GUIDELINES FOR PRACTICAL PART OF CA200 (2012-2013) MODULE

NOTE: You may find it convenient to bring a memory stick or other storage device to practicals to save your work. Alternatively, it may be that you have been allocated space on some server drive - this you should be aware of yourself!

The survey questionnaire and the survey results (first set of data) should be available from 1st practical session.

To import the survey results into R:

(a) It is first necessary to save the file, from within Excel, in comma separated (csv) format. But this has already been done for you, that is the circulated file is of .csv type.
(b) Start R and read the file in R using "read.csv". For example, if the file is on the "C" drive then use:
results<-read.csv("C:/CA200_Questionnaire_Summary.csv", header=T)
this will read the data from the file into the data frame called "results".
(c) Once the survey data has been imported **you should save it** in R data frame. You can check if data have been read correctly by using command:
results$Q1
[1] 1 1 1 1 1 2 1 1 2 1 1 2 1 2 2 1 1 1 1 1 1 2 2 1 2 2 NA 1 2 1 NA 1 1 2 2 1 1 2 2 1 1 1 1 1 1 2
this will give you data for the column Q1 in the file.

To import the survey results into SPSS:

(a) Survey results file typically excel format. e.g. CA200_survey_2012.xls"), save these in a folder on your local disk space (e.g. "C"drive). Then close the file.
(b) Start SPSS and from within it open the file that you have just saved (say, "survey_2011.xls"). To do this, select Open -> Data ->

survey_2011.xls (from the folder you saved to)

(c) Once the survey data have been imported **you should save them** as an SPSS ".sav" file (e.g. survey_2011.sav).

After you have successfully imported the data to SPSS you should carry out the activities indicated below.

(A) Most of the steps to be carried out are specified explicitly below so performing these should be largely a mechanical exercise. Still, after carrying them out, you should have a good idea of the basics of using SPSS and will have done a good deal of analysis of the survey data.

(B) You are encouraged to devise extra investigations of your own and to use SPSS to perform them.

Notes:

(1) "Results 3" go beyond descriptive statistics and talk about viable statistical analyses.

(2) These guidelines were prepared for a slightly earlier (legacy) version of SPSS so that minor adjustments are sometimes required.

Results1 - Simple analyses of individual questions

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
1	q1 - gender	Pie-chart to show breakdown between genders	t <- table(results$Q1) lbls <- paste(c("Male:", "Female:"), t, sep="") pie(t, labels=lbls, main="Gender breakdown")	<graphs><pie> <summaries for groups of cases> <define> <define slices by [q1 in this case]> [Also, add a title] (*used "Legacy Dialogs"*) [***Save the results in a file of your own naming & leave the file open for insertion of results of #2, etc***]
2	q2 - height	Summary statistics of height data overall	summary(results$Q2)	<analyze> <descriptive statistics> <descriptives> <select q2 as variable in this case>
3	q3 - weight	Summary statistics of weight data overall	summary(results$Q3)	As previous with q3 in place of q2
4	q4 - left/right	Pie-chart to show breakdown between right and left "handedness"	t <- table(results$Q4) lbls <- paste(c("Right:", "Left:"), t, sep="") pie(t, labels=lbls, main="Handedness breakdown")	As #1 with q4 in place of q1
5	q5 - residence in term	Pie or other suitable graph to show breakdown between categories	t <- table(results$Q5) lbls <- paste(c("At home:", "In digs:", "Alone in bedsit/apartment:", "Sharing house/apartment:", "DCU residences:", "Other:"), t, sep="") pie(t, labels = lbls, main="Residence breakdown")	As #1 with q5 in place of q1
6	q6 - weekly rent	Summary statistics of rental data (also useful to display	As #2 with Q6 in place of Q2	As #2 with q6 in place of q2

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
		graphically)		
7	q7 - transport mode	Pie or other suitable graph to show breakdown between categories	As #5 with Q7 options in place of Q5 options	As #1 with q7 in place of q1
8	q8 - distance	Summary statistics of travel distance data	As #2 with Q8 in place of Q2	As #2 with q8 in place of q2
9	q9 - travel cost	Summary statistics of travel cost data	As #2 with Q9 in place of Q2	As #2 with q9 in place of q2
10	q10 - food cost	Summary statistics of food cost data	As #2 with Q10 in place of Q2	As #2 with q10 in place of q2
11	q11 - residence out of term	Pie or other suitable graph to show breakdown between categories	As #5 with Q7 options in place of Q5 options	As #1 with q11 in place of q1
12	q12 - grant	Pie or other suitable graph to show breakdown between categories	As #5 with Q12 options in place of Q5 options	As #1 with q12 in place of q1
13	q13 - summer work	Pie or other suitable graph to show breakdown between categories	As #5 with Q13 options in place of Q5 options	As #1 with q13 in place of q1
14	q14 - part-time work	Pie or other suitable graph to show breakdown between categories	As #5 with Q14 options in place of Q5 options	As #1 with q14 in place of q1
15	q15a, q15b	Line graph showing number of times abroad in past year and past 3 years	plot(results$Q15a, type="o", col="blue", ylim=c(0, 20), xlab="Student No.", ylab="Times abroad"); lines(results$Q15b, type="o", col="red"); legend(1, 19.5, c("last 12 months", "last 3	<graphs><line> <multiple> & “Values of individual cases” <define - lines represent q15a, q15b in this case> [Also, add a title] “*Legacy Dial*”

