was determined by the value v_j that person j attached to voting and by the time z_j , measured in hours, that was required for the person to travel to the polling location and vote. In particular, he assumes that the voting utility of person j was

$$u_j(v_j, z_j) = v_j - z_j.$$

Thus, the researcher assumes that $y_j = 1$ if $v_j - z_j > 0$ and $y_j = 0$ if $v_j - z_j \leq 0$.

Let $P(y, z, v)$ denote the distribution of voting (decisions, times, values) in the population. Suppose that voting decisions y and times z are observable, but voting values v are unobservable. In particular, the researcher observes that, for $\zeta \in [0, 5]$,

$$P(z \leq \zeta) = \zeta/5, \quad P(y = 1|z = \zeta) = 9/(10 + 10\zeta).$$

1. The researcher combines the empirical evidence with the utility model to learn about the distribution $P(v|z)$ of voting value conditional on voting time.

A. (7 points) Let $M(v|z)$ denote the median of v conditional on z . What can the researcher conclude about $M(v|z = \zeta)$ for $\zeta \in [0, 5]$?

B. (7 points) Assume that v is statistically independent of z . Now what can the researcher conclude about $M(v|z = \zeta)$ for $\zeta \in [0, 5]$?

2. The researcher wants to predict the voting decisions that would have occurred under a hypothetical policy that makes the voting time of every person equal to t hours for a specified $t \in [0, 5]$. Assume that the utility model would continue to hold under this policy. Thus, the voting utility function of person j is

$$u_j(v_j, t) = v_j - t$$

and the voting decision with voting time t is $y_j(t) = 1$ if $v_j - t > 0$ and $y_j(t) = 0$ if $v_j - t \leq 0$.

A. (7 points) For $\zeta \in [0, 5]$, let $P[y(t) = 1|z = \zeta]$ denote the fraction of persons with realized voting time ζ who would have voted if they had faced voting time t . Combining the empirical evidence with the utility model, what can the researcher conclude about $P[y(t) = 1|z = \zeta]$?

B. (7 points) Assume that v is statistically independent of z . Now what can the researcher conclude about $P[y(t) = 1|z = \zeta]$?

3. (8 points) Return to the empirical evidence. Suppose that $P(y = 1|z = \zeta) = \zeta/(\zeta + 1)$ rather than the evidence stated earlier. Thus, the empirical probability of voting increases with voting time rather than decreases. A critic of the utility model $u_j(v_j, t) = v_j - t$ asserts that the empirical evidence refutes the model, which assumes that the utility of voting decreases with voting time. Does the empirical evidence refute the model? Explain.

4. As above let the voting utility function of person j be $u_j(v_j, t) = v_j - t$. However, suppose that j must decide whether to vote without knowing the time t required to travel to the polling place and vote. He only knows that t will take some value in the interval $[0, 5]$.

A. (7 points) What is j 's voting decision if he uses the maximin criterion?

B. (7 points) What is j 's voting decision if he uses the minimax-regret criterion?

1. An online firm considers choice between a status quo strategy and an innovation for marketing subscription products to a population J consisting of 1000 persons. These are

Strategy A: The firm currently markets subscription product A to the population, selling it at price $p_A > 0$ per year. The product has constant marginal cost per subscriber, $c_A > 0$, where $p_A > c_A$. For each $j \in J$, let $q_{Aj}(A) = 1$ if person j purchases a subscription and $q_{Aj}(A) = 0$ otherwise. The firm observes the number of persons who currently purchase a subscription, namely $Q_A(A) \equiv \sum_{j \in J} q_{Aj}(A)$. The firm's profit is $(p_A - c_A)Q_A(A)$.

