

**THE ROYAL STATISTICAL SOCIETY  
2016 EXAMINATIONS – SOLUTIONS  
HIGHER CERTIFICATE – MODULE 3**

The Society is providing these solutions to assist candidates preparing for the examinations in 2017.

The solutions are intended as learning aids and should not be seen as "model answers".

Users of the solutions should always be aware that in many cases there are valid alternative methods. Also, in the many cases where discussion is called for, there may be other valid points that could be made.

While every care has been taken with the preparation of these solutions, the Society will not be responsible for any errors or omissions.

The Society will not enter into any correspondence in respect of these solutions.

1. (i) (a) The significance level of a test is the probability of rejecting the null hypothesis when it is true. It is the probability of making a Type I error.

1 (prob of Type I error), 1 (rejecting null when true).

(b) The power of a test is the probability of correctly rejecting the null hypothesis when it is false. If the probability of making a Type II error is denoted by  $\beta$ , then the power is  $1 - \beta$ .

1 (rejecting null when false), 1 (1 - prob of Type II error).

TOTAL 4

(ii) (a)  $H_0 : \mu = 200$      $H_1 : \mu \neq 200$ .

1 (null hyp), 1 (alternative hyp)

(b)  $\bar{X} \square N\left(200, \frac{15^2}{9}\right)$  i.e.  $\bar{X} \square N(200, 25)$ .

1 (Normal), 1 (mean 200), 1 (var 25 o.e.)

(c) Significance level =  $P(\text{reject } H_0 | H_0 \text{ true}) =$

$$P(\bar{X} < 191 \text{ or } \bar{X} > 209 | \mu = 200) =$$

1 (rej region correct)

$$P\left(Z < \frac{191 - 200}{15/\sqrt{3}}\right) + P\left(Z > \frac{209 - 200}{15/\sqrt{3}}\right)$$

1 (method)

$$= P(Z < -1.8) + P(Z > 1.8)$$

1 (1.8 and/or -1.8)

$$= 2(1 - \Phi(1.8)) = 2(1 - 0.9641) = 0.0718.$$

1 (0.0718)

Since this is greater than 0.05, a test at the 5% significance level would conclude that the null hypothesis should not be rejected i.e. there is insufficient evidence to suggest that the machine is not working correctly.

1 (conclusion correct)

(d) Power when  $\mu = 205 = P(\text{reject } H_0 | \mu = 205)$

$$= P(\bar{X} < 191 \text{ or } \bar{X} > 209 | \mu = 205)$$

1 (method)

$$= P\left(Z < \frac{191 - 205}{5}\right) + P\left(Z > \frac{209 - 205}{5}\right)$$

$$= P(Z < -2.8) + P(Z > 0.8)$$

1 (-2.8), 1 (0.8)

$$= 1 - \Phi(2.8) + 1 - \Phi(0.8)$$

1 (method)

$$= 2 - 0.9974 - 0.7881$$

$$= 0.2145.$$

1 (0.2145)

When  $\mu = 210$  we would expect the power to be higher than this value since 210 is further away from the null value of 200 and so the probability of detecting the incorrect value will be higher.

1 (correct comment)

TOTAL 16

2. (i) If the failure rate is still 12% then the number of failed seeds in a pack follows a Bin(10,0.12) distribution. 1 (Binomial), 2 (correct parameters)

The binomial distribution assumes that failure/non-failure of individual seeds are independent of one another and that  $p$  is constant from one pack to another. This may not be true if (e.g.) a set of seeds is sent out with damaged packaging. Also the probability of survival for different packs may not be constant for packs with longer/shorter journeys to make.

1 (indep), 1 (p constant)  
2 (1 each valid comment) TOTAL 7

- (ii)  $H_0$  : The underlying distribution of failed seeds is B(10, 0.12) 1 (correct null)  
 $H_1$  : Stalwart failure rate is not 12%. 1 (correct alternative)  
1 (if make a wider statement about the model as a whole)

Goodness-of-fit test statistic is  $\sum \frac{(O_i - E_i)^2}{E_i}$  but require all  $E_i$  values to be sufficiently large.

1 (test statistic), 1 (require E sufficiently large, explicit or suggested)  
1 (correct pooling)

So pool final two categories, giving

$O_i$	25	28	5	2
$E_i$	16.71	22.79	13.98	6.52

Then test statistic is  $\frac{(25-16.71)^2}{16.71} + \frac{(28-22.79)^2}{22.79} + \frac{(5-13.98)^2}{13.98} + \frac{(2-6.52)^2}{6.52}$   
 $= 4.113 + 1.191 + 5.768 + 3.133$   
 $= 14.21.$

1 (14.21)

We compare this with the upper 5% value from  $\chi^2_3$  which is 7.815. 1 (3 df), 1 (7.815)

Our test statistic is thus significant, indicating that we reject the null hypothesis in favour of the alternative i.e. the failure rate of the Stalwart plants appears not to be the same as that for the previous variety.

