Assignment 3 (Week 5 - 6)

STAT 2601 - Business Statistics (2024 Fall)
SCHOOL OF MATHEMATICS AND STATISTICS, CARLETON UNIVERSITY

Due Date and Time: Wednesday 30 October 2024, before 10:00 am
Total Marks: 34

Q1: [10] Canadian Household Debt

Based on a report of Canada Mortgage Housing Corporation (CMHC), Canada had the highest household debt among G7 countries in 2023. The household debt of Canadian households excluding mortgages in 2023 followed an approximately normal distribution with a mean of $\mu = \$41,500$ and a standard deviation of $\sigma = \$3,900$. Let X represent the debt of Canadian household excluding mortgages in 2023.

(a) [2.5] What percentage of Canadian households had burden of debts above $50,000 in 2023?

(b) [2.5] What is the probability that a randomly selected Canadian household had household debt excluding mortgages between $35,000 and $45,000 in 2023?

(c) [2.5] What was the household debt of the top 20% Canadian household in 2023?

(d) [2.5] Suppose a Canadian household debt was at 65th percentile in 2023. How much was the debt?

Q2: [6] Real Estate

The data set **RealEState** contains information on the listings of 1,047 real estate properties.

(a) [1] Treating this data set as the population, use EXCEL to calculate the population mean of the Price variable. Report this value and include the EXCEL output.

(b) [1] Use EXCEL to produce a histogram of the Price variable and include the output.

(c) [2] Following Lab 3, use EXCEL to draw twenty (20) samples of size n = 30 from the Price population. For each sample, use EXCEL to calculate t-based 95% confidence interval estimate for the population mean assuming that population standard deviation (σ) is unknown. Report only the confidence interval for each sample and fill in the following table.

Samples	Confidence Interval for μ	Covered μ (Y/N)?
Sample 1		
Sample 2		
$\vdots$		
Sample 20		

(d) [2] Count the number of confidence intervals out of your twenty that contain the true value of the population mean from part a). Does this value match the confidence level specified in part c)? Should this always be the case? Justify.

<div style="text-align: center;">

EXCEL Instructions

</div>

- **Descriptive Statistics:**
  Select data "A2:A1048" > Data > Data Analysis > Descriptive Statistics > OK > Select the Data (without the variable name) by clicking on "↑" of the Input Range (or type $A$2:$A$1048) > "Summary Statistics" > OK.

- **Histogram:**
  Select data "A2:A1048" > Insert > Insert Statistics Chart > Histogram > Double click on Chart Title and type "Distribution of Real Estate Price Population.

- **Generation of Samples:**
  Create Sample Labels: Sample 1 - Sample 20 (Column C - V).
  Sample 1: Data > Data Analysis > Sampling > OK > Select the Data (without the variable name) by clicking on "↑" of the Input Range (or type $A$2:$A$1048) > Random (Number of Samples) (30) > Output Range ($C$2:$C$31) > OK.
  Samples 2 - 20: Replicate the above steps to generate samples Sample 2 – 20 in columns D – V. When repeating the above steps, only output range needs to be updated for outputting the specified sample as other fields are auto populated from the last action. For instance, in case of sample 2, the output range will be ($D$2:$D$31).

- **Confidence Intervals:**
  Sample 1:

  (i) Type "=AVERAGE(C2:C31)" in cell C33 and hit Enter. (Sample Mean)

  (ii) Type "=STDEV.S(C2:C31)" in cell C34 and hit Enter. (Sample SD)

  (iii) Type "=30" in cell C35 and hit Enter. (Sample Size).

  (iv) Type "=TINV(0.05,C35-1)" in cell C36 and hit Enter. (95% t CV).

  (v) Type "=C33 + (C36*C34/SQRT(C35))" in cell C37 and hit Enter. (95% UCL for Population Mean)

  (vi) Type "=C33 - (C36*C34/SQRT(C35))" in cell C38 and hit Enter. (95% LCL for Population Mean).

  Samples 2 - 20:
  Replicate the above steps to generate confidence intervals for the remaining samples 2 – 20 in columns D – V. Now, select the cells "C33:C38", hover over the bottom right corner of the cell until a tiny "+" (called "fill-handle") appears and drag the formula down "V33:V38".

Q3: [5] Online Ordering

J & K, a retail drug store in Ottawa, is assessing the efficacy of a new online ordering system. A random sample of 120 customers is chosen and 85 out of 120 customers responded favourably to the new online ordering system.

(a) [2] What sample size would be required to estimate the population proportion of customers who favour the online ordering system with 90% confidence and 2% margin of error?

(b) [1.5] Calculate 98% confidence interval for the population proportion of customers who favour the online ordering system.

(c) [1.5] What assumptions and conditions are required for the confidence interval in (b) to be valid? Explain.

Q4: [8] Airport Wait Time for Check In

The wait time to check in for the passengers at Ottawa International Airport follows a Normal distribution with a mean of $\mu = 5$ minutes and a standard deviation of $\sigma = 2$ minutes. Let $\bar{X}$ represent the average wait time to check in of 35 randomly selected passengers.

(a) [1.5] What is the sampling distribution of $\bar{X}$?

(b) [2.5] What is the probability that the average wait time to check in of these 35 passengers is between 4 and 5.5 minutes?

(c) [2.5] If the distribution of wait time is positively skewed, would the answer in part (b) still be valid? Justify.

(d) [1.5] If the distribution of wait time is positively skewed, what is the probability that the average wait time to check in of 15 randomly selected passengers is between 4.5 and 6 minutes?

Q5: [5] Air Canada Flight Delay

Passengers of Air Canada in Ottawa have lodged several complaints of flight delays recently, so the CEO plans to conduct a study to measure the average amount of time Air Canada planes are delayed in Ottawa. He has also learned that the delay time of Air Canada flights follows a Normal distribution with standard deviation of 10 minutes.

(a) [2] What sample size would be required to estimate the average delay time of Air Canada flights in Ottawa to be within 2 minutes with 95% confidence?

(b) [1.5] A random sample of 15 Air Canada flights in Ottawa shows an average delay of 25 minutes. Calculate 95% confidence interval for the population averge delay time of Air Canada flights in Ottawa.

(c) [1.5] What assumptions and conditions are required for the confidence interval in (b) to be valid? Explain.