Statistics 3

Edexcel IAL S3 (WST03/01); combinations, sampling, estimators, intervals and tests, chi-squared, Spearman

Formulae to memorise

Family	**Must know**
combinations	$aX \pm bY$: variance $a^2\sigma_x^2 + b^2\sigma_y^2$
sample mean	$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
unbiased variance	$s^2 = \frac{\sum x^2 - n\bar{x}^2}{n-1}$
chi-squared	$\sum \frac{(O-E)^2}{E}$
Spearman	$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$

Differences of Normals

$$P(|C_1 - C_2| > 0.02) = 2(1 - \Phi(0.4714)) = 0.638$$

- Means add or subtract.
- Variances always add.

Worked: a sum against a multiple

1. **Three bags** $\sum S \sim N(54, 0.0027)$: $P = 0.0274$
2. **Rearrange** $\sum_{28} S + \sum_{5} C - 29 P_{1} < 190$
3. **Moments** mean 194, var 33.67: $P = 0.245$

Decision

Square every coefficient on a variance.

Four ways to sample

> Simple random gives every sample of size n an equal chance.

- Know one advantage and one drawback.
- Systematic, 20 from 360: every 18th.

Worked: bias and unbiased estimators

1. **R** $E(R) = \frac{5}{4}\mu$: **bias** $\frac{\mu}{4}$
2. **S and T** $a - b = 2,\ 14a + 6b = 20$
3. **Solve** $a = 1.6,\ b = -0.4$

Decision

Unbiased: the expectation equals the parameter.

The standard error

$$\mathrm{SE} = \frac{\sigma}{\sqrt{n}}$$

- Same centre, narrower spread.
- s^2 divides by $n - 1$.

Worked: recovering the parameters

1. **z-values** -0.4 and 2.1
2. **Equations** $\mu - 0.4\,\mathrm{SE} = 45.2$, $\mu + 2.1\,\mathrm{SE} = 75.2$
3. **Solve** SE 12, $\mu = 50$: $P(\overline{X} > 59) = 0.2266$

Decision

$\sigma = 12\sqrt{n}$

What 95% confidence means

> The procedure captures μ in 95% of samples.

- $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$.
- 99%: 2.5758.

Two samples, with a gap

$z = 2.46 > 1.6449$: reject H_0

- H_0: $\mu_x - \mu_y = 200$, H_1: > 200, 5%.
- Large n: $\sigma^2 = s^2$, CLT.

Worked: the two-sample statistic

1. **Red mean** $\bar{y} = \frac{23200}{80} = 290$
2. **Red variance** $s_y^2 = \frac{6741351 - 80(290)^2}{79} = 169$
3. **Statistic** $z = \frac{210 - 200}{\sqrt{1156/80 + 169/80}} = \frac{10}{4.0698}$

Decision

$z = 2.46 > 1.6449$: reject H_0

Does a binomial fit?

$X^2 = 10.95 > 9.488$ on 4 df: reject

- Combine cells with E below 5 first.
- Estimating p costs a degree of freedom.

Worked: a contingency table

1. **Expected** $E = \frac{35 \times 36}{250} = 5.04$, contributes 4.88
2. **Total** $X^2 = 4.88 + 7.444 = 12.325$
3. **Compare** 3×3: 4 df, $12.325 > 9.488$

Decision

Evidence of association between engine type and timing.

Spearman's test

$0.6 < 0.7455$: do not reject H_0

- $H_0 \colon \rho = 0$, $H_1 \colon \rho > 0$, 1%.
- $n = 10$, $\sum d^2 = 66$.

Check yourself

Question	**Answer**
Variance of $X - Y$?	$\sigma_x^2 + \sigma_y^2$
SE of $\overline{X}$?	$\frac{\sigma}{\sqrt{n}}$
Expected value below 5?	combine cells first
df, binomial fit, p estimated?	cells -2
Spearman without ties?	$1 - \frac{6\sum d^2}{n(n^2-1)}$