PRIFYSGOL ABERYSTWYTH UNIVERSITY

ADRAN MATHEMATEG / DEPARTMENT OF MATHEMATICS

ARHOLIADAU SEMESTER 2 / SEMESTER 2 EXAMINATIONS

MAI / MAY 2022

MA26620 - Applied Statistics

The questions on this paper are written in English.

If you have questions about the paper during the exam, contact the module co-ordinator, Dr Adam Vellender, on asv2@aber.ac.uk.

You should write out solutions to the paper and upload them to Blackboard as a single PDF file.

Amser a ganiateir - 3 awr
Mae'n rhaid cyflwyno eich atebion erbyn 17:00 (amser y DU).

- Gellir rhoi cynnig ar bob cwestiwn.
- Rhoddir mwy o ystyriaeth i berfformiad yn rhan B wrth bennu marc dosbarth cyntaf.
- Mae tablau ystadegol ar gael ar Blackboard.
- Mae modd i fyfyrwyr gyflwyno atebion i'r papur hwn naill ai yn y Gymraeg neu'r Saesneg.

Time allowed - 3 hours
Submission must be completed by 17:00 (UK time).

- All questions may be attempted.
- Performance in section B will be given greater consideration in assigning a first class mark.
- Statistical tables are available on Blackboard.
- Students may submit answers to this paper in either Welsh or English.

Section A

1. A location tracking device is fitted to each of 24 sea-ducks, an equal number of each of four different species: eider, goldeneye, common scoter and velvet scoter. For each duck, the complete distance travelled in hundreds of km over the first three days of migration is recorded, leading to the following (incomplete) ANOVA table:

Source	SS	DF	MS	F-ratio
				3
Within species				
Total (corr.)	58			

(a) Copy and complete the above ANOVA table. [10 marks]

(b) State hypotheses for the one-way ANOVA test. Conducting the test at the 5% significance level, would you believe that the distance travelled over the first three days of migration was the same for each species? Give reasons. [5 marks]

2. An animal behaviourist conducts an experiment concerning the scavenging behaviour of mice. A mouse enters a chamber through a door. Small piles of food (labelled A, B, C, D, and E) are placed in a straight line at respective distances of 10cm, 20cm, 30cm, 40cm, and 50cm from the door, with respective masses of 5g, 10g, 15g, 20g, and 25g.

The behaviourist will observe which pile is first visited by the mouse. She hypothesises pile A is twice as likely as B, B is twice as likely as C, C is twice as likely as D, and D is twice as likely as E.

She then conducts this experiment 200 times (using a different mouse each time) and records the following frequencies:

A	B	C	D	E
127	42	20	8	3

Clearly stating any assumptions made, conduct an appropriate statistical test at the 5% significance level to determine whether the behaviourist's hypothesis is supported by the data. [10 marks]

3. Wild members of the critically endangered marsupial species Gilbert's potoroo (*Petorous gilbertii*), whose diet consists mainly of multiple types of truffle-like fungi, have an average weight of 1050g. Conservationists aiming to replicate this diet in captive breeding programs raise seven baby Gilbert's potoroos on a diet of truffles that is less varied than that found in the creatures' natural habitat. They are concerned that the reduced variety in diet may lead to a reduction in the animals' weights.

Upon reaching adulthood, the seven captive-bred Gilbert's potoroos have a mean weight of 970g, with sample variance 23,000g^2.

Clearly stating your hypotheses and any assumptions you make, test whether the Gilbert's potoroos raised on the reduced-variety diet have a significantly lower weight than their wild counterparts. [10 marks]

4. Quails' eggs are collected and their diameters measured, along with the weight of the quail that laid them. The data from six such eggs are given below:

x: Weight of laying quail (g)	130	120	100	105	135	135
y: Diameter of egg (mm)	30	27	22	24	28	31

(a) Showing your working, calculate the equation of the least squares regression line of y on x. [9 marks]

(b) Showing your working, calculate the value of R^2 and give an interpretation of what it tells you about the data. [3 marks]

(c) Estimate the egg diameter from quails that weigh:

(i) 125g;

(ii) 200g. [3 marks]

(d) Quails in one barn are on average 20g heavier than those in another. 100 eggs are gathered from each barn. What is the expected difference between the mean egg diameters from the two barns? [2 marks]

5. Usually, herring gulls attempt to steal food from students leaving Aberystwyth's student union building at a rate of three per hour over the lunchtime period. Recently, student union staff have become concerned that this rate is increasing. Over several days' lunchtimes, they observe the main entrance for a total of 10 hours, during which they witness 42 gull attacks.

