OLLSCOIL NA hEIREANN, CORCAIGH
The National University of Ireland, Cork

COLAISTE NA hOLLSCOILE, CORCAIGH
University College, Cork

2014/2015

Semester 2 - Summer 2015

Second University Examination in Science
Financial Mathematics and Actuarial Science; Mathematical Sciences;
Higher Diploma in Statistics

ST2054 - Probability and Mathematical Statistics

Mr. P. Cooper
Professor F. O'Sullivan
Dr. M. Cronin
Mr. A. O'Leary

Time allowed: Three hours.

Statistical tables are available. A calculator may be used provided that it does not contain any information stored by any person prior to this examination.

Fifteen minutes of reading time are permitted prior to this examination.

PLEASE ANSWER ANY NINE QUESTIONS
All questions carry equal marks

PLEASE DO NOT TURN THIS PAGE UNTIL INSTRUCTED TO DO SO

PLEASE
ENSURE THAT YOU HAVE THE CORRECT EXAM PAPER

Question 1

(i) State the Axioms of Probability. [2 marks]

Axioms of probability:

1. $\forall \omega \in \Omega,\ P(\omega) \geq 0$
2. $P(\Omega) = 1$ **2 marks.**
3. $\forall \omega_i$ **disjoint:** $P(\cup_{i=1}^{n} \omega_i) = \sum_{i=1}^{n} P(\omega_i)$

(ii) Using the Axioms of Probability, prove that for any two events A and, B, the probability that exactly one of those two events will occur is given by the expression:

$$P(A) + P(B) - 2P(A \cap B)$$

[3 marks]

(ii) Expression for $P((A \backslash B) \cup (B \backslash A)) = P(A \backslash B) + P(B \backslash A) = P(A) + P(B) - 2P(A \cap B)$**:** **1 mark for each step.**

(iii) Consider 12 teams of two people (couples) at a particular time, including two particular teams designated as team A and team B. Suppose that at a later time six people (out of this group of 24) have been promoted, and it may be assumed that these promotions have occured randomly among the 24 people.

1. Find the probability that both people in team A have been promoted at the later time. [3 marks]
2. Find also the probability that the four people in teams A and B have been promoted at the later time. [2 marks]

Solution: Each of the 24 **people has an equal probability of** $\frac{6}{24} = 0.25$ **of being promoted. Thus:**

1. **P[both members of team A were promoted]**$= \frac{\binom{2}{2}\binom{22}{4}}{\binom{24}{6}} = \frac{5}{92} = 0.0543?$ **1 mark for each correct** $\binom{n}{x}$**, 1 mark for correct answer.**
2. **P[all members of teams A and B were promoted]**$= \frac{\binom{4}{4}\binom{20}{2}}{\binom{24}{6}} = \frac{5}{3542} = 0.00141$ **1 mark for correct** $\binom{n}{x}$**, 1 mark for correct answer.**

Question 2

Each of three different students, Joe, Hugh and Rachel are given the same problem to solve. They work on the problem independently and have probabilities 0.8, 0.7 and 0.6 of solving it, respectively.

(i) What is the probability that none of the students solve the problem? [4 marks]

Probability not solving also independent. **[2 marks]**

P[none solve problem]$=0.2 * 0.3 * 0.4 = 0.024$ **[2 marks]**

(ii) What is the probability that the problem will be solved by one or more of the students? [2 marks]

P[at least one solves problem]=1-P[none solve problem] **[1 mark]**

$=1 - 0.024 = 0.976$ **[1 mark]**

(iii) Given the problem was solved, what is the probability that the solution is due to Rachel alone? [4 marks]

P[Rachel solves problem alone | problem solved]=P[Rachel solves problem | problem solved]*P[others didn't solve problem] **[2 marks]**

P[Rachel solves problem | problem solved]=P[Rachel solves problem]/P[problem solved]$=\frac{0.6}{0.976} = 0.615$ **[1 mark]**

