STAT 337 S21 Assignment 1   Due: 11:59pm EDT Friday, June 11ᵗʰ 2021

Notes for Submission: Upload your assignment directly to Crowdmark via the link you received by email. It is your responsibility to make sure your solution to each question is submitted in the correct section, that the pages are rotated correctly, and that everything is legible. Typed solutions are preferred.

Question 1 (12 marks):

The tables below describe the experience of two independent closed populations of 7 people susceptible to two different diseases. The studies begin on January 1 and observation ends on December 31. At the start of each month, the subjects' health status is assessed and coded as per the tables listed below.

For each disease (1) and (2) calculate the following:

(a)   (2 marks) The period prevalence for the interval January 1 to December 31.
(b)   (2 marks) The point prevalence on July 15?
(c)   (2 marks) The cumulative incidence of the disease over the year?
(d)   (4 marks) Calculate the incidence density for this disease using a person-time calculation in the denominator. Assume these diseases are recurrent (i.e. after recovery survivors re-enter the population at risk, see subject 2 of disease 1).
(e)   (2 marks) Calculate the annual crude death/ mortality rate for these populations.

Table 1: Disease 1 – Very Dangerous Disease

Subject	Diagnosis	Recovers	Diagnosis	Death
1	< Jan 1			May 31
2	May 1	June 30	Oct 1	Oct 31
3	July 1			
4	Apr 1	Aug 31		
5	May 1			June 30
6	na			
7	< Jan 1			

Table 2: Disease 2 – Rapidly Fatal Disease

Subject	Diagnosis	Recovers	Diagnosis	Death
1	June 1			June 30
2	na			
3	na			
4	Feb 1			Feb 28
5	na			
6	na			
7	Mar 1			Apr 30

Question 2 (15 marks):

The data in Table 3 gives the number of people age >=18 with Diabetes and the population counts for men and women in Ontario for the year 2019. The table also provides the number of people with Diabetes and the population size in Canada in 2019. For the following parts you are free to do your calculations in any software program (SAS, R, Excel, etc.) but your solutions should clearly explain the steps you used in the computation and given the formulas that you used.

Table 3: Number of Diabetes Cases by Age Group for Females and Males in Ontario in 2019

	Ontario Females		Ontario Males		Canada	
Stratum	**Diabetes**	**Population**	**Diabetes**	**Population**	**Diabetes**	**Population**
18 - 34	16600	1499360	32100	1588773	93200	7706705
35 - 49	61300	1438551	62500	1379456	323200	7401067
50 - 64	170200	1519258	223100	1479350	845400	7769520
>=65	233600	1373033	279900	1136265	1217900	6595070

Source: Statistics Canada. Table 13-10-0096-07 Diabetes, by age group and Table 17-10-0005-01 Population estimates on July 1st, by age and sex

(a) (2 marks) Compare the population stratum distributions of the men versus the women in Ontario.

(b) (3 marks) Calculate the stratum-specific and overall prevalence rates for the men and women.

(c) (4 marks) Use Direct Standardization to compute the Directly Adjusted Rates (DAR) for the men and women using the given national reference population.

(d) (4 marks) Use the reference population rates and Indirect Standardization to compute Standardized Prevalence Ratios (SPR) and Indirectly Adjusted Rates (IAR) for the men and women.

(e) (2 marks) Provide a few sentences comparing the diabetes prevalence of the men and women of Ontario in 2019. Make specific reference to which measures (i.e. DAR, SPR, IAR) you are using in your comparison(s).

Question 3 (10 marks):
For each study described below determine the risk factor(s) or exposure(s) under study, the outcome(s) under study, and indicate what medical study design is being described.

(a) (2 marks) Abete et al (2018) assessed the health histories and various lifestyle-related factors in a group of 51 stroke survivors and 51 similar subject matched according to sex, age, and BMI. Patients suffering a stroke showed higher prevalence of diabetes (30% vs. 7.7%) and hypertension (74.5% vs. 40.3%) and were less physically active (36.7% vs. 66.6%) than those who did not suffer a stroke.

(b) (2 marks) Pereira-Morales et al. (2018) investigated the effect of computer-assisted cognitive training in healthy older adults with subjective memory complaints. Forty older adults with subjective memory complaints were randomly assigned to an integrated psychostimulation program, computerized cognitive training (CCT) or a control group. Effectiveness was evaluated with a neuropsychological assessment battery, used before and after the implementation of the cognitive training. A decrease in anxiety symptoms was observed after the cognitive intervention in the IPP and CCT groups.

(c) (2 marks) Jiang et al. (2018) conducted a study to determine whether early food allergy and/or respiratory allergy symptoms are associated with the prevalence of ADHD in Chinese school-age children. Parents completed a questionnaire which assessed the diagnosis his tory of ADHD and allergic diseases (food allergy, allergic rhinitis, and bronchial asthma). The analysis showed that having a food allergy was associated with an increased prevalence of ADHD.