Clearly stating the meaning of any notation introduced, along with any assumptions and hypotheses used, test whether the rate of gull attacks has significantly increased. [9 marks]

6. On a large sheep farm, a vet scans 70 pregnant ewes before lambing season. All sheep scanned are either expecting a single lamb, or twins, and a total of 126 foetuses are detected. Clearly stating any assumptions made, calculate a 90% confidence interval for the proportion of ewes expecting twins. [9 marks]

(Section B begins overleaf)

Section B

7. Kangaroos of two different species, red kangaroos (*Osphranter rufus*) and eastern grey kangaroos (*Macropus giganteus*), are observed hopping. Their maximum speed is recorded and is summarised below:

	Red kangaroos	Eastern grey kangaroos
Number observed	22	17
Mean of maximum speeds (km/h)	40	35
Standard deviation of maximum speeds (km/h)	7.0	6.0

Showing your working and assuming equal variances for both samples, conduct a two-sample t-test to assess whether red kangaroos are significantly faster than eastern grey kangaroos. [11 marks]

8. King cobras kill their prey via injection with deadly venom. A toxicologist wishes to construct a 95% confidence interval of the amount of venom delivered per bite, measured in milligrams (mg). The amount of venom delivered per bite is believed to be distributed Normally, but the variance is not known.

A small pilot study, using a dummy arm that five king cobras bite, has sample standard deviation $S = 15$mg of venom. The toxicologist wishes to conduct a larger data collection exercise in order to reduce the width of the resulting confidence interval to at most 5mg. Based on the data from this pilot study, what is the minimum number of king cobra bites that her sample should contain in order to achieve this aim? [12 marks]

9. An ornithologist is interested in the proportion of nests of dunnocks (a small garden bird species) which contain a cuckoo chick. He finds the results of a study from 1991, which surveyed a large number of dunnock nests and concluded that 6% of them contained cuckoo chicks.

He and his team conducted a similar survey in the same geographical area in spring 2021, visiting 350 nests, of which 10 contained a cuckoo chick.

Test, making use of a suitable approximation, whether the probability of a dunnock nest containing a cuckoo chick has significantly declined. [10 marks]

(Paper continues overleaf)

10. Skye terriers (ST), West Highland white terriers (WHWT), border collies (BC), and rough collies (RC) compete in a dog agility competition. For each breed, an equal number of dogs completes an agility course. The time taken to do so, in seconds, is recorded. Some R output is presented below:

```
> tapply(Time,Breed,mean)
      BC       RC       ST     WHWT
48.00000 42.71429 56.14286 53.71429
> summary(aov(Time~Breed))
            Df Sum Sq Mean Sq F value Pr(>F)
Breed        3  759.7  253.24   3.818 0.0228 *
Residuals   24 1591.7   66.32
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

(a) How many dogs of each breed competed in the competition? Which part of the output informs your answer? [2 marks]

(b) Is there a significant time difference between breeds? Which part of the output informs your answer? [2 marks]

(c) Fully justifying your answer, construct suitable contrasts to determine whether there is there a significant time difference between:

(i) the two terrier breeds;

(ii) the two collie breeds;

(iii) terriers and collies. [13 marks]

You may find the supplementary table on the next page helpful.

END OF QUESTIONS

Supplementary material follows

Supplementary material

The following extends the first table on page 15 of the statistical tables.

The table shows the upper 5% points of the F-distribution. For example, for the F-distribution on 3 and 25 degrees of freedom (abbreviated below as d.o.f.), $0.05 = P(F > 2.9912)$.

Denominator d.o.f. \ Numerator d.o.f.	1	2	3	4	5	6	7	8	9	10
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646