2001 M246 Exam

Q.2

1. The relationship is a positive one - the larger the distance, the higher the climb.
2. The relationship is roughly a linear one, up to about 10 miles distance and around 4500 ft. Beyond that it is harder to say due to so few data points.

Q.3

(0.54,1.28,2.8,7.31,9.95)

Q.4

Median = 0.5

$1 - \exp(-y^2) = 0.5$

$\exp(-y^2) = 0.5$

$-y^2 = \ln 0.5$

$y = 0.833$ (to 3 d.p)

Interquartile range = 0.75 - 0.25:

$1 - \exp(-y^2) = 0.25$ $\quad y = 0.536$ (to 3 d.p.)

$1 - \exp(-y^2) = 0.75$ $\quad y = 1.177$ (to 3 d.p,)

thus interquartile range = 1.177 - 0.536 = 0.641 (to 3 d.p.)

Q.5

(a) Continuous uniform distribution U(0,200)

(b) f(x) = 1/ b-a = 1/ 200 - 0 = 1/200

P(X > 150) work out using integration. Integrate 1/200 between limits of 200 and 150 to give: 200/200 - 150/200 = 1/4

(c) Graph has highest value at low x and high x values, and dips steeply to a constant low value in the mid x range.

Q.6

1. Need a continuous distribution to model this; binomial is discrete.
2. Probability of breakdowns over time will not be the same over the lifetime of the equipment. Binomial assumes probability of failure to be the same each time.

Q.7

(a) Poisson distribution - because the event of finding a defective bulb is relatively rare and occurs singly and at random intervals in time. X ~ Po(0.001).

(b) Mean E(X) = 1/1000 = 0.001

Variance V(X) = E(X) = 0.001,
so standard deviation = $\sqrt{(0.001)}$ = 0.032 (to 3 d.p.)

Geometric distribution mean 1000 , sd = 999.5
Agreed Alan though I have mean=sd=1000

count upto 1st faulty lightbulb ; then count reset. i.e counts the waiting times between failures.

Q.8
I would not believe this statement. The bar chart shows a negatively skewed distribution, so the mean should be less than -2.5.

Q.9
(a)
1. There are a finite number of experiments.
2. The experiments are independent.
3. The probability of a success is the same for each experiment.
4. The outcome of each experiment is either a Success or a Failure.

(d) P(X=2) = 0.04

0.04 = 2! / (2! 0!) x p^2 x $(1 - p)^{2-2}$
= 1 x p^2 x 1
= p^2

thus p = 0.2

Q.10
(a) n=4 p = 0.3 q =1-0.3 = 0.7

P(X > 0) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)
= 0.4116 + 0.2646 + 0.0756 + 0.0081
= 0.7599
= 0.760 (to 3 d.p.)

(b)
P(X = 2) = 4! / (2! 2!) x 0.3^2 x 0.7^2
= 6 x 0.09 x 0.49
= 0.265 (to 3 d.p.)

Q.11
(a) The total sum must be 1.0 for the distribution to be valid. Hence the value of p(0) is 1/4.

(b) mean E(X) = (-2 x 1/8) + (-1 x 1/4) + (0 x 1/4) + (1 x 1/4) + 2 x 1/8) = 0

variance - not sure how to do this.

for variance see p.4 H/B ch 3 Ans 1.5

Q.12
(a)
P(Z = 0) = exp(-u)
P(Z = 1) = u exp(-u)

If P(Z = 0) = P(Z = 1) then u must = 1.0

(c) (not sure if I'm right here!)

P(Z > 3) = 1 - P(Z < 4)
= 1 - (P(Z = 0) + P(Z = 1) + P(Z = 2) + P(Z = 3))
= 1 - exp(-1) + 1 x exp(-1) + exp(-1) x 1^2/2! + exp(-1) x 1^3/3!
= 1 - 0.2452
= 0.7548 (to 4 d.p.)

P(Z>=3) = 1 - P(0)-P(1)-P(2) = 0.08

Q.13
(a) $V(3X) = 3^2 \sigma^2 = 9 \sigma^2$

(b) Sum $S = X_1 + X_2 + X_3 = 3 \sigma^2$

Q.14
Using exponential distribution, mean = 2 wks u = 1/lambda thus lambda = 0.5

P(X >= 3) = 1 - P(X < 3)
= 1 - (P(X = 0) + P(X = 1) + P(X = 2))
= 1 - (1 - exp(-0.5 x 0) + 1 - exp(-0.5 x 1) + 1 - exp(-0.5 x 2))
= 1 - (0 + 0.3934 + 0.6321)
= - 0.0255

Think I've gone wrong somewhere here - shouldn't have a negative answer.

this is a continuous distribution!

*P(X>=3)=1-P(X<3) = 1 - (1-EXP(-LAMBDA*X)*

Ans 0.2231

good annotation tip. P(X>=x) = -exp^-lamda*x saves working it out on the day

Q.15
Let X be a Poisson random variable with mean 2 and Y, independently, be a Poisson random variable with mean 2.5. What is the probability that the sum, X*Y, is less than 3?

Didn't know how to tackle this one.

