ASE, Autumn Class exercises.

Pedro Gomes: p.gomes@bbk.ac.uk

2019-20

1 Introduction

These exercises will be discussed each week in the tutorial classes. You can only learn statistics by doing it, these excercises are designed to give you practice at doing it.

Try the class exercises before the class or you will not know what your difficulties are.

2 Week 2 Using Numbers

In the table below Gross Domestic Product is a measure of the output of a country. These data are available in a spreadsheet on Moodle. When doing calculations, you must be careful about units. These variables are all measured in different units and ratios will depend on the units of the numerator and denominator. Expressing the units as powers of 10 is often useful. $0.1 = 10^{-1}; 1 = 10^0$; $10 = 10^1$; $100 = 10^2$; $1,000,000 = 10^6$. The power gives you the number of zeros after the one. For ratios $10^a/10^b = 10^{a-b}$.

Gross Domestic Product, GDP, (billions of constant 1995 US$), Population (millions), Military Expenditure (billions of constant 1995 US$) and the number in the armed forces (thousands) in Developed and Developing Countries, 1985 and 1995.

	1985	1995
GDP		
Rich	21190	23950
Poor	4184	7010
Population		
Rich	1215.7	1151.5
Poor	3620.8	4520.0
Military Expenditures		
Rich	1100.8	667.8
Poor	230.0	196.7
Number in Armed Forces		
Rich	11920	7667
Poor	16150	15120

Source World Military Expenditures and Arms Transfers, US Arms Control and Disarmament Agency. The Rich group includes 33 high per-capita income countries, the Poor group the rest of the world.

(a) Express the units of each of the four variables as powers of 10.

From this table **calculate**:

(b) Per-capita GDP (divide GDP by population) for rich and poor countries in 1985 and 1995. What units is this measure in?

(c) The growth rate of per-capita GDP 1985-1995 for rich and poor countries. What units is this measure in?

(d) The percentage share of military expenditure in GDP for rich and poor countries in 1985 and 1995? Comment on the change.

(e) The number of people in the armed forces per 1000 population for rich and poor countries in 1985 and 1995.

(f) Military expenditure per member of the armed forces for rich and poor countries in 1985 and 1995.

3 Week 3, Descriptive Statistics

- (1) In a speech, *Why Banks failed the stress test*, February 2009, Andrew Haldane of the Bank of England provides the following summary statistics for the "golden era" 1998-2007 and for a long period. Growth is annual percent GDP growth, inflation is annual percent change in the RPI and for both the long period is 1857-2007. FTSE is the monthly percent change in the all share index and the long period is 1693-2007.

	Growth		Inflation		FTSE	
	98-07	long	98-07	long	98-07	long
Mean	2.9	2.0	2.8	3.1	0.2	0.2
SD	0.6	2.7	0.9	5.9	4.1	4.1
Skew	0.2	-0.8	0.0	1.2	-0.8	2.6
Excess Kurtosis	-0.8	2.2	-0.3	3.0	3.8	62.3

(a) Explain how the mean; standard deviation, SD; coefficient of skewness and coefficient of kurtosis are calculated, and what they measure.

(b) What values for the coefficients of skewness and kurtosis would you expect from a normal distribution. Which of the series shows the least evidence of normality.

(c) Haldane says "these distributions suggest that the Golden Era" distributions have a much smaller variance and slimmer tails" and "many risk management models developed within the private sector during the golden decade were, in effect, pre-programmed to induce disaster myopia.". Explain what he means using these statistics.

- (2) Consider the set of observations on a variable x_i, $i = 1, 2, ..., N$ and

$$w_i = a + bx_i.$$

Define the mean of x_i as

$$\bar{x} = \sum_{i=1}^{N} x_i/N$$

and similarly for $\bar{w}$. Show that each of the following three expressions are true:

$$\begin{aligned}
\sum_{i=1}^{N} w_i &= Na + b\sum_{i=1}^{N} x_i; \\
\bar{w} &= a + b\bar{x} \\
\sum_{i=1}^{N}(x_i - \bar{x}) &= 0.
\end{aligned}$$

4 Week 4 Variances and covariances

Consider the set of observations on variables x_i, y_i; $i = 1, 2, ..., N$. Define the variance, $V(x)$, and standard deviation, $S(x)$, of x_i as

$$V(x) = \sum(x_i - \bar{x})^2/N; \quad S(x) = \sqrt{V(x)}$$

and similarly for y_i. Define the covariance between x_i and y_i as

$$Cov(x, y) = \sum_{i=1}^{N}(y_i - \bar{y})(x_i - \bar{x})/N$$

and the correlation r between x_i and y_i as

$$r(x, y) = \frac{Cov(x, y)}{S(X)S(Y)}.$$

Let

$$z(y_i) = \frac{y_i - \bar{y}}{S(y)}, \quad z\left(x_i\right) = \frac{x_i - \bar{x}}{S(x)}.$$

- (1) Show that each of the following four expressions are true

$$\begin{aligned}
\sum_{i=1}^{N} z(y_i)/N &= 0 \\
\sum_{i=1}^{N}[z(y_i)]^2/N &= 1 \\
r(x, y) &= \sum z(y_i)z(x_i)/N \\
r(x, x) &= 1
\end{aligned}$$

From this show that $-1 \leq r \leq +1$.

