**BIA – 5201, Fall 2021**

**Individual Assignment 1**

**Instructions:**

* Complete the tasks outlined for each question
* Submit:
  + a written report (Word file or .pdf) which includes answers to all the questions/tasks.
  + Any SPSS files you’ve updated (data and output files)
  + Any Excel files you’ve updated
* Submit via email to [marc.stuyver@humber.ca](mailto:marc.stuyver@humber.ca)
* Subject line should indicate your name, student number and the title of the assignment (Individual Assignment 1)
* Each submitted file should indicate your last name ex. Stuyver\_HockeyPlayers.sav

**Question 1.**

In his book “Outliers”, Malcom Gladwell put forth a theory that Canadian-born professional hockey players are disproportionately born in the earlier months of the year. The rationale goes as such: kids start off playing house league hockey when they are young, and then move on to rep hockey at about 8 or 9 years of age. Rep hockey involves a tryout and selection process. The kids that are born earlier in the year have an advantage of being a year larger, stronger and faster than kids born closer to the end of the year, which at 8 or 9 years of age can make a big difference. The bigger, stronger, faster kids get selected first for rep hockey, which involves more ice time (practices and games), so those kids selected for rep hockey advance to higher levels than those kids who were not selected for rep hockey. By virtue of being born earlier in the year, kids stand a better chance of being selected for rep hockey, and therefore have a greater chance of making it to the professional level.

You have conducted a survey of all NHL players from the 1990/91 season through the 2007/08 season. The survey includes the year of the survey, the birth month of each player, as well as other details included in Appendix 1.

**Tasks (all to be conducted in SPSS):**

1. Using the data set provided, create an appropriate graph to show whether Mr. Gladwell’s theory is true.

**Working Steps:**

1. We need to filter out the Canadian born players for following analysis. We need to select ‘Birth\_Country’ and filter out Canadian born players.

To do that, Click on Data -> Select Cases -> if condition is satisfied and drag Birth\_Country and select Canada. When we are applying filter on Strings, it should be opened and closed with single quotes. ( Birth\_Country = ‘ “Canada”’)

1. Now, go to Graphs -> Chart Builder and select simple Line chart/Bar graph from Gallery view.
2. Now, Drag ‘Month of Birth’ on x-axis and ‘Birth\_country (“Canada = 1”) on filter variable and click OK. We can add data labels by clicking on graph and select show data labels option under Elements tab.

**Now, for other question we need to turn off the filter so that it won’t effect our data. To do that go to Data -> Select Cases and select all cases and click OK.**

1. What position is, on average, the highest paid position?

**Working Steps:**

1. Now, go to Graphs -> Chart Builder and select simple Bar graph from Gallery view.
2. Drag ‘Position’ variable on X-axis and ‘Salary’ on Y-axis. It will automatically display the mean values for Salary variable for each position.
3. Are Canadian born players, on average, paid higher than players born in other countries?

**Working Steps:**

1. Now, go to Graphs -> Chart Builder and select simple Bar graph from Gallery view.
2. Drag ‘Country of Birth variable on X-axis and ‘Salary’ on Y-axis. It will automatically display the mean values for Salary variable for each country.

**From the chart, we can conclude that ‘Canadian’ born players on average are earning $1226435.40 and it is lower when compared with other players who were born in other countries.**

1. Analyze and comment on the distribution of salaries (i.e. are they normally distributed?).

**Working Steps:**

1. Go to Analyze -> Descriptive statistics -> Explore and select drag ‘Salary’ variable under Dependent List and Label cases by ‘Month of birth’.
2. Now we have to enable

a) Select outliers and descriptives under Statistics option.

b) Select only Histogram and choose Normality plots under Plots option.

c) Select ‘Exclude cases pairwise’ under Options tab as it will select all missing values data as well

d) Now choose ‘Both’ option and click on OK

From the histogram we can conclude that the salary variable is not normally distributed.

**Question 2.**

You work for a consulting engineering firm which has been hired by the City of Toronto to conduct a traffic survey on Highway 401. You have spent a year recording the speeds of vehicles and the conditions in which the speed was recorded (is the road wet? Is it snowing? Is the visibility good? Is there traffic congestion?).

Your data is now ready for analysis. The data is contained on the “Sample 2” tab in the “Speeding Data.xlsx” spreadsheet.

**Tasks (all to be conducted in Excel):**

1. Using one formula (no filters or pivot tables), find the average speed on the days when the roads are dry, there is no snow, the visibility is good, and there is not congestion.

**=AVERAGEIFS(B8:B10007,C8:C10007,"=0",D8:D10007,"=0",E8:E10007,"=0",F8:F10007,"=0")**

1. Set up appropriate bins for grouping the speed data

We have considered intervals of 15 between each bin.

|  |  |
| --- | --- |
| Bins | Frequemcy |
| 1 | 58 |
| 16 | 331 |
| 32 | 872 |
| 48 | 1468 |
| 64 | 1927 |
| 80 | 1959 |
| 96 | 1601 |
| 112 | 1053 |
| 128 | 588 |
| 144 | 138 |
| 160 | 4 |
| 176 | 1 |

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

1. Using an array function, find the frequency for speeds in each of the bins you’ve created, and create a histogram of the results.

**Select K2 to K13 and Enter FREQUENCY(B2:B10001,J2:J13) in K2 cell and press ctrl+shift+enter at same time to calculate frequencies.**

|  |  |
| --- | --- |
| Bins | Frequemcy |
| 1 | 58 |
| 16 | 331 |
| 32 | 872 |
| 48 | 1468 |
| 64 | 1927 |
| 80 | 1959 |
| 96 | 1601 |
| 112 | 1053 |
| 128 | 588 |
| 144 | 138 |
| 160 | 4 |
| 176 | 1 |

1. Using the Histogram function (Data >> Data Analysis >> Histogram), find the frequency for speeds in each of the bins you’ve created. Comment on any differences between the results of d) and c).