P[Rachel solves problem alone | problem solved]$=0.615 * 0.2 * 0.3 = 0.037$ **[1 mark]**

Question 3

The random variable X has an exponential distribution with *pdf* as follows:

$$f(y) = \lambda e^{-\lambda x}, \;\; x > 0$$

$$= 0, \;\; otherwise$$

(i) Showing your workings, find $P(Y > s|Y > t)$, for $s \geq t$. [3 marks]

Conditional - use Bayes. $P(Y > s|Y > t) = \frac{P(Y>t|Y>s)P(Y>s)}{P(Y>t)} = \frac{P(Y>s)}{P(Y>t)}$**.** **[2 marks]**

$\frac{P(Y>s)}{P(Y>t)} = \frac{e^{-\lambda s}}{e^{-\lambda t}} = e^{-\lambda(s-t)}$**.** **[1 marks]**

(ii) Derive an expression an expression for the *pdf* of Y, assuming that $Y \leq 200$. [3 mark]

$\frac{CDF(Y)}{CDF(200)}$ **for** $Y \leq 200$ **and** $0, \;\; o.w.$**.** **[2 marks]**

$\frac{CDF(Y)}{CDF(200)} = \frac{1-e^{-\lambda y}}{1-e^{-\lambda 200}} \implies pdf = \frac{d(\frac{1-e^{-\lambda y}}{1-e^{-\lambda 200}})}{dy} = \frac{\lambda e^{-\lambda y}}{1-e^{-\lambda 200}}$ **[1 mark]**

Suppose we have a Poisson process, $N(t)$, with rate λ.

(iii) Show that the waiting time (X) until the first event follows a negative exponential distribution. *(Hint: Consider the probability of there being 0 events in time t.)* [2 marks]

$P(N(t) = 0) = \frac{e^{-\lambda t}\lambda^0}{0!} = e^{-\lambda t}$**. This is the probability of the waiting time being greater than time** t**.** **[1 mark]**

$1 - CDF(X) = e^{-\lambda t} \implies CDF(X) = 1 - e^{-\lambda t}$**, cdf of a negative exponential distribution.** **[1 mark]**

(iv) Explain the relationship between the mean of the Poisson Distribution with rate λ and the mean of the associated distribution for the waiting time. [2 mark]

Suppose $\lambda = 3$ **every hour, then we'd expect 3 observations every hour and the average waiting time between them would be 20 minutes, i.e.,** $\frac{60}{3}$**.[2 mark]**

Question 4

Suppose we observe a random variable X which has a Multinomial distribution with 6 events. Each event has an equal probability of being observed, i.e., $p_i = \frac{1}{6}$, $i = 1, \ldots, 6$. Let X_i be the number of observations of event i.

We make 10 observations of the random variable. Assuming that all observations are independent:

(i) Calculate the probability of the following event: $X_1 = 1$, $X_3 = 3$ and $X_6 = 6$. [2 marks]

$\frac{10!}{1!0!3!0!0!6!}(\frac{1}{6}^{10}) = 0.0000389$

(ii) State the form of the marginal distribution for X_i. State the form of the conditional distribution for X_3, X_4, X_5 and X_6, given $X_1 = x_1$ and $X_2 = x_2$. [3 marks]

Marginal distribution: $X_i \sim Bin(10, \frac{1}{6})$ [1 mark]

Conditional distribution: $f(X_3, X_4, X_5, X_6 | X_1 = x_1, X_2 = x_2) = \frac{(10-x_1-x_2)!}{x_3!x_4!x_5!x_6!} \frac{p_3^{x_3} p_4^{x_4} p_5^{x_5} p_6^{x_6}}{(1-p_1-p_2)^{n-x_1-x_2}}$ [2 marks]

Each member of a group of n players roll a fair, 6-sided die. Stating any assumptions you make, find an expression for the following:

For any pair of players who throw the same number, the group scores one point.