Strategy AB: The firm contemplates introduction of a second subscription product B , whose marginal cost per subscriber is $c_B > 0$, and to sell it at price $p_B > c_B$. For each $j \in J$, let $[q_{Aj}(AB), q_{Bj}(AB)]$ denote the subscription choice of person j for the two products if both were to be available for purchase. Thus, $[q_{Aj}(AB), q_{Bj}(AB)]$ takes one of the four values $(0, 0)$, $(1, 0)$, $(0, 1)$, $(1, 1)$. Let $Q_A(AB) \equiv \sum_{j \in J} q_{Aj}(AB)$ and $Q_B(AB) \equiv \sum_{j \in J} q_{Bj}(AB)$ be the numbers of subscriptions that would be sold. This would yield profit $(p_A - c_A)Q_A(AB) + (p_B - c_B)Q_B(AB)$.

The firm observes the realized product demand $Q_A(A)$ under strategy A and hence knows its realized profit. Suppose that $Q_A(A) = 500$.

To provide empirical evidence on profit under strategy AB , the firm sends an email to all members of the sub-population who currently purchase product A and informs them of the availability of product B . That is, the firm emails all persons for whom $q_{Aj}(A) = 1$. Assume that all of them read the email. Following the email notification, the firm observes that 200 of these persons purchase only product A , 200 purchase only product B , and 100 purchase both A and B .

The firm cannot send the email to the 500 persons with $q_{Aj}(A) = 0$, because it does not know their email addresses. These persons remain unaware that B is available. They continue to not purchase A .

When answering the questions below, maintain the following assumptions:

- (i) The firm has the empirical evidence described above.
- (ii) If the firm were to fully implement a new marketing strategy, it would advertise widely and all members of the population would be aware of the new strategy.
- (iii) Treatment response is individualistic.

A. (5 points) In the absence of further information, what can the firm conclude about the profit it would earn if it were to fully implement strategy AB ?

B. (5 points) Suppose the firm assumes that persons who do not purchase product A under Strategy A would not purchase product A under full implementation of Strategy AB . Now what can the firm conclude about profit if it were to fully implement Strategy AB ?

C. (5 points) Suppose the firm considers an alternative marketing strategy, being

Strategy B: Discontinue sale of product A and market only product B .

Let $q_{Bj}(B)$ be the purchase decision that person j would make if only product B were available and $Q_B(B) \equiv \sum_{j \in J} q_{Bj}(B)$. Given the maintained assumptions and no further information, what can the firm conclude about profit under Strategy B ?

D. (5 points) Suppose the firm assumes that persons who would purchase product B under Strategy AB would also purchase it under Strategy B . Now what can the firm conclude about profit under strategy B ?

E. (5 points) Suppose that the firm considers yet another marketing strategy, being

Strategy π_A : Continue to sell only product A, but change its price from p_A to π_A , where $p_A > \pi_A > 0$. Suppose the firm assumes that every person who purchases product A under Strategy A would do so under Strategy π_A . What can the firm conclude about profit under Strategy π_A ?

2. Let (y, x, δ) be a trivariate random variable. It is known that $y = x^\delta$, δ is statistically independent of x , and δ is distributed Bernoulli with parameter p , where $0 < p < 1$. x has logical domain $(0, \infty)$.

A. (5 points) What is the best predictor of y given x under square loss, considered as a function of x ?

B. (5 points) What is the best predictor of y given x under absolute loss, considered as a function of x ?

C. (5 points) What is the best predictor of x given y under square loss, evaluated at $y = 2$?

Final Examination: Thursday December 3 (50 points total)

A planner must assign one of two treatments $T = \{a, b\}$ to each member of population J . Each $j \in J$ has a response function $u_j(\cdot): T \rightarrow \mathbb{R}$ mapping treatments into welfare outcomes. The population is a probability space (J, Ω, P) . The population is "large;" formally J is uncountable and $P(j) = 0, j \in J$.

Members of the population are observationally identical to the planner. A statistical treatment rule (STR) maps sample data into a treatment allocation. Let Q denote the sampling distribution generating the available data and let Ψ denote the sample space. An STR is a function $\delta(\cdot): \Psi \rightarrow [0, 1]$ that allocates fraction $\delta(\psi)$ of J to treatment b and $1 - \delta(\psi)$ to treatment a when the data are $\psi \in \Psi$.

The planner wants to maximize additive population welfare. Let S denote the state space. Thus, $\{(P_s, Q_s), s \in S\}$ is the set of (P, Q) pairs that the planner deems possible. Given data ψ , the population welfare that would be realized in state s if the planner were to choose rule δ is

$$U(\delta, P_s, \psi) \equiv E_s[u(a)] \cdot [1 - \delta(\psi)] + E_s[u(b)] \cdot \delta(\psi) \equiv \mu_{sa} \cdot [1 - \delta(\psi)] + \mu_{sb} \cdot \delta(\psi),$$

where $\mu_{sa} \equiv E_s[u(a)] \equiv \int_J u_j(a) dP_s(j)$ and $\mu_{sb} \equiv E_s[u(b)] \equiv \int_J u_j(b) dP_s(j)$ are assumed to be finite. Assume that S contains at least one state such that $\mu_{sa} > \mu_{sb}$ and another such that $\mu_{sa} < \mu_{sb}$.

The planner measures the performance of rule δ in state s by its expected welfare

$$W(\delta, P_s, Q_s) = \mu_{sa} \cdot \{1 - E_s[\delta(\psi)]\} + \mu_{sb} \cdot E_s[\delta(\psi)],$$

where $E_s[\delta(\psi)] \equiv \int_\Psi \delta(\psi) dQ_s(\psi)$ is the mean (across samples) fraction of persons assigned to treatment b . Define rule δ to be ε -optimal for a specified $\varepsilon > 0$ if $W(\delta, P_s, Q_s) \geq \max(\mu_{sa}, \mu_{sb}) - \varepsilon$ for all $s \in S$.

1. (10 points) Suppose that rule δ has been determined to be ε -optimal. What can you conclude about its maximum regret?

2. (10 points) It has been common in medical applications to employ STRs that use the outcome of a hypothesis test to choose between two treatments. Construction of a test rule begins by partitioning the state space into disjoint subsets S_a and S_b , where S_a contains all states in which treatment a is optimal and S_b contains all states in which b is optimal. Thus, $\mu_{sa} > \mu_{sb} \Rightarrow s \in S_a$, $\mu_{sa} < \mu_{sb} \Rightarrow s \in S_b$, and the states with $\mu_{sa} = \mu_{sb}$ are somehow split between the two sets. Let s^* denote the unknown true state. The two hypotheses are $[s^* \in S_a]$ and $[s^* \in S_b]$.

A test rule δ partitions Ψ into disjoint acceptance regions $\Psi_{\delta a}$ and $\Psi_{\delta b}$. When the data ψ lie in $\Psi_{\delta a}$, the rule accepts hypothesis $[s^* \in S_a]$ by setting $\delta(\psi) = 0$. When ψ lies in $\Psi_{\delta b}$, the rule accepts $[s^* \in S_b]$ by setting $\delta(\psi) = 1$. (We use the word "accepts" rather than the traditional term "does not reject" because treatment choice is an affirmative action.) Let $R_s(\delta)$ be the state-dependent probability that δ yields an error, choosing the inferior treatment over the superior one. That is,

$$\begin{aligned} R_s(\delta) &= Q_s[\delta(\psi) = 0] \quad \text{if } \mu_{sa} < \mu_{sb}, \\ &= Q_s[\delta(\psi) = 1] \quad \text{if } \mu_{sa} > \mu_{sb}, \\ &= 0 \quad \text{if } \mu_{sa} = \mu_{sb}. \end{aligned}$$

Let $\varepsilon > 0$. What conditions on $[\mu_{sa}, \mu_{sb}, R_s(\delta)], s \in S$ are necessary and sufficient for rule δ to be ε -optimal? Be specific.

3. (10 points) The classical theory of hypothesis testing views the two hypotheses asymmetrically, calling $[s^* \in S_a]$ the null hypothesis and $[s^* \in S_b]$ the alternative hypothesis. Choice of treatment b when the null is correct is called a Type I error and choice of treatment a when the alternative is correct is called a Type II error. A classical test rule satisfies two criteria: (i) the maximum probability of a Type I error across all states in S_a equals a specified value $\alpha \in (0, 1)$ and (ii) the probability of a Type II error equals a specified value β

$\in (0, 1)$ at a specified state in S_b . Consider the special case of a simple hypothesis, in which each of sets S_a and S_b contains one element. For what values of ϵ is a test with error probabilities (α, β) ϵ -optimal?

