

Question 1A : Introductory Definitions

A researcher wants to know if the Irish public are in support of a new government policy. The researcher believes that less than 40% of the public support it, and wishes to test this hypothesis by carrying out a survey. From a list of all individuals living in the local area, she draws a random sample of 200 individuals and finds that 70 are in favour.

- (a) What is the population? (b) What is the sample? (c) What is the value of n ?
(d) Identify the parameter and statistic - give the symbol and value for both. (e) What type of data was collected on each of the individuals in the sample? (f) Identify any potential bias.

Question 1B : Discrete and Continuous Variables

Determine if the following are discrete or continuous variables:

- (i) The number of defective items in a consignment of 100 items.
- (ii) The heights of your classmates.
- (iii) The distance traveled by a randomly chosen business consultant last year.
- (iv) The number of books on your shelves.
- (v) The weights of aubergines.
- (vi) The age of a person.
- (vii) The number of words in a book.

Question 5 : SPSS Output

1. What is the mean Science score for all Males?
2. What is the mean Science score for all Females?
3. What is the mean Science score for everybody in the study?
4. How many respondents are there altogether?
5. How many people are in the 15-20 age group altogether?
6. Which Age group has the highest Score?

Probability Distributions

- The discrete uniform distribution
- The continuous uniform distribution
- The binomial distribution
- The poisson distribution
- The exponential distribution
- the Normal distribution
- What is the probability that the server breaks down three times in a quarter?
- What is the probability that a server breaks down exactly five times in one year?

Assume that the number of weekly study hours for students at a certain university is approximately normally distributed with a mean of 22 and a standard deviation of 6.

- Find the probability that a randomly chosen student studies less than 12 hours.
- Estimate the percentage of students that study more than 37 hours.

In an examination the scores of students who attend schools of type A are normally distributed about a mean of 55 with a standard deviation of 6. The scores of students who attend type B schools are normally distributed about a mean of 60 with a standard deviation of 5. Which type of school would have a higher proportion of students with marks above 70?

The heights for a group of forty rowing club members are tabulated as follows;

141	148	149	149	155	156	167	169	169	170
171	173	175	176	177	179	182	182	183	183
183	184	184	184	185	185	185	186	186	189
191	191	191	191	192	192	192	193	194	199

1. (6 marks) Summarize the data in the above table using a frequency table. Use 6 class intervals, with 140 as the lower limit of the first interval.
2. (6 marks) Draw a histogram for the above data.
3. (4 marks) Comment on the shape of the histogram. Based on the shape of the histogram, what is the best measure of centrality and variability?
4. (12 marks) Construct a box plot for the above data. Clearly demonstrate how all of the necessary values were computed.

1 KB Tutorial 2

Question 4

Consider a RAID (redundant array of inexpensive disks) system where multiple hard disks are used simultaneously.

Let's assume that we have two hard disks that work *independently* of each other. Define the events H_1 = "hard disk one works" and H_2 = "hard disk two works" and also assume that $\Pr(H_1) = \Pr(H_2) = 0.9$.

- RAID-0 is a system which increases performance but only works if *both* hard disks work.
- RAID-1 is a system which does not increase performance but still works with only one working hard disk.

(a) Calculate $\Pr(\text{RAID-0 works})$ and $\Pr(\text{RAID-0 fails})$.

(b) Calculate $\Pr(\text{RAID-1 works})$ and $\Pr(\text{RAID-1 fails})$. (c) Calculate $\Pr(H_1^c)$ and $\Pr(H_2^c)$.

(d) Cheap hard disks exist with $\Pr(H) = 0.6$. Consider a RAID-1 system with 3 of these hard disks - calculate $\Pr(\text{RAID-1 fails})$ in this case.

(e) In part (a) we found that $\Pr(\text{RAID-1 fails}) = 0.01$. How many cheap disks would be required to match this level of reliability?

