Exam-style practice: A Level

1 a Let X_m denote the distribution of males and X_f denote the distribution of females.
Then define the random variable
$Y = \sum_{i=1}^{5} X_{mi} + \sum_{j=1}^{2} X_{fj}$.
The expected value of Y is:

$$E(Y) = E\left(\sum_{i=1}^{5} X_{mi} + \sum_{j=1}^{2} X_{fj}\right)$$
$$= E\left(\sum_{i=1}^{5} X_{mi}\right) + E\left(\sum_{j=1}^{2} X_{fj}\right)$$
$$= 5E\left(X_m\right) + 2E\left(X_f\right)$$
$$= 5\times 90 + 2\times 60$$
$$= 570$$

The variance is:

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^{5} X_{mi}\right) + \text{Var}\left(\sum_{j=1}^{2} X_{fj}\right)$$
$$= \sum_{i=1}^{5} \text{Var}(X_{mi}) + \sum_{i=1}^{2} \text{Var}(X_{fj})$$
$$= 5\,\text{Var}(X_m) + 2\,\text{Var}(X_f)$$
$$= 5\times 10^2 + 2\times 5^2$$
$$= 550$$

So $Y \sim N(570, 550)$.
In order to calculate $P(Y > 560)$ we standardise to

$$P\left(Z > \frac{560-570}{\sqrt{550}}\right) = P\left(Z > -0.43\right)$$

where $Z = \dfrac{Y-570}{\sqrt{550}}$

In order to use the statistic tables we reformat our calculation to:

$$P\left(Z > -0.43\right) = P(Z < 0.43)$$
$$= 1 - P(Z > 0.43)$$

Thus using the tables we get:

$$P(Z < -0.43)$$
$$= 1 - P(Z > 0.43)$$
$$= 1 - 0.335$$
$$= 0.665$$

1 b Using the same notation as the previous question, define a new random variable as $W = 1.4X_f - X_m$.

$$E(W) = E(1.4X_f - X_m)$$
$$= 1.4E(X_f) - E(X_m)$$
$$= 1.4\times 60 - 90$$
$$= -6$$

$\text{Var}(W) = 1.4^2 \times 5^2 + 10^2 = 149.$
So we have $W \sim N\left(-6, 149\right)$.
We want to find $P(W > 0)$ and we standardise then use the tables in order to get:

$$P\left(V > \frac{0 - -6}{\sqrt{149}}\right) = P\left(V > 0.49\right) = 0.312.$$

2 a Let $\sigma_s^{\,2}$ denote the variance of the saplings with fertiliser and $\sigma_l^{\,2}$ denote the variance of the saplings without fertiliser.
Our hypotheses are:
$H_0 : \sigma_s^{\,2} = \sigma_l^{\,2}$
and
$H_1 : \sigma_s^{\,2} \neq \sigma_l^{\,2}$.
The significance level is 10% (5% at each tail) with degrees of freedom:
$v_l = 13 - 1 = 12,$
$v_s = 10 - 1 = 9.$
$s_l^{\,2} = 46.7856,$
$s_s^{\,2} = 27.9841.$
From the table, we find the critical value of $F_{12,9}\left(0.05\right) = 3.07.$

The test statistic is $\dfrac{s_l^{\,2}}{s_s^{\,2}} = 1.67$.

$1.67 < 3.07$
So there is insufficient evidence to reject H_0 and we may assume that the two populations have equal variance.

2 b Let μ_f denote the mean height of saplings with fertiliser and μ_o denote the mean height of saplings without fertiliser.
The null hypothesis is that the difference between the means is 0. The alternative hypothesis is that the difference is non-zero:

$H_0 : \mu_f - \mu_o = 0$

$H_1 : \mu_f - \mu_o \neq 0.$

We have standard deviations and sample sizes of:

$s_f = 5.29,$

$n_f = 10,$

$s_o = 6.84,$

$n_o = 13.$

We calculate an unbiased estimate of the population variance σ^2 using both samples:

$$s_p{}^2 = \frac{(n_f - 1)s_f{}^2 + (n_o - 1)s_o{}^2}{(n_f - 1) + (n_o - 1)}$$

$$= \frac{(10-1)\times 5.29^2 + (13-1)\times 6.84^2}{(10-1)+(13-1)}$$

$$= \frac{9\times 5.29^2 + 12\times 6.84^2}{21}$$

$$= 38.73$$

So:

$$t = \frac{\bar{x}_f - \bar{x}_o - (\mu_f - \mu_o)}{s_p\sqrt{\frac{1}{n_f} + \frac{1}{n_o}}}$$

$$= \frac{23.36 - 19.96 - (0)}{\sqrt{38.73}\times\sqrt{\frac{1}{10} + \frac{1}{13}}}$$

$$= 1.299 \text{ (3 d.p.)}$$

(Note that we used the null hypothesis of $H_0 : \mu_f - \mu_o = 0$ in this calculation.)

