INSTRUCTIONS: Calculations need not be exact.Each question is worth xx points. Your xx best answers will be counted.

1. A company reported the following data pertaining to the disability of a random sample of n = 100 of its employees:

number (x) of absences of 1 day or more as result of disability	number (f) of employees	Expected number (f_{exp}) [for part x below]
0	46	___
1	31	___
2	13	___
3	7	___
4	3	___
>4	0	___
	$\sum f = 100$	$\sum fx = 90$

$\sum fx^2 = 194$

$90^2 / 100 = 81$

difference 113

a Fill in the blanks:

modal x = _____ mean x = _____
median x = _____ range of x = _____
variance of x = _____
standard deviation (sd) of x = _____

b If you added the data for a 101st employee, which of the statistics of central tendency would change?

c If you analysed data for a sample of 200 employees, would you expect the sd to be larger than the sd calculated for the 100 shown here? Explain.

d How could you use some of the statistics calculated in (a) above to check whether x was distributed as a Poisson random variable? [this is a necessary but not a sufficient condition]

e In the column labeled "Expected number (f_{exp})", fill in the numbers "expected" in a Poisson distribution with the same mean. [see Table of Poisson Probabilities]

f Regardless of the "fit", give one reason why a Poisson distribution might not be suitable.

2. Refer to the figures from an article on dementia of the Alzheimer type.

Fig1 Is the median age of onset less than 68.4? _____
Fig2 Sketch the rough shape of the distribution.
Fig4 How could the presentation be changed to make the comparison of familial and non-familial cases easier?

3. In a recent review article on dementia in a neurology journal, the authors state

"The validation of clinical diagnosis of Alzheimer's disease continues to be based on the "gold standard" of pathological findings. Sensitivities between 55-82% have been confirmed pathologically, while the specificities ranged from 78-84% [refs]."

Their table giving the sensitivities and specificities found in various studies included footnotes defining the terms.

a Fill in the blanks in their definitions:

_____ is defined as the number of patients with both clinical and pathological diagnoses of a given "disease A", _____ by the total number of pathological cases with "disease A".

_____is defined as the ____ negative rate, i.e. the number of patients whose clinical and pathological diagnoses were "not A", _____ by the total number of patients with pathological diagnoses of "not A".

b Using the letters C+ and C- for the two "Clinical" diagnoses, and P+ and P- for the two "Pathological" states, write sensitivity and specificity in algebraic notation, i.e.
sensitivity = Prob(|)
specificity = Prob(|)

4. A variable X has a Gaussian distribution with a mean of 120 and a standard deviation of 10.

What is the -
a value of X corresponding to z = 0.5 ? _____
b value of z corresponding to X = 135 ? _____
c area under the curve between X = 120 and X = 125 ? _____

5. If we draw a simple random sample $y_1, y_2, \ldots, y_n$ of n "0/1" observations from a population with Prob(Y=1) = π, [exactly or approximately] what distribution(s) does $\sum y$ follow :

regardless of the size of n or π? _____

when π is small and n is large? _____

when π is near 0.5, and n > 10? _____

6. In which of the following situations would you <u>not</u> expect X to have a binomial distribution? Why?

X = #, out of 60 occupants of 30 randomly chosen cars, wearing seatbelts.

X = #, out of 60 occupants of 60 randomly chosen cars, wearing seatbelts.

X = #, out of a department's 10 microcomputers and 4 printers, that are going to fail in their first year.

X = #, out of simple random sample of 100 individuals, that are left-handed.

X = #, out of 5000 randomly selected from mothers giving birth each month in Quebec, who will test HIV positive.

7. You performed a t-test to compare the mean score on the Blessed-Tomlinson-Roth dementia severity index of familial cases of Alzheimer's disease with the mean score of non-familial cases (see Fig3 and Fig4 above). Your collaborator objects, saying that the data need to be normally distributed in order to use the t-test.

   a Defend your choice of the t-test.

   b Use symbols to state the null/alternative hypotheses:-
   H_0: H_{alt}:

   You find that the t-test (or since the df are 96, the z-test) gives a P-value of 0.15. Your collaborator, who is preparing the first draft of the manuscript writes:-

   "we found that the mean scores on the dementia index did not differ between the two types of cases".

   c Make any alterations/corrections to this presentation of the results.

8. Explain in a few sentences to someone who has never had a course in statistics

   what a confidence interval is: what a P-value is: