1999 paper – my answers ***(comments of another student)***

1)
(a) % recorded of BSC crimes / % reported = % recorded of number reported. Inthis case 27/47 = 57.
The first entry (3090) is calculated from the sum of the figures above.
2410 + 459 + 221 = 3090.

(b) Each type of offence is broken down into its component parts (eg. burglary is broken down into attempts and no loss and burglary with loss). These are then grouped into one of the three classes of crime (acquisitive crime, vandalism and violence). The figures generated further down the column do not match the total comparable figures. In fact there is a short fall of 459 which could be accounted for by various unlisted crimes falling into one of these three categories.

Extra-entry from FC:

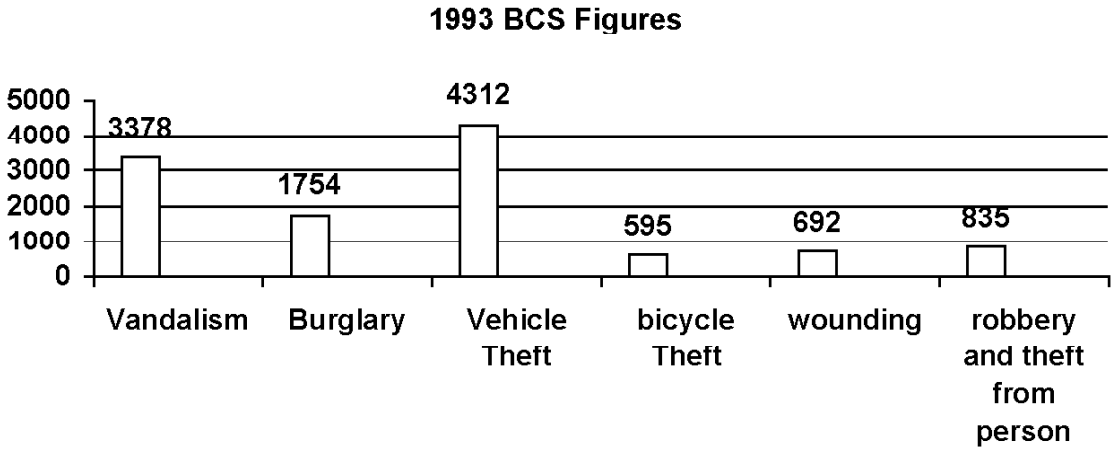
>Has anyone got any idea how they've managed to get the figures in part
>(b) of this question. I thought it would just be a case of adding up the
>various sections but they've obviously done something else as it doesn't
>tally that way.
>
>In the 1993 police column, figures shown above the total column are the
>same as the figures shown below the total column, but they are
>catagorized differently. It is likely the figures below the total line
>are BSC catagories. Within the totals below the total column, the
>catagories of burglary and all vehicle thefts are futher subdivided. So
>the total of crimes below the total line is 2630, while the total of all
>offenses is 3090. There are at least two possibilities for the
>differences. It could be that the difference is caused by some crimes in
>the total column being offences which are not BSC offences and therefore
>they are not able to be put in the subcatagories of BSC offences.
>Another possibility is that the police considered some of the crimes
>which are below the line as being more then one offence, a combination of
>aquisition crime, vandalism, and violence.

Total re 1993 = 3090 and below this 2630

The difference is approx 459 ie vandalism category

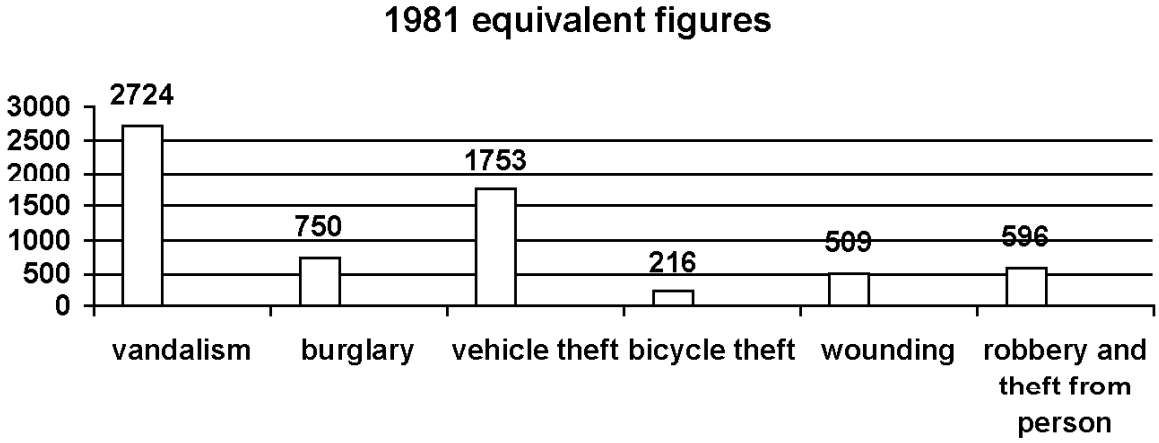
The BCS total re vandalism (3378) appears in first two entries below TOTAL line.
No such figs for police 1993 column.

