Economics 257 – Assignment 2

1. (This is a classic problem on which doctors are often said to do poorly, suggesting many mis-interpreted test results; however to be fair to them I think there may be semantic issues in that many understand a different meaning of the term 'false positive'). A certain test for a genetic disorder has a false positive rate of 2%, meaning that in cases where the condition is absent, the test will falsely indicate that it is present 2% of the time. If an individual does have the disorder, the test correctly identifies this 99% of the time. The disorder occurs in 1 of every 1000 people. If someone gets a positive result, what is the probability that the individual has the disorder?

Denote by $+$ the event of a positive result. Denote by H the event that a tested individual is healthy, and by D the event that the individual has the disorder. The information we are given is:

$$P(+|H) = 0.02, \quad P(+|D) = 0.99, \quad P(D) = 0.001.$$

By the definition of conditional probabilities, these pieces of information are equivalent to

$$P(+\cap H) = P(H)P(+|H), \quad P(+\cap D) = P(D)P(+|D), \quad P(H) = 0.999.$$

Thus, numerically,

$$P(+\cap H) = 0.999 \times 0.02 = 0.01998, \quad P(+\cap D) = 0.001 \times 0.99 = 0.00099$$

In addition, since $H \cup D = \Omega$, $P(+) = P(+\cap H) + P(+\cap D) = 0.01998 + 0.00099 = 0.02097$.

We seek the probability that the tested individual has the disorder conditional on a positive test result, that is, $P(D|+)$. Compute as follows:

$$P(D|+) = \frac{P(+\cap D)}{P(+)} = \frac{0.00099}{0.02097} = 0.04719,$$

where the answer is rounded to five digits after the decimal point. The answer is surprising, in that it shows that there is a rather small probability of the individual having the disorder even in the presence of a positive test result.

(10 marks)

2. An equity analyst claims to be able to predict which of a set of stocks will gain the most in the coming year. He is given a list of eight stocks, and asked to predict, in order, the top three. What is the probability of getting the top three right, in order, by chance? (That is, assume that he in fact has no useful knowledge to bring to bear on this question, so that any stock is as likely as any other to get randomly picked for any place in his top three.)

Show two different ways to get this answer.

There is one chance in eight of selecting the top stock. Assuming that it is chosen correctly, there remain seven stocks from which to choose the second top, and, after that, six stocks from which to choose the third top. If the choices are all by chance, then the probability of getting the right top three is $1/8 \times 1/7 \times 1/6 = 1/336$.

Alternatively, we may ask how many ways there are of choosing three stocks in order from a set of eight. the answer is ${}_3P_8 = 8 \cdot 7 \cdot 6 = 336$. Only one of these is right, and so, if all the choices are equally probable, the probability of getting it right is $1/336$, as before.

(4 marks)

3. Obtain data from the web site of the Federal Reserve Bank of St. Louis, at

`http://research.stlouisfed.org/fred2/data/`

on the following US variables:

Y : real GDP, quarterly, seasonally adjusted;

X : initial jobless claims, seasonally adjusted (SA) if available, otherwise unadjusted (NSA), available monthly, but convert to quarterly average either via web site or do it yourself from monthly data

P : consumer price index (CPI – all items), available monthly, again convert to quarterly average either via web site or do it yourself from monthly data

In your assignment, indicate the day on which you downloaded the data, since the data sets may be updated.

I downloaded my data on November 1st, 2016. It was necessary to begin the sample in the 2nd quarter of 1967, since, although GDP and the CPI were available from the late forties, the initial jobless claims (UC) were not available until later. The CPI was available monthly, and so was averaged over the three months of each quarter. The UC data were weekly, and so, in order to get quarterly data, it was necessary to add up the claims in each week of a given quarter. I ended up with a sample of 198 observations. Students may well have handled the data differently, which is fine, provided they say what they did – and it makes sense!

a. For each series Z, compute the quarterly percentage change $(Z_t - Z_{t-1})/Z_{t-1} \times 100$. Call these by the same variable name but lower case: y, x, p.

b. Compute the correlation matrix (a 3×3 matrix) of the percentage changes y, x, p. The correlation matrix has nine entries, but the three on the main diagonal are 1, and of the other 6, each is a repeat of another entry (the matrix is symmetric), and so there are only three numbers to calculate: $\text{corr}(y, x)$, $\text{corr}(y, p)$, and $\text{corr}(x, p)$. Interpret these results briefly.

