OCR Further Maths A 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 $x = 1, 2, \ldots n$
$E(X) = \frac{n+1}{2}$
$Var(X) = \frac{n^2 - 1}{12}$

Geometric Distribution

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

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

Continuous Random Variables

Probability Density Function

For a probability function $f(x)$
$\int_{-\infty}^{\infty} f(x)\,dx = 1$
$f(x) \geq 0$
$P(a < X < b) = \int_a^b f(x)\,dx$

Mean and Variance of a Continuous Probability Distribution

$\mu = E(X) = \int_{-\infty}^{\infty} x\,f(x)\,dx$
$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$ $= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

Mean and Variance of a Function of a Continuous Variable

$E(aX + c) = aE(X) + c$
$Var(aX + c) = a^2 Var(X)$
$E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x) f(x)\,dx$

Cumulative Distribution Function

$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\,dt$
Probability density function $= f(x) = \frac{d}{dx}F(x)$

Median, Mode, and Percentiles of a Continuous Probability Distribution

Median (m)	$\int_{-\infty}^{m} f(x)\,dx = \frac{1}{2}$
Mode	Value of x when $f(x)$ is at the maximum
Lower Quartile (Q_1)	$\int_{-\infty}^{Q_1} f(x)\,dx = \frac{1}{4}$
Upper Quartile (Q_3)	$\int_{-\infty}^{Q_3} f(x)\,dx = \frac{3}{4}$

Continuous Uniform Distribution

If X follows a uniform distribution between $[a, b]$
$f(x) = \frac{1}{b-a}$ for $a < x < b$
$E(X) = \frac{a+b}{2}$
$Var(X) = \frac{(b-a)^2}{12}$

Exponential Distribution

For $X \sim Exp(\lambda)$
$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$
$E(X) = \frac{1}{\lambda}$
$Var(X) = \frac{1}{\lambda^2}$
$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$

Normal Distribution

For $X \sim N(\mu, \sigma^2)$
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
$E(X) = \mu$
$Var(X) = \sigma^2$

Linear Combinations of Random Variables

Linear Combination of Any Independent Variables

Where X and Y are independent random variables
$E(aX + bY + c) = aE(X) + bE(Y) + c$
$Var(aX + bY + c) = a^2Var(X) + b^2Var(Y)$

Hypothesis Tests and Confidence Intervals

Expectation and Variance of the Sample Mean

$E(\bar{X}) = \mu$
$Var(\bar{X}) = \frac{\sigma^2}{n}$

Sampling Distribution

For $X \sim N(\mu, \sigma^2)$
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Unbiased Estimate of the Population Mean and Variance

$\bar{x} = \frac{\sum x}{n}$
$s^2 = \frac{n}{n-1}\left(\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2\right)$

Central Limit Theorem

The mean of any distribution with $E(X) = \mu, Var(X) = \sigma^2$ and $n > 25$ can be approximated to have a normal distribution
$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Confidence Intervals

For a population with known variance and normally distributed sample mean	
$c\%$ confidence interval for population mean (μ)	$\left(\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}}\right)$ Where $z = \Phi^{-1}\left(0.5 + \frac{\frac{1}{2}c}{100}\right)$
Width of confidence interval	$2z\frac{\sigma}{\sqrt{n}}$

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 $v = 1$
$\chi^2_{\text{Yates}} = \frac{\sum \lvert(O_i - E_i\rvert - 0.5)^2}{E_i}$

Goodness of Fit Test

$v = \text{number of bins} - \text{number of constraints}$
Expected frequency in bin $A =$ $\text{probability of being in bin } A \times \text{total observed frequency}$

Non-parametric Tests

Single Sample Sign Test

For a sample of size n
1. Count the number of observed values above the median stated in H_0. 2. Find the probability of this value or higher using $B(n, 0.5)$.

Single Sample Wilcoxon Signed Rank Test

1. Assigned ranks according to the differences from median in the order of increasing size.
2. Add a negative sign to the ranks for values below median.
3. Test statistic T is the smaller of W_+ (sum of positive ranks) and W_- (sum of negative ranks).

Wilcoxon Rank Sum Test

For two samples of sizes m and n where $m \leq n$
$R_m =$ sum of ranks of items in sample of size m where a lower rank is given to smaller values
Test statistic $= W =$ smaller of R_m and $m(m+n+1) - R_m$

Approximate Distribution for Large Samples

Wilcoxon Signed Rank test	$T \sim N\left(\frac{1}{4}n(n+1), \frac{1}{24}n(n+1)(2n+1)\right)$
Wilcoxon Rank Sum test (sample of sizes m and n, where $m \leq n$)	$W \sim N\left(\frac{1}{2}m(m+n+1), \frac{1}{12}mn(m+n+1\right)$

Correlation

Pearson's Product Moment Correlation Coefficient

For a set of bivariate date with variables X and Y

$$r = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right)\left(\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right)}} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

$$S_{xx} = \sum (x_i - \bar{x})^2 = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

$$S_{yy} = \sum (y_i - \bar{y})^2 = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

$$S_{xy} = \sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{n}$$

Spearman's Rank Correlation Coefficient

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where d = difference in ranks and n = number of data pairs

Linear Regression

Least Squares Regression Line

$$\boldsymbol{y = ax + b}$$

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$