(c)



(d) Equivalent values in 1981 were:

vandalism	2724
burglary	750
vehicle theft	1753
bicycle theft	216
wounding	509
robbery & theft from person	596



Between 1981 and 1993, the placing of vandalism and vehicle theft have swapped over. In 1981, vandalism was the most noted offence with vehicle theft in second place but by 1993, the two had swapped over. There are still roughly three times as many burglaries as there are bicycle thefts in both years. There has, however, been a drop in the levels of wounding compared to robbery & theft from persons between the two years.

2 m= (26 + 1)/2 = 13.5^{th} position
= 1.555

Q_L = (26 + 1)/ 4 = $x_{6.75}$
= 1.27 + 0.1/ 4
= 1.295

3. There appears to be a positive association between cholesterol levels and age. Cholesterol levels appear to rise with age.

4. The distribution is bimodal with extreme lack of data in the middle of the range.

5. The median and the quartiles for the girls are situated at a higher level than for the boys. The distribution of data within the box appears to be symmetrical about the median whereas for the girls it is slightly left skew.

6. (unsure about this one – though I was under the impression that assumptions were correct for the exponential distribution)

7. because 2 – 3/1 = -1 and you can't have a negative cumulative probability.

8. (i) Binomial (40,0.125)
(ii) Assume that any reaction is classed as a success.
(iii) p(hat) = 5/40 = 0.125

9. Its not appropriate to model my weekly shopping bill using a normal distribution as what I buy each time is dependent upon what needs replacing. To use a normal distribution, the items on it would need to be independent of what was indoors.

10. E(X) = 0.375 + (2*0.125) + (3*0.125) = 1
V(X) = -1^2(0.375) + 1^2 (0.125) + 2^2 (0.125) = 1

11 E(X) = 1/ (1/100) = 100
V(X) = q/p^2 = 0.99/0.01^2 = 9900

12. (i) continuous uniform distribution
(ii) mean = 5
variance = 100/12 = 8.33
(iii) p(leak under house) = 3/10 = 0.3

13. $0.9 = 1-\exp^{-\lambda x}$
$\exp^{-\lambda x} = 0.1$
$\lambda x = -\log(0.1)$
$x = 5.3$

14. p(x = 1) = 0.5
p(y = 1) = 0.9
If x and y were independent then p(x=1,y=1) would be the same as p(x=1)p(y=1).
p(x=1,y=1) = 0.45

p(x=1)p(y=1) = 0.5 * 0.9 = 0.45
therefore they are independent.

15. Po(0.5)
p(x *f* 2) = 1- p(x □ 1)
= 1- (p(x=1) + p(x=0))
= 1- (0.6065 + 0.3033)
=1 – 0.9098
= 0.0902

16. The first is the histogram of binomial(10,0.6) since the maximum number in any sample is 10 which isn't exceeded in this graph.

17. No it isn't since the variance and the mean of the distribution should be the same for the model to be valid.

18. ②= 5/60 mins
so in 30 mins, ③ = 2.5
$p(x=0) = (e^{-2.5}2.5^{0})/ 0!$ = 0.08208

E(X) = 24② = 120
V(X) = 24②= 120 since mean and variance are the same.

19. $N(10,5^2)$

(i) $p(x \square 15) = p(z \square 1.0) = 0.8413$

(I know this should really be a t distribution but we haven't got the probability tables for that.).

(ii) $\alpha = 0.9$, $q_{\alpha} = 1.282$

$z = 1.288$

$(X-10) / 5 = 1.288$

$X = 5(1.288) + 10$

$X = 16.44$

20. ③ = (6*5 + 6*4)

= 54 / year

p(xμ 65 / year) approximating to normal

Po(54) σ N(54,54)

$(65 - 54) / \sqrt{54} = 1.50$

$p(z \geq 1.50) = 1 - 0.9332$

$= 0.0668$

21 since distribution occurs within 3 s.d's of the mean, then weight will be

$T_n \approx N(n\mu, n\sigma^2) = N(1500, 100^2)$

between 1200 and 1800 grams.

ii) $(1450 \quad 1500) / 100 = -0.5$

$p(z \leq -0.5) = 1 - p(z \leq 0.5)$

$= 1 - 0.6915$

$- 0.3085$

I got a dfifferent answer then you

i. I got Tn N(n, n2) = N(1500, 100) as opposed to N(1500, 100^2). I can see why we got different answers, and I am not sure who is correct.

the formula said v= nσ^2. I calculated σ^2 as 1 (1^2 = 1) and then multilied by 100. you did (n*σ)^2, so you got 100^2.*

ii because of the difference in 1, I got z=(1450-1500)/sqtr(100) = -5, so my answer was p is close to 0.