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Bins* | *Frequency* |  |
|  | 1 | 58 |  |
|  | 16 | 331 |  |
|  | 32 | 872 |  |
|  | 48 | 1468 |  |
|  | 64 | 1927 |  |
|  | 80 | 1959 |  |
|  | 96 | 1601 |  |
|  | 112 | 1053 |  |
|  | 128 | 588 |  |
|  | 144 | 138 |  |
|  | 160 | 4 |  |
|  | 176 | 1 |  |
|  | More | 0 |  |

No difference

**Question 3.**

You have been hired to investigate the relationship between a student’s grade goal and self-efficacy (both ‘specific’ and ‘general’) at the outset of an MBA programme and their perceptions of social support and their actual average grade at the end of the programme (approximately 18 months later). Age and gender of the students were obtained as control variables. Details of the variables and their measurement are provided in Appendix 2.

**Tasks:**

1. Create a variable **ssupport** that is a summated scale, comprising seven items (see Appendix 2). Undertake a reliability analysis for this scale and comment on the acceptability of the Cronbach’s alpha coefficient.

**Working steps:**

1. First we need to identify the negative impacting questions mentioned in survey under Social support and reverse the ratings.
2. Out of 7, we can identify that 1, 2 and 6 are negatively impacted. So, we need to reverse the survey results for these comments.
3. To reverse the results for these comments, go to Transform -> Recode into Different variables and drag 1, 2 and 6 question variables and label them by clicking on ‘change’ option.
4. Once all the reversed variables are labelled, click on ‘old and new variables’ option and mention both the old and new values and click on ‘Add’. We are replacing old values with new values (1 ->5, 2 ->4, 3 ->3, 4 ->2, 5 ->1).
5. Now, press continue and click on OK. It will reverse the survey results and also we can see that three new variables have been added with updated values.
6. Now, we need to create summated scale comprising of all 7 items. To do that go to Transform -> Compute variable and add all of them by drag 3, 4, 6 and 7 and newly created 3 items under Numeric Expression tab and select ‘+’ sign between each item to create the summated scale.

**Reliability Analysis working steps:**

1. Go to Analyze -> Scale -> Reliability Analysis and drag the summated scale variable which was created from above steps.
2. Make sure that Model is set to Alpha. Also, we need to select few options here.
3. Select Scale under ‘Descriptives for’ tab
4. Select Co-relations under ‘inter-item’ and ‘summaries’ tabs and click OK.

**Analysis:** Based on reversing variables, our Cronbach’s alpha coefficient value was 0.697 and if we did not reverse the variables then Cronbach’s alpha coefficient would be 0.711

If Cronbach’s alpha coefficient value is greater than or equal to 0.7 for efficient analysis. In the ‘Item – statistics’ summary we can see the impact on Cronbach’s alpha coefficient value if we delete one specific item. Based on following summary, we can see that Cronbach’s alpha coefficient value would be 0.699 which is closer to 0.7 if we delete ‘reversed stressful’ variable. But as there is no significant difference between each variable even after deleted, we can keep the following data as it is without any modifications.

**Appendix 1: Variable Definitions**

**Hockey Players - Survey Dataset**

Variable Name Description

Time 0= 1990/91 season; 1= 1991/92 season ... to 17=2007/08 season

Birthmonth 1= January; 2= February ... to 12= December

SalaryAdj Inflation-adjusted annual earnings (in U.S. dollars)

Birth\_Country Country in which the player was born:

Canada Czech Republic

United States Slovakia

Russia Former Soviet Republic

Sweden Rest of Europe

Finland Rest of World

Age Age in years

Height The player’s height in inches

Weight The player’s weight in pounds

Seasoninleague Number of years of experience as a professional hockey player

Captain Binary variable equal to 1 if player is the team captain; 0 otherwise

Position C = Centre;

R = Right Wing;

L = Left Wing;

D = Defence;

G = Goalie;

F = Forward;

W=Winger (plays either wing)

**Appendix 2: Variable Definitions**

**Academic Performance – MBA survey**

1. Age (**age**): measured as the calendar age of the student, in years, at the start of the MBA programme.
2. Gender (**gender**): 0=female, 1=male
3. Overall Grade (**ovgrade**): measured as the weighted average percentage mark of the student at the end of the programme.
4. Self-efficacy strength (**sestren**): students were asked at the outset of the programme to indicate whether they could achieve a particular grade goal and to indicate their level of confidence in achieving that grade goal on a 1-10 scale. The measure of self-efficacy strength was the sum of all confidence ratings for those grade self-efficacy levels that were answered ‘yes’ (Lee and Bobko, 1994).
5. Grade goal (**gradegl**): measured by a student’s grade goal, from C to A, at the outset of the programme, (with A scoring 3), based on Mone (1994).
6. Generalised Self-Efficacy (**gse**): measured at the outset of the programme by the 17-item Sherer et al. (1982) GSE scale.
7. Social support (**ssupport**): refers to the provision of emotional, appraisal, informational and instrumental support (House, 1981), which in an academic context may prevent or buffer students from the negative effects of stress (Jacobi, 1991). This construct was measured by a 7-item scale with the following questions:

* People on the MBA often think about quitting N
* I frequently think of quitting the MBA N
* I enjoy my studies on the MBA Programme P
* I feel satisfied with my relations with faculty and tutors P
* I get sufficient support from the university P
* The MBA is very stressful N
* The stress involved with the MBA Programme doesn’t interfere with my academic achievement P