The 5% (two-tailed) critical value for t with 21 degrees of freedom is $t_{21} = 2.080$.
So our test statistic value is not significant enough to reject H_0. Thus we assume that the mean height of plants in both populations are the same.

2 c The test in part **b** requires that both the variances are equal. The test in part **a** established that this was reasonable.

3 a A 95% confidence interval with 5 degrees of freedom has a t value of $t_5 = 2.571$.
So the confidence interval is of the form:

$$\bar{x} \pm 2.571 \times \frac{\hat{\sigma}}{\sqrt{n}}$$

If we take the higher value of the confidence interval, 223.5, we may solve for $\hat{\sigma}$.

$$223.5 = \bar{x} + 2.571 \times \frac{\hat{\sigma}}{\sqrt{n}}$$

$$= \frac{223.5 + 206.2}{2} + 2.571 \times \frac{\hat{\sigma}}{\sqrt{6}}$$

$$= 214.85 + 1.0496 \times \hat{\sigma}$$

$$\hat{\sigma} = 8.24$$

$$\hat{\sigma}^2 = 67.9 \text{ (3 s.f.)}$$

b The percentage points are

$\chi_5^2(0.975) = 0.831$

and

$\chi_5^2(0.025) = 12.832.$

We have calculated $\hat{\sigma}^2 = 67.9$ and so can calculate that the critical points are

$$\frac{(6-1)\hat{\sigma}^2}{\chi_5^2(0.975)} = \frac{5\times 67.9}{0.831} = 408.5$$

and

$$\frac{(6-1)\hat{\sigma}^2}{\chi_5^2(0.025)} = \frac{5\times 67.9}{12.832} = 26.5.$$

So the 95% confidence interval for **variance** is (26.5, 408.5). The 95% confidence interval for the **standard deviation** has the square root of the limits of this interval as its limits. i.e. (5.14, 20.2).

3 c Let S denote the span of an adult male's hand. We want to find $P(S > 230)$ which can be standardised to $P\left(Z > \frac{230-\mu}{\sigma}\right)$.

In order to maximise the probability, we need $\frac{230-\mu}{\sigma}$ to be as small as possible.

So we choose the biggest value for μ and σ that are in our confidence intervals.

$$P\left(Z > \frac{230-223.5}{20.2}\right) = P(Z > 0.3218)$$
$$= 1 - P(Z < 0.3218)$$
$$= 1 - 0.626$$
$$= 0.374$$

Thus, the highest estimate of the proportion of adult males with hand span greater than 230 mm is 0.37 (2 d.p.).

4 a $$S_{hh} = \sum h_i^2 - \frac{\left(\sum h_i\right)^2}{n}$$
$$= 272094 - \frac{1562^2}{9}$$
$$= 1000.22$$

$$S_{cc} = \sum c_i^2 - \frac{\left(\sum c_i\right)^2}{n}$$
$$= 2878966 - \frac{5088^2}{9}$$
$$= 2550$$

and

$$S_{hc} = \sum h_i c_i - \frac{\sum h_i \sum c_i}{n}$$
$$= 884484 - \frac{1562 \times 5088}{9}$$
$$= 1433.33$$

b We calculate

$$b = \frac{S_{hc}}{S_{hh}} = \frac{1433.33}{1000.22} = 1.433$$
$$a = \bar{c} - b\bar{h}$$
$$= \frac{5088}{9} - 1.433 \times \frac{1562}{9}$$
$$= 316.626$$
$$c = 317 + 1.43h \text{ (3 s.f.)}$$

4 c b tells us the rate at which the confidence increases with increased height. In this case for every centimetre of height, confidence increases by 1.433 units.

d It would not make sense to have an interpretation of a since it would imply that it is possible to have 0 cm height.

e i

h	c	$c = 317 + 1.43h$	e
179	569	572.97	−3.97
169	561	558.67	2.33
187	579	584.41	−5.41
166	561	554.38	6.62
162	540	548.66	−8.66
193	598	592.99	5.01
161	542	547.23	−5.23
177	565	570.11	−5.11
168	573	557.24	15.76

ii The incorrect value is 573 as the magnitude of its residual is significantly greater than the others.