1 (significant), 1 (conclusion)  
TOTAL 11

- (iii) There are more Stalwart packs with 0 or 1 failures than expected and correspondingly fewer Stalwart packs with 2 or more failures than expected. This indicates that the new variety may be more robust and thus be a more successful variety to send out in the future.

1 (direction of differences), 1 (valid comment).  
TOTAL 2

3. (i) For the remodelling costs, sample mean  $\bar{x} = 5.918$  and sample standard deviation  $s = 2.1637$ . 1 (for both 5.918 and 2.1637)  
 Sample size is 11 so use  $t$  distribution on 10 degrees of freedom, the upper 2.5% point of which is 2.228. 1 (10 df), 1 (2.228)

The confidence interval for the mean is given by  $\bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$  1 (method)

which is  $5.918 \pm 2.228 \times \frac{2.1637}{\sqrt{11}}$  i.e. (4.464, 7.372). 1 (4.464 & 7.372)

(Accept if values given to nearest £50)

So this gives a 95% interval from £4464 to £7372. 1 (units applied)

Assumes underlying Normal distribution for remodelling costs. 1 (assumption)

TOTAL 7

(ii) The differences (valuation increase – cost) are:

1.35   1.55   0.7   0.7   0.8   1.9   1.05   1.1   0.7   0.9   0.4

1 (diffs correct)

The values are paired (not independent) so we use a paired  $t$ -test. 1 (correct test)

Testing  $H_0 : \mu = 1$  against  $H_1 : \mu > 1$ .

The mean of the 11 observed differences is 1.0136 and their standard deviation is 0.43880.

Test statistic is  $\frac{\bar{d} - 1}{s_d / \sqrt{n}} = \frac{1.0136 - 1}{0.43880 / \sqrt{11}} = 0.103$ . 1 (method), 1 (0.103)

This should be tested against the upper 5% point of the  $t_{10}$  distribution which is 1.812.

1 (1.812)

It is clearly not significant, we do not reject  $H_0$  in favour of  $H_1$ , and so the website's statement is not supported. 1 (not sigt), 1 (no evidence to reject H0 in favour of H1),  
1 (statement not supported)

Assumption: the differences follow a normal distribution. 1 (assumption correct)

TOTAL 9

(iii) Sign Test: advantage is that it is quick to apply. 1 (Sign), 1 (quick)

Wilcoxon Signed Rank Test: takes into account magnitudes of differences and is more powerful. 1 (Wilcoxon SR), 1 (more powerful)

TOTAL 4

4. (i) Mean is midpoint =  $(48.388 + 56.012) / 2 = 52.2$  hours. 1 (52.2)  
 90% interval is  $52.2 \pm t \times \text{s.e.}$  where  $t$  is the upper 5% point of  $t_{14}$  which is 1.761. 1 (14 df)  
1 (1.761)

So  $(56.012 - 52.2) = 1.761 \times \text{s.e.}$  and so the standard error is  $\frac{3.812}{1.761} = 2.1647$  hours 1 (2.1647)

(ii) For the 95% interval we require  $\bar{x} \pm 2.145 \times \text{s.e.}$  which is  $52.2 \pm 2.145 \times 2.1647$ , giving (47.56, 56.84) hours. 1 (correct interval)

This is wider than the previous interval as we are more confident that the true mean burning time lies between the limits. Both intervals are based on the assumption that the burning times are normally distributed. 1 (wider, more confident), 1 (normal)  
TOTAL 7

(iii) From (a)(i),  $\frac{s_1}{\sqrt{15}} = 2.1647$ , so the standard deviation is 8.3838 hours. 1 (8.3838)

Test  $H_0 : \sigma_1^2 = \sigma_2^2$  against  $H_1 : \sigma_1^2 \neq \sigma_2^2$ . Test statistic is  $F = \frac{s_2^2}{s_1^2} = \frac{10.2^2}{8.3838^2} = 1.48$ . 1 (1.48)

Comparing to tables, this is not significant (upper 2.5% point of  $F_{18,14}$  is 2.88), so the variances of the burning times do not differ significantly. 1 (18, 14 df)  
1 (correct value magnitude)  
1 (correct conclusion)

(iv) Test  $H_0 : \mu_1 = \mu_2$  against  $H_1 : \mu_1 < \mu_2$ . 1 (hypotheses correct)

Pooled sample variance is  $\frac{(14 \times 8.3838^2) + (18 \times 10.2^2)}{32} = 89.27 \Rightarrow s = 9.448$ .

1 (9.448 or 89.27)

Test statistic is  $t = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{52.2 - 58.4}{9.448 \sqrt{\frac{1}{15} + \frac{1}{19}}} = -1.900$ . 1 (method), 1 (-1.900)

This is compared with the lower 5% point of  $t_{32}$  which is -1.694. We have a significant result, so the perfumed candles burn more quickly on average 1 (32 df)  
1 (correct  $t$  value)  
1 (significant), 1 (correct conclusion)

(NOTE: give marks for follow through of arithmetic errors, except negative variance without comment).

TOTAL 13