Mathematics Test

Question 1:

An HR officer of local software development company is going through the application for promotion over last year and he found that ‘in the last year, there were 450 female applicants for promotion, of whom 40 were successful and 410 were unsuccessful. There were 760 male applicants for promotion of whom 124 were successful and 636 unsuccessful.

Using the above information:

1. present the information in a suitable tabular format.
2. summarise the table in such a way that comparisons between women and men can be made more easily.
3. comment on the result obtained.

Answer

This is a 2 x 2 contingency table that can be used to compare the success rates during the application processes for promotion between males and females.

	Successful	**Unsuccessful**	**Totals**
Male	124	636	*760*
Female	40	410	*450*
Totals	*164*	*1046*	

Success rates for male candidates were 124/760*100 = 16.3%

Success rates for female candidates were 40/450*100 = 8.8%

Statistical tests that can be performed on these tables include the Fisher’s exact test and Chi-square test. P-values can be two or one-tailed.

Both these tests are used to assess associations between categorical data such as those presented in the contingency table. Fisher’s exact test is more appropriate if we had smaller sample sizes. In this case our sample size is large and therefore we can use Chi-square test. Two-tailed test is recommended here because we are not predicting a direction of the effect.

In this case the Chi-squared equals 13.306 with 1 degrees of freedom.
The two-tailed P value equals 0.0003.
We should reject our null hypothesis.
The association between rows (male vs females) and columns (successful outcome vs unsuccessful outcome) is therefore considered to be extremely statistically significant.

Mathematics Test
Question 2:
A bank is studying the number of times their automatic cash point located in the High Street is used. The following data set details how many times it was used on each of the last 30 days.

83	64	84	76	84	54	75	59	70	61
63	80	84	73	68	52	65	90	52	77
95	36	78	61	59	84	95	47	87	60

Using the above information:
1. produce a frequency distribution for the number of times the cash point was used
2. what was the smallest and largest number of times that the machine was used
3. around what values did the number of times the machine was used tend to cluster?
4. from the distribution, how many times would you say the machine was used on typical day?

Answer

1. First, I categorised the data into intervals and counted the number of occurrences within each interval:

Number if times	Frequency
35-44	0
45-54	3
55-64	5
65-74	6
75-84	9
85-94	7

Next I plotted this on a histogram:

2. Smallest number of times = 36 and largest number of times = 95

3. The most frequent interval was 75-84 times per day. -

4. You can see from the histogram that the data is not normally distributed, therefore a median and IQR are appropriate to examine the centre and variance of the distribution: Median – 71.5; IQR = 24 (83.5-59.5)

Mathematics Test

Question 3:

At a London hospital, there is concern about the high turnover of nurses. A survey was conducted to find out how long (in months) nurses had been in their current positions. The responses of 20 nurses are given below:

23	2	5	14	25	36	27	42	12	8
7	23	29	26	28	11	20	31	8	36

Another survey was done at that hospital to find out how long (in months) health care assistants had been in their current positions. The responses of 20 health care assistants are given below:

25	22	7	24	26	31	18	14	17	20
31	42	6	25	22	3	29	32	15	72

From the above dataset:

1. Calculate the five-figure summary for both sets of data.
2. Does the turnover of nursing staff appear to be different from that of health care assistants?

Answer

1. A five-number summary is the preliminary investigation of a dataset used 1n descriptive statistics. It normally consists of five values: The minimum, first quartile (Q1), median, third quartile (Q3) and the maximum.

The first dataset (nurse responses):

Minimum = 2
First quartile (Q1) = 9.5
Median = 23
Third quartile (Q3) = 28.5
Maximum = 42

The second dataset (healthcare assistants):

Minimum = 3
First quartile (Q1) = 16
Median = 23
Third quartile (Q3) = 30
Maximum = 72

2. Both datasets have similar turnover with the same median value of 23. The IQR in the nursing set (measure of variance) is 19, and the IQR in the

healthcare assistant set is 14. So, the levels of variance are also not too dissimilar, despite the healthcare assistant set having a much higher maximum as an outlier.

Mathematics Test

Question 4:

Review the output produced by R for a t-test for the cholesterol levels among a group of adults undertaking regular exercise and a control group.

Two Sample t -test

Data: Cholesterol level measuremen

t=3.776, df=19, p-value= 0.001278

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.2274728   0.7932545

Sample estimates:

Mean in group control - 5.064000

Mean in group exercise - 4.553636

1. State the null hypothesis and the alternative hypothesis from the R output.
2. By examining the R output, what can you conclude about the differences in the mean control and exercise group?
3. Interpret the 95% confidence interval of the difference.
4. Interpret the findings based on p-value.

Answer

3. The null hypothesis (H0) is that there is no significant difference in the mean cholesterol levels in adults between the two groups, the alternative hypothesis (H1) is that exercise does have an effect on mean cholesterol levels in adults.
4. The level of cholesterol in the adults that were in the exercise group was lower than the cholesterol levels of the adults in the control group.
5. The 95% confidence intervals do not cross ‘0’, the value which indicates no difference in mean. Because the confidence interval provides a range of values that is believed to contain the true value with a certain level of confidence, we can say that there is a statistically significant result with less than 5% chance of the result being due to random variation alone. In this case the true value lies somewhere between (roughly) 0.227 and 0.793 indicating that the exercise really does have an effect at lowering the levels of cholesterol in adults.
6. The p-value is less than 0.05 and this corroborates the above statement about the 95% confidence intervals. There is a less than 5% chance of the result being due to chance/random variation alone. In any case, from both the 95% confidence intervals and P values we can reject the null hypothesis and conclude that exercise does have a real effect on cholesterol levels in adults.

Mathematics Test

Question 5:

Look at the output produced by R for a Paired t-test, for the weight among obese students in a local school before and after 12 weeks a low-calorie diet treatment.

```
Paired t-test
Data: Obese student weight


t=3.0737, df=45, p-value=0.003585
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.1469699      5.6285366
Sample estimates:
Mean of the differences
2.4165
```

1. State the null hypothesis and the alternative hypothesis from the R output.
2. Interpret the 95% confidence interval of the difference.
3. Interpret the findings based on p-value.
4. Explain why this method is appropriate in this case.

Answer

7. The null hypothesis (H0) is that the12 week low calorie diet treatment has no effect on the weight among obese students, i.e. there is no difference in the weight of the students before and after the intervention, and the alternative hypothesis (H1) is that it does have an effect.
8. In this case, the 95% confidence interval indicates that the mean difference is from roughly 0.14 to 5.63 and indicates that we can be 95% confidence that the true value lies between these, therefore it is likely that the intervention led to some weight loss, especially as even the lower limit is greater than zero.
9. The p-value is less than 0.05 indicating a statistically significant difference. Therefore, we can reject the null hypothesis. In combination with the 95% confidence interval, we can conclude that the mean weight loss is about 2.42 kg (units assumed to be kilograms) over the 12-week low calorie diet treatment,
10. This method is appropriate in this case because it is working with the same group of participants before and after an intervention, rather than between two different groups with different exposures (such as the previous example in question 4). The paired t-test in this case is used as there are dependent samples.