Stats 2 (AQA) Formula

(everything in blue in formula book)

Expectation (mean) **& Variance of Random Variable** from Probability Distribution

$$E(X) = \mu = \sum x_i p_i \qquad Var(X) = \sigma^2 = \sum {x_i}^2 p_i - \mu^2$$

$$= E(X^2) - \mu^2$$

($\mu \neq \bar{x}$. μ is *theoretical mean*, $\bar{x}$ is *sample mean*.)

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

Poisson distribution (a discrete distribution)

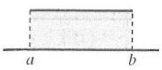
$$X \sim Po(\lambda) \qquad P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad \mu = \lambda,\ \ \sigma^2 = \lambda$$

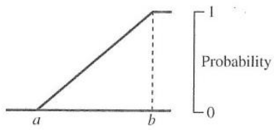
$$p_x = \frac{\lambda}{x} p_{x-1}$$

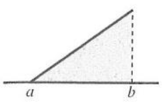
If $X \sim Po(\lambda_1)$ and $Y \sim Po(\lambda_2)$, then the distribution of $(X + Y) \sim Po(\lambda_1 + \lambda_2)$.

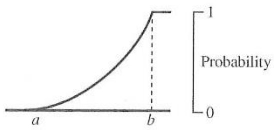
Continuous Random Variables

Probability Density Function (pdf) = f Cumulative Distribution Function (cdf) = F









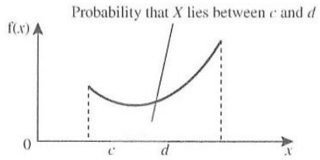
$$f(x) = F(x)\frac{d}{dx} \qquad F(x) = \int_{-\infty}^{x} f(t)\, dt$$

(replace lower limit with lower limit of function when calculating)

$$F(-\infty) = 0 \qquad F(+\infty) = 1$$

Three methods to find specific probabilities of X:

1. $P(c < X < d) = \int_c^d f(x)\,dx$ (ie integrate to find area between two values)



2. Having integrated, evaluate between limits (similar to finding normal distribution probabilities).

$$P(c < X < d) = F(d) - F(c)$$
$$F(X > d) = 1 - F(d)$$

3. Use geometry of graph of $f(x)$ (where possible, splitting it into triangles, trapezia etc).

To find median value, m, solve for m either:

$$0.5 = \int_{-\infty}^{m} f(x)dx \quad \text{or} \quad 0.5 = F(m)$$

To find, for example, the value at 95th percentile, solve for d;

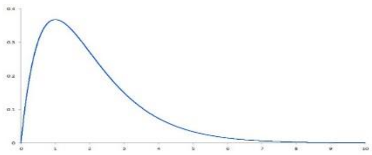
$$0.95 = \int_{-\infty}^{d} f(x)dx \quad \text{or} \quad 0.95 = F(d)$$

Mean and variance…

$$E(X) = \mu = \int_{-\infty}^{\infty} xf(x)\,dx \qquad E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx \qquad E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

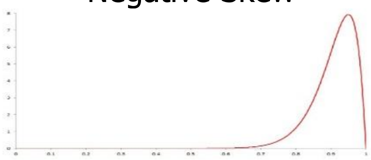
If distribution is symmetrical then $E(X) = central\ value$.

Positive Skew



Mode < Median < Mean

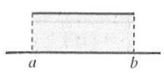
Negative Skew



Mean < Median < Mode

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

Rectangular (Uniform) Distribution



$$f(x) = \frac{1}{b-a} \qquad F(x) = \frac{x-a}{b-a}$$

$$E(X) = \frac{1}{2}(a+b) \qquad Var(X) = \frac{1}{12}(b-a)^2$$

Estimation

The t distribution (two random variables, $\bar{X}$ and S, i.e. unknown population variance).

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

$$T \sim t_\nu \text{ or } T \sim t_{n-1}$$

$$\nu = number\ of\ degrees\ of\ freedom = n - 1$$

(parameter ν pronounced 'nu')

Confidence intervals given by

$$\bar{x} \pm \text{"c"} \frac{s}{\sqrt{n}} \qquad \bar{x} \pm \text{"c"} \sqrt{\frac{s^2}{n}}$$

Hypothesis Testing

Null Hypothesis	Alternative Hypothesis	
$H_0: \mu = 435$	$H_1: \mu < 435$ One tailed	$H_1: \mu \neq 435$ Two tailed

Acceptance of null hypothesis does not mean it is true, rather that the data provides no evidence to prefer the alternative.

Type 1 error – to reject H_0 (and accept H_1) when H_0 is actually true.
Type 2 error – to reject H_1 (and accept H_0) when H_1 is actually true.

$$P(Type\ 1\ error) = significance\ level$$

Test Procedure:

1. Write down the two hypotheses.
2. Identify an appropriate test statistic and the distribution of the corresponding random variable.
3. Identify the significance level (usually given). This is also $P(Type\ 1\ error)$.
4. Determine the critical region (should be done before collecting data).
5. Calculate the value of the test statistic.
6. Determine *and clarify in context* the outcome of the test.

χ^2 Contingency Tables Tests

Part A – The χ^2 Distribution

$$X^2 \sim \chi^2_v$$

$$\mu = v \qquad \sigma^2 = 2v$$

(where $v = m - 1$)

$$X^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i}$$

(where m = number of different possible outcomes)

All expected frequencies must be greater than 5

Part B – Contingency Tables

Associated vs independent

To calculate expected frequencies from observed frequencies use

$$\frac{row\ total \times column\ total}{grand\ total}$$

$$v = (r-1)(c-1)$$

Yates's correction for a 2x2 table ($v = 1$):

$$X_C^2 = \sum_{1}^{4} \frac{(|O_I - E_i| - 0.5)^2}{E_i}$$
$$= \frac{N}{mnrs}\left(|ad - bc| - \frac{N}{2}\right)^2$$

Where...

a	b	m
c	d	n
r	s	N