1. For each of the following, **i)** name the variables measured and **ii)** state the type of variables that have been used in the experiment. Where more than one variable has been studied indicate which, if any, is the response versus the explanatory variable. Provide only short answers, full sentences are not necessary. (9 marks)
(Example answers: Variable: femur length – numerical-continuous, response variable
Variable: eye colour – categorical-nominal, explanatory variable).

a) You count the number of spots on 40 leopards.

Number of spots: numerical-discrete

b) You measure the pH of the soil from 12 random samples taken from a field.

pH: numerical-continuous

c) You wish to determine the effect of water temperature on swimming speed of goldfish. You randomly sample 30 fish, 15 of which are placed in a tank with cold water, while 15 fish are placed in a tank with warm water. You then measure the swimming speed of each fish in cm/second.

Water temperature: Categorical-nominal Explanatory
Swimming speed: Numerical-continuous Response

d) You wish to determine whether the rate of photosynthesis increases as a function of leaf area. You randomly sample 20 plants, and quantify both the total leaf area (cm^2) of each plant and the rate of photosynthesis of each plant (ppm of O_2 released).

Leaf area: Numerical-continuous Explanatory
Rate photosynthesis: Numerical-continuous Response

e) You explore two factors that might increase the size of tumours in rats. You randomly assign 10 rats to the following combinations of treatments and measure the size of the largest tumour (in mm) after 4 months.
10 are exposed to second hand smoke, and consume a common pesticide
10 are NOT exposed to second hand smoke, and consume a common pesticide
10 are exposed to second hand smoke, and do NOT consume a common pesticide
10 are NOT exposed to second hand smoke, and do NOT consume a common pesticide

Tumour size: Numerical-continuous Response
Smoke exposure: Categorical-nominal Explanatory
Pesiticide consumed: Categorical-nominal Explanatory

2. Estimate the descriptive statistics indicated below for the following two data sets. Identify all extreme values, if there are any:
(8 marks)

a) Data: 10, 8, 19, 1, 5, 7 sorted 1, 5, 7, 8, 10, 19 n = 6

Mean =__8.3___

Median =___7.5__ given by (7 + 8)/2

Inter Quartile range =Q3-Q1 = 10 – 5 = 5.0

Standard error of mean =2.5 (note variance is 36.67, s = 6.06)

List extreme values 19 (since 19 exceeds 10+1.5 x IQR = 17.5)

b) Data: 1.6, -2.1, 1.1, 1.5, 1.3 sorted -2.1, 1.1, 1.3, 1.5, 1.6

Mean =0.68_

Median =1.30

Inter Quartile range = 2.05 (q1 = -0.5 q3 = 1.55)

Coefficient of variation = 230.28

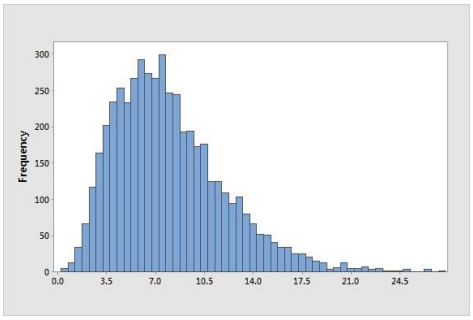
List extreme points NONE

3. For the histogram shown below:

a) What statistical term best describes this histogram?
Skewed Right (or positively skewed)

b) What term would best describe the shape of the sampling distribution of means (based on n = 30) derived from the histogram below? Normal

c) If the true variance of the population is $\sigma^2 = 16$, what is the approximate standard error of the mean based on a sample size of n = 9? 1.33 (or 4/3)
(3 marks).



4. A biologist estimates the variation of body mass for a population of ducks and finds and estimate of $s^2 = 25.0$, while the mean body mass is 5 kg. In their study they obtain a standard error of the mean equal to 0.5. What is the sample size, n, upon which their mean is based (2 marks)?
Since $SE = s/n^{0.5}$ or $SE = (s^2/n)^{0.5}$,
rearranging, $n = s^2/SE^2$
So $n = 25.0 / 0.5^2$
n = 100

So, their sample size was n = 100

5. In a population of lupines, the proportion of plants with purple flowers is known to be 0.3 and all the other plants have pink flowers. What is the probability of random sampling 3 plants, all of which have pink flowers? What assumption did you make, if any, to determine the probability? (2 marks)
Prob(pink) = 1 – 0.3 = 0.7.
Prob of all 3 pink is 0.7 x 0.7 x 0.7 = 0.343
Assumption is that plants sampled are independent of each other

6. What is the probability of tossing a coin that ends up as tails, and at the same time rolling one die that shows up as a 3? (2 marks)

Prob(tails) = 0.5 or ½ , Prob (of 3) = 1/6

Prob(tails and of 3) = ½ x 1/6 = 1/12 or 0.083

7. You are told that hawks have a mean flying speed of 100 km/hour. You are skeptical of this claim so you obtain the flying speeds of 5 randomly sampled hawks in km/hour (below). **Estimate the approximate 95% confidence interval for mean flying speed**. Does the 95% confidence interval support the claim that the flying speed is 100 km/hour? **Explain briefly** in one sentence why it does or does not? (3marks)

Flying speed data in km/hour: 95, 105, 93, 100, 90

Mean = 96.6
Standard Error mean = 2.657

Upper 95% Confidence limit = 96.6 + 2 x 2.657 = 101.9
Lower 95% Confidence limit = 96.6 - 2 x 2.657 = 91.3

The confidence interval supports the claim because we are 95% certain that the true mean speed falls between 101.9 and 91.3 so we have no reason to doubt that hawks fly at 100 km/hour

_____

8. For each of the following, state the null (Ho) and alternative (Ha) hypotheses **clearly STATING if Ha is one-sided or two-sided**. (6 marks)

a) You wish to determine whether listening to Karaoke increases one's blood pressure. Forty students are subjected to Karaoke for 4 hours, while 40 others wear noise cancelling headphones and so are not exposed. Blood pressure of each student is measured immediately after exposure (or non-exposure).

Ho: Blood pres Karaoke = Blood pres. nonKaraoke

Ha: Blood pres Karaoke > Blood pres. nonKaraoke 1-sided

b. You wish to determine whether listening to classical music while writing a test effects performance on the test. Ten students listen to classical music and ten others do not, while writing a math test. The scores on each test are recorded and compared.

Ho: Testscore classical = testscore no music

Ha: Testscore classical $\neq$ testscore no music 2-sided

c. You wish to determine if dead armadillos along roadsides in Texas are longer than living ones. You randomly sample and measure 10 dead and 10 living armadillos.

Ho: length dead armadillos = length living armadillos

Ha: length dead armadillos > length living armadillos 1-sided

9) Write a single complete SAS program to obtain descriptive statistics (for example, the mean, median, Q1, Q3) and a histogram (for all variables) for EACH of the two bird species below (Sparrow versus Pigeon). For each individual bird, both beak width and beak length are listed in that order, and were measured in mm.
Have the SAS program compute the square root of beak length. Include the data in your program exactly as they are written below (6 marks).

```
DATA QUEST9;
INPUT BIRDSP $  WIDTH  LENGTH ;
SQLENGTH = SQRT(LENGTH);
DATALINES; (OR CARDS;)

Sparrow     3.0  4.7
Pidgeon      5.4  5.8
Sparrow     3.6  4.8
Pidgeon      5.2  5.9
Sparrow     3.1  4.5
Pidgeon      5.5  5.0

PROC SORT;
   BY BIRDSP;

PROC UNIVARIATE;
HISTOGRAM / VSCALE = COUNT;
  BY BIRDSP;

RUN;
```