IAQS
INSTITUTE OF ACTUARIAL
& QUANTITATIVE STUDIES

Subject: **Probability and Statistics II**

Chapter: **Unit 2**

Category: **Practice Questions**

1. **IFoA CT3 April 2010 Question 4**

   It is assumed that the numbers of claims arising in one year from motor insurance policies for young male drivers and young female drivers are distributed as Poisson random variables with parameters λ_m and λ_f, respectively.

   Independent random samples of 120 policies for young male drivers and 80 policies for young female drivers were examined and yielded the following mean number of claims per policy in the last calendar year: $\bar{x}_m$= 0.24 and $\bar{x}_f$=0.15.

   Calculate an approximate 95% confidence interval for $\lambda_m - \lambda_f$, the difference between the respective Poisson parameters.

2. **IFoA CT3 April 2010 Question 7**

   An employment survey is carried out in order to determine the percentage, p, of unemployed people in a certain population in a way such that the estimation has a margin of error less than 0.5% with probability at least 0.95. In a similar study conducted a year ago it was found that the percentage of unemployed people in the population was 6%.

   Calculate the sample size, n, that is required to achieve this margin of error, by constructing an appropriate confidence interval (or otherwise).

3. **IFoA CT3 April 2011 Question 9**

Claims on a certain type of policy are such that the claim amounts are approximately normally distributed.

i. A sample of 101 such claim amounts (in £) yield a sample mean of £416 and sample standard deviation of £72. For this type of policy:

a) Obtain a 95% confidence interval for the mean of the claim amounts.
b) Obtain a 95% confidence interval for the standard deviation of the claim amounts.

The company makes various alterations to its policy conditions and thinks that these changes may result in a change in the mean, but not the standard deviation, of the claim amounts. It wants to take a random sample of claims in order to estimate the new mean amount with a 95% confidence interval equal to:

sample mean ±£10.

ii. Determine how large a sample must be taken, using the following as an estimate of the standard deviation:

a) The sample standard deviation from part (I).
b) The upper limit of the confidence interval for the standard deviation from part (i)(b).

iii. Comment briefly on your two answers in (ii)(a) and (ii)(b).

4. **IFoA CT3 October 2011 Question 9**

In a recent study of attitudes to a proposed new piece of consumer legislation ("proposal X") independent random samples of 200 men and 200 women were asked to state simply whether they were "for" (in favour of) , or "against", the proposal.

The resulting frequencies, as reported by the consultants who carried out the survey, are given in the following table:

	Men	*Women*
For	138	130
Against	62	70

i. Carry out a formal chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X.

Note: in this and any later such tests in this question you should state the P-value of the data and your conclusion clearly.

At a subsequent meeting to discuss these and other results, the consultants revealed that they had in fact stratified the survey, sampling 100 men and 100 women in England and 100 men and 100 women in Wales. The resulting frequencies were as follows:

	England		*Wales*	
	Men	*Women*	*Men*	*Women*
For	82	66	56	64
Against	18	34	44	36

A chi-squared test to investigate whether or not an association exists between gender and attitude to proposal X in England gives $\chi 2 = 6.653$, while an equivalent test for Wales gives $\chi 2 = 1.333$.

ii. a) Find the P-value for each of the chi-squared tests mentioned above and state your conclusions regarding possible association between gender and attitude to proposal X in England and in Wales.

b) Discuss the results of the survey for England and Wales separately and together, quoting relevant percentages to support your comments.

iii. A different survey of 200 people conducted in each of England, Wales, and Scotland gave the following percentages in favour of another proposal:

	England	*Wales*	*Scotland*
% in favour of proposal	62%	53%	58%

A chi-squared test of association between country and attitude to the proposal gives $\chi 2 = 3.332$ on 2 degrees of freedom, with P-value 0.189.

