1i Since most data is expected to be within three standard deviations of their mean (assuming normality), I assume that a mean:sd ratio of less then 3:1 gives a high standard deviation relative to the mean.
In this case, both measles and whooping cough meat that criteria with ratios of 1.95:1 and 1.19:1. Since the notifications can not be 0, this suggests that the normal model is not a good approximation for the diseases. The median and interquartile range would describe the data better then the mean and standard deviation.

ii. year = x, notification of disease = y
Both measles and whooping cough show wild fluctuations. Measles seams to peak every 2-3 years and whooping cough seems to peak once every 5 years.

iii:year = x, notification of disease = y

iv. Turburculosis appears to be having a downward trend, while measles seems to have a cyclicar pattern.

v. Measles appears to have a 2 to 3 year cycle.
Whooping cough appears to have a 5 year cycle
Food poisoning has risen over the 10 year period.
Turburculosis has fallen over the 10 year period.
Dyssentry has not changed much over the 10 year periodm

2. mean = 13.2
Iqr = 5.325

3. As the log of the area goes up, the number of species increases at an increasing rate. There is a positive correlation between the number of species and the log of the area. Correlation is not causation.

4. The fertile data set has a higher 'box' then the sterile data set and since the box shows the lower and upper quartiles (as well as the median). one can infer the fertile plants are taller then the sterile ones. Although both sets have outliers, the fertile set has upper outliers, while the sterile set has a lower one.

5. The boxplot shows many lower outliers. Since the mean is more affected by outliers, the mean should be lower then the median, not higher then it. Then median is about 20, so the mean can not be 21.4.

6. Since the data is positivly skewed, the mean should be greater then the median.

7i Poisson process
X~ Poisson(λt) Since t=52 days,
X~ Poisson(λ52)

ii $\lambda = 1 / \mu$
Therefore, λ = 1 / sample mean

iii. A two sample t-test requires normailty. A poisson distribution can only be usefully aproximated by a nomal distribution when $\mu > 30$.

8. The toal of all values of f(x) equal 1 for a probability density function. In this case, the sum of all f(x) will be greater then 1.

9. For N ~ Geometric(.8), p(0)=.8. Therefore, the first diagram is Geometric (.8).
For N ~ Geometric(.4), p(0)=.4. Therefore, the second diagram is Geometric (.4).

10 mean = 1, variance = 1

11 .6988

12 N~Geometric (1/6)
E(N) = 6
V(N) = 30

13 If the projects under £50,000 were awarded close to 50% of the awards, then projects over £50,000 were awarded close to 50% of awards. Assuming those projects over £50,000 are considered large, then both large and small projects have similiarily benefited.

14 exact = B(1000,.002). Since np<5, the normal aproximation is not useful.

The Poisson approximation is useful. X~Poisson(2)
prob = .1429

15 i X ~ Poisson(270)
ii(Y ~ M (18)

16 .7788

17 N(64,64)
.9599

18
r of x=3, r of y = 2 mean = r, vare = 2r for chi sq
E(X) = 3, V(X) = 6
E(Y) = 2, V(Y) = 4

E(X-Y) = 1, V(X-Y) = 10

19. .3446

20 multiplying the p(x)'s
(1/m)^3 * (1-5/m)^3

max likelihood for discrete uniform distribution = Xmax = at least 5

21i (.49305, .59445)

ii If the experiment were repeated 100 times, 99 of the 100 confidence intervals produced would included the true p for the population.

iii Using the result in part i, at the 99% level of significance, the null hypothesis that p=.33 is rejected in favour of the alternative hypothesis that p is not equal to .33. In fact, since .33 is below the confidence interval it appears that p > .33.

22. (52.9, 56.9)

23. With an SP of .38, there is no evidence to reject the null hypothesis that the two means are equal. An SP of .38 means 38% of experiments would give less support for the null hypothesis then the current sample.

24.The normal distribution can model the variation in either population
The variances of the two populations are the same
The samples are independent of each other.

25. Find the mean of the sample.

Find the expected frequesncies of 0,1,2,3,4,5,and >=6 for Poisson(mean)
Group the observed and expected data into groups where the expected frequency is at least 5.
Find X^2 using the formula sum ((obs-exp)^2/exp)
The degrees of freedom = catagories - 2 (as one parameter is estimated)
Use a computer of tables to find the probability of X^2 for the degrees of freedom.
If the SP is small, there is enough evidence to reject the null hypothesis that the data can be modelled by the Poisson distribution

26.N ~(23,380.5,1,243,063.25)
Z=6.339
Sp aprox 0
With an SP of almost 0, there is enough evidence to reject the null hypothesis that the eman of A and B are the same.

27i . It does appear as if a straight line model would approximate most of the data. It would not go through the origin. There appear to be two outliers which should be investigated. One at income 2, prestige 90 and the other at income8, prestige 100. If they were removed it would give a better straight line model.

ii y = 39.97.
Clergymen is an outlier. Being a clergyperson is a proffession, so there is prestige, but unlike most proffessions, it is not entered into to make money, but is entered into to improve society. It should be removed when fitting the least squares line as it is a special case.

28 The plot indicates that there are two outliers. One at the lower income range and one at the higher income range. If there were removed (if appropriate), it would lead to a better straight line fit. The erst of the data points are reasonably scattered over the line.

29i .3002
ii. .2857
iii a .3733 b .4907
iv 35.43

30 a perfect negative correlation (show y decreaseing as x increases and all data points on line

b reasonable positive correlation show y increasing with x and points scattered a bit across the line.

31 A Pearson cxorrelation coefficient measures the linear association of the data. The data set does not look linear, so the Pearson correlation does not really say much about the data. A Spearman correlation should have been used.

However, assuming that the Pearson correlation coefficient was appropriate,
with a Pearson coefficient of .322 implies there is a weak positive association between the two variables/ The test of zero association gave an SP of .048 which means we can reject the theorey of no association in favour of a theory of association. So there is a weak association between earthquakes and the log number of deaths. Correlation is not causation.

32.X^2 for o drink, o packs = 75.85
Degrees of freedom = 4
SP for X^2 824.8, df=4 is almost 0.
There is enough evidence to reject the null hypothesis of no association in favor of a theory of association. The number of packs a day is correlated with the number of drinks per day. In fact, those who don't drink are more likely not to smoke.

33.i alpha = .3 beta = .5294
ii. runs = 19
iii SP = .2076
With an SP of .2076, there is not enough evidence to reject the hypotheisis that the realization was generated by the Bernoulli process.