The itallic version should be correct acc. to one's tutor.

22. $\alpha^6 (1- \alpha - \alpha^2)^6$

ii) $\alpha \varphi 0.31$

23 (i) p= 541/1023

q= 482/1023

$(p_-, p_+) = (0.529 - 1.96\sqrt{(0.529 \times 0.471)/1023},\ 0.529 + 1.96\sqrt{(0.529 \times 0.471)/1023})$
$= (0.498, 0.560)$

(ii) In this context, it means that there is 95% confidence that the true value of p for those in favour of joining a single European currency lies within the given interval.

(iii) Since p=0.5 lies within the interval above, there is no reason to reject the null hypothesis that p=0.5. However it is possible that a type 2 error has been made.

24 There was strong evidence against equality between the two means.

25. Volunteers don't constitute a random sample. Therefore the sample will be biased and not representative.

26. The pairs are independent and normally distributed.

I also had the populations have the same variance.
Also, just to clarify, I could read your comment of the pairs are independant two ways, one of which is right and the other which is not. It could mean that each pair is independant of each other pair (which is correct) or each sample is independant of each other sample, which is not correct. For example in a medical study, to determine the difference before and after taking the drug. Each item in a pair is linked (to the same patient), but each pair is independant of each other pair (different patients).
Then again, maybe I am the only one who read it funny.

27. Command to perform a one sample t-test on the null hypothesis that there is zero difference between the vectors english and greek.

The test statistic (t) is -2.211 and is tested against 31 degrrees of freedom. The SP of 0.03457 indicates that there us sufficient evidence to reject the null hypothesis of a zero mean difference.

You forgot to mention that the sizes of both samples were 32.

28. Use frequency as observed values and calculate expected values using $n \times p$
Calculate mean and use formula $n \frac{e^{-\lambda}\lambda^x}{x!}$
$\sum (O-E)^2/E$
is the test statistic to be tested against 12 degrees of freedom.

You forgot to mention a few things here, here is my answer. I may be giving more detail then the exam wants.
Find the mean of the sample.
Find the expected frequesncies of 0,1,2,3,4,5, 6, and >=7 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

29. Data is skewed. Try using transformation of x^1/3, x^1/2 or log (x) since they have the affect of greter reduction on higher numbers than on lower ones.

30 $w_+ = (151 - ((32*33)/4))/\sqrt{(32 \cdot 33 \cdot 65)/24}$
$= -2.113$

yes it would as approximated result of -2.113 gives probability of 1- 0.9826 = 0.0174 which is strong evidence against the null hypothesis.

I think that the test is two sided giving as SP of .0348. Otherwise I got the same answer.

31 Data doesn't follow a straight line as it is curved. Interpolation will therefore be inaccurate.

(ii) wouldn't expect fitted line to pass through origin as that would imply ability to speak recognisible words at birth. The estimate of the intercept implies that at one year of age, a child has a vocabulary of - 201.93 words. From the given data point of 100 at x=1, the intercept is clearly wrong.

I aded that x was only >= 0 from about 1.36 onwards.

32 There is some degree of pattern either side of the residual line which shows that the assumption oif a straight line has been violated.

33 (i) 12063 / 23756 = 0.508 (3dp)
(ii) 0.201 *I got .2687 (5687+696)/23756*
(iii) 0.799
(iv) 19.278. The expected amount is a lot lower than observed.

34 There were 4 df.
Test statistic is above upper value contained in tables for relevant df so reject the null hypothesis. Conclude that the variables are dependent upon each other.

I added correlation is not causation, but it probably wasn't nccessary.

35 just three sketches. First going downwards from left to right with some scatter, 2nd going straight across graph and third rising from left to right without scatter.

graph 2 where p =0, i am not sure what you meant about "going straight across graph", but I put a graph where dots where completely randomly scattered about as r=0 means no association.

I just put the same value for y regardless of the value for x so that the line was parallel to the x-axis.

36. $M(\text{hat}) = \begin{matrix} 23/32 & 9/32 \\ 9/17 & 8/17 \end{matrix}$

$= \begin{matrix} 0.719 & 0.281 \\ 0.529 & 0.471 \end{matrix}$

(ii) runs = 19

(iii) p= 1- 0.9222 = 0.0778. There is insufficient evidence that the data doesn't fit a Bernoulli model.

Again, I think this test is two sided, so I think the SP is .1556. I also added that the normal model is assumed to be ok when n0 and n1 are greater than 20. Since n=17. the normal model may not be appropriate. This is on page 489 of the book; however, maybe they mean when n0+n1 is greater then 20.