- (2) Show that the variance equals the mean of the squares minus the square of the mean:

$$N^{-1}\sum_{i=1}^{N}(x_i-\overline{x})^2 = N^{-1}\sum_{i=1}^{N}x_i^2-(\overline{x})^2$$

where $\overline{x}=\sum x_i/N$.

5 Week 5 Index Numbers

- (1) A firm buys two goods A and B in two years. The table below shows the price and quantity of each purchased in year zero and year one.

	Year	0	Year	1
Good	P_{i0}	Q_{i0}	P_{i1}	Q_{i1}
A	20	10	10	20
B	10	20	20	10

(a) What was total expenditure $\sum_i P_{it}Q_{it}$, in each year $t=0,1$, where $i=A,B$?

(b) What would have been its total expenditure in each year if in both years it had bought (i) year zero quantities (ii) year one quantities?

(c) Use the estimates in (b) to calculate two measure of inflation (i) using year zero quantities (ii) using year one quantities.

(d) Comment on your results.

(e) Consider the three indexes below. Explain which one is a price index, which one a quantity index and which one is an expenditure index. Interpret the relationship between them.

$$\frac{\sum_i P_{i1}Q_{i1}}{\sum_i P_{i0}Q_{i0}} = \frac{\sum_i P_{i1}Q_{i0}}{\sum_i P_{i0}Q_{i0}} \times \frac{\sum_i P_{i1}Q_{i1}}{\sum_i P_{i1}Q_{i0}}$$

- (2) UK GDP in current market prices in 1995 was £712,548m, while in 1997 it was £801,972m. GDP at constant 1995 market prices in 1997 was £756,144m.

(a) Construct index numbers, 1995=100 for: current price GDP; constant price GDP; and the GDP deflator in 1997.

(b) From these numbers calculate the average annual rate of inflation between 1995 and 1997.

(c) From these numbers calculate the average annual rate of growth between 1995 and 1997.

(d) If the interest rate on two year bonds in 1995 was 10% per annum what would the real per annum interest rate over this period be.

(e) Explain what Gross Domestic Product measures. What limitations does it have as a measure of the economic wellbeing of a nation.

6 Week 6 Reading Week

7 Week 7 Probability

- (1) Suppose you toss a fair coin three times in a row. What is the probability of:

(a) three heads in a row;
(b) a tail followed by two heads.
(c) at least one tail in the three throws.
Either write out the 8 (2^3) possible outcomes and count how many are involved in each case or look at the qustion in week 8..

- (2) Students take two exams A and B. 60% pass A, 80% pass B, 50% pass both.

(a) Fill in the remaining five elements of the joint and marginal distributions below, where PA indicates pass A, FB fail B, etc.
(b) What is the probability of a student passing B given that they passed A?
(c) Are the two events passing A and passing B (i) mutually exclusive (ii) independent?

	PA	FA	B
PB	50		80
FB			
A	60		100

- (3) Define $P(A)$ as the probability of event A happening; $P(B)$ the probability of event B happening; $P(A \cap B)$ the probability of both A and B happening; $P(A \cup B)$ the probability of either A or B happening; and $P(A \mid B)$ the probability of A happening conditional on B already having happened.

(a) What is $P(A \cap B)$ if A and B are mutually exclusive.
(b) What is $P(A \cap B)$ if A and B are independent?
(c) What is $P(A \cup B)$?
(d) Show that

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

- (4) You are in a US quiz show. The host shows you three closed boxes in one of which there is a prize. The host knows which box the prize is in, you do not. You choose a box. The host then opens another box, not the one you chose, and shows that it is empty. He can always do this. You can either stick with the box you originally chose or change to the other unopened box. What should you do: stick or change? What is the probability that the prize is in the other unopened box?

- (5) **(Optional)**. Calculate the probability that two people in a group of size N will have the same birthday. What size group do you need for there to be a 50% chance that two people will have the same birthday? Ignore leap years.

Use a spreadsheet for this and work it out in terms of the probability of not having the same birthday. In the first row we are going to put values for N (the number of people in the group), in the second row we are going to put the probability that no two people in a group of that size have the same birthday.

In A1 put 1, in B1 put =A1+1, copy this to the right to Z1.

In A2 put 1. Now in B2 we need to calculate the probability that two people will NOT share the same birthday. There are 364 possible days, i.e. any day but the first person's birthday, so the probability is 364/365. So put in B2 =A2*(365-A1)/365. Copy this right. Go to C2, the formula will give you $1 \times (364/365) \times (363/365)$. The third person, has to have birthdays that are different from the first and the second. Follow along until the probability of no two people having the same birthday falls below a half.

8 Week 8 Random Variables

- (1) In 1968, an accountant named Charles Reep published an article in the Journal of the Royal Statistical Society. After analysing more than 2200 football matches he concluded that 80 percent of goals were scored from plays of fewer than four passes. He claimed that three passes – a long ball, knockdown and strike – was the path for success. Taken from the book "Game changers" by Joao Medeiros.

