SIT191 Problem solving task 3
Due: Friday 26th January 2024

Total marks: 80

Weighting: 30%

Q3.1 Smarties (sugar coated chocolate confectionary) come in 8 colours – green, yellow, red, orange, pink, purple, blue and brown. You buy a bag containing 120 smarties to investigate the distribution of colours, and count 12 green, 14 yellow, 17 red, 15 orange, 16 pink, 17 purple, 11 blue and 18 brown smarties.

a) If smarties are packaged in equal proportions, how many of each colour would you expect in the bag?
b) To see if these results are unusual, should you perform a goodness-of-fit test or a test of independence?
c) State your hypotheses.
d) How many degrees of freedom are there?
e) Find χ^2 and the P-value.
f) State your conclusion (use α = 0.05) in the context of the question.

[1+1+1+1+3+1 = 8 marks]

Q3.2 The following table shows data on randomly selected data on security level of staff and position held. Are security level and the position held in the company independent?

	Very Secure	Secure	Insecure
Non manager	22	55	15
Manager	45	35	15

a) Write appropriate hypotheses.
b) How many degrees of freedom are there?
c) Find χ^2 and the P-value.
d) State your conclusion (use α = 0.05).

[2+1+5+1 = 9 marks]

Q3.3 During an angiogram, heart problems can be examined via a small tube (a catheter) threaded into the heart from a vein in the patient's leg. It is important that the company that manufactures the catheters maintains a diameter of 2.00 mm. A random sample of 36 catheters is taken and the average diameter is 2.03 mm with standard deviation 0.05 mm.

a) Create a 95% confidence interval for the mean diameter of catheters produced by the company.
b) Explain in context what your interval means.
c) Perform a hypothesis test to find out if the mean diameter of the catheters is significantly different to the required 2.00 mm. Use a significance level of 5% and give your conclusion.

[4+2+ 6= 12 marks]

Q3.4 Many drivers believe that they can get better gas mileage by using premium rather than regular gas. To test this we use 10 cars from a company fleet in which all the cars run on regular gas. Each car is filled first with either regular or premium gasoline decided by a coin toss, and the mileage for that tankful is recorded. Then the mileage is recorded again for the same cars for a tankful of the other kind of gasoline. We do not let the drivers know about this experiment. The results are provided in the table below:

Miles per gallon

Car#	1	2	3	4	5	6	7	8	9	10
Regular	16	20	21	22	23	22	24	28	27	28
Premium	18	21	23	24	25	23	26	26	28	29

a) Do the use of premium gasoline differ significantly in the mileage? Carry out a hypothesis test (preferably with technology such as SPSS) using α=0.05 and write your conclusion.
b) Calculate a 90% confidence interval for the difference in mileage for the two different gasoline and interpret the interval.

[5+5 = 10 marks]

Q3.5 A researcher wanted to see whether there is a significant difference in the blood pressure of men and women. He collects data from 23 men and 35 women. The average systolic blood pressure reading for women was 105 with a standard deviation of 12 whereas for the men the average was 115 with a standard deviation of 15. The data are summarised below:

	Mean	Std.Dev.
Men	115	15
Women	105	12

a) Are systolic blood pressure readings for the men significantly higher than those of women? Carry out a hypothesis test. Use a significance level of 5% and use df = n_1+n_2-2.
b) Create a 90% confidence interval for the difference in mean systolic blood pressure readings and interpret the interval.
c) Does the confidence interval confirm your answer to a)? Explain.

[6+5+1 = 12 marks]

Q3.6 A horticulturist was investigating the phosphorus content of tree leaves from three different varieties of apple trees. Random samples of five leaves from each of the three varieties were analysed for phosphorus content. The data are given in the following table:

Variety	mean Phosphorus Content
1	0.460
2	0.776
3	0.708

SPSS output of ANOVA analysis and Bonferroni's multiple comparisons are given below:

ANOVA

PHOSPHOR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.277	2	.138	16.972	.000
Within Groups	.098	12	.008		
Total	.374	14			

Multiple Comparisons

Dependent Variable: PHOSPHOR

Bonferroni

(I) VARIETY	(J) VARIETY	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.3160*	.05710	.000	-.4747	-.1573
	3	-.2480*	.05710	.003	-.4067	-.0893
2	1	.3160*	.05710	.000	.1573	.4747
	3	.0680	.05710	.770	-.0907	.2267
3	1	.2480*	.05710	.003	.0893	.4067
	2	-.0680	.05710	.770	-.2267	.0907

*. The mean difference is significant at the .05 level.

a) Do the phosphorous content differed between any of the varieties. State the appropriate hypotheses for this test.

b) Give the values of the test statistic and P-value, and your conclusion for the test. Use $\alpha = 0.05$.

c) Would it be appropriate to perform a Bonferroni test to see if apple tree varieties show a significant difference in mean phosphorus content of leaves. Explain.

[2+3+3 = 8 marks]

Q3.7 Wild bears were caught and anesthetised so that various measurements could be made. In particular, the usefulness of a bear's chest circumference to predict its weight was of interest. A random sample of 10 bears was used, with the chest and weight measurements shown below, as well as the linear regression analysis.