Sum of poisson distributed variables is itself poisson distributed with parameter mu1+mu2(in this case 4.5) Call the sum S then S follows Poisson(4.5)

work out P(s<3) as standard

X+Y - Poisson(4.5)

P(X+Y)<=2) = P(0)+P(1)+P(2)

Ans 0.1736

Q.16
P(X >= 72) = P(Z >= (72 - 69.2) / 5.76)
= P(Z >= 0.486)
= 1 - phi(0.486)
= 1 - 0.6844
= 0.316 (to 3 d.p)

As per Barbara but using sd = 5.76^0.5
Ans 0.123

since tail probbility gives the proportion

Q.17
Approx to a normal model with corresponding mean u = np = 25 x 3/4 = 18.75

And variance: σ^2 = npq = 25 x 3/4 x 1/4 = 4.6875

So, X ~ N(18.75, 4.6875)

P(X >= 21) = P(Z >= (20.5 - 18.75) / √ 4.6875) using continuity correction
= P(Z >= 0.808)
= 1 - phi(0.808)
= 1 - 0.7910
= 0.209

I added the continuity correction.... why did you both take it away??

I'm only going by the examples I found in my A level book - (it's been a real Godsend!):
e.g.
P(X > 4) becomes P(X > 4.5)
but P(X >= 4) becomes P(X > 3.5)

similarly: P(X < 4) becomes P(X < 3.5)
but P(X <= 4) becomes P(X < 4.5)

and P(X = 4) becomes P(3.5 < X < 4.5)

Drawing a sketch with bars for the discrete values and a curve passing through them for the normal curve may make this clearer.

Q.18
The variance cannot be the same in the sample as in the parent population, as in the sample the standard deviation is much wider compared to the parent population. Thus the variance is wider, too.

sample should be selected randomly and cover the population.

expected value of sample mean = population mean

expected value of sample variance = population variance

Q.19
(a) $(1/2\ (1 - r))^{147}\ (1/2\ r)^{65}\ (1/2\ r)^{58}\ (1/2\ (1 - r))^{133}$

$= (1/2\ (1 - r))^{280}\ (1/2\ r)^{123}$

$= ((1 - r)^{280}\ r^{123}) / 2^{403}$

(b) r approx = 0.31

Q.20
1. The $\chi^2(2)$ distribution is strictly positive; the graph above shows negative values.

2. The mean should equal the number of degrees of freedom, thus mean should equal 2. The graph doesn't extend as far as 2, so it cannot be that of $\chi^2(2)$.

Q.21
A researcher used the same sample to calculate a 90% confidence interval and a 95% confidence interval for the overall mean mu of some normal data. He obtained the 90% confidence interval (7.5, 10.7) and the 95% confidence interval (8.5, 10.9). Give two reasons why at least one of these confidence intervals must be wrong.

Look at the values of t coresponding to the percintiles requited. As the level of confidance gets higher the value of t rises so a 95% CI should

i) be wider than a corresponding 90% CI and
ii)contain the 90% CI

WHOOPS on both of the above

1. The means of these intervals should be equal, but are not.
   (7.5, 10.7) mean is 9.1. (8.5, 10.9) mean is 9.7.

2. If the 90% interval is correct, the 95% values should fall inside those of (7.5, 10.7), but they do not.

If the 90% interval is correct, the 95% values should be outside those of (7.5,10.7) ie wider interval than 90%

Q.22
(u-, u+) = (6754 - (1.645 x 1142/11), 6745 + (1.645 x 1142/11))
= (6538, 6925) (to 4 s.f.)

as per Barbara but using t=1.812 ; std error = 11^0.5

Q.23
The data used for the survey seems to be only taken from customers who made a saving on their insurance. Customers who did not were not included. This introduces bias into the sample and so the claim made by Lloyd's is not appropriate.

Q.24
(a)
The means of 2 data sets are calculated: mean(old) = 56, mean(new) = 68.14

The variances of both data sets are calculated: var(old)= 74.67, var(new)= 136.5

The variances were determined to see whether a two sample t-test would be valid. The variances are within a factor of 3 of each other, so test can proceed.

Two-sample t-test is carried out to compare the means of both old and new. The t statistic was -2.211 with 12 degrees of freedom. The total significance probability was 0.0472.

(c) The SP is reasonably small, so I would conclude that the means of the two samples old and new are not the same, ie reject the null hypothesis that u(old) = u(new), and accept the alternative hypothesis, u(old) is not equal to u(new).

Q.25

(a) The Students two-sample t-test to compare the means could be used.

(b) Pooled sample SD, both sample means.

(c) $t(n_1 + n_2 - 2) = 8 + 8 - 2 = 14$ degrees of freedom

(d) 5% significance level, t-quantile required is: 2.145 (14 df)

$T = (x_N - x_G) / s_P \sqrt{(1/n_1 + 1/n_2)}$

$= (459 - 455) / 29.56 \times \sqrt{(1/8 + 1/8)}$

$= 0.271$ (to 3 d.p.)

$0.271 < 2.145$ so accept H_0. Conclude there is no evidence to reject H_0, so there is no difference in the corneal thickness of a normal eye compared with an eye with glaucoma.