4 f $S_{hh2} = \sum h_i^2 - \frac{\left(\sum h_i\right)^2}{n}$

$= 272094 - 168^2 - \frac{(1562-168)^2}{8}$

$= 965.5$

$S_{cc2} = \sum c_i^2 - \frac{\left(\sum c_i\right)^2}{n}$

$= 2878966 - 573^2 - \frac{(5088-573)^2}{8}$

$= 2483.875$

and

$S_{hc2} = \sum h_i c_i - \frac{\sum h_i \sum c_i}{n}$

$= 884484 - 168 \times 573$

$- \frac{(1562-168)\times(5088-573)}{8}$

$= 1481.25$

We calculate

$b_2 = \frac{S_{hc2}}{S_{hh2}} = \frac{1481.25}{965.5} = 1.534$

$a_2 = \bar{c} - b_2\bar{h}$

$= \frac{5088-573}{8} - 1.534 \times \frac{1562-168}{8}$

$= 297.044$

$c = 297.044 + 1.534h$

g $c = 297.044 + 1.534 \times 172$

$= 561$

5 a Let x_{rank} denote the place in which a person came in the actual race and y_{rank} denote the rank in which a person came in the qualifying lap-times.

Name	x_{rank}	y_{rank}	d	d^2
Carl	1	3	−2	4
Paula	2	2	0	0
Sarah	3	4	−1	1
David	4	1	3	9
Dhruv	5	5	0	0
Amy	6	7	−1	1
Jake	7	6	1	1
Ali	8	8	0	0

Now we have the data, we sum all d^2 values and get

$\sum d^2 = 16$

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

$= 1 - \frac{6 \times 16}{8(8^2-1)}$

$= 0.810$ (3 s.f.)

b $H_0 : \rho = 0$

$H_1 : \rho > 0$

We have a sample size of 8 and the significance level in the tail is 0.05. From the table, the critical value of r for a 0.05 significance level with a sample size of 8 is $r = 0.6429$, so the critical region is $r > 0.6429$. The observed value of $r = 0.810$ is inside of the critical region, so we reject H_0. There is sufficient evidence at the 5% level of significance that there is a positive association between qualifying lap-times and actual race results.

c Race ranks are not measurable on a continuous scale.

5 d Data will have 4 values with tied rank. Assign a rank equal to the mean of the tied ranks. Calculate the PMCC directly from the ranked data rather than using the formula.

6 a First we find the probability density function by differentiating the cumulative distribution function. We find that:

$$f(x)=\begin{cases} 2x^3+\frac{3}{2}x^2 & 0\leqslant x\leqslant 1 \\ 0 & \text{otherwise}\end{cases}$$

In order to find $E(X)$, we calculate:

$$E(X)=\int_{-\infty}^{\infty}(x\times f(x))dx$$
$$=\int_0^1\left(x\left(2x^3+\frac{3}{2}x^2\right)\right)dx$$
$$=\int_0^1\left(2x^4+\frac{3}{2}x^3\right)dx$$
$$=\left[\frac{2}{5}x^5+\frac{3}{8}x^4\right]_0^1$$
$$=\left[\frac{2}{5}+\frac{3}{8}\right]-[0+0]$$
$$=\frac{31}{40}$$

$$E(X^2)=\int_{-\infty}^{\infty}(x^2\times f(x))dx$$
$$=\int_0^1\left(x^2\left(2x^3+\frac{3}{2}x^2\right)\right)dx$$
$$=\int_0^1\left(2x^5+\frac{3}{2}x^4\right)dx$$
$$=\left[\frac{1}{3}x^6+\frac{3}{10}x^5\right]_0^1$$
$$=\left[\frac{1}{3}+\frac{3}{10}\right]-[0+0]$$
$$=\frac{19}{30}$$

So,

$$\text{Var}(X)=E(X^2)-E(X)^2$$
$$=\frac{19}{30}-\left(\frac{31}{40}\right)^2$$
$$=\frac{157}{4800}=0.033$$

6 b We see that the maximal point occurs when $x=1$ (since we have an increasing function) and so we have that the mode is 1.

c The mode > mean and so we say that it is negatively skewed.

d $P(k<X<3k)=P(X<3k)-P(X<k)$

$$=\frac{27k^3}{2}(3k+1)-\frac{k^3}{2}(k+1)$$
$$=40k^4+13k^3$$

7 Let $A=\frac{X(6-X)}{2}$ denote the area enclosed by the framework and the ground.

$$E(A)=E\left(\frac{X(6-X)}{2}\right)$$
$$=\frac{1}{2}E(X(6-X))$$
$$=\frac{1}{2}E(6X-X^2)$$
$$=\frac{1}{2}\int_0^6\frac{(6x-x^2)}{6}dx$$
$$=\frac{1}{12}\left[3x^2-\frac{x^3}{3}\right]_0^6$$
$$=\frac{3\times 6^2-\frac{6^3}{3}}{12}$$
$$=\frac{6^2}{12}$$
$$=3\text{ m}^2$$