EXAMINATIONS OF THE ROYAL STATISTICAL SOCIETY

HIGHER CERTIFICATE IN STATISTICS, 2017

MODULE 1 : Data collection and interpretation

Time allowed: One and a half hours

Candidates should answer THREE questions.

Each question carries 20 marks.
The number of marks allotted for each part-question is shown in brackets.

Graph paper and Official tables are provided.

Candidates may use calculators in accordance with the regulations published in
the Society's "Guide to Examinations" (document Ex1).

The notation $\log$ denotes logarithm to base $e$.
Logarithms to any other base are explicitly identified, e.g. $\log_{10}$.

Note also that $\binom{n}{r}$ is the same as $^nC_r$. 

This examination paper consists of 8 printed pages.
This front cover is page 1.
Question 1 starts on page 2.

There are 4 questions altogether in the paper.

© RSS 2017
1. The table below shows how support by voters for the main political parties in England in 2015 varied with a range of demographic factors. Voters are classified by sex, age group, socio-economic group, location, and ethnicity. Each figure given is an index showing support by a group for a party. A figure of 100 indicates that support is at the national average level; a figure of 105 shows that support is 5% higher than the national average level; a figure of 90 indicates that support is 10% less than the national average level. (The socio-economic group AB is upper middle class and middle class, C1 is lower middle class, C2 is skilled working class, DE is working class and those at the lowest level of subsistence.)

<table>
<thead>
<tr>
<th></th>
<th>Green</th>
<th>Labour</th>
<th>Liberal Democrat</th>
<th>Conservative</th>
<th>UKIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>82</td>
<td>96</td>
<td>106</td>
<td>97</td>
<td>117</td>
</tr>
<tr>
<td>Female</td>
<td>122</td>
<td>105</td>
<td>92</td>
<td>103</td>
<td>79</td>
</tr>
<tr>
<td>18–24</td>
<td>227</td>
<td>120</td>
<td>92</td>
<td>92</td>
<td>36</td>
</tr>
<tr>
<td>25–34</td>
<td>122</td>
<td>140</td>
<td>119</td>
<td>72</td>
<td>36</td>
</tr>
<tr>
<td>35–44</td>
<td>104</td>
<td>129</td>
<td>111</td>
<td>76</td>
<td>77</td>
</tr>
<tr>
<td>45–54</td>
<td>78</td>
<td>99</td>
<td>91</td>
<td>91</td>
<td>128</td>
</tr>
<tr>
<td>55–64</td>
<td>76</td>
<td>88</td>
<td>93</td>
<td>111</td>
<td>114</td>
</tr>
<tr>
<td>65+</td>
<td>75</td>
<td>65</td>
<td>91</td>
<td>130</td>
<td>139</td>
</tr>
<tr>
<td>AB</td>
<td>98</td>
<td>83</td>
<td>138</td>
<td>117</td>
<td>79</td>
</tr>
<tr>
<td>C1</td>
<td>118</td>
<td>98</td>
<td>88</td>
<td>105</td>
<td>92</td>
</tr>
<tr>
<td>C2</td>
<td>84</td>
<td>109</td>
<td>87</td>
<td>90</td>
<td>122</td>
</tr>
<tr>
<td>DE</td>
<td>98</td>
<td>116</td>
<td>75</td>
<td>80</td>
<td>117</td>
</tr>
<tr>
<td>Urban</td>
<td>102</td>
<td>108</td>
<td>100</td>
<td>93</td>
<td>97</td>
</tr>
<tr>
<td>Rural</td>
<td>98</td>
<td>72</td>
<td>100</td>
<td>123</td>
<td>113</td>
</tr>
<tr>
<td>White</td>
<td>102</td>
<td>93</td>
<td>102</td>
<td>104</td>
<td>106</td>
</tr>
<tr>
<td>Non-White</td>
<td>59</td>
<td>195</td>
<td>64</td>
<td>56</td>
<td>19</td>
</tr>
</tbody>
</table>

*Source: Populus (an opinion polling company)*

(i) Summarise how sex is related to political party preference in the data. Draw a suitable graph, or graphs, to illustrate your answer. (4)

(ii) The political parties are shown in order from what are generally considered to be the most left wing (Green) to most right wing (UKIP). Identify the main patterns in political preferences as they vary by age. Draw a suitable graph, or graphs, to illustrate your answer. (6)
(iii) Of the sixteen categories in the table identify that with profile most like the national average, and that with profile least like the national average. Discuss briefly why these groups might be expected to have such profiles. (4)

(iv) Show that about 60% of Green supporters are female. Find the corresponding percentage for UKIP supporters. State any assumptions that you have made in answering those questions. (4)

(v) A newspaper article noted the figures of 92 for the level of support among people in the 18–24 age range for the Liberal Democrats and for the Conservatives. The article stated that support in this age-group was evenly split between these two parties. Explain why that is the wrong conclusion to draw from the data, and state the correct conclusion. (2)
2. Each of the following is a possible but unsatisfactory item on a questionnaire. In each case state the problem (or problems) with the item and give or describe a new version that avoids the problem (or problems).

