1

The data below give the concentrations (mM) of thiol in the lysate of the blood of two groups of volunteers, the first group being normal and the second group being sufferers from rheumatoid arthritis.

Thiol concentration (mM)

Normal	Rheumatoid
1.84	1.81
2.32	3.06
1.94	2.62
1.92	3.66
1.65	3.07
1.91	2.76
2.07	1.58

The data were analysed using a t-test in Minitab. The output obtained appears in the box below.

Two Sample T-Test and Confidence Interval

```
Two sample T for Normal vs Rheuma

         N      Mean      StDev   SE Mean
Normal   7     1.950      0.207     0.078
Rheuma   7     2.651      0.733      0.28

95% CI for mu Normal - mu Rheuma: ( -1.329,  -0.07)
T-Test mu Normal = mu Rheuma (vs not =): T = -2.44  P = 0.031  DF = 12
Both use Pooled StDev = 0.539
```

(a) Which type of t-test was used? Explain why this was chosen and the assumptions underlying the test.

(b) Write a brief interpretation of the parts of the output in italics/bold type.

(c) What conclusion can be drawn from the output about whether or not the mean thiol concentration of the two groups differ?

Two methods of determining the percentage fibre content of the raw materials used in an oat-based breakfast cereal were compared. Nine samples of the cereal were used. Each was divided into two parts, one part analysed by the old method and one by the new. The results were as follows:

% fibre	Sample								
	1	2	3	4	5	6	7	8	9
New Method	12.7	13.6	11.6	12.7	14.2	14	11.8	12.5	13.2
Old method	12.2	12.7	11.6	12.3	13.9	13.8	11.9	12.0	13.4

A paired t-test and an unpaired t-test have been carried out using MINITAB and the analyses are given below.

(i) Which analysis is correct for the above data? Explain your reasoning.

(ii) What assumptions have to be made about the data for your chosen test to be valid?

(iii) Explain what null and alternative hypotheses have been tested.

(iv) What conclusion can be drawn from the test? Your answer should be one sentence suitable for a scientific report.

```
Paired T- Test and Confidence Interval

Paired T for new - old

                N      Mean     StDev    SE Mean
new             9    12.922     0.909     0.303
old             9    12.644     0.856     0.285
Difference      9     0.278     0.346     0.115

95% CI for mean difference: (0.012, 0.543)
T-Test of mean difference = 0 (vs not = 0): T-Value = 2.41  P-Value = 0.042


Two Sample T-Test and Confidence Interval

Two sample T for new vs old

           N      Mean     StDev   SE Mean
new        9    12.922     0.909      0.30
old        9    12.644     0.856      0.29

95% CI for mu new - mu old: ( -0.61,  1.17)
T-Test mu new = mu old (vs not =): T = 0.67  P = 0.51  DF = 15
```

3

Two assay methods for measuring the level of vitamin B12 in healthy red blood cells were compared. Ideally they should give the same results. Blood was taken from 10 healthy adults and the level of B12 was found to be as follows:

Subject	1	2	3	4	5	6	7	8	9	10
Method 1	224	238	209	277	197	227	207	205	131	282
Method 2	204	238	198	253	180	209	217	204	137	250

The methods have been compared using both an unpaired t-test and a paired t-test. The MINITAB analysis is given below.

(i) Explain which is the correct analysis giving your reasoning.

(ii) State the null and alternative hypotheses for the correct test.

(iii) What statistical conclusion can be drawn?

(iv) Explain the implications of this in the context of the question.

Two Sample T-Test and Confidence Interval

```
Two sample T for Method 1 vs Method 2

           N      Mean     StDev   SE Mean
Method 1  10     219.7      42.7        14
Method 2  10     209.0      34.5        11

95% CI for mu Method 1 - mu Method 2: ( -26,  47)
T-Test mu Method 1 = mu Method 2 (vs not =): T = 0.62  P = 0.55  DF = 17
```

Paired T-Test and Confidence Interval

```
Paired T for Method 1 - Method 2

                  N      Mean     StDev   SE Mean
Method 1         10     219.7      42.7      13.5
Method 2         10     209.0      34.5      10.9
Difference       10     10.70     13.85      4.38

95% CI for mean difference: (0.79, 20.61)
T-Test of mean difference = 0 (vs not = 0): T-Value = 2.44  P-Value = 0.037
```

