AQA Further Maths A Level

Statistics

Formula Sheet

Provided in formula book	Not provided in formula book

Discrete Random Variables and Expectation

Measures of Average and Spread for a Discrete Random Variable

Expectation	$E(X) = \sum_i x_i p_i$
Variance	$Var(X) = \sum_i x_i^2 p_i - \left(E(X)\right)^2$ $= E(X^2) - (E(X))^2$
Standard Deviation	$\sigma = \sqrt{Var(X)}$
Mode	The value of X which has the largest probability
Median	$P(X \leq M) = 0.5$ $P(X \geq M) = 0.5$

Functions of Discrete Random Variables

$E(aX + b) = aE(X) + b$
$Var(aX + b) = a^2 Var(X)$

Expectation of a General Function of a Discrete Random Variable

$$E(g(X)) = \sum g(x_i) p_i$$

Discrete Uniform Distribution

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

Poisson Distribution

$X \sim Po(\lambda)$	$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$
	$Mean = \lambda$
	$Variance = \lambda$

Sum of Independent Poisson Distribution

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

Type I and Type II Errors

Defining Type I and Type II Errors

Type I Error	H_0 is rejected when it is true
Type II Error	H_0 is not rejected when it is false

In hypothesis testing:
$\text{significance level} = P(\text{type I error}) = P(\text{rejecting } \mathrm{H}_0 \mid \mathrm{H}_0 \text{ is true})$

Power of a Test and Using Type II Errors

In hypothesis testing:
$P(\text{type II error}) = P(\text{not rejecting } \mathrm{H}_0 \mid \text{ a specific alternative to } \mathrm{H}_0)$
$\text{Power} = 1 - P(\text{type II error})$

Continuous Random Variables

Probability Density Function (PDF)

For continuous random variable X with pdf $f(x)$:
$P(a < x < b) = \int_a^b f(x)\ dx$
$\int_{-\infty}^{\infty} f(x)\ dx = 1$
$f(x) \geq 0$ for all x

Median and Quartiles of a Given Probability Density Function

Lower quartile (Q_1)	$\int_{-\infty}^{Q_1} f(x)\ dx = \frac{1}{4}$
Median	$\int_{-\infty}^{M} f(x)\ dx = \frac{1}{2}$
Upper quartile (Q_3)	$\int_{-\infty}^{Q_3} f(x)\ dx = \frac{3}{4}$

Expectation and Variance of a Continuous Random Variable

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

Functions of a Continuous Random Variable

Linear Transformations	$E(aX+b) = aE(X) + b$ $Var(aX+b) = a^2 Var(X)$
General Function	$E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x)f(x)\ dx$

Expectation and Variance of the Sum of Two Independent Random Variables

$E(X+Y) = E(X) + E(Y)$
$Var(X+Y) = Var(X) + Var(Y)$

Cumulative Distribution Function

For a continuous distribution:
$P(X \leq x) = F(x) = \int_{-\infty}^{x} f(t)\, dt$
$f(x) = \frac{d}{dx} F(x)$

Rectangular Distribution

For the variable X following a rectangular distribution over $[a, b]$
$f(x) = \frac{1}{b-a}$ for $a < x < b$
$E(X) = \frac{a+b}{2}$
$Var(X) = \frac{(b-a)^2}{12}$

Chi squared Tests for Association

Chi squared Values and Degrees of Freedom

Expected frequency	$E_i = \frac{row\ total \times column\ total}{overall\ total}$
Chi squared value	$\chi^2_{calc} = \sum_i \frac{(O_i - E_i)^2}{E_i}$
Degrees of freedom	$v = (no.of\ rows - 1)(no.of\ columns - 1)$

Condition on Expected Frequency

For $E_i > 5$:	$\chi^2_{calc} \sim \chi^2_v$

Yates' Correction

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

Exponential Distribution

Probability Density Function and Cumulative Distribution Function

For $X \sim \exp(\lambda)$
$f(x) = \lambda e^{-\lambda x}$ for $x > 0$
$F(x) = 1 - e^{-\lambda x}$

Mean and Variance of an Exponential Distribution

Mean	$E(X) = \frac{1}{\lambda}$
Variance	$Var(X) = \frac{1}{\lambda^2}$

Inference – One Sample t-Distribution

Testing for the Mean of a Normal Distribution with Unknown Variance

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Confidence Intervals

Confidence interval for population mean μ	$(\underline{x} - z\frac{\sigma}{\sqrt{n}}, \underline{x} + z\frac{\sigma}{\sqrt{n}})$
Width of confidence interval	$2z\frac{\sigma}{\sqrt{n}}$

Symmetric Confidence Intervals from Small Samples Using the t-Distribution

For normally distributed sample with a small sample size and estimated variance s^2, the $c\%$ CI for population mean:
$\bar{x} - t\frac{s}{\sqrt{n}} < \mu < \bar{x} + t\frac{s}{\sqrt{n}}$ where $P(t_v < t) = 0.5 + \frac{\frac{1}{2}c}{100}$