I found the following matrix, with y in the first row and column, x in the second, p in the third.

$$\text{correlation matrix} = \begin{bmatrix} 1 & -0.396302 & 0.494971 \\ -0.396302 & 1 & 0.155095 \\ 0.494971 & 0.155095 & 1 \end{bmatrix}.$$

It appears that changes in GDP are negatively correlated with changes in unemployment claims – reasonable as unemployment is counter-cyclical – while changes in GDP and the CPI are positively correlated – again reasonable as prices are usually pro-cyclical.

c. Do the same for the the original variables Y, X, P.

For the original variables, what I found was:

$$\text{correlation matrix} = \begin{bmatrix} 1 & 0.914427 & 0.995082 \\ 0.914427 & 1 & 0.893265 \\ 0.995082 & 0.893265 & 1 \end{bmatrix}.$$

d. Interpret c. The result is actually nonsense: why? If you're not sure, try to download data on some clearly unrelated trending series such as cumulative quarterly rainfall in Tierra del Fuego, or the cumulative quarterly number of UFOs reported to NASA; compute correlations with Y, X, P and think about the result

The result is nonsense because almost all economic time series trend upwards, and this is what gives rise to the large positive correlations. Note however that unemployment is less highly correlated with the other two variables than they are taken together. The series is not the unemployment *rate*, but the number of first-time claims in the period, that is, the number of transitions from employment to unemployment, or perhaps from full-time education to unemployment. As the population grows, so does the number of potentially unemployed people.

If you manage to download an unrelated trending series, you will also find strong positive correlation.

(16 marks: we can't expect everyone to get the same numbers, or to treat the data in the same way)

4. Newspaper subscribers are polled on whether they read the business section, and whether they invest in the stock market. The responses can be broken down as follows: invest and read bus. regularly, 20%; invest and read occasionally, 11%; invest and never read, 6%; don't invest and read regularly, 15%; don't invest and read occasionally, 26%; don't invest and never read it, 22%.

a) Find the probability that a randomly-chosen subscriber never reads the business section.

b) Find the probability that a randomly chosen subscriber invests in stocks.

c) Find the probability that a subscriber who never reads the business section invests in stocks.

d) Find the probability that a subscriber who invests in stocks never reads the business section.

e) Find the probability that a subscriber who does not regularly read (that is, occasionally or never) the business section does invest in stocks.

Define the following events: RR: reads regularly; RO = reads occasionally; NR = never reads; I = invests; NI = don't invest. Our information is:

$$P(I \cap RR) = 0.20, \quad P(I \cap RO) = 0.11, \quad P(I \cap NR) = 0.06$$
$$P(NI \cap RR) = 0.15, \quad P(NI \cap RO) = 0.26, \quad P(NI \cap NR) = 0.22$$

Note that these events are mutually exclusive and exhaustive.

a) Since $I \cup NI = \Omega$, the event 'Never reads' is $NR = (NR \cap I) \cup (NR \cap NI)$ and the probability of this is $P(NR) = 0.06 + 0.22 = 0.28$.

(2 marks)

b) $P(I) = P(I \cap RR) + P(I \cap RO) + P(I \cap NR) = 0.20 + 0.11 + 0.06 = 0.37$.

(2 marks)

c) $P(I|NR) = P(I \cap NR)/P(NR) = 0.06/0.28 = 3/14$.

(2 marks)

d) $P(NR|I) = P(I \cap NR)/P(I) = 0.06/(0.20 + 0.11 + 0.06) = 0.06/0.37 = 6/37$.

(2 marks)

e) First, note that $P(RO) = P(I \cap RO) + P(NI \cap RO) = 0.11 + 0.26 = 0.37$, and $P(NR) = P(I \cap NR) + P(NI \cap NR) = 0.06 + 0.22 = 0.28$. Then

$$\begin{aligned} P(I|(RO \cup NR)) &= P(I \cap (RO \cup NR))/P(RO \cup NR) \\ &= \big[(P\big(I \cap RO) \cup (I \cap NR)\big)\big]/\big(P(RO) + P(NR)\big) \\ &= \big[(0.11 + 0.06)\big]/(0.37 + 0.28) = 0.17/0.65 = 17/65. \end{aligned}$$

(4 marks)

(Total for question 4: 12 marks)

(Total for assignment: 10+4+16+12 = 42)