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
			years"), col=c("blue", "red"), pch=21:21, lty=1:1);	
16	q16 exam grades	Pie or other suitable graph to show breakdown between categories	As #5 with Q16 options in place of Q5 options	As #1 with q16 in place of q1
17	q17 - smoke?	Pie or other suitable graph to show breakdown between categories	As #5 with Q17 options in place of Q5 options	As #1 with q17 in place of q1
18	q18 - drink alcohol?	Pie or other suitable graph to show breakdown between categories	As #5 with Q18 options in place of Q5 options	As #1 with q18 in place of q1
19	q19 - sport/ train	Pie or other suitable graph to show breakdown between categories	As #5 with Q19 options in place of Q5 options	As #1 with q19 in place of q1

Results2 - Filtering of data & analysis of filtered data

2.1 Analysis of summer work

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
0	q13	To analyse 13a to 13d (i.e. #1 to #4, following), we need to filter out those who didn't do summer work	summer_work <- subset(results, results$Q13 == 1)	<data><select cases><set if condition "q13=1">
1	q13a - how many weeks?	Summary statistics of data on number of weeks worked.	summary(summer_work$Q13a)	<analyze> <descriptive statistics> <descriptives> <select q13a as variable in this case> [***Save the results in a file of your own naming & leave the file open for insertion of results of #2, etc and those for sections 2.2 to 2.5***]
2	q13b - weekly earnings	Summary statistics of weekly earnings	As #1 with Q13b in place of Q13a	As #1 with q13b in place of q13a
3	q13c - savings	Summary statistics of savings	As #1 with Q13c in place of Q13a	As #1 with q13c in place of q13a
4	q13d - location	Pie to show breakdown between categories	t <- table(summer_work$Q13d) lbls <- paste(c("Ireland:", "UK:", "US:", "Elsewhere:"), t, sep="") pie(t, labels = lbls, main="Summer work location")	<graphs><pie> <summaries for groups of cases> <define> <define slices by [q13d in this case]> [Also, add a title] "Legacy dialogues"

2.2 Analysis of current part-time work

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
0	q14	To analyse 14a and 14b (i.e. #1 and #2, following), we need to filter out those who are not doing part-time work.	part_time_work <- subset(results, results$Q14 == 1)	<data><select cases><set if condition "q14=1">
1	q14a - how many hours?	Summary statistics of data on number of hours worked per week.	summary(part_time_work$Q14a)	<analyze> <descriptive statistics> <descriptives> <select q14a as variable in this case>
2	q14b - hourly rate	Summary statistics of hourly pay rate	As #1 with Q14b in place of Q14a	As #1 with q14b in place of q14a

2.3 Analysis of smoking

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
0	q17	To analyse 17y (i.e. #1, following), we need to filter out those who are not smoking.	smoking <- subset(results, results$Q17 == 1)	<data><select cases><set if condition "q17=1">
1	q17y - how many cigarettes?	Summary statistics of data on number of cigarettes smoked daily.	summary(smoking$Q17y)	<analyze> <descriptive statistics> <descriptives> <select q17y as variable in this case>

2.4 Analysis of weekly alcohol consumption

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
0	q18	To analyse 18y (i.e. #1, following), we need to filter out those who are not consuming alcohol.	drinking <- subset(results, results$Q18 == 1)	<data><select cases><set if condition "q18=1">
1	q18y - how many units?	Summary statistics of data on number of alcohol units consumed weekly.	summary(drinking$Q18y)	<analyze> <descriptive statistics> <descriptives> <select q18y as variable in this case>

2.5 Analysis of sport training "intensity"

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
0	q19	To analyse 19y (i.e. #1, following), we need to filter out those who do not engage in sport or training.	playing_sports <- subset(results, results$Q19 == 1)	<data><select cases><set if condition "q19=1">
1	q19y - how many hours?	Summary statistics of data on number of hours spent weekly.	summary(playing_sports$Q18y)	<analyze> <descriptive statistics> <descriptives> <select q19y as variable in this case>

Results3 - Some illustrative statistical investigations

3.1 Are exam marks unaffected by gender?

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of SPSS syntax (how to proceed)
1	q1 and q16	Produce a contingency table of gender versus exam results, to which the chi-squared test can be applied.	install.packages("gmodels") library("gmodels") CrossTable(results$Q16, results$Q1)	<analyze> <descriptive statistics> <Cross-tabs> <select q16 as row and q1 as column (say) in this case; also make sure chi-square is ticked under statistics>

Note: Relevant result is the value for the "Pearson chi-square".

3.2 What is the association between height and weight?

#	Question/Topic	What to produce	Detail of R syntax (how to proceed)	Detail of syntax
1	q2, q3	Regression of weight against height (male and female) Useful also to do scatter plot to see if there is any reason to anticipate a linear relationship.	reghw <- rm(results$Q2~results$Q3) plot(results$Q3, results$Q2, xlab="weight(kg)", ylab="height(cm)"); abline(reghw)	<analyze> <regression> <linear> <select dependent and independent variables>
2	q2, q3	Regression of weight against height (male only)	Same as #1 except first filter out females (refer to section 2 (above) for how to do this).	Same as #1 except first filter out females (refer to section 2 (above) for how to do this).
3	q2, q3	Regression of weight against height (female only)	Same as #1 except first filter out males (refer to section 2 (above) for how to do this).	Same as #1 except first filter out males (refer to section 2 (above) for how to do this).