Question 5

A software company examined blocks of code written by its employees. Each block of code was tested for bugs and, in addition, the skill level of the employee was also recorded. See table:

		Skill Level			Total
		High	Average	Low	
Bug in	No	140	600	100	840
Code	Yes	5	70	40	115
Total		145	670	140	955

In answering the following questions use appropriate probability notation.

Let B = “bug” and, hence, B^c = “no bug”.

Also let S_H = “skill: high”, S_A = “skill: average” and S_L = “skill: low”.

(a) Calculate the probability that the programmer has: (i) high skill, (ii) average skill and (iii) low skill. (b) Calculate the probability of a bug. (c) Calculate the probability of a bug *given that* the code was written by a programmer with: (i) high skill, (ii) average skill and (iii) low skill. (d) Comment on the above conditional (i.e., updated) probabilities compared with $\Pr(B)$ calculated in part (b). Is the presence of bugs independent of the skill level? (e) Show that $\Pr(S_A | B) > \Pr(S_L | B)$. Explain the reason for this.

MA4603 and MA4505 Tutorial 2 (Week 3)

Remark : No SPSS related questions this week

Question 1

A laptop manufacturer wishes to test a particular brand of CPU. A sample of 60 CPUs were selected and used to perform an intensive task for 1 hour. The temperature of each one was then recorded.

17.1	17.6	18.3	20.8	20.8	25.1	26.3	27.3	28.8	31.2
34.2	36.7	36.8	37.9	37.9	38.1	38.3	39.2	40.7	41.6
42.7	43.7	44.0	45.2	45.3	47.8	48.9	50.3	50.9	51.5
52.5	53.6	53.8	55.0	56.3	57.2	58.2	59.4	59.7	60.9
62.6	62.8	63.9	64.7	66.8	67.2	67.4	68.1	68.4	69.7
69.7	70.2	70.5	71.0	81.6	82.1	82.5	82.8	85.1	89.8

- (i) Construct a simple frequency table with 8 bins (or class intervals) (use 10 as the first break-point, and use “decades”).
- (ii) Draw the histogram (use relative frequency).
- (iii) Calculate the median.
- (iv) Calculate the quartiles.
- (v) Using the lower/upper fences, identify any outliers.
- (vi) Draw the boxplot.
- (vii) Is the data symmetric, left-skewed or right-skewed?

Question 2

The masses of 30 human males and 30 arabian stallions were observed. Their masses (in lbs) are given below

Humans

106, 120, 130, 138, 145, 151, 156, 161, 166, 171
176, 180, 185, 189, 194, 198, 203, 208, 212, 217
223, 228, 234, 240, 247, 255, 264, 276, 290, 313

Stallions

808, 824, 835, 843, 851, 857, 862, 868, 872, 877
881, 886, 890, 894, 898, 902, 906, 910, 914, 919
923, 928, 932, 938, 943, 949, 957, 965, 976, 992

- a) Draw histograms for these samples and compare them with respect to shape, centrality and relative dispersion.
- b) Calculate the medians of these samples (from the raw data).

Question 3

(MA4102 Exam Question from 2011)

Image: Week3Boxplotquestion

1.1 Boxplots

4	6	8	9	17	17	18	19	20	22
22	27	28	29	31	35	38	39	40	46
48	56	56	57	57	58	58	60	61	62
64	66	68	69	74	75	78	79	80	82

lower fence? Upper fence?

Any values above or below fences?

2 Tutorial C - Probability

Question 1

Assume that there are three different routes to get to a particular location: R_1 , R_2 and R_3 . You take R_1 75% of the time, R_2 20% of the time and R_3 the rest of the time. If you take R_1 , there is a 90% chance that you will be on time; if you take R_2 , there is a 50% chance that you will be on time and, if you take R_3 , there is a 70% chance that you will be on time.

Let T represent on time.

(a) If T represents “on time”, what notation would we use for “late”? (b) What is the value of $\Pr(R_1 \cap R_2)$? (c) Calculate the probability of being on time. (d) *Given that you are on time*, calculate the probabilities of having used each of the routes. (e) *Given that you are late*, what is the probability that you used R_1 ?