(iii) Find the mean and variance of the total score of the group. [2 marks]

Assume pairwise independence. Then S is binomially distributed with $n = \binom{n}{2}$ and $p = \frac{1}{6}$. $E[S] = \binom{n}{2}(\frac{1}{6})$, $V[S] = \binom{n}{2}(\frac{1}{6})(\frac{5}{6})$. [2 marks]

For any pair of players who throw the same number, the group scores that number of points (e.g. player 1 rolls a 6 and player 2 rolls a 6 - the group scores 6).

(iv) Find the mean and variance of the total score of the group. [3 marks]

S is multinomially distributed with $n = \binom{n}{2}$ and $p_i = \frac{1}{36}$, $i = 1, \ldots, 6$, each being marginally distributed as a binomial with $n = \binom{n}{2}$ and $p = \frac{1}{36}$ and $p_0 = \frac{5}{6}$ the probability that they don't have the same number when they roll and score 0. [1 mark]

$E[S] = \sum_{i=1}^{6} E[X_i] = \binom{n}{2}(\frac{1}{36}) \sum_{i=1}^{6} i = \binom{n}{2}(\frac{3.5}{6})$ **[1 mark]**

$V[S] = \sum_{i=1}^{6} V[X_i] = \binom{n}{2}(\frac{35}{36^2}) \sum_{i=1}^{6} i^2 = \binom{n}{2} 2.4576$. **[1 mark]**

Question 5

The continuous random variable X has finite mean and variance. Furthermore, the function $g(x)$ takes positive values only, and $E[g(X)]$ exists.

(i) Show that for any $c>0$,

$$P[g(X) \geq c] \leq \frac{1}{c}E[g(X)]$$

[3 marks]

$E[g(x)] = \int g(x)f(x)dx = \int_{all\ x:\ g(x)\geq c} g(x)f(x)dx + \int_{all\ x:\ g(x)<c} g(x)f(x)dx \geq \int_{all\ x:\ g(x)\geq c} g(x)f(x)dx$ **[1 mark]**

$\int_{all\ x:\ g(x)\geq c} g(x)f(x)dx \geq \int_{all\ x:\ g(x)\geq c} cf(x)dx = c\int_{all\ x:\ g(x)\geq c} f(x)dx = cP[g(x) \geq c]$ **[1 mark]**

$\therefore \frac{1}{c}E[g(x)] \geq P[g(x) \geq c]$ **[1 mark]**

(ii) Use the above to deduce Chebyshev's Inequality:

$$P[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2},$$

where μ, σ^2 and the mean and variance of X. [2 marks]

Let $g(x) = (X-\mu)^2$ **and** $c = k^2\sigma^2 \implies g(x), c \geq 0$**.** **[$\frac{1}{2}$ mark]**

$E[g(x)] = E[(X-\mu)^2] = \sigma^2$ **[$\frac{1}{2}$ mark]**

$P[(X-\mu)^2 \geq k^2\sigma^2] \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$ **[$\frac{1}{2}$ mark]**

$\implies P[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}$ **[$\frac{1}{2}$ mark]**

(iii) Explain what is meant by convergence in probability for a sequence of random variables. [2 marks]

A sequence $\{X_n\}$ **of random variables converges in probability towards** X **if for all** $\boldsymbol{\varepsilon} > 0$**:** **[1 mark]**

$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1$ **[1 mark]**

(iv) Consider a sequence of n Bernoulli trials, with probability of success denoted p, and let X_n denote the number of successes. Show that the sequence $(\frac{X_n}{n})$ converges in probability to p. [3 marks]

Take $\varepsilon = k\sigma$**:** $P[|\frac{X_n}{n} - p| \leq k\sigma] \geq 1 - \frac{1}{k^2}$ **[1 mark]**

$X_n \sim Bin(n,p) \implies E[\frac{X_n}{n}] = \frac{np}{n} = p$ **and** $Var[\frac{X_n}{n}] = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$**.** **[1 mark]**

$\varepsilon = \frac{k}{\sigma} = \frac{k\sqrt{p(1-p)}}{\sqrt{n}} \implies \varepsilon^2 = k^2\frac{p(1-p)}{n} \implies \frac{1}{k^2} = \frac{p(1-p)}{\varepsilon^2 n}$