(d) (2 marks) Lidon-Moyano et al. (2017) assessed the correlation between the implementation of tobacco control policies and tobacco consumption, particularly rolling tobacco, electronic cigarettes (e-cigarettes) users and the intent to quit smoking in 27 countries of the European Union. Data on tobacco control activities were measured by the Tobacco Control Scale (TCS), in 27 European countries, in 2010, and the prevalence of tobacco consumption data from the Eurobarometer of 2012. Analysis showed that the level of smoke-free legislation was correlated with a decrease in the prevalence of smoking of conventional cigarettes and an increase in the intent to quit smoking within the past 12 months.

(e) (2 marks) Rai et al. (2017) examined the relation between the Dietary Approaches to Stop Hypertension (DASH) and Western diets and risk of gout in men. Participants were 44,444 men with no history of gout at baseline. Food frequency questionnaires were used to assign each participant a DASH dietary pattern score (based on high intake of fruits, vegetables, nuts and legumes, low fat dairy products, and whole grains, and low intake of sodium, sweetened beverages, and red and processed meats) and a Western dietary pattern score (based on high intake of red and processed meats, French fries, refined grains, sweets, and desserts). During 26 years of follow-up, 1731 confirmed cases of incident gout were documented. A higher DASH dietary pattern score was associated with a lower risk for gout while a higher Western dietary pattern score was associated with an increased risk for gout.

Question 4 (13 marks):
This question is based on the following paper, the paper has been added to the course reserves and the link can be found under the Assignment 1 folder on Learn.

Gillborg, S., Akerman, S., Ekberg, E. (2020) Tooth wear in Swedish adults – A cross-sectional study. Journal of Oral Rehabilitation, 47: 235-245.

a) (2 marks) In your own words state the scientific research question the authors were trying to address in this study.
b) (3 marks) Other than the Daily intake of acidic drinks, list one of the other exposures of interest in this study. In addition, discuss how the sample was selected. Be specific as to any inclusion/exclusion criteria used.
c) (2 marks) What were the primary outcome(s) of interest in this study? How were the outcomes ascertained in the study participants?
d) (2 marks) The Table below summarises some results given in Table 4 in the paper. Use the data to calculate the prevalence of the outcome in the exposed and unexposed groups.

Table 4: Frequency table of Tooth wear status vs. Exposure to acidic drinks

	Not Diseased (No or low risk)	Diseased (Medium or High risk)
Exposed (Daily intake of acidic drinks)	90	77
Unexposed (Almost never)	265	239

e) (2 marks) Without doing any formal statistical tests, do you think there is an association between the exposure and outcome given in part d)? Write 1-2 sentences describing the results of this study.
f) (2 marks) Describe how you would run a cohort study to assess the same potential association.

Question 5 (10 marks)
Cho, A. et al (2018) conducted a cohort study to investigate the potential association between cigarette smoking and thyroid cancer risk. The paper, listed below, has been added to the course reserves and the link can be found under the Assignment 1 folder on Learn.

Cho, A. Chang, Y. et al (2018) Cigarette smoking and thyroid cancer risk: a cohort study. British Journal of Cancer, 119: 638 – 645.

Table 5 summarises the incidence of disease (thyroid cancer) by the exposure (smoking status). The information is obtained from Tables 2 and 3 of the paper.

Table 5: Incidence of Thyroid Cancer by smoking status

	Not Diseased	Diseased
Exposed (ever or current smoker)	40866	272
Unexposed (Never smoked)	54739	978

(a) (3 marks) Calculate the Relative Risk (RR) and Risk Difference (RD). Give an interpretation of each statistic in the context of this study.

(b) (2 marks) ~~Calculate the Attributable Risk Percent (AR%) for smokers vs. non-smokers. Give an interpretation of this statistic in the context of this study.~~ In the course notes we discuss the notion of attributable risk percent, which relates the excess risk in exposed vs. unexposed as the percentage of the absolute risk. In certain applications it makes more sense to solve for the Preventative fraction expressed as a percentage. This is defined as:

$$(1 - RR) \times 100\%$$

And captures the percentage of disease reduction in the exposed group that can be attributed to the exposure.

Use the data to solve for the Preventative fraction expressed as a percentage. Give an interpretation of this statistic in the context of this study.

(c) (3 marks) Using the data in Table 5 calculate an appropriate measure of association for this data and test its significance using a χ^2 test. Give the formula for the test statistic, calculate its value and find the p-value. What is the conclusion of the test?

(d) (2 marks) Write 1-2 sentences summarizing the results of your analysis in parts (a) to (c).