Question 5

In your favourite RPG game, let's assume that in selecting your character there are 5 character classes and 2 genders. Let's also assume there are 3 levels of difficulty for this game.

(a) How many possible ways can you play this game? (b) What if you always choose the “warrior” class? (c) What if you always choose a female character? (d) What if you always play on the highest difficulty setting? (e) Let's assume the game has a two-player mode. How many possible ways can you play this game? (hint: you cannot play the game on different difficulty levels). (f) What if your friend chooses a different character class to you?

Question 6

Assume that you are going to an exam and you can only bring 3 items. You have the following items: {mobile phone, pen, ruler, calculator, laptop, apple}.

- (i) In order to make your decision, you first *arrange* these 6 items on your desk. How many possible arrangements are there?
- (ii) How many possible groups of three items can you bring with you?
- (iii) What if you decide that the pen is essential?
- (iv) What if the pen is essential and you also decide that you won't bring an apple or a laptop?

Question 7

A team of 5 people is required to perform a particular task. We are selecting from a group of 7 women and 3 men.

How many selections are there:

(a) Altogether? (b) If one of the men is an expert and must be on the team? (c) If two of the individuals do not get along and cannot be on the team together? (d) If the group must contain 3 women and 2 men? (e) If the group must contain more women than men? (f) If the group must contain more men than women?

3 Tutorial G - Normal Distribution

Question 1

Assume that a character in a game is programmed to have an attack power according to $X \sim \text{Normal}(\mu = 40, \sigma = 3)$.

(a) What is the probability that the attack is greater than 45? (b) What is the probability that the attack is between 32 and 42? (c) Let X_1 and X_2 be the first and second attacks. What is the probability that the *sum* of these two attacks is greater than 85 units? (d) Calculate 99% limits for the sum of two attacks. (e) What is the probability that the *difference* in attacks is more than 5 units? Note that attack 2 can be 5 units more than attack 1 or attack 1 can be 5 units more than attack 2, i.e., $\Pr(|D| > 5) = \Pr(D < -5) + \Pr(D > 5)$.

Question 2

A character in a game deals a standard attack 75% of the time and a critical attack the rest of the time (call these events S and S^c). Given that it is a standard attack, the attack power is $X|S \sim \text{Normal}(\mu = 40, \sigma = 3)$. When the character deals a critical attack, a random fluctuation is added to this according to a $\text{Normal}(\mu = 5, \sigma = 1)$ distribution.

(a) What is the distribution of $X|S^c$? (b) Calculate $\Pr(X < 43|S)$ and $\Pr(X < 43|S^c)$. (c) Calculate $\Pr(X < 43)$. (hint: law of total probability) (d) If the character deals less than 43 damage points, what is the probability that the attack was a critical attack?

Question 3

The income of a technician (in thousands) is $X_1 \sim \text{Normal}(\mu = 30, \sigma = 2)$. The income of an engineer is $X_2 \sim \text{Normal}(\mu = 40, \sigma = 3.5)$.

(a) Calculate the probability that an engineer earns more than a technician. (b) Calculate 90% limits for the difference in their income. (c) For a group of 25 technicians, calculate the probability that the average wage is less than 30500, i.e., $\Pr(\bar{X}_1 < 30.5)$. (d) In a group of 10 engineers, what is the probability that *at least two* of them earn more than 45000? (hint: binomial with $p = \Pr(X_2 > 45)$) (e) For a sample of 30 technicians and 35 engineers, calculate the 80% limits for the difference in their average wages.

Question 4

Let $X \sim \text{Exponential}(\lambda = 0.02)$. Calculate the following:

(a) $\Pr(\bar{X} > 55)$ in a group of 100. (b) $\Pr(\bar{X} < 53)$ in a group of 40. (c) The value of \bar{x} such that $\Pr(\bar{X} > \bar{x}) = 0.1$ when $n = 65$. (d) The value of n if $\Pr(\bar{X} < 49) = 0.1$.