Thus: $P[|\frac{X_n}{n} - p| \leq k\sigma] \geq 1 - \frac{1}{k^2} \equiv P[|\frac{X_n}{n} - p| \leq \frac{k\sqrt{n}}{\sqrt{p(1-p)}}] \geq 1 - \frac{p(1-p)}{\varepsilon^2 n}$ **[1 mark]**

Taking the limit: $\lim_{n\to\infty} P[|\frac{X_n}{n} - p| \leq k\frac{\sqrt{n}}{\sqrt{p(1-p)}}] \geq 1$

Since a probability can't be greater than one: $\implies \lim_{n\to\infty} P[|\frac{p_n}{n} - p| \leq k\frac{\sqrt{n}}{\sqrt{p(1-p)}}] = 1$ **[1 mark]**

Question 6

The continuous random variables X and Y have a joint pdf $f(x, y)$ where:

$$f(x, y) = k(x^2 + 2xy) \ \ , for \ \ 0 < x < 1, \ \ 0 < y < 1$$

(i) Find k. [2 marks]

$k = \frac{1}{\int_0^1 \int_0^1 (x^2+2xy)dxdy} = \frac{6}{5}$ **[2 marks]**

(ii) Show that the marginal pdf of X is $\frac{6}{5}(x^2 + x)$. [2 marks]

$\int_0^1 \frac{6}{5}(x^2 + 2xy)dy = \frac{6}{5}(x^2 + x)$ **[2 marks]**

(iii) Show that the conditional pdf of Y is $\frac{x+2y}{1+x}$. [2 marks]

$f_{Y|X}(y|x) = \frac{f(x,y)}{f_x(x)}$ **[1 mark]**

$f_{Y|X}(y|x) = \frac{\frac{6}{5}(x^2+2xy)}{\frac{6}{5}(x^2+x)} = \frac{x+2y}{1+x}$ **[1 mark]**

(iv) Find $P[Y < 0.5|X = 0.5]$. [2 marks]

$\int_0^{0.5} f_{Y|X}(y|x = 0.5)dy = \int_0^{0.5} \frac{0.5+2y}{1.5}dy = \frac{1}{3}$ **[2 marks]**

(v) Find $P[X < Y]$. [2 marks]

$\int_0^1 \int_0^y f(x, y)dxdy = \frac{2}{5}$ **[2 marks]**

Question 7

(i) Show that for any random variables X and Y, $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$. [3 marks]

$Var(Y) = E[Y^2] - E[Y]^2$ **[1 mark]**

$E[E[Y^2|X]] - E[E[Y|X]]^2$**. Since** $E[E[Y^2|X]] = E[Var(Y|X) + E[Y|X]^2]$**, it follows that:** $E[E[Y^2|X]] - E[E[Y|X]]^2 = E[Var(Y|X)] + E[E[Y|X]^2] - E[E[Y|X]]^2$**[1 mark]**

Hence$Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$ **[1 mark]**

(ii) The number N of claims occuring on a collection of insurance policies has a Poisson distribution with parameter m. Let Y_i be the claim amount and assume the random variables $\{Y_i\}$ are independent and identically distributed as $N(\mu, \sigma^2)$. Derive the moment generating function of X_N, the total claim amount of the aggregate claims. [4 marks]

$E[e^{X_N s}] = E_N[E[e^{X_N s}|N]] = E_N[e^{N(s\mu + \frac{s^2\sigma^2}{2})}]$ **[2 marks]**

$E_N[e^{Ns'}]$, **where** $s' = s\mu + \frac{s^2\sigma^2}{2}$**, hence** $E_N[e^{Ns'}] = e^{m(e^{s'}-1)}$ **[2 marks]**

(iii) Calculate $Var(X_N)$.(*Hint: use (i).*) [3 marks]

