**PRINCIPLES OF ECOLOGY INTRODUCTORY STATISTICS ASSESSMENT – B179570**

**QUESTION 1**

* A: What is the standard deviation for species richness in each habitat type? [2 marks]

Mean standard deviation for burned: 43.75923

Mean standard deviation for spruce: 36.44880

Mean standard deviation for unburned: 40.38587

Mean standard deviation for wind: 43.57326

* B: What is the mean species richness for each habitat type? [2marks]

Mean species richness for burned: 117.20000

Mean species richness for spruce: 108.46875

Mean species richness for unburned: 90.27273

Mean species richness for wind: 78.14286

* C: What is the standard error for species richness in each habitat type? [4 marks]

Standard error for burned: 7.396661

Standard error for spruce: 6.443299

Standard error for unburned: 12.176797

Standard error for wind: 9.508464

**QUESTION 2**

* Present the R code you used to generate the values above. [2 marks]

#SD

SD <- aggregate(richness~habitat, data = STREAMS, sd)

#Mean Richness

mean\_richness <- aggregate(richness~habitat, data = STREAMS, mean)

#Standard error

length <- aggregate(richness~habitat, data=STREAMS, length)

sqrt\_length <- sqrt(length$richness)

st\_error <- SD$richness / sqrt\_length

**QUESTION 3**

* Explain the difference between the standard deviation and the standard error. [6 marks]

Standard deviation measures how far the data points (in this case species richness measurements) deviate from the mean. The standard error estimates how far the sample mean is from the true mean.

The difference between standard deviation and standard error can be attributed to the spread of the data. A high standard deviation value means the data points are widely spread around the mean (the data exhibits high variation). The wide distribution of samples results in a small standard error (the measured mean is fairly close to the true mean).

**QUESTION 4**

* A: Conduct an ANOVA on the stream invert richness data, and present the F value, the degrees of freedom and the p value. [6 marks]

The among habitats degrees of freedom: 3

The within habitats degrees of freedom: 95

F – value: 4.4906

p - value: 0.005423

* B: If the conventional threshold for statistical significance is p=0.05, does your data indicate whether land use regime has an effect stream invertebrate species richness? [2 marks]

The p-value is 0.0054, below the 0.05 threshold – making it statistically significant, indicating that land use regime has an effect on stream invertebrate species richness.

**QUESTION 5**

* A: Create a plot to illustrate the effect of land use regime on stream invertebrate species richness. [10 marks]

Chart, scatter chart

Description automatically generated

* B: What additional information does the plot provide that is not provide by the ANOVA output (if any)? [2 marks]

The plot shows the differences in the mean species richness and species richness data point distributions for each habitat. The ANOVA table only tells us that there are differences between the means and the distributions.

**QUESTION 6**

* Your data set only contains information on land management regime. What other variables are likely to affect invertebrate species richness? [6 marks]

Depth, temperature, predation, competition, the speed of the stream, vegetation type (and if vegetation present at all), substrate type, oxygen availability, …

**QUESTION 7**

* A: Calculate the variance in weight for each of the deer categories. [2 marks]

Variance for Red deer: 804.9965 kg^2

Variance for Sika deer: 90.9393 kg^2

Variance for Unknown deer: 504.1597 kg^2

* B: What units is the variance calculated in? [2 marks]

Kg^2

* C: Suggest what the ‘unknown’ deer might be\*. [2 marks]

It is possible that the unknown deer are either Roe or Fallow deer, both are present in Scotland. It is known that fallow deer and red deer are both, on average, larger than roe deer. All three are, on average, heavier than sika deer.

The average weight of a roe deer ranges between 10 and 35kg, and the average weight of a fallow deer ranges between 46 and 80kg.

I would thus hypothesize that the unknown deer species is the Fallow deer species due to the mean weight.

**QUESTION 8**

* Can you generate this figure using only the data provided in the table? Explain your answer. [4 marks]

Technically, **with only the data in the table, you cannot reproduce this graph** because the sample size is not given in the table (you need this to calculate standard error), but only as the background information.

If background information can be used, then it can be graphed. To find standard error, you need standard deviation (in the table) and the sample size (in the background info), both of which we know. You can then add error bars to the data plot:

Chart, scatter chart

Description automatically generated

**QUESTION 9**

* Mean antler weight is 7.6573 kg, with a standard deviation of 3.867095.
* The covariance between body weight and antler weight is 52.4265
* Given this, calculate the following (show your working):
  + A: the slope of the relationship between body weight and antler weight. [4 marks]

m is given by covariance / variance = covariance / (sd (mean carcass weight))^2

m = 52.4265 / 804.9965 = 0.0651

* + B: the intercept of the relationship between body weight and antler weight [6 marks]

y = mx + b

mean antler weight= 3.5058 \* mean body weight + b

7.6573 – (0.065 \* 64.19934) = 3.4779

**QUESTION 10**

* A: What are the maximum and minimum worm burdens? [2 marks]

The maximum burden in 383, the minimum is 0.

* B: Present the R code you used to obtain this. [2 marks]

max(GROUSE$burden)

min(GROUSE$burden)

**QUESTION 11**

* A: Conduct a linear regression on the grouse data, and present the F value, the degrees of freedom and the p value. [6 marks]

The regression (density) degree of freedom: 1

The burden degrees of freedom: 980

The F – value: 132.73

The p-value: 2.2. x 10^(-16)

* B: If the conventional threshold for statistical significance is p=0.05, what does your data set indicate about the effect of population density on worm burden? [2 marks]

The p-value for this data set is very small (2.2 x 10^(-16)), which is below the 0.05 threshold – making it statistically significant. This means that population density has an effect on worm burden.

**QUESTION 12**

* A: your model gives you an intercept. Can you explain what this represents in this case? Give both a mathematical explanation and suggest an ecological scenario under which this is plausible. [6 marks]

The intercept is 78.3272.

The intercept gives the y-value at which the x-value is equal to zero. In this case, it tells us the worm burden at 0 grouse density.

This could possibly occur due to the cycles that grouse and worm populations go through. In the case that the grouse population decreased enough to go below the effective population limit, it is possible that the worm population could drive the species to extinction, at which point, the grouse density would be 0.

* B: What does your model tell you about the effect of adding 1 grouse to 1km2 of moor? [2 marks]

Adding one grouse increases the worm burden (by 0.26 eggs/gram faeces).

**QUESTION 13**

* What percentage of the variation in worm burden is explained by your regression line? [3 marks]

The r squared value is 0.1192825, so approximately 11.93% of the variation in worm burden is explained by my linear regression.

**QUESTION 14**

* Plot the relationship between grouse population density and worm burden, including appropriately labelled axes and a regression line. [8 marks]

Chart, scatter chart

Description automatically generated

**QUESTION 15**

* A: How would you experimentally determine whether worms had any effect on grouse population cycles? [4 marks]

Design an experiment in which there is a control grouse population with no presence of worms and compare the density data to the data of a population with worms present. Observing the control vs the experimental population density values could determine whether worms have an effect on grouse population cycles.

* B: what other variables are likely to affect grouse population density? [3 marks]

Competition within species and with other species, predation, food availability, disease, human interference (hunting, pollution, habitat fragmentation), …