4 KB Tutorial 3

Question 1

Assume that there are three different routes to get to a particular location: R_1 , R_2 and R_3 . You take R_1 75% of the time, R_2 20% of the time and R_3 the rest of the time. If you take R_1 , there is a 90% chance that you will be on time; if you take R_2 , there is a 50% chance that you will be on time and, if you take R_3 , there is a 70% chance that you will be on time.

Let T represent on time.

(a) If T represents “on time”, what notation would we use for “late”? (b) What is the value of $\Pr(R_1 \cap R_2)$? (c) Calculate the probability of being on time. (d) *Given that* you are on time, calculate the probabilities of having used each of the routes. (e) *Given that* you are late, what is the probability that you used R_1 ?

5 KB tutorial 4

Question 1

You develop a random number generator which assigns a value to the random variable X according to the following probability distribution:

x	0.0	0.5	1.0	2.0	3.0
$\Pr(X = x)$	0.4	0.2	0.15	0.15	?

(a) What is value the value of $\Pr(X = 3.0)$? (b) Calculate $E(X)$ and $Sd(X)$. (c) You produce a gambling game where the player wins (in euro) the value of X generated, e.g., if a 2.0 appears, €2 is won. How much should you charge for a play of this game so that that *you* (the programmer) make a profit of €0.10 on average per game? (i.e., the player *loses* €0.10 on average) (d) Using your answer to part (c), determine the probability that you make a profit when somebody plays this game. (e) If 10 people play this game, what is the probability that you make a profit 8 times?

Question 2

You flip three coins. Let X = “the number of heads” and Y = “the number of unique faces”.

(a) What is the sample space for this experiment? (b) Construct the *joint distribution* for X and Y . (c) Based on this joint distribution, construct the *marginal* distribution for X and for Y . (d) Are X and Y independent? (e) Calculate $E(Y)$ and $Sd(Y)$. (f) Calculate $\Pr(Y = 2 | X = 2)$ and interpret its value (compare with $\Pr(Y = 2)$).

Question 3

Let X = “the attack power of player 1” and let Y = “the attack power of player 2”.

Let the probability distributions for X and Y be:

x	0	100	300
$\Pr(X = x)$	0.2	0.75	0.05

y	0	80	200
$\Pr(Y = y)$	0.1	0.6	0.3

(e.g., p1 misses 20% of the time, deals 100 points of damage 75% of the time and performs a critical blow 5% of the time.)

(a) What is the average attack power of each player? (b) If both players have 1000 hit-points, how many attacks does it take for player 1 to defeat player 2 and vice versa? Which player will win on average? (c) Let's now assume that player 1 uses his/her *first* turn to cast a spell (and therefore does not attack in this turn). The result of the spell is that player 2 can no longer perform a critical blow, i.e., $\Pr(Y = 200) = 0$, *from turn two onwards*. Since setting $p(200) = 0$ leads to $\sum p(y) \neq 1$, assume that the remaining probability (= 0.3) is distributed evenly between $p(0)$ and $p(80)$. What is the outcome of the battle now?

Question 4

You flip a coin 10 times - let X = “the number of heads”. Using the binomial probability function, calculate the following:

(a) $\Pr(X = 2)$. (b) $\Pr(X = 0)$. (c) $\Pr(X > 2)$. (d) $\Pr(X \leq 3)$. (e) $\Pr(5 \leq X \leq 7)$. (f) $E(X)$ and $Sd(X)$. (g) Using the binomial tables, calculate $\Pr(X \leq 10)$ in the case where the coin is flipped 20 times. (h) If the coin is flipped 50 times, what is $E(X)$?

Question 5

Repeat Question 4 (a) - (e) but now using the binomial tables.