Suppose that in a typical football match there are on average 900 plays. Out of those only 100 have 4 or more passes. There are 5 goals per match.

a) Complete the blank cells in following table given the information above

Number of plays in a typical football match, G=ending in goal, NG=not ending in goal.

	G	NG	Total
<4 passes			
⪰4 passes			
Total			900

b) What is the probability of scoring a goal, conditional on a play with less than 4 passes: $P(G \mid< 4)$?

c) What is the probability of scoring a goal, conditional on a play with 4 or more passes: $P(G \mid\succeq 4)$, ?

d) From the definitions of conditional probability show in algebra and with the numbers above that

$$P\left(G \mid< 4\right) = \frac{P(< 4 \mid G)P(G)}{P(< 4)}$$

e) Reep's conclusion, that three passes was the path for success, was mistaken. Explain the mistake in his analysis.

.

- (2) If a random variable follows a binomial distribution with probability of success p on a particular trial, then the probability of getting k successes in n trials is given by

$$\left(\frac{n!}{k!(n-k)!}\right) p^k (1-p)^{n-k}$$

(a|) Suppose that you toss a fair coin 5 times. Give the probability of getting, $0, 1, 2, 3, 4, 5$ heads.

(b) Calculate the expected value and variance for the number of heads. Show that these are np and $np(1-p)$.

- (3) In Excel type: =RAND() into cell A1. This will give you will get a uniformly distributed random number over zero-one. Copy this to B1:O1; In cell P1 type =AVERAGE(A1:O1). In Cell Q1 type =STDEV.S(A1:O1). In Cell R1 type =STDEV.P(A1:O1). Copy the line A1:R1 to lines 2:50. You now have 50 samples of 15 random numbers plus estimates of the mean and sample and population standard deviations. In cell P51, type =AVERAGE(P1:P50). In cell Q51 type =AVERAGE(Q1:Q50) In cell R51 type =AVERAGE(R1:R50) . In cell P52 type =STDEV.S(P1:P50).

Since the uniform distribution has expected value 0.5 and variance 1/12, standard deviation 0.288675, the theoretical numbers you should expect in the cells are: P51= 0.5; Q51=0.288675; R51=0.288675(49/50) = 0.282902; P52=0.288675/($\sqrt{15}$) = 0.0745. They will differ from this because of sampling variation. Construct a histogram for the values of the mean of the 15 observations which are in cells P1:P50. Does the distribution look normal?

9 Week 9, Normal Distribution

For this and subsequent questions you need to use values for the cumulative normal distribution which you can get from statistical tables (e.g. the ones in Barrow) or from the web or a spreadsheet.

Marks on an exam x_i are distributed $N(60, 10^2)$.

(a) Ii) Explain in words what is meant by $N(60, 10^2)$, For a random variable Z, explain what is meant by $P(Z < z)$.

(b) What proportion of the students get

(i) less than 40;

(ii) between 40 and 50;

(iii) between 50 and 60

(iv) between 60 and 70

(v) between 70 and 80

(vi) 80 and over.

(c) What mark does a student need to get into the top 10%.

10 Week 10 Properties of estimators and distributions

- (1) Suppose you have a sample of data, Y_i, $i = 1, 2, .., N$, where $Y \sim IN(\mu, \sigma^2)$.

(a) Explain what $Y \sim IN(\mu, \sigma^2)$ means.
(b) How would you obtain unbiased estimates of μ and σ^2? Explain what unbiased means.
(c) How would you estimate the standard error of your estimate of μ?
(d) Suppose that the distribution of your sample was not normal but highly skewed. Explain what this means and discuss what other measures of central tendency that you might use.

- (2) Marks on an exam, $Y_i \sim IN(50, 10^2)$, are independently normally distributed with expected value 50 and standard deviation 10. In a class of 16, what is the probability that the average mark is greater than 53?

11 Week 11 Confidence Intervals

- (1) In a developed country with data for a long period, $t = 1, 2, ..., T$, the real return on equities,Y_t seems to have been independently distributed with an average real return of 5% with a standard deviation of 10%. These can be taken as known.

(a) Assuming that the distribution is normal, what is your forecast for returns in $T + 1$ and your (i) 66%, (ii) 90% (ii) 95% confidence interval for the forecast.
(b) Again assuming that the distribution is normal, What is your forecast for average returns over the next 25 years: $T + 1, .., T + 25$, and your (i) 68%, (ii) 90% (ii) 95% confidence interval.
(c) Suppose the distribution was non-normal how would your answers to (a) and (b) change.

- (2) In an emerging market with only 9 years of data the estimated mean is 5% and estimated sample standard deviation is 10%. However, the data does seem independently normally distributed.

(a) What is the formula used for calculating the sample standard deviation in this case.
(b) What is your forecast for returns in $T + 1$ and your (i) 90% (ii) 95% confidence interval for the forecast.
(c) What is your forecast for average returns over the next 25 years: $T + 1, .., T + 25$, and your (i) 90% (ii) 95% confidence interval.