$Var(X_N|N) = N^2\sigma^2$ **[1 mark]**

$E[Y|X] = N\mu$ **[1 mark]**

Hence$Var(X_N) = E[N^2\sigma^2] + Var(N\mu) = (m^2 + m)\sigma^2 + m\mu^2$ **[1 mark]**

Question 8

The random variable Z has a the Standard Normal distribution, and the random variable U has a Chi-square distribution with n degrees of freedom, and Z, U are independent. The random variable T is constructed as follows:

$$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

(i) Write down the joint probability density function (pdf) of Z and U. [3 marks],

(ii) Find the joint pdf of T and U. [3 marks]

(iii) Use this to find the pdf of T. Name the distribution T. [4 marks]

Solution:

$Z \sim N(0,1)$, $U \sim \chi^2_n$

$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ **[$\frac{1}{2}$ mark]**

$f(u) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} e^{-\frac{u}{2}}$ **[$\frac{1}{2}$ mark]**

Since Z and U are independent the joint pdf is the product of the pdf of each. [1 mark]

$f(z,u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} e^{-\frac{u}{2}}$ **[1 mark]**

(ii) Find the joint pdf of T and U. [3 marks]

Solution:

$t = h(z) = \frac{z}{\sqrt{\frac{U}{n}}} \implies z = h^{-1}(t) = t\sqrt{\frac{u}{n}}$ **[1 mark]**

$\therefore \frac{dh^{-1}(t)}{dt} = \sqrt{\frac{u}{n}}$

Change of variable:

$f(t,u) = f(h^{-1}(t),u)\frac{d(h^{-1}(t))}{dt} = (\frac{1}{\sqrt{2\pi}} e^{-\frac{(t\sqrt{\frac{u}{n}})^2}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} u^{\frac{n}{2}-1} e^{-\frac{u}{2}})\sqrt{\frac{u}{n}}$ **[1 mark]**

$f(t,u) = \frac{1}{\sqrt{n\pi}} \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n}{2})} u^{(\frac{n}{2}+\frac{1}{2})-1} e^{-\frac{u}{2}(1+\frac{t^2}{n})}$ **[1 mark]**

(iii) Use this to find the pdf of T. Name the distribution T. [4 marks]

Solution:

$f_t(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n+1}{2})} u^{\frac{n+1}{2}-1} e^{-\frac{u}{2}(1+\frac{t^2}{n})} du$ **[1 mark]**

Let $x = u(1 + \frac{t^2}{n})$**. Then** $dx = (1 + \frac{t^2}{n})du$ **and** $u = \frac{x}{1+\frac{t^2}{n}}$**.**

$f_t(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{t^2}{n})^{-\frac{n+1}{2}} \int_0^\infty \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n+1}{2})} x^{\frac{n+1}{2}-1} e^{-\frac{x}{2}} dx$

$\int_0^\infty \frac{1}{2^{\frac{n+1}{2}}\Gamma(\frac{n-1}{2})} x^{\frac{n+1}{2}-1} e^{-\frac{x}{2}} dx$ **is the pdf of a Chi-square distribution with** $n + 1$ **degrees of freedom and integrates to 1.** **[1 mark]**

$f_t(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1 + \frac{t^2}{n})^{-\frac{n+1}{2}}$ **[1 mark]**

This is the pdf of a Student's T distributed random variable with n **degrees of freedom.** **[1 mark]**

Question 9

In a sample of 320 families with 5 children, the numbers of families with a specific number of boys were as follows:

Number of Boys R	0	1	2	3	4	5
Number of Families	8	40	88	110	56	18

Assume the number of boys in a family with 5 children follows a Binomial Distribution, with the probability of a child being born a boy being p.