Suppose a second survey of the same size is conducted in the three countries and results in the same percentages in favour of the proposal as in the first survey. The results of the two surveys are now combined, giving a survey based on the attitudes of 1,200 people.

a) State (or find) the results of a second chi-squared test for an association between country and attitude to the proposal, based on the overall survey of 1,200 people.

b) Comment briefly on the results.

5. **IFoA CT3 April 2012 Question 12**
Consider a random sample $X_1, \dots, X_4$ of size k = 400 . Statistician A wants to use a χ2 -test to test the hypothesis that the distribution of X_i is a binomial distribution with parameters n = 2 and unknown p based on the following observed frequencies of outcomes of X_i:

Possible realisation of X_i	0	1	2
Frequency	90	220	90

i) Estimate the parameter p using the method of moments.

ii) Test the hypothesis that X_i has a binomial distribution at the 0.05 significance level using the data in the above table and the estimate of p obtained in part (i).

Statistician B assumes that the data are from a binomial distribution and wants to test the hypothesis that the true parameter is p_0 = 0.5.

iii) Explain whether there is any evidence against this hypothesis by using the estimate of p in part (i) and without performing any further calculations.

Statistician C wants to test the hypothesis that the random variables X_i have a binomial distribution with known parameters n= 2 and p = 0.5.

iv) Write down the null hypothesis and the alternative hypothesis for the test in this situation.

v) Carry out the test at the significance level of 0.05 stating your decision.

vi) Explain briefly the relationship between the test decisions in parts (ii), (iii) and (v), and in particular whether there is any contradiction.

6. IFoA CT3 October 2013 Question 6

A researcher obtains samples of 25 items from normally distributed measurements from each of two factories. The sample variances are 2.86 and 9.21 respectively.

i. Perform a test to determine if the true variances are the same.

ii. For each factory calculate central 95% confidence intervals for the true variances of the measurements.

iii. Comment on how your answers in parts (i) and (ii) relate to each other.

7. IFoA CT3 April 2014 Question 6

In an opinion poll, a sample of 100 people from a large town were asked which candidate they would vote for in a forthcoming national election with the following results:

Candidate	A	B	C
Supporters	32	47	21

i) Determine the approximate probability that candidate B will get more than 50% of the vote.

A second opinion poll of 150 people was conducted in a different town with the following results:

Candidate	A	B	C
Supporters	57	56	37

ii) Use an appropriate test to decide whether the two towns have significantly different voting intentions.

8. **IFoA CT3 September 2014 Question 6**

In a medical study conducted to test the suggestion that daily exercise has the effect of lowering blood pressure, a sample of eight patients with high blood pressure was selected. Their blood pressure was measured initially and then again, a month later after they had participated in an exercise programme. The results are shown in the table below:

Patient	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Before	155	152	146	153	146	160	139	148
After	145	147	123	137	141	142	140	138

i. Explain why a standard two-sample t-test would not be appropriate in this investigation to test the suggestion that daily exercise has the effect of lowering blood pressure.

ii) Perform a suitable t-test for this medical study. You should clearly state the null and alternative hypotheses.

9. **IFoA CT3 April 2015 Question 9**

An insurance company has calculated premiums assuming that the average claim size per claim for a certain class of insurance policies does not exceed £20,000 per annum. An actuary analyses 25 such claims that have been randomly selected. She finds that the average claim size in the sample is £21,000 and the sample standard deviation is £2,500. Assume that the size of a single claim is normally distributed with unknown expectation α and variance σ^2.

i) Calculate a 95% confidence interval for α based on the sample of 25 claims.

ii) Perform a test for the null hypothesis that the expected claim size is not greater than £20,000 at a 5% significance level.

iii) Discuss whether your answers to parts (i) and (ii) are consistent.

iv) Calculate the largest expected claim size, α_0, for which the hypothesis $\alpha \leq \alpha_0$ can be rejected at a 5% significance level based on the sample of 25 claims.

The insurer is also concerned about the number of claims made each year. It is found that the average number of claims per policy was 0.5 during the year 2011. When the analysis was repeated in 2012 it was found that the average number of claims per policy had increased to 0.6. These averages were calculated on the basis of random samples of 100 policies in each of the two years. Assume that the number of claims per policy per year has a Poisson distribution with unknown expectation λ and is independent from the number of claims in any other year or for any other policy.

v) Perform a test at 5% significance level for the null hypothesis that $\lambda = 0.6$ during the year 2011.

vi) Perform a test to decide whether the average number of claims has increased from 2011 to 2012.

10. IFoA CT3 October 2015 Question 9

A survey team is using satellite technology to measure the height of a mountain. This is an established technology and the variability of measurements is known. On each satellite pass over a mountain, they get a measurement that they know lies within ±5m of the true height with a 95% probability. The survey height is given by the mean of the measurements. They assume that all measurements are independent and follow a normal distribution with mean equal to the true height.

i) a) Show that the standard deviation of a single measurement is 2.551m.

b) Determine how many satellites passes over a mountain are required to have a 95% confidence interval for the true height with width less than 1m.

In a mountain range there are two summits which appear to have a similar height.

The team manages to get 20 measurements for each summit and finds there is a difference of 1.6m between the mean survey height of the two summits.

ii) Perform a statistical test of the null hypothesis that the summits' true heights are the same, against the alternative that they are different.

At the same time the team is testing a new system on these two summits. They again get 20 measurements on each summit with an estimated standard deviation on the first summit of 2.5m and on the second of 2.6m. This system also measures the difference in survey heights between the two peaks to be 1.6m.

iii) Perform a statistical test of the same hypotheses as in part (ii) when heights are measured by the new system, justifying any assumptions you make.

iv) Comment on your answers to parts (ii) and (iii).

11. IFoA CT3 April 2016 Question 10

Consider a large portfolio of insurance policies and denote the claim size (in £) per claim by X. A random sample of policies with a total of 20 claims is taken from this portfolio and the claims made for these policies are reported in the following table:

Claim i	1	2	3	4	5	6	7	8	9	10
Claim size x_i	130	164	170	173	173	175	177	183	183	184

Claim i	11	12	13	14	15	16	17	18	19	20
Claim size x_i	185	186	197	202	208	213	215	229	233	272

For these data: $\sum x_i$ = 3,852 and $\sum x_i^2$ = 759,348.

i) Calculate the mean, the median and the standard deviation of the claim size per claim in this sample.

ii) Determine a 95% confidence interval for the expected value E[X] based on the above random sample, stating any assumptions you make.

iii) Determine a 95% confidence interval for the standard deviation of X based on the above random sample.

iv) Explain briefly why the confidence interval in part (iii) is not symmetric around the estimated value of the standard deviation.

An actuary assumes that the number of claims from each policy has a Poisson distribution with an unknown parameter λ. In a new sample of 50 policies the actuary has observed a total of 80 claims yielding an estimated value of $\hat{\lambda}$ =1.6 for the parameter λ.

v) Determine a 95% confidence interval for λ using a normal approximation.

vi) Determine the smallest required sample size n for which a 95% confidence interval for λ has a width of less than 0.5. You should use the same normal approximation as in part (v), and assume that the estimated value of λ will not change.

Now assume that the true value of λ is 1.6 and the values calculated in part (i) are the true values. Also assume that all claims in the portfolio are independent and the claim sizes are independent of the number of claims.

vii) Determine the expected value and the standard deviation of the total amount of all claims from a portfolio of 5,000 insurance policies.

12. IFoA CT3 October 2016 Question 4

Consider two portfolios, A and B, of insurance policies and denote by X_A the number of claims received in portfolio A and by X_B the number of claims received in portfolio B during a calendar year. The observed numbers of claims received during the last calendar year are 134 for portfolio A and 91 for portfolio B. X_A and X_B are assumed to be independent and to have Poisson distributions with unknown parameters β_A and β_B.

Determine an approximate 99% confidence interval for the difference$\beta_A - \beta_B$. You may use an appropriate normal distribution.