4. (10 points) Let welfare outcomes be binary, taking the value zero or one. Let a be a status quo treatment with known mean outcome μ_a^* and let b be an innovation with unknown mean outcome. Thus, $(\mu_{sa}, s \in S) = \mu_a^*$ and $(\mu_{sb}, s \in S) = [0, 1]$. Let the sampling process draw a positive finite number N persons at random from J and assign treatment b to each sampled person. Let the data ψ be n , the number of sample members whose binary outcome equals one; thus, n is an integer in the set $(0, 1, 2, \dots, N - 1, N)$. Consider the empirical success rule in which $\delta(n) = 1$ if $n/N > \mu_a^*$ and $\delta(n) = 0$ if $n/N \leq \mu_a^*$. Let $\epsilon > 0$. What criterion must N satisfy in order that the empirical success rule be ϵ -optimal? (Note: you will not be able to solve this problem analytically. You should aim to characterize the solution to the problem as clearly as possible).

5. (10 points) Again let welfare outcomes be binary, taking the value zero or one. Let a be a status quo treatment with known mean outcome μ_a^* . Let b be an innovation with mean outcome known to lie in the interval $[L_b, U_b]$, where $0 \leq L_b < \mu_a^* < U_b \leq 1$. Thus, $(\mu_{sa}, s \in S) = \mu_a^*$ and $(\mu_{sb}, s \in S) = [L_b, U_b]$. Suppose that no sample data on outcomes under treatment b are available. Let $\epsilon > 0$. What conditions on (μ_a^*, L_b, U_b) are necessary and sufficient for there to exist an ϵ -optimal treatment rule?

Midterm Examination: Tuesday October 25 (40 points total)

A week before the American presidential election, a survey firm draws a sample of registered voters and asks each sample member how they intend to vote in the upcoming election. Let $z = 1$ if a sample member answers the voting question and $z = 0$ otherwise. The possible responses to the voting question are:

($y = c$) vote for Clinton; ($y = t$) vote for Trump, ($y = a$) vote for another candidate, ($y = n$) will not vote.

The survey firm knows the gender of every sample member, regardless of the value of z . Let $x = m$ if male and $x = f$ if female.

Consider the sample to be the population of interest. The survey yields these findings:

$$P(x = m) = 0.5, P(z = 1|x = m) = 0.4, P(y = c|x = m, z = 1) = 0.3, P(y = t|x = m, z = 1) = 0.4, \\ P(y = a|x = m, z = 1) = 0.1, P(y = n|x = m, z = 1) = 0.2.$$

$$P(x = f) = 0.5, P(z = 1|x = f) = 0.5, P(y = c|x = f, z = 1) = 0.5, P(y = t|x = f, z = 1) = 0.3, \\ P(y = a|x = f, z = 1) = 0.1, P(y = n|x = f, z = 1) = 0.1.$$

1. (12 points) Using the data alone, what is the identification region for $P(y = c) - P(y = t)$?
2. (4 points) Assume that nonrespondents never vote; thus, $P(y = n|z = 0) = 1$. What is the identification region for $P(y = c) - P(y = t)$?
3. (12 points) Assume that $P(y = c|x = m) = (0.5)P(y = c|x = f)$. What is the identification region for $P(y=c)$?
4. (12 points) Assume that $P(y = c|z = 0) = P(y = c|z = 1)$. What is the identification region for $P(y = c|x = m, z = 0)$?

Final Examination: Thursday December 1 (50 points total)

Consider the problem of best prediction of an outcome under square loss. Each member j of a large population has a binary outcome y_j that takes the value 0 or 1. The set of feasible predictors is the $[0, 1]$ interval. The square loss function $(y - t)^2$ expresses the loss from choosing predictor t when the outcome is y . The objective is to choose a predictor that minimizes mean loss $E(y - t)^2$. When answering each question below, be as explicit as possible.