(i) Use the data to provide the Method of Moments estimator of the probability of a child being a boy. [2 marks] [2 marks]

$p = \frac{\#Boys}{\#Children} = \frac{860}{1600} = 0.5375$ **[2 marks]**

(ii) Evaluate the expected frequencies of the number of boys in a family with 5 children, assuming your estimated probability in **(i)** and also those expected frequencies if the probability of a boy is assumed to be 0.5. [4 marks]

Number of Boys	**Expected Frequency,** $p = 0.5375$	**Expected Frequency,** $p = 0.5$
0	$\binom{5}{0}(0.5375)^0(1-0.5375)^5 * 320$ =6.771871	$\binom{5}{0}(0.5)^0(1-0.5)^5 * 320$ =10
1	39.35006	50
2	91.4623	100
3	106.294	100
4	61.76545	50
5	14.35629	10

[4 marks]

(iii) Stating clearly your hypotheses, conduct Chi-Square tests to test the hypotheses that the binomial model is correct for each of the following cases, assuming that all observations are independent [4 marks]

1. The probability of a boy being 0.5 and
2. Using the estimate in (ii).

Number of Boys R	$(O-E)^2/E$ **with** $p = 0.5375$	$(O-E)^2/E$ **with** $p = 0.5$
0	0.22273038	0.40
1	0.01073498	2.00
2	0.13106529	1.44
3	0.12920990	1.00
4	0.53817124	0.72
5	0.92479295	6.40

[2 marks]

1. $H_0 : \#Boys \sim Bin(5, 0.5)$ vs H_1: $\#Boys\ not\ Bin(5, 0.5)$. Degrees of freedom = 5, Test statistics = $\sum(O-E)^2/E = 11.96$: compare with table to see greater than critical value $\implies$ we reject H_0. **[1 mark]**
2. $H_0 : \#Boys \sim Bin(5, 0.5375)$ vs H_1: $\#Boys\ not\ Bin(5, 0.5375)$. Degrees of freedom = 4, as p is estimated from data, Test statistics = $\sum(O-E)^2/E = 1.9567$: compare with table to see less than critical value $\implies$ we do not reject H_0. **[1 mark]**

Question 10

A random sample of size n from a $N(\mu, \sigma^2)$ distribution is to be used to test the following null and alternative hypotheses:

$$H_0 : \mu \leq 2000 \qquad H_A : \mu > 2000$$

Assuming the usual z-test will be applied, it is required to decide on the sample size n so that the risk of making a type 1 error is at most 1% and also so that if the value of μ is at least 2300, there is a 90% chance of rejecting the null hypothesis.

(i) Taking the population standard deviation to be 500, find the value of n. [4 marks]

$P[Reject|H_0] \leq .01 \implies P[\frac{\bar{X}-2000}{600}\sqrt{n} > c] \leq 0.01$, **if** $c = 2.32$. **[1 mark]**

Now, $P[Reject|\mu > 2300] = 0.1$. **Hence:** $P[Z + \frac{\mu-\mu_0}{\sigma}\sqrt{n} > c] = 0.9$, **so we obtain:** **[1 mark]**

$P[Z > 2.32 - \frac{\mu-\mu_0}{\sigma}\sqrt{n}] = 0.9$, **and thus:** $2.32 - \frac{300}{600}\sqrt{n} = -1.28$. **[1 mark]**

It follows that $\sqrt{n} = 7.2$, **hence** $n = 52$. **[1 mark]**

(ii) Compute the width of the 95% confidence interval for μ that would result from the study in **(i)**. [1 mark]

Width $= 2 * 1.96\frac{\sigma}{\sqrt{n}} = 232.3$ **[1 mark]**

Consider the one-way analysis of varaince (ANOVA) model with equal numbers of observations per treatment group:

$$Y_{ij} = \mu + \tau_i + e_{ij}, \begin{cases} i = 1, 2, ...n \quad j = 1, 2, ..., m \end{cases}$$

(iii) Give an explanation of each of the terms in the model and state the usual model assumptions. [2 marks]

Y_{ij} **is the** j^{th} **observation in the** i^{th} **category of the model. .**

μ **is the overall grand mean.**

τ_i **is the mean additional effect of the** i^{th} **category compared to the grand mean.**

e_{ij} **is the error term for the difference between the predicted and observed values for** Y_{ij}**.** **[1 mark]**