4

Two media are to be compared to determine which one is the better for detecting the presence of *Salmonella.* Medium 2 is the current standard; medium 1 is a new medium which is said to be more effective in detecting *Salmonella*. Eight variants of *Salmonella* were chosen, labelled A to H here. (It is important the medium chosen is effective over the whole spectrum of *Salmonella* variants.) Each variant was grown on both media. The logarithms of the counts were as follows:

	Log (count)	
strain	**Medium 1**	**Medium 2**
A	2.598	2.365
B	3.120	3.320
C	4.325	4.120
D	2.890	2.780
E	3.640	2.904
F	2.630	2.156
G	4.450	4.320
H	3.780	3.652

The data have been analysed using both a paired t-test and an unpaired t-test. The MINITAB printout is given below.

a) Which is the correct analysis? Explain your choice clearly.

b) For your chosen test state the null and alternative hypotheses which have been used in the MINITAB printout.

c) State the conclusion from the test in statistical terms and in a manner suitable for a scientific report.

Paired T-Test and CI: Medium 1, Medium 2

```
Paired T for Medium 1 - Medium 2
Log Count
                  N      Mean     StDev   SE Mean
Medium 1          8     3.429     0.730     0.258
Medium 2          8     3.202     0.790     0.279
Difference        8    0.2270    0.2767    0.0978

95% lower bound for mean difference: 0.0417
T-Test of mean difference = 0 (vs > 0): T-Value = 2.32  P-Value = 0.027
```

Two-Sample T-Test and CI: Medium 1, Medium 2

```
Log(count)
Two-sample T for Medium 1 vs Medium 2

          N      Mean     StDev   SE Mean
Medium 1  8     3.429     0.730      0.26
Medium 2  8     3.202     0.790      0.28

Difference = mu Medium 1 - mu Medium 2
Estimate for difference:  0.227
95% lower bound for difference: -0.443
T-Test of difference = 0 (vs >): T-Value = 0.60  P-Value = 0.280  DF = 14
Both use Pooled StDev = 0.761
```

5

A botanist hopes to demonstrate that the plants of the species *Thlaspi alpestre* found in Derbyshire are a different subspecies from the plants found in North Yorkshire. Both are virtually confined to lead mining spoil heaps. The claimed morphological difference is in the height of the plants, those in Derbyshire being taller. The heights (cm) of 8 flowering stems from each area were measured and were as follows:

Plant	1	2	3	4	5	6	7	8
N. Yorks	14.6	15.2	17.1	10.4	11.6	13.8	14.2	16.1
Derbyshire	17.3	12.4	23.6	18.0	19.6	16.2	17.3	13.4

Using Minitab, the data has been analysed to compare heights as both a paired and an unpaired design and the analyses are given below.

a) Explain whether these data should be analysed as a paired or unpaired experiment

b) What are the conditions required for
   i. an unpaired t-test
   ii. a paired t-test?

c) State, in words, the null and alternative hypotheses for the appropriate analysis.

d) What conclusion should be drawn from the analysis?

Paired T-Test and CI: NYorks, Derbys

```
Paired T for NYorks - Derbys

             N      Mean    StDev  SE Mean
NYorks       8   14.1250   2.2199   0.7848
Derbys       8   17.2250   3.5021   1.2382
Difference   8  -3.10000  4.22442  1.49356


95% CI for mean difference: (-6.63170, 0.43170)
T-Test of mean difference = 0 (vs not = 0): T-Value = -2.08  P-
Value = 0.077
```

Two-Sample T-Test and CI: NYorks, Derbys

```
Two-sample T for NYorks vs Derbys

         N   Mean  StDev  SE Mean
NYorks  8  14.13   2.22     0.78
Derbys  8  17.23   3.50      1.2


Difference = mu (NYorks) - mu (Derbys)
Estimate for difference:  -3.10000
95% upper bound for difference:  -0.51795
T-Test of difference = 0 (vs <): T-Value = -2.11  P-Value = 0.026
DF = 14
Both use Pooled StDev = 2.9320
```

6

It is claimed that giving children high levels of omega 3 fish oils aids concentration. A sample of 10 children was tested to measure the length of time they could concentrate on doing a jigsaw puzzle. They were then given a fish oil supplement for two months and were tested again. Their concentration times were as follows:

Time (min)										
Child	1	2	3	4	5	6	7	8	9	10
Before	10	5	6	11	8	2	1	8	5	3
After	13	4	8	12	10	3	2	9	5	3

MINITAB output is given below for an analysis of this experiment as a paired t-test and as an unpaired t-test.

a) Which is the correct analysis? Explain your reasoning.

b) For your chosen test state the null and alternative hypotheses and any necessary assumptions about the data.

c) State your conclusion in one sentence in a form suitable for a report.