1. (8 points) Suppose that you have no knowledge of $P(y = 1)$. What value of t minimizes maximum regret? What is the value of minimax regret?
2. (8 points) Let the outcome y_j of person j be observable if $z_j = 1$ and unobservable if $z_j = 0$. Suppose that you know $P(y = 1|z = 1)$ and $P(z = 1)$, with $0 < P(z = 1) < 1$. You have no knowledge of $P(y = 1|z = 0)$. What value of t minimizes maximum regret? What is the value of minimax regret?
3. (8 points) Consider again the setting of question 2. Suppose that someone chooses $t = P(y = 1|z = 1)$ as the predictor. In what cases, if any, does $t = P(y = 1|z = 1)$ have larger maximum regret than the minimax regret value obtained in question 1, which used no distributional knowledge?
4. (10 points) Suppose that you have no knowledge of $P(y = 1)$, but you do have data ψ taking a value in a sample space Ψ indexing all possible data realizations. Let Q denote the sampling distribution generating the realization of ψ . Let $\delta(\cdot): \Psi \rightarrow [0, 1]$ be a predictor function; that is, a statistical decision function that maps the observed data into a predictor value. Let S denote the state space; thus, $(P_s, Q_s, s \in S)$ is the set of feasible values of (P, Q) . The risk of predictor function δ in state s is a function of certain moments of the distribution of $[y, \delta(\psi)]$ in state s . What is this function?
5. (8 points) Consider again the setting of question 4. Using risk to measure performance, consider the regret of predictor function δ in state s . Regret is a function of certain moments of the distribution of $[y, \delta(\psi)]$ in state s . What is this function?
6. (8 points) Suppose as in question 2 that the outcome y_j of person j is observable if $z_j = 1$ and unobservable if $z_j = 0$. Suppose as before that you know $P(z = 1)$, with $0 < P(z = 1) < 1$, and that you have no knowledge of $P(y = 1|z = 0)$. In contrast to question 2, suppose that you do not know $P(y = 1|z = 1)$, but do know that $P(y = 1|z = 1) = P(y = 1|z = 0)$. You also observe data $\psi = (y_n, n = 1, \dots, N)$, this being a random sample of N outcomes drawn from the distribution $P(y|z = 1)$. Let $\delta(\cdot)$ be the sample average of the observed data; thus, $\delta(\psi) = (1/N) \sum_n y_n$. Using risk to measure performance, what is the maximum regret of $\delta(\cdot)$?

Midterm Examination: Tuesday October 24 (40 points total)

Each member j of a study population J is characterized by a vector $[y_j(A), y_j(B), y_j, z_j, w_j]$. A and B are mutually exclusive and exhaustive treatments. z_j denotes the treatment actually received by j . $y_j(A)$ and $y_j(B)$ are binary potential outcomes, with value 0 or 1. $y_j = y_j(A) \cdot 1[z_j = A] + y_j(B) \cdot 1[z_j = B]$ is the outcome realized by j . w_j is a binary covariate, with value 0 or 1.

Let P denote the population distribution of $[y(A), y(B), y, z, w]$. Let $P(y = 1|z = A) = \alpha$, $P(y = 1|z = B) = \beta$, $P(z = B) = \gamma$, and $P(w = 1) = \delta$. When answering the questions below, suppose that you know the values of $(\alpha, \beta, \gamma, \delta)$. You have no other empirical evidence. When answering, consider the assumptions in each question separately.

1. (8 points) Let M denote the median operator. What is the identification region for $M[y(B)]$, expressed as a function of $(\alpha, \beta, \gamma, \delta)$?
2. (8 points) Consider the event $[y(A) = y(B)]$. What is the identification region for $P[y(A) = y(B)]$, expressed as a function of $(\alpha, \beta, \gamma, \delta)$?
3. (8 points) Assume that $P[y(B) = 1] = P[y(A) = 1]$. What is the identification region for $P[y(B) = 1]$, expressed as a function of $(\alpha, \beta, \gamma, \delta)$?
4. (8 points) Let treatments A and B be ordered, with $B > A$. Assume that $y_j(B) \geq y_j(A)$ for all $j \in J$. What is the sharp bound on $P[y(B) = 1]$, expressed as a function of $(\alpha, \beta, \gamma, \delta)$?
5. (8 points) What is the identification region for $P[y(B) = 1|w = 1]$, expressed as a function of $(\alpha, \beta, \gamma, \delta)$?