Question 6

Let's assume that a sequence of bits (binary numbers) is transmitted and, at the other end, decoded; the decoder has a 10% chance reading a bit incorrectly (i.e., reading a 0 as 1 or vice versa). Let X be the number of errors in the sequence received (i.e., the decoded sequence). Calculate the probability that there are:

- (a) No errors in a 20-bit string. (b) Less than three errors in a 10-bit string. (c) More than 10 errors in (i) a 50-bit string and (ii) a 100-bit string (hint: use tables). (d) Calculate the average number of errors in a 100-bit string. Calculate the standard deviation also.

Question 7

We follow on from Question 6 but now consider the case where, to reduce the probability of error, each bit is sent *three* times and then a “majority vote” approach is used to determine the value of each received bit. The following example explains the situation:

Sent	0	1	1	0
	$\underbrace{000}$	$\underbrace{111}$	$\underbrace{111}$	$\underbrace{000}$
Received	$\underbrace{001}$	$\underbrace{111}$	$\underbrace{010}$	$\underbrace{000}$
	0	1	0	0

\Rightarrow there is one error in decoding the first 000, but since the majority result is taken, this bit is correctly identified as a 0. There are two errors in decoding the second 111, so this bit is misread as a 0. It is clear that a character is misread if the decoder makes *two or three errors* in these blocks of three replicates.

- (a) Show that sending each bit 3 times reduces the error probability from 10% to 2.8%.
 (b) Using this reduced value, $p = 0.028$, calculate the probability that there are no errors in a 20-bit string. Compare this result to Q6(a). (c) Now assume that each bit is sent 5 times and, again, the majority vote approach is used. Calculate the probability that there are no errors in a 20-bit string in this case.

6 KB Tutorial 6

Question 4

(Note: this is not a queueing theory question. It is a generalisation of a question which appears on Tutorial2)

There are two possible routes to a particular location. You take R_1 80% of the time and R_2 20% of the time. We assume that travel time has an exponential distribution and, furthermore, the average travel time is 0.25 hours if you take R_1 and 0.5 hours if you take R_2 .

(a) Calculate the probability that the journey takes more than 0.5 hours for each of the routes, i.e., $\Pr(T > 0.5 | R_1)$ and $\Pr(T > 0.5 | R_2)$ respectively. (b) Calculate $\Pr(T > 0.5)$. (hint: law of total probability) (c) Given that $T > 0.5$ hours, what is the probability that you used R_1 ? (i.e., calculate $\Pr(R_1 | T > 0.5)$) (d) Derive a general expression for $\Pr(R_1 | T > t)$ and evaluate it at $t = 0.25$, $t = 1$ and $t = 2$ respectively. Interpret the results.

Question 5

Let $X \sim \text{Normal}(\mu = 10, \sigma = 2)$. Calculate the following:

(a) $\Pr(X > 10)$. (b) $\Pr(X < 3)$. (c) $\Pr(X > 8.4)$. (d) $\Pr(6 < X < 14)$. (e) The value of x such that $\Pr(X > x) = 0.3$. (f) The value of x such that $\Pr(X > x) = 0.8$.

Question 6

Assume that speeds of cars on a motorway have a normal distribution with mean 115km/hr and standard deviation 4km/hr.

(a) Draw a rough sketch of the distribution. (b) $\Pr(X > 120) = ?$ (c) $\Pr(X < 100) = ?$ (d) $\Pr(100 < X < 110) = ?$ (e) 1% of drivers travel above what speed?

Question 7

For *any* normal variable $X \sim \text{Normal}(\mu, \sigma)$:

(a) Show that $\Pr(\mu - 3\sigma < X < \mu + 3\sigma) = 0.997$. (b) Find a value for k such that $\Pr(\mu - k\sigma < X < \mu + k\sigma) = 0.95$. (c) Find k such that $\Pr(\mu - k\sigma < X < \mu + k\sigma) = 0.99$. (d) Show that $\Pr(X > \mu + 1.64\sigma) = 0.05$.