Paired T-Test and CI: before, after

```
Paired T for before - after

               N      Mean     StDev  SE Mean
before        10   5.90000  3.34830  1.05883
after         10   6.90000  4.01248  1.26886
Difference    10  -1.00000  1.15470  0.36515


95% upper bound for mean difference: -0.33064
T-Test of mean difference = 0 (vs < 0): T-Value = -2.74  P-Value =
0.011
```

Two-Sample T-Test and CI: before, after

```
Two-sample T for before vs after

         N  Mean  StDev  SE Mean
before  10  5.90   3.35      1.1
after   10  6.90   4.01      1.3


Difference = mu (before) - mu (after)
Estimate for difference:  -1.000000
95% upper bound for difference:  1.865726
T-Test of difference = 0 (vs <): T-Value = -0.61  P-Value = 0.276
DF = 18
Both use Pooled StDev = 3.6953
```

7

An experiment was conducted to study the effect of tamoxifen on patients with cervical cancer. One of the measurements made, both before and after tamoxifen was given, was microvessel density (MVD). MVD, *number of vessels per* mm^2 is a measurement that relates to the formation of blood vessels that feed a tumour and allow it to grow and spread. Thus, small values of MVD are better than large values.
Data for 18 patients are shown.

Patient	MVD before	MVD after	Patient	MVD before	MVD after
1	98	75	**10**	70	60
2	100	60	**11**	60	65
3	82	25	**12**	88	45
4	100	55	**13**	45	36
5	93	78	**14**	159	144
6	119	102	**15**	65	27
7	70	58	**16**	98	90
8	78	70	**17**	66	16
9	104	90	**18**	67	53

MINITAB output is given in the table below for an analysis of this experiment as a paired t-test and as an unpaired t-test.

Paired T-Test and CI: MVD before, MVD after

```
Paired T for MVD before - MVD after

              N   Mean  StDev  SE Mean
MVD before  18  86.78  26.14     6.16
MVD after   18  63.83  30.83     7.27
Difference  18  22.94  17.71     4.17


95% lower bound for mean difference: 15.68
T-Test of mean difference = 0 (vs > 0): T-Value = 5.50  P-Value = 0.000
```

Two-Sample T-Test and CI: MVD before, MVD after

```
Two-sample T for MVD before vs MVD after

              N  Mean  StDev  SE Mean
MVD before  18  86.8   26.1      6.2
MVD after   18  63.8   30.8      7.3


Difference = mu (MVD before) - mu (MVD after)
Estimate for difference:  22.94
95% lower bound for difference:  6.84
T-Test of difference = 0 (vs >): T-Value = 2.41  P-Value = 0.011  DF = 34
Both use Pooled StDev = 28.5810
```

a) Which is the correct analysis? Explain your reasoning.
b) For your chosen test state the null and alternative hypotheses and any necessary assumptions about the data.
c) State your conclusion in one sentence in a form suitable for a report.

8

A UWE student measured the biomass (milligrams) produced by bacterium A and bacterium B in shake flasks containing glucose as substrate.

Bacterium A	410	447	448	449	454	467	482	507
Bacterium B	511	480	466	460	459	455	435	416

a) Explain which of the two tests in the Minitab output below is using the correct analysis approach.

b) State the hypotheses which are being tested in your chosen analysis.

c) What conclusion can be drawn from the data about her hypothesis?

d) State the assumptions needed for the test to be valid.

Paired T-Test and CI: BacteriumA, BacteriumB

```
Paired T for BacteriumA - BacteriumB

            N   Mean  StDev  SE Mean
BacteriumA  8  458.0   28.5     10.1
BacteriumB  8  460.3   28.3     10.0
Difference  8   -2.3   56.5     20.0


95% CI for mean difference: (-49.5, 45.0)
T-Test of mean difference = 0 (vs not = 0): T-Value = -0.11
P-Value = 0.913
```

Two-Sample T-Test and CI: BacteriumA, BacteriumB

```
Two-sample T for BacteriumA vs BacteriumB

                             SE
            N   Mean  StDev  Mean
BacteriumA  8  458.0   28.5    10
BacteriumB  8  460.3   28.3    10


Difference = mu (BacteriumA) - mu (BacteriumB)
Estimate for difference:  -2.2
95% CI for difference:  (-32.7, 28.2)
T-Test of difference = 0 (vs not =): T-Value = -0.16
P-Value = 0.876  DF = 14
Both use Pooled StDev = 28.4147
```

9

An overdose of paracetamol can cause fatal hepatic necrosis. Rapid and accurate determination of plasma paracetamol is necessary for the administration of antidotal treatment.
To measure paracetamol concentrations, a standard High Performance Liquid Chromatograph (HPLC) method can be used in the laboratory. An alternative commercial method (PCP) has been proposed for use in Accident and Emergency departments. To measure the efficacy of PCP, samples of blood were taken from 10 suspected overdose patients. For those 10 samples the plasma concentrations (mg/l) were measured using HPLC and PCP. The measurements and two analyses of these data using MINITAB are shown below below.

Conc. (mg/l)

HPLC	**PCP**
0	2
48	53
36	46
3	7
3	3
5	6
12	14
53	50
65	61
98	107

```
Two-Sample T-Test and CI: HPLC, PCP
Two-sample T for HPLC vs PCP
       N  Mean  StDev  SE Mean
HPLC  10  32.3   33.3    11
PCP   10  34.9   34.5    11

Difference = mu (HPLC) - mu (PCP)
Estimate for difference:  -2.6
95% CI for difference:  (-35, 29)
T-Test of difference = 0 (vs not =): T-Value = -0.17  P-Value = 0.866  DF =
17


Paired T-Test and CI: HPLC, PCP
Paired T for HPLC - PCP
             N   Mean  StDev  SE Mean
HPLC        10   32.3   33.3     10.5
PCP         10   34.9   34.5     10.9
Difference  10  -2.60   4.58     1.45

95% CI for mean difference: (-5.87, 0.67)
T-Test of mean difference = 0 (vs not = 0): T-Value = -1.80  P-Value =
0.106
```

a) Explain which of the two tests above is the correct analysis giving your reasoning.

b) Using the appropriate output given above,

(i) State the hypotheses which are being tested in your chosen analysis.
(ii) Give a 95% confidence interval for the difference between the mean paracetamol concentrations from the two methods and explain what it means. What statistical conclusion can be drawn?
(iii) Explain the implications of this in the context of the question.
(iv) State the assumptions needed for the test to be valid.

10

In an investigation of possible brain damage due to alcoholism, an X-ray procedure known as a computerized tomography (CT) scan was used to measure brain densities of 11 chronic alcoholics with a view to determining whether alcoholism reduces brain density.

For each alcoholic, a non-alcoholic control was selected who matched the alcoholic on age, sex, education, and other factors. Their brain densities were as follows:

Alcoholics	**Control**
40.1	41.3
38.5	40.2
36.9	37.4
41.4	46.1
40.6	43.9
42.3	41.9
37.2	39.9
38.6	40.4
38.5	38.6
38.4	38.1
38.1	39.5

Two t-tests, paired and unpaired, for the above data using MINITAB are given below.

Two-Sample T-Test and Confidence Interval:

```
Two-sample T for Alcoholics vs Control

             N   Mean   StDev    SE Mean
Alcoholics   11  39.15   1.72     0.52
Control      11  40.66   2.56     0.77

Difference = mu (Alcoholics) - mu (Control)
95% CI for difference:  (-3.460, 0.424)
T-Test of difference = 0 (vs <): T-Value = -1.63  P-Value = 0.059  DF = 20
Both use Pooled StDev = 2.1833
```

Paired T-Test and Confidence Interval:

```
Paired T for Alcoholics - Control

             N    Mean  StDev  SE Mean
Alcoholics  11  39.145  1.722    0.519
Control     11  40.664  2.563    0.773
Difference  11  -1.518  1.580    0.476

95% CI for mean difference: (-2.579, -0.457)
T-Test of mean difference = 0 (vs < 0): T-Value = -3.19  P-Value = 0.005
```

a) Which is the correct analysis? Explain your reasoning.*Note that matched-controls are used here. Why?*

b) For your chosen test state the null and alternative hypotheses and any necessary assumptions about the data.

c) State your conclusion in one sentence in a form suitable for a report.