Bear	Chest (inches)	weight (pounds)
1	26	80
2	45	344
3	54	416
4	49	348
5	35	166
6	41	220
7	41	262
8	49	360
9	38	204
10	31	144

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant	-251.948	33.814		-7.451	.000
	CHEST	12.380	.810	.983	15.277	.000

a Dependent Variable: WEIGHT

a) Identify the explanatory (independent) and response (dependent) variables.
b) State the linear regression equation.
c) Can you interpret the meaning of the *y*-intercept in the context of this study? Explain.
d) Interpret the slope of the regression equation in the context of these variables.
e) Perform the hypothesis test for the slope, and give your conclusion in words, in the context of the variables. Use $\alpha = 0.05$.

[2+2+2+2+5 = 13 marks]

Q3.8

Forensic scientists investigated the thigh temperature of deer carcasses as a possible predictor for the time since death. The thigh temperature and time since death were recorded for 10 deer.

Below are the data for each deer:

temp (°C)	Time since death (hrs)
27	6
38	2
35	3
21	8
24	7
36	1
26	5
32	4
29	4
37	2

Use SPSS or other technology to produce the output necessary to answer the following questions.

a) Produce a scatterplot and use it to describe the apparent relationship between the thigh temperature and time since death.

b) State the regression equation.

c) Is it meaningful to interpret the *y*-intercept for this study? Explain.

d) Predict the time since death if the thigh temperature for a deer carcass is 25 °C. Is this prediction reliable? Explain

[3+2+1+2 = 8 marks]

Marking Scheme:

Questions		Maximum marks
Question 1	Part(a) – Calculation of expected value for each colour (0.125 marks each)	1
	Part(b) – Identifying the type of test (1 mark)	1
	Part(c) – Correct null and alternative hypotheses (0.5 marks each)	1
	Part(d) – Calculation of degrees of freedom (1 mark)	1
	Part(e) –Correct Chi-square statistic value (2 marks) and correct P-value (1 mark)	3
	Part(f) – Reasonable comparison of P-value and significance level (0.5 marks) and correct conclusion (0.5 marks)	1

Question 2	Part(a) – Correct null and alternative hypotheses (1 mark each)	2
	Part(b) – Calculation of degrees of freedom (1 mark)	1
	Part(c) – Correct expected values (2 marks), correct Chi-square statistic value (2 marks) and correct P-value (1 mark)	5
	Part(d) – Reasonable comparison of P-value and significance level (0.5 marks) and correct conclusion (0.5 marks)	1
Question 3	Part(a) –correct formula (1 mark) and correct intervals (1.5 marks each)	4
	Part(b) – Reasonable interpretation of confidence intervals in context to the question (2 marks)	2
	Part(c) – Correct null and alternative hypotheses (0.5 marks each), correct formula (1 mark), correct test statistic value (1 mark), correct P-value (1 mark), Reasonable comparison of P-value and significance level (1 mark) and correct conclusion (1 mark)	6
Question 4	Part(a) – Correct null and alternative hypotheses (0.5 marks each), correct formula (1 mark), correct test statistic value (1 mark), correct P-value (1 mark), Reasonable comparison of P-value and significance level (0.5 marks) and correct conclusion (0.5 marks)	5
	Part(b) – correct formula (1 mark), correct intervals (1.5 marks each) and reasonable interpretation of confidence intervals in context to the question (1 mark)	5
Question 5	Part(a) – Correct null and alternative hypotheses (0.5 marks each), correct formula (1 mark), correct test statistic value (1 mark), correct P-value (1 mark), Reasonable comparison of P-value and significance level (1 mark) and correct conclusion (1 mark)	6
	Part(b) – Correct formula (1 mark), correct intervals (1.5 marks each) and reasonable interpretation of confidence intervals in context to the question (1 mark)	5
	Part(c) – Reasonable comparison of part(a) and (b) conclusion with respect to null hypotheses (1 mark)	1
Question 6	Part(a) – Correct null and alternative hypotheses (1 mark each)	2
	Part(b) – Correct test statistic value (1 mark), correct P-value (1 mark), Reasonable comparison of P-value and significance level (0.5 marks) and correct conclusion (0.5 marks)	3

	Part(c) – Reasonable interpretation of Bonferroni test (1 mark) and comparison of different groups (2 marks)	3
Question 7	Part(a) – Identifying the type of variables (1 mark each)	2
	Part(b) – Correct equation (2 marks)	2
	Part(c) – Reasonable interpretation of *y*-intercept (2 marks)	2
	Part(d) – Reasonable interpretation of slope (2 marks)	2
	Part(e) – Correct null and alternative hypotheses (0.5 marks each), correct test statistic value (1 mark), correct P-value (1 mark), Reasonable comparison of P-value and significance level (1 mark) and correct conclusion (1 mark)	5
Question 8	Part(a) – plotting correct scatterplot (2 marks) and identifying the type of linear relationship (1 mark)	3
	Part(b) – Correct equation (2 marks)	2
	Part(c) – Reasonable interpretation of *y*-intercept (1 mark)	1
	Part(d) – Correct calculation (1 mark) and relevant explanation (1 mark)	2