Final Examination: Thursday November 30 (50 points total)

Consider a population J of travelers, each of whom chooses between two travel modes, labeled 0 and 1. The utility of mode c to person j is $u_{jc\gamma} = \alpha x_{jc\gamma} + \beta z_{jc} + \varepsilon_{jc}$. Here x is travel time in minutes, z is travel cost in dollars, (α, β) are parameters and γ denotes the degree of travel congestion. ε is comfort, which is measured on a 0 to 1 scale. The presence of a γ -subscript on $x_{jc\gamma}$ indicates that travel time may vary with γ . The absence of a γ -subscript on z_{jc} and ε_{jc} indicates that travel cost and comfort do not vary with the degree of congestion. Travelers do not know γ when they make their decisions. Let d_j denote the travel mode chosen by person j .

A researcher wants to learn the values of (α, β) . He knows a priori that $\alpha < 0$ and $\beta < 0$. For each person $j \in J$, the researcher observes (d_j, z_{j0}, z_{j1}) .

1. (8 points) Assume that each person j chooses a mode that maximizes subjective expected utility $\int u_{jc\gamma} dQ_{j\gamma}$, where $Q_{j\gamma}$ is the subjective distribution that j places on γ . Thus, the expected utility of mode c to person j is $\alpha v_{jc} + \beta z_{jc} + \varepsilon_{jc}$, where $v_{jc} \equiv \int x_{jc\gamma} dQ_{j\gamma}$.

The researcher does not observe $(v_{j0}, v_{j1}, \varepsilon_{j0}, \varepsilon_{j1})$ but he can place credible bounds on their values. Specifically, he assumes that $(v_{j0}, v_{j1}, \varepsilon_{j0}, \varepsilon_{j1}) \in [v_{j0L}, v_{j0U}] \times [v_{j1L}, v_{j1U}] \times [0, 1] \times [0, 1]$. Here L and U denote known person-specific lower and upper bounds on subjective expected travel times. Characterize the identification region for (α, β) . Be as explicit as possible.

2. (8 points) Maintain the above assumptions except that, instead of observing (d_j, z_{j0}, z_{j1}) for every person in the population, the researcher observes these values for a sample of N persons. The researcher does not know the process generating the sample; for example, it need not be a random sample. What do the observed data reveal about the identification region for (α, β) ?

3. (8 points) Maintain the assumptions of question 1 except that the researcher believes that persons do not have subjective distributions on γ and, hence, cannot maximize expected utility. Instead, the researcher assumes that each person j knows that $(x_{j0}, x_{j1}) \in [x_{j0L}, x_{j0U}] \times [x_{j1L}, x_{j1U}]$ and chooses a mode that maximizes minimum utility. Here L and U denote known person-specific lower and upper bounds on actual travel times. Now characterize the identification region for (α, β) . Be as explicit as possible.

4. (10 points) Consider again the informational setting of question 3 but assume now that each person minimizes maximum regret. Now characterize the identification region for (α, β) . Be as explicit as possible.

5 (8 points). Are the three models of decision making posed in questions 1, 3, and 4 refutable? If so, give the condition in which each model is refuted. If not, explain why the model is not refutable.

6. (8 points) Questions 1, 3, and 4 supposed that, for each person $j \in J$, the researcher observes (d_j, z_{j0}, z_{j1}) . Suppose that the researcher also observes the actual degree of congestion γ^* and the resulting travel times. Thus, the researcher observes $(d_j, x_{j0\gamma^*}, x_{j1\gamma^*}, z_{j0}, z_{j1})$ for each person j . How does possession of these additional data affect the identification regions obtained in questions 1, 3, and 4. Be as explicit as possible.