**QMS210: Applied Statistics for Business**

**SPSS GROUP PROJECT**

**WINTER 2021**

1. **No handwritten reports will be considered for marking purposes.**
2. **There will be penalties for the inclusion of unnecessary information.**
3. **must be included with proper labelling of graphs and charts.**
4. **Read this entire document carefully.**

**You will analyze a data set that comes from a random sample of the National Collision Database for 2017.** This data set was retrieved from the Government of Canada website. It includes all motor vehicle collisions in Canada on public roads in 2017 which have been reported to the police. This data set contains a variety of data variables which are summarized in the data dictionary posted on the Government of Canada website.

Source:

Government of Canada. (2017). *National Collision Database.*

**Description of Data Variables:**

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Year |  |
| Month | 01-January up until 12-December |
| Day of week | 01-Monday up until 07-Sunday |
| Collision hour | 00:01 to 23:59 |
| Collision severity | 1 – fatal or 2 – non-fatal |
| Number of vehicles involved in collision |  |
| Weather condition | 1 – clear and sunny, 2 – cloudy, 3 – raining, 4 – snowing, 5 – freezing rain, 6 – limited visibility, 7 – strong wind |
| Road surface | 1 – dry, 2 – wet, 3 – snow, 4 – slush, 5 – icy, 6 – sand/gravel/dirt, 7 – muddy, 8 – oil, 9 – flood |
| Road alignment | 1 – straight and level, 2 – straight with gradient, 3 – curved and level, 4 – curved with gradient, 5 – top of hill, 6 – bottom of hill |
| Vehicle ID |  |
| Vehicle type | **See Below.** |
| Vehicle model year |  |
| Person ID |  |
| Person sex | 01 – Female, 02 – Male |
| Person age |  |
| Person position | **See Below.** |
| Medical treatment required | 1 – no injury, 2 – injury, 3 – fatality |
| Safety device used | 01 – none, 02 – safety device used, 09 – helmet worn, 10 – reflective clothing worn, 11 – helmet and reflective clothing, 12 – other safety device used, 13 – no safety device equipped |
| Road user class | 1 – Motor Vehicle Driver, 2 – Motor Vehicle Passenger, 3 – Pedestrian, 4 – Bicyclist, 5 - Motorcyclist |

**Vehicle Type.**

|  |  |  |
| --- | --- | --- |
| **Code** | **Description** | |
| 01 | Light Duty Vehicle (Passenger car, Passenger van, Light utility vehicles and light duty pickup trucks) | |
| 05 | Panel/cargo van <= 4536 KG GVWR | Panel or window type of van designed primarily for carrying goods. |
| 06 | Other trucks and vans <= 4536 KG GVWR | Unspecified, or any other types of LTVs that do not fit into the above categories (e.g., delivery or service vehicles, chip wagons, small tow trucks etc.) |
| 07 | Unit trucks > 4536 KG GVWR | All heavy unit trucks, with or without a trailer |
| 08 | Road tractor | With or without a semi-trailer |
| 09 | School bus | Standard large type |
| 10 | Smaller school bus | Smaller type, seats < 25 passengers |
| 11 | Urban and Intercity Bus |  |
| 14 | Motorcycle and moped | Motorcycle and limited-speed motorcycle |
| 16 | Off road vehicles | Off road motorcycles (e.g. dirt bikes) and all terrain vehicles |
| 17 | Bicycle |  |
| 18 | Purpose-built motorhome | Exclude pickup campers |
| 19 | Farm equipment |  |
| 20 | Construction equipment |  |
| 21 | Fire engine |  |
| 22 | Snowmobile |  |
| 23 | Street car |  |

**Person Position.**

|  |  |  |
| --- | --- | --- |
| **Code** | **Description** | |
| 11 | Driver | |
| 12 | Front row, center | |
| 13 | Front row, right outboard, including motorcycle passenger in sidecar | |
| 21 | Second row, left outboard, including motorcycle passenger | |
| 22 | Second row, center | |
| 23 | Second row, right outboard | |
| 31 | Third row, left outboard | |
| 32 | Third row, center | |
| 33 | Third row, right outboard | |
| etc. |  | |
| 96 | Position unknown, but the person was definitely an occupant | |
| 97 | Sitting on someone’s lap | |
| 98 | Outside passenger compartment | e.g. riding in the back of a pick-up truck |
| 99 | Pedestrian |  |

**How the Project is graded**

Your submission will be graded based upon the following factors: substance, presentation, accuracy, grammar and clarity. A demonstration of effort is the driving force of this assignment. Assignments will be compared to discern levels of effort and excellence.

As a minimum, your report must include the following:

1. Cut and paste all relevant outputs in the write-up section at the bottom of your answer to each question.
2. A complete write up of your chosen hypothesis test must include your assumptions, analysis of results and your conclusions. **You must use both approaches (critical value and p-value approaches) to make your statistical decision.**
3. Not using the **exact** data set assigned to your group will result in getting a zero mark for the project.
4. **All data analyses must be done with SPSS and/or EXCEL. Only critical values can be found using the recommended calculator or EXCEL.**
5. **Number the pages of your report.**

**Solutions for some potential problems with your group:**

1. Any detection of plagiarism in the report will be charged with academic misconduct and all groups will receive a zero mark. You must use the Turnitin when you submit your project on D2L.

Each GROUP HAS ITS OWN unique Data Set. The data set assigned to your group

consists of 600 records of the motor vehicle collisions.

**Question 3 (10 marks)**

1. Use the variable "**Person age**" to construct the confidence intervals for the estimate of population mean “**Person age**” involved in a collision on a “**cloudy day”** (Code=2), at both 92% and 96% levels. Interpret your confidence intervals.
2. Did you make any assumptions when constructing your confidence intervals? If yes, which assumptions; if not, why?
3. Based on your finding, is it warranted for insurance companies to charge drivers higher premiums based on their collision records? Explain.

**SPSS PROJECT HINTS: avoid these pitfalls**

* You misread the question.
* You used the wrong test (e.g., Using a Z test instead of a t test).
* Your hypothesis was in the wrong direction (or H1 has > or < instead of ≠).
* The null hypothesis or the alternative hypothesis (or both) was wrong.
* You came to a wrong conclusion.
* You used the wrong data (or Incorrect inputs).
* A hypothesis with a μ or p or has the wrong one.
* Ha contains one of {= ,≤ or ≥} OR Ho contains one of {> , < , or ≠}.
* You used sample statistics in your hypotheses.
* Failed to check the requirements to use a test
* Misread p-values or comparison of p to α is wrong.
* Reaching a wrong conclusion, i.e., rejecting H0, when p > α.
* There is no statistical decision (or a wrong one).
* There is no managerial conclusion (or a bad one).
* The test is a one-sided test (not 2-sided).
* Not taking ½ of the Sig value from SPSS for a 1-sided test.
* You failed to state the problem and/or define the variables.
* A printout of your DATA SET IS MISSING! It had to be included!
* Missing LEVENE TEST of homogeneity.
* Forget to discuss or check for normality.