OCR Further Maths AS-level

Statistics

Formula Sheet

Provided in formula book | Not provided in formula book

Probability

Permutations and Combinations

Number of permutations of n distinct objects	$n! = n \times (n-1) \times (n-2) \ldots \times 2 \times 1$
Number of combinations when choosing r objects from n objects	$\binom{n}{r} = {}^nC_r = \frac{n!}{r!\,(n-r)!}$
Number of permutations of a subset of size r from a set of n distinct objects	${}^nP_r = {}^nC_r \times r! = \frac{n!}{(n-r)!}$
Number of permutations from n objects with r_A of type A, r_B of type B etc.	$\frac{n!}{r_A!\,r_B!\ldots}$

Probability Problems

$n(A)$	The number of ways of making a choice about A
Product principle	$n(A \text{ and } B) = n(A) \times n(B)$
Addition principle (given that A and B are mutually exclusive)	$n(A \text{ or B}) = n(A) + n(B)$
Counting principles in probability (given that all outcomes are equally likely)	$P(A) = \frac{\text{number of outcomes in which } A \text{ occurs}}{\text{total number of possible outcomes}}$

Discrete Random Variables

Average and Spread of Discrete Random Variables

For the random variable taking the values x_i with $P(X = x_i) = p_i$	
Expectation	$\mu = E(X) = \sum x_i p_i$
Variance	$\sigma^2 = Var(X) = \sum (x_i - \mu)^2 p_i$ $= \sum x_i^2 p_i - \mu^2$ $= E(X^2) - (E(X))^2$

Linear Coding with the Mean and Variance

For $Y = aX + b$:

$E(Y) = aE(X) + b$

$Var(Y) = a^2 Var(X)$

Binomial Distribution

For $X \sim B(n, p)$
$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
$E(X) = np$
$Var(X) = np(1-p)$

Discrete Uniform Distribution

For $X \sim U(n)$
$P(X = x) = \frac{1}{n}$ for $x = 1, 2, \dots n$
$E(X) = \frac{n+1}{2}$
$Var(X) = \frac{n^2 - 1}{12}$

Geometric Distribution

For $\boldsymbol{X \sim Geo(p)}$
$P(X = x) = p(1-p)^{x-1}$ for $x = 1, 2, 3 \ldots$
$P(X > x) = (1-p)^x$
$E(X) = \frac{1}{p}$
$Var(X) = \frac{1-p}{p^2}$

Poisson Distribution

For $\boldsymbol{X \sim Po(\lambda)}$
$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2 \ldots$
$E(X) = \lambda$
$Var(X) = \lambda$

$X \sim Po(\lambda)$ $Y \sim Po(\mu)$	When $Z = X + Y$, $Z \sim Po(\lambda + \mu)$

Chi-squared Tests

Contingency Tables

Expected value in cell i	$E_i = \frac{\text{row total } \times \text{column total}}{\text{overall total}}$
Chi-squared value	$\chi^2_{\text{calc}} = \frac{\sum(O_i - E_i)^2}{E_i}$
Degrees of freedom in an $n \times m$ contigency table	$v = (n-1)(m-1)$

Hypothesis Testing

If variables are independent and $E_i > 5$ for all i

$$\chi^2_{\text{calc}} = \frac{\sum(O_i - E_i)^2}{E_i} \approx \chi^2_v$$

Yates' Correction

When $v = 1$

$$\chi^2_{\text{Yates}} = \frac{\sum |(O_i - E_i| - 0.5)^2}{E_i}$$

Goodness of Fit Test

$$v = \text{number of bins} - \text{number of constraints}$$

Correlation

Pearson's Product Moment Correlation Coefficient

For a set of bivariate date with variables X and Y
$r = \dfrac{\sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \dfrac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \dfrac{(\sum y_i)^2}{n}\right)}} = \dfrac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$
$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \dfrac{(\sum x_i)^2}{n}$
$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \dfrac{(\sum y_i)^2}{n}$
$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \dfrac{\sum x_i \sum y_i}{n}$

Spearman's Rank Correlation Coefficient

$r_s = 1 - \dfrac{6\sum d^2}{n(n^2 - 1)}$
where d = difference in ranks and n = number of data pairs

Linear Regression

Least Squares Regression Line

$\boldsymbol{y = ax + b}$
$a = \bar{y} - b\bar{x}$
$b = \dfrac{S_{xy}}{S_{xx}} = \dfrac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$