Assumptions:

1. **Independence of observations – this is an assumption of the model that simplifies the statistical analysis.**
2. **Normality – the distributions of the residuals are normal.**
3. **Equality of variances — the variance of data in groups should be the same.**
4. $\sum \tau_i = 0$ **[1 mark]**

(iv) Derive the least squares estimators for μ and τ_i. [3 marks]

Sum of Squares estimation:

$e_{ij} = Y_{ij} - (\mu + \tau_i) \implies e^2 = \sum_i \sum_j (Y_{ij} - \mu - \tau_i)^2$ **[1 mark]**

Sum of Squares estimation for $\hat{\mu}$: $\frac{\partial e^2}{\partial \mu} = -2\sum_i \sum_j (Y_{ij} - \mu - \tau_i) = 0 \implies \hat{\mu} = \frac{\sum_i \sum_j (Y_{ij})}{n} - \frac{\sum_j 1 \sum_i \tau_i}{n} = \frac{\sum_i \sum_j (Y_{ij})}{n} - 0 = \frac{\sum_i \sum_j (Y_{ij})}{n}$ **[1 mark]**

Sum of Squares estimation for $\hat{\tau}_i$: $\frac{\partial e^2}{\partial \tau_i} = -2\sum_j (Y_{ij} - \mu - \tau_i) = 0 \implies \hat{\tau}_i = \frac{\sum_j (Y_{ij})}{\sum_j 1} - \frac{\sum_j 1}{\sum_j 1}\hat{\mu} = \frac{\sum_j (Y_{ij})}{\sum_j 1} - \hat{\mu}$ **[1 mark]**

Formulae for ST2054

Law of Total Probability: $P(A) = \sum_{i=1}^{k} P(A|E_i)P(E_i)$ when $\cup_{i=1}^{k} E_i = \Omega$ and all E_i are mutually exclusive.

$X \sim Binomial(n, p)$	$E(X) = np$	$Var(X) = np(1-p)$
$U \sim Chi\text{-}Square(n)$	$E(X) = n$	$Var(X) = 2n$
$X \sim N(\mu, \sigma^2)$	$E(X) = \mu$	$Var(X) = \sigma^2$
$X \sim Poisson(m)$	$E(X) = m$	$Var(X) = m$

Probability Distributions and MGFs:

Exponential pdf $f(x) = \lambda e^{-\lambda x}$ **Uniform pdf** $f(x) = \frac{1}{b-a}$

Binomial pmf $P[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$

Multinomial pmf $P[X_1 = x_1, \ldots X_k = x_k] = \frac{n!}{x_1!, \ldots, x_k!} p_1^{x_1} \ldots p_k^{x_k}$, $\sum_{i=1}^{k} x_i = n$

Poisson pmf $P[X = k] = \frac{m^k e^{-m}}{k!}$ $MGF = e^{m(e^s - 1)}$

Gamma pdf $f(x) = \frac{1}{\Gamma(\alpha)} (\frac{x}{\beta})^{\alpha-1} e^{\frac{-x}{\beta}} \frac{1}{\beta}$ $MGF = \frac{1}{(1-\beta s)^\alpha}$

Normal pdf $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$ $MGF = e^{s\mu + \frac{1}{2}s^2\sigma^2}$

Chebyshev's Inequality:

$p[|X - \mu| > k\sigma] \leq \frac{1}{k^2}$, where μ and σ^2 are the mean and variance of X.

Linear Regression

$\hat{\beta}_1 = \frac{SXY}{SXX} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}$ $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

Analysis of Variance:

$SST = (\sum_i \sum_j (y_{ij} - \bar{y})^2) = \sum_i \sum_j (y_{ij})^2 - \frac{(y_{..})^2}{n}$ $SSB = [\sum_i \frac{(y_{i.})^2}{n_i}] - \frac{(y_{..})^2}{n}$

Where $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$ and $y_{..} = \sum_i \sum_j y_{ij}$