Q.26

(a) H_0: $u_{now} = u_{1990}$ the mean number of tuna fish has not decreased since 1990.

H_1: $u_{now} < u_{1990}$ the mean number of tuna fish has decreased since 1990.

(b) Rejecting H_0 when H_0 is true, i.e. rejecting that the mean number of tuna fish has not decreased since 1990 when in fact that statement is true.

(c) Accepting H_0 when H_0 is false, i.e. accepting that the mean number of tuna fish has not decreased since 1990 when if fact the mean number of tuna fish HAS decreased since 1990.

Q.27

The data is positively skewed, and so a normal model is not suitable for modelling these data. However, a suitable transformation of the data to reduce the skew would make a normal model appropriate. This could be done by using a transformation such as: $x^{1/2}$, $x^{1/3}$ or log(x). These transformations have a greater reducing effect on higher x-values such as occur here, and will reduce skewness.

Q.28

For the model to be appropriate, a straight line should fit these data. Whilst the data points below about $x = 0.7$ are reasonably linear, the later points are not so close to the line. Thus, the exponential model may not be appropriate here, as the points do not lie on the straight line, especially at higher values of x.

Q.29

(a) The MWW test can be used as a test for zero difference analogous to the Student's t-test, but does not require the assumption of normality so it can be used on data where normality is not assumed.

(b) When applied to a random sample of size 20 (sample A) and a random sample of 30 (sample B), the Mann-whitney-wilcoxon test produced a test statistic U subscript A = 636. Would this justify rejecting the hypothesis that the2 samples arose from the same population? Show any calculations that you carry out.

i)MWW No assumption of normility in the data sets.

ii) Look what you have to calculate for the 2 sample test. Means, variances, pooled variances and then bung the lot in a bloody unfrieandly formila for the test stat. Now look at MWW. All you have to do is sort out a list and then do some adding. What would you rather do given the choice?

$U_A = 636$ $n_A = 20$ $n_B = 30$

the expected value of U_A under H_0 that the 2 samples are from identical populations is:

$$(n_A(n_A + n_B + 1)) / 2 = (20(20 + 30 + 1)) / 2 = 510$$

observed U_A of 636 is larger than the expected value of 510.

Variance under H_0: $(n_A n_B(n_A + n_B +1)) / 12 = (20(20 + 30 + 1)) / 12 = 2550$

Observed U_A has corresponding z-value:

$$Z = (636 - 510) / \sqrt{2550} = 2.495$$

SP(total) = 2 x phi(2.495)
= 2 x 0.9936
= 1.9872

Conclude that there is some evidence to reject the null hypothesis that the 2 samples came from the same population.

Q.30.

(a) A straight line could be fitted (not passing through the origin) but many points would lie some distance from the line, also the final two data points would lie a considerable distance from it. The lower region of the line gives a better fit that the later data points.

(b) Height = 2.79 + 0.268(20) = 8.15 metres.

(c) The line is not reliable at this height because the last data points fall well away from the line, so an accurate age cannot be estimated. A different model could be fitted, e.g. a curved line which passes close to the last points.

(d) At one year of age:

Height = 2.79 + 0.268(1) = 3.058 metres.

The line doesn't pass through the origin and so over-estimates the height of a one-year old tree.

Q.31

(a) The fluctuation around zero is fairly random so the model assumptions have been fulfilled. (?? Not sure)

(b) The variance is not constant, as shown by this plot.

Q.32

1. The random terms wi are independent with mean 0 and constant variance.
2. The random terms wi are normally distributed.

Q.33

(a)

	Improvement	No improve
Drug	34	26
Sugar Pill	34	26

(c) Observed - expected values:

	Improvement	No improve
Drug	41 - 34 = 7	19 - 26 = -7
Sugar Pill	27 - 34 = -7	33 - 26 = 7

Calculate χ^2:

$\chi^2 = 7^2/34 + -7^2/26 + -7^2/34 + 7^2/26 = 6.65$ (to 2 d.p.)
Falls in 0.99 quantile in Table 6.

Q.34

Df = ((row - 1)(column -1)) = ((4 -1)(5 - 1)) = 12 degrees of freedom

Q.35

(a)

Counting number of transitions:

$$N = \begin{matrix} 0 \\ 1 \end{matrix} \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} = \begin{matrix} 0 \\ 1 \end{matrix} \begin{bmatrix} 22 & 10 \\ 9 & 8 \end{bmatrix} \begin{matrix} 32 \\ 17 \end{matrix}$$

$$\hat{M} = \begin{matrix} 0 \\ 1 \end{matrix} \begin{bmatrix} 0.688 & 0.312 \\ 0.529 & 0.471 \end{bmatrix}$$

(b) r = n01 + n10 + 1 = 10 + 9 + 1 = 20

Q.36

The Kolmogorov test statistic D is the maximum vertical distance between the two distribution functions, (the empirical distribution function and the hypothesised model). The c.d.f. of the hypothesised model is superimposed on the graph of the empirical distribution. (A straight line is drawn from the origin through the stepped lines, and the largest difference is measured)

D = 0.6 - 0.16 = 0.44