Medical Statistics (MATH38071) Exercise Sheet 2
(Basic Analyses of Continuous Outcome Measures)

1. A randomised controlled trial was conducted comparing two treatments for pain relief during the recovery period following wisdom tooth extraction surgery. A new dual-therapy (Acetaminophen + Ibuprofen) was compared against a mono-therapy (Ibuprofen alone). Post-operative pain was measured at 15 minute intervals using a 100mm visual analogue scale with "100" representing the worst imaginable pain and "0" representing no pain. The table below gives the summary statistics for each follow-up time point.

a. Calculate the point estimate of the treatment effect of acetaminophen + ibuprofen as compared to ibuprofen at that time point.

Answer

(i) The table below summarizes the results parts (a),(b), (c) & (d)

Data provided							Answers							
	Ibuprofen			Acetaminophen + Ibuprofen								95%	c.i.		Reject
Time (mins)	Mean (mm)	S.D. (mm)	N	Mean (mm)	S.D. (mm)	N	Pooled diff	S.D.	S.E.	df	$t_{0.975,df}$	low	upper	T	H_0
15	27.9	14.8	24	18.2	13.1	24	-9.7	13.98	4.03	46	2.0129	-17.821	-1.579	-2.404	Yes
30	32.6	24.4	25	25.3	20.9	25	-7.3	22.72	6.43	48	2.0106	-20.219	5.619	-1.136	No
45	35.5	23.2	22	28.7	23.3	20	-6.8	23.25	7.18	40	2.0211	-21.316	7.716	-0.947	No
60	31.3	18.9	19	25.1	22.8	23	-6.2	21.13	6.55	40	2.0211	-19.442	7.042	-0.946	No
75	29.9	18.8	24	14.9	13.8	24	-15	16.49	4.76	46	2.0129	-24.582	-5.418	-3.151	Yes
90	23.8	17.9	22	15	14.2	24	-8.8	16.07	4.74	44	2.0154	-18.361	0.761	-1.855	No
105	22.7	16.4	21	13.7	12.8	19	-9	14.80	4.69	38	2.0244	-18.489	0.489	-1.920	No
120	20.9	17.2	24	15.2	14.4	23	-5.7	15.89	4.64	45	2.0141	-15.040	3.640	-1.229	No
AUC	27.9	13.6	25	19.5	12.3	26	-8.4	12.95	3.63	49	2.0096	-15.691	-1.109	-2.315	Yes

S.D.=standard deviation, T= test statistic, A.U.C.= Area under the curve.

(e) The assumptions required for a t-test analysis to be plausible are (i) the subjects are independent and identically distributed, (ii) the data are normally distributed within each group, and (iii) the populations standard deviations should be the same. The data cannot be strictly normal as the visual analogue scale has a minimum and a maximum whereas the normal distribution is unbounded. There is some evidence also of skewness as the mean - 2 s.d. is negative in all cases except the AUC. In practice the t-test is quite robust in sample of this size to departures from normality.

(ii) Area under the curve (A.U.C) is a summary measure that was calculated for all patients. It is a weighted average of the time-point pain scores. Repeat the calculation of question 1 for A.U.C.

Answer

See table

(iii) The p-values for the nine possible analyses computed using Excel are

Time (mins)	15	30	45	60	75	90	105	120	A.U.C.
p-value	0.0203	0.2616	0.3495	0.3497	0.0029	0.0703	0.0624	0.2254	0.0248

Why might it be inadvisable to carry out all 9 analyses? What type of bias might this cause?

Answer

It would be inadvisable to carry out all 9 analysis as this will increase the probability of obtaining one or more statistically significant result where the null hypothesis holds, that is a type I error. It would be better to carry out a single analysis using data from all 8 time points to test the specific hypothesis of interest.

(iv) Which analysis do you feel gives the best estimate of the treatment effect?

Answer

If the hypothesis of interest was "does the addition of Acetaminophen to standard the medication Ibuprofen reduce pain over the follow-up period" then the analysis using the A.U.C is the best of the 9 analyses, as it is a summarizes the average pain experience over the follow-up period. What is more it's distribution will be closer to a normal distribution, being an average of other measures.

Note that with 8 data points for each person there are other statistical analyses that could address different hypothesis. For example the investigator might be concerned with (i) the maximum pain experienced of patients or (ii) the rate at which pain reduced during the two hour recovery period. A hypothesis related to (i) could be tested by calculating the maximum of the eight measurements for each patient then apply a t-test to these maxima. A hypothesis related to (ii) could be tested by calculating the rate of change of pain for each measurement using a linear regression on each patient's data to calculate the gradient, before comparing these gradients for each patient, again with a t-test.