(a) Rate the speed and accuracy with which you work, compared with your colleagues, on the following scale.

   Above average   Average   Below average

(b) How far is it, in miles, from your home to your place of work?

(c) What is your favourite leisure activity?

   Watching TV  Using a computer  Playing games  Other

(d) Where do you like to go on holiday?

   In Britain  Somewhere in Europe  Elsewhere in the world

(e) Please state your annual income.

(f) How often do you exercise?

   Every day  Two or three times a week  Two or three times a month  Never

(g) To what extent do you agree with the statement that we should reduce welfare spending in this country?

   Completely agree  Somewhat agree  Somewhat disagree  Completely disagree

(h) What type of computer operating system do you use?

   MacOS  Windows  Linux

(i) What percentage of your household budget do you spend on food and drink?

(j) Addressed to a pregnant woman: Do you smoke?

   Yes  No
3. (a) The chart below shows, for 22 countries, the level of happiness for citizens aged 15 to 74 and for those aged 75 and over. The countries are, from left to right, in increasing order of GDP per capita.

*Experienced Happiness by GDP Per Capita and Age*
Countries are sorted by GDP per capita (from the lowest to the highest)

- 15 to 74
- 75+

Surveys conducted between 2005 and 2010. Results for countries where sample sizes drop below 100 for those 75 and older are not reported.

(i) Comment briefly on the strengths and weaknesses of the way in which the data are presented.  

(ii) Summarise in words the main messages in these data.  

(iii) Identify the additional information you would wish to have in order to be able to interpret these data better. 

(4)  

(4)  

(2)  

Question 3 continued on next page
(b) The chart below shows, for a group of heart disease patients, the distributions of their weights, their diastolic blood pressures, and their systolic blood pressures. It also shows scatter plots for each pair of variables.

(i) Comment on the strengths and weaknesses of the ways in which the data are displayed.

(4)

(ii) Identify the additional information you would wish to have in order to interpret these data better.

(2)

(iii) Discuss what useful conclusions, if any, can be drawn from

(A) the histograms,

(B) the scatter plots.

(4)
4. The table below shows road traffic accident statistics for a European country from 2004–2013. Four of the columns are labelled (A) to (D) for later reference.

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of accidents (A)</th>
<th>Number involving death or injury (B)</th>
<th>Number of persons killed (C)</th>
<th>Number of persons injured (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2004</td>
<td>537 000</td>
<td>77 000</td>
<td>4427</td>
<td>136 000</td>
</tr>
<tr>
<td>2005</td>
<td>621 000</td>
<td>87 000</td>
<td>4505</td>
<td>154 000</td>
</tr>
<tr>
<td>2006</td>
<td>729 000</td>
<td>96 000</td>
<td>4633</td>
<td>169 000</td>
</tr>
<tr>
<td>2007</td>
<td>826 000</td>
<td>107 000</td>
<td>5007</td>
<td>189 000</td>
</tr>
<tr>
<td>2008</td>
<td>950 000</td>
<td>104 000</td>
<td>4236</td>
<td>184 000</td>
</tr>
<tr>
<td>2009</td>
<td>1 053 000</td>
<td>111 000</td>
<td>4324</td>
<td>201 000</td>
</tr>
<tr>
<td>2010</td>
<td>1 106 000</td>
<td>117 000</td>
<td>4045</td>
<td>211 000</td>
</tr>
<tr>
<td>2011</td>
<td>1 229 000</td>
<td>132 000</td>
<td>3835</td>
<td>238 000</td>
</tr>
<tr>
<td>2012</td>
<td>1 297 000</td>
<td>154 000</td>
<td>3750</td>
<td>268 000</td>
</tr>
<tr>
<td>2013</td>
<td>1 207 000</td>
<td>161 000</td>
<td>3685</td>
<td>275 000</td>
</tr>
</tbody>
</table>

(i) Describe the trends shown in each of columns (A) to (D). Explain why these trends, taken individually, give only limited useful information.

(6)

(ii) Draw a graph to show the trend in the number of persons killed per thousand accidents. Describe this trend in words, say what it indicates about road safety, and discuss possible underlying causes.

(8)

(iii) Investigate the trend in the number of persons killed relative to the number of persons injured. Discuss what, if anything, this trend adds to your conclusions in part (ii).

(6)