13. IFoA CT3 April 2017 Question 6

We consider the impact that different types of cars have on the amount spent on fuel per month. Three different types of cars are considered: small, medium and large. For each type of car a group of 15 drivers are asked about the amount of money (in £) spent on fuel per month. The results are summarised in the following table

Type of car	*Small*	*Medium*	*Large*
Sample mean	70	75	83
Sample standard deviation	16	19	16

For example, the 15 drivers of medium sized cars spent on average £75 per month with a sample standard deviation of £19.

(i) Perform a one-way analysis of variance to test the hypothesis that the type of car has no impact on the monthly amount spent on fuel.

For some further investigation, only the difference between small and large cars is considered.

(ii) Determine a 95% confidence interval for the difference between the average amount spent on fuel for small cars and large cars, stating any assumptions you make.

(iii) Test the null hypothesis that the average fuel costs for small and large cars are the same at a 5% significance level against the alternative that the fuel costs for small and large cars are different.

14. IFoA CT3 April 2017 Question 9

A statistician is examining the survey methodology of a country's national statistics department. It conducts much of its data collection by telephoning individuals selected at random and asking them questions.

i) Comment on whether this methodology will give a random sample.

ii) Comment on whether this methodology will give a representative random sample of the population.

The department has been experimenting with surveying in person by visiting randomly selected individuals in their homes. To make this economical the department will only conduct surveys in a limited number of areas. It has asked the statistician to validate the effectiveness of its process.

For its first trial it conducts a small survey in two locations on the daily time spent accessing social media and gets the following results (in minutes).

Type of car	*Small*	*Medium*	*Large*
Sample mean	70	75	83
Sample standard deviation	16	19	16

The statistician assumes that the underlying population is normally distributed.

iii) a) Determine a 95% confidence interval for the ratio of sample variances.

b) Determine whether it is reasonable to assume that the variances are equal.

iv) Perform a test at a 5% significance level to investigate whether the means are the same against a two sided alternative.

The statistician then learns that there is an expectation that the mean of Area 2 is larger than the mean of Area 1.

v) Perform a test to investigate whether the means are the same against an appropriate alternative at the same significance level as in part (iv).

vi) Comment on the results of parts (iv) and (v).

15. IFoA CT3 October 2017 Question 5

In an election between two candidates A and B in a large district, a sample poll of 100 voters chosen at random, indicated that 55% were in favour of candidate A.

i) Calculate a 95% confidence interval for the proportion of all voters in favour of candidate A based on the above sample.

A candidate is elected if they win more than 50% of the votes. We want a test in which the alternative hypothesis is that support for candidate A is such that she will win the election.

ii) a) Write down the hypotheses for this test in terms of a suitable parameter.

b) Explain whether or not the confidence interval in part (i) can be used to test the hypothesis in part (ii)(a) at the 5% level of significance.

It has been reported in the news that a new poll estimates support for candidate A at 52%, with a margin of error of no more than ±2% with confidence 95%.

iii) Determine the minimum size of the sample of voters that was taken in this new poll.

16. IFoA CT3 April 2018 Question 10

A large pension scheme regularly investigates the lifestyle of its pensioners using surveys. In successive surveys it draws a random sample from all pensioners in the scheme and it obtains the following data on whether the pensioners smoke.

Survey 1: Of 124 pensioners surveyed, 36 were classed as smokers.
Survey 2: Of 136 pensioners surveyed, 25 were classed as smokers.

An actuary wants to investigate, using statistical testing at a 5% significance level, whether there have been significant changes in the proportion of pensioners, p, who smoke in the entire pension scheme.

i) Perform a statistical test, without using a contingency table, to determine if the proportion p has changed from the first survey to the second.

When a third survey is performed it is found that 26 out of the 141 surveyed pensioners, are smokers.

ii) Perform a statistical test using a contingency table to determine if the proportion p is different among the three surveys.

iii) a) Calculate the proportion of smokers in the third survey.
b) Comment on your answers to parts (i) and (ii).