Statistics Paper 1

1. A gardener wanted to investigate whether sunflowers planted in her garden would grow to a height of 4 metres as claimed on the seed packet. She collected data on the heights, in metres, of a sample of 15 sunflowers grown in her garden. The results are summarised as follows:

   N = 15, sample mean $\bar{x}$ = 3.60, sample standard deviation s = 1.18

   Calculate a 95% confidence interval for the population mean of sunflowers grown in her garden

2. Suppose that the blood plasma nicotine levels in smokers may be modelled by a random variable T which is normally distributed with mean 315 and standard deviation 131.

   a) According to the model, what proportion of smokers have nicotine levels above 450?
   b) What nicotine level is such that the nicotine levels of only 4% of smokers are higher?

3. A man fired a gun at four targets. He took three shots to hit the first target, one shot to hit the second target, and two shots for each of the other two targets, suppose that the number of shots he takes to hit a target can be modelled by a geometric distribution with parameter θ. Given these data, find an expression for L(θ), the likelihood of θ.

4. When testing data relating to the regression of abrasion loss on tensile strength, the following output is obtained:

Source	d.f	s.s	m.s	v.r	F pr.
Regression	1	20035	20035	2.74	0.109
Residual	28	204977	7321		
Total	29	225011	7759		

Percentage variance accounted for 5.7

Standard error of observations is estimated to be 85.6

Message: The following units have large standardised residuals.

Unit	Response	Residual
1	372	2.19

What does this output tell you about the regression of abrasion loss on tensile strength?

5. Data was analysed to assess the strengths of plywood produced with six different glues. An ANOVA test was performed to test the hypothesis that all treatment means are equal. The output generated as below:

Source of variation	d.f	s.s	m.s	v.r	F pr.
glue	5	115280.4	23056.1	35.08	<.001
Residual	54	35486.9	657.2		
Total	59	150767.2			

What can you conclude from the output? Are the treatment means equal? Justify your answer

6. The below table shows yields of apples before (x, in bushels) and during (y, in pounds) an experiment on ground-cover treatments

Treatment	Blocks							
	I		II		III		IV	
	x	y	x	y	x	y	x	y
O	7.6	222	10.1	301	9	238	10.5	357
A	8.2	287	9.4	290	7.7	254	8.5	307
B	8.2	271	6	209	9.1	243	10.1	348
C	6.8	234	7	210	9.7	286	9.9	371
D	5.7	189	5.5	205	10.2	312	10.3	375
E	6.1	210	7	276	8.7	279	8.1	344

a) Identify the response variable and the explanatory variables. In each case, say whether the variable is continuous or categorical.
b) What model might it be appropriate to start with for your analysis of these data?

7. This question relates to an investigation of the effect of vitamin D supplementation for the prevention of low levels of calcium in new-born babies. The data in the below table came from a clinical trial in which a sample of babies who were breast-fed were compared with a sample of babies who were bottle-fed: the measured quantity was the level of calcium in the baby's blood ('serum calcium') at 1 week of age.

	Sample size	Sample Mean	Sample standard deviation
Breast-fed	64	2.45	0.292
Bottle-fed	169	2.3	0.274

Carry out a two-sample z-test to investigate whether the mean serum calcium level of week-old babies was the same whether they were breast-fed or bottle-fed.