Midterm #2

10/25/2012

Instructions: The exam is divided into two parts. In Part I, your score is entirely based on your numerical answer, and not any accompanying explanation. In Part II, your grade for each question will be based primarily on your written explanation, and unsupported answers will receive no points. For both parts, you should refer to the provided probability tables as necessary. You may use a calculator, and one sheet of notes.

Part I (7 questions)

Each short question is worth 7 points. Write your answer in the space provided.

Problem 1 (7 points) Consider the following hypothesis test:

$$H_0 : \mu = 40$$
$$H_1 : \mu \neq 40$$

You collect the following data: $\overline{x} = 40.5$, $\sigma = 2$, and $n = 64$. What is the p-value of the test?

.0455

Problem 2 (7 points) Consider the following hypothesis test:

$$H_0 : \mu = 100$$
$$H_1 : \mu > 100$$

You collect the following data: $\overline{x} = 104$, $\sigma = 21$, and $n = 49$. What is the p-value of the test?

.091

Problem 3 (7 points) Referring back to problem 2, interpret the P-value *in plain English.* What does it tell you?

The P-value is the lowest value of α for which the null hypothesis could be rejected. That is, we can reject H_0 only if we are comfortable with at least a 9.1% chance of a type I error.

Problem 4 (7 points) Consider the following hypothesis test:

$$H_0 : \mu = 150$$
$$H_1 : \mu > 150$$

Suppose that σ is known to be 12. You wish to test your hypotheses with a 1% chance of Type I error. You are about to collect a sample of size 144. What is the probability of type II error if, in reality, $\mu = 154$?

.0475

Problem 5 (7 points) Consider the following hypothesis test:

$$H_0 : \mu = 40$$
$$H_1 : \mu \neq 40$$

Suppose that σ is known to be 10. You wish to test your hypotheses with a 5% chance of Type I error. You are about to collect a sample of size 36. What is the probability of type II error if, in reality, $\mu = 39$?

.908

Problem 6 (7 points) Consider the following hypothesis test:

$$H_0 : \mu = 230$$
$$H_1 : \mu < 230$$

Suppose that σ is known to be 4. You determine that type I error is very costly, and so you wish to test your hypotheses with a 0% chance of Type I error. You are about to collect a sample of size 225. What is the probability of type II error if, in reality, $\mu = 200$?

1

Problem 7 (7 points)

$$H_0 : \mu = 63$$
$$H_1 : \mu \neq 63$$

Suppose that σ is known to be 9, and you will have access to data from a sample of size $n = 100$. You determine that type II error is very costly, and so you design a test with a 0% chance of type II error. What is the probability of type I error for this test? Suppose that, in reality, $\mu = 62$, although you don't know this.

A 0% chance of type II error is only possible if you always reject H_0. This means that, in the event H_0 is true, there is a 100% chance of rejecting. Therefore, the answer is $\alpha = 1$

Part II (3 questions)

Questions 8, 9, and 10 are worth 15, 25, and 10 points, respectively. Thoroughly support all of your answers. Credit will not be given without complete explanations.

Problem 8 (15 points) A researcher runs a regression of monthly crime rate (CRIME) on number of police (POLICE), where CRIME is measured as the number of felonies reported per 100,000 people committed that month, and POLICE is the number of policemen in active duty in the state that month. The researcher uses a data set of 6 years worth of state-level observations (one observation is the crime rate and the number of police in a given state in a given month), and obtains the following results:

Regression	Statistics
Multiple R	.321827
R Square	.14816
Adjusted R Square	.121732
Standard error	6.653
Observations	3,600

	coefficients	Standard error	t stat	P-value	Lower 95%	Upper 95%
Intercept	41.3546	12.5760	3.2884	.0005038	16.70704	66.00496
POLICE	4.2849	1.0746	3.9874	3.3395 E-005	2.1786	6.3911

a. Interpret the regression results. What is the relationship between number of police and the crime rate? Is number of police a statistically significant variable in explaining variation in the crime rate?

The regression predicts that one more policeman is correlated with 4.2849 additional crimes per 100,000 state residents. The variable is a statistically significant explained of the crime rate. 12.2% of the overall variation in the crime rate can be explained by the number of police.

b. Do these results cause you to conclude that police cause crimes? If not, what alternative explanations are there for the regression results? How would you more accurately measure the impact of hiring more police on the crime rate?

The result of the regression is that more police are correlated with higher crime rates. It is very unlikely that more police working cause more crimes (indeed, the opposite is almost certainly true). So why the result? One potential explanation is that when crime increases, more police are hired to deal with the higher crime rate. In this case, the causality is that higher crime causes there to be more police, not the other way around. Another potential explanation is that a higher percentage of crimes are reported when more police are hired. In this case, regardless of whether crime declines appreciably in response to there being more police, it is possible that the data would show an increase in *reported* crimes following the hiring of more police.

If you think that police are hired in response to more crime (the first problem), you could look for variation in the number of policeman that has nothing to do with the crime rate. One example used in the literature is that spending on police prior to elections tends to go up, so that the incumbent mayor can claim that he is tough on crime as he campaigns for reelection. This particular empirical strategy has been used successfully to measure the decrease in crime resulting from more police. If you are worried about crimes being reported more often after more police are hired, one strategy might be to look at only murders (i.e. the effect of more police on murders), as presumably a very high percentage of murders are reported, regardless of the number of police. This might not be true for property crimes.

Problem 9 (25 points) In a 2003 paper in the *American Law and Economics Review*, Katz, Levitt, and Shustorovich study the deterrent effects of the death penalty and poor prison conditions on crime rates. They are interested in testing two hypotheses:

1. The death penalty deters crime
2. Poor prison conditions deter crime

Their data set consists of 2050 observations, where each observation is a collection of information from one state in one year. Their measurement of prison conditions is the number of deaths per 1,000 prisoners that occur in a year. They have access to the following data:

- CRIMERATE: violent crime per 100,000 residents (mean: 356)
- PRISON_DEATHS: prison deaths per 1,000 prisoners (mean: 3.1)
- EXECTUIONS: executions per 1,000 prisoners (mean: .11)
- PRISONERS: prisoners per 100,000 residents (mean: 126)
- INCOME: real per capita income (mean: 13,724)
- UNEMPLOYMENT: unemployment rate (mean: 3.37)
- BLACK: fraction of the state that is black (mean: .11)
- URBAN: fraction of the state that lives in urban areas (mean: .72)
- YOUNG: fraction of the state that is 24 years or younger (mean: .42)

The researchers consider the following linear regression model:

$$\begin{aligned} CRIMERATE = \beta_0 &+ \beta_1 * PRISON\_DEATHS + \beta_2 * EXECUTIONS + \beta_3 * PRISONERS \\ &+ \beta_4 * INCOME + \beta_5 * UNEMPLOYMENT + \beta_6 * BLACK + \beta_7 * URBAN \\ &+ \beta_8 * YOUNG + \epsilon \end{aligned}$$

a. What sign do you expect β_1, β_2, ..., β_8 to have, and why? Give 8 separate answers. You may use the next page if necessary.

Points will depend on whether or not you are able to say something useful about the anticipated sign of each coefficient. Possible answers:

- β_1: negative. Harsher prison conditions raise the cost of committing a crime.
- β_2: negative. The possibility of being executed raises the cost of committing a violent crime.
- β_3: negative. More prisoners means fewer criminals outside of prison to commit crime, and may also mean that the criminal justice system is more efficient, raising the cost of committing a crime.
- β_4: negative. Higher income means more legitimate opportunities, raising the opportunity cost of committing a crime.
- β_5: positive. Unemployed people have more free time and a lower opportunity cost of committing crime.
- β_6: positive. The crime rate is much higher among black men in particular than it is for the rest of the US population.
- β_7: positive. No strong prior belief, actually, but one hypothesis would be that concentrated urban populations will have more crime than rural areas, where people live far apart from one another.

- β_8: positive. Young people commit far more crimes than old people.

The researchers run a regression of violent crime on the other 8 variables, and obtain the following results:

variable	coefficients	standard error	t-stat	P-value
Intercept	324	123.1	2.63	.0085
Prison deaths/1,000 prisoners	-3.4	1.5	-2.27	.024
Executions/1,000 prisoners	-4.1	3.0	-1.37	.172
Prisoners/100,000 residents	.21	.27	.78	.436
Real per capita income	.42	.16	2.63	.008
Unemployment rate	-.15	.08	1.875	.061
Fraction black	37.1	16.1	2.30	.021
Fraction urban	-7.3	6.9	1.06	.290
fraction under 24	2.6	10.2	.255	.799

b. Which variables are significant predictors of the crime rate and which are not?

Prison deaths, income, and the fraction of the population that is black are significant at the 5% level. Additionally, the unemployment rate is significant at the 10% level. The other variables are insignificant.

c. Is there more evidence to say that the death penalty deters crime or that poor prison conditions deter crime?

Based on these results, prison conditions appear to do more to deter crime than executions, as only the former has a statistically significant effect on the crime rate.

Problem 10 (10 points) A friend shows you the following headline:

> "A new poll of likely voters in Nevada shows that Obama has 50% support and Romney has 48%. The poll's margin of error is +/- 4.5%."

a. Your friend asks you for clarification. Is Obama winning in Nevada? Use statistical concepts studied in Eco 391 to explain to your friend, *in plain English*, what is going on.

There is not enough evidence to say that Obama is winning. Any poll is subject to statistical noise, meaning that, just by chance, more Obama voters or more Romney voters could turn up than in the general voting population. So it is possible that a small (2 percentage point) Obama lead is attributable only to this statistical noise, and that if the polling company re-ran the poll, they would get a different result. Indeed, the polling company thinks that their results should only be trusted within 4.5 percentage points, which means that in their view, most of their polls will be correct within 4.5 percentage points of the actual outcome, but should not be relied upon to be able to distinguish differences smaller than 4.5 percentage points.

b. Some polls have larger margins or error than others. What is one factor that determines the size of margin of error? Explain.

One answer is sample size. As the number of people polled increases, the margin of error goes down. Polling companies weight the benefit of a smaller margin of error against the cost of a larger sample.