1. **(10 marks)** The number of hours per day that students spend on their smartphones follows a normal distribution with mean μ and standard deviation σ. Two students, Rachel and Ross, each collected two independent and random samples of six observations. Ross's sample produced $\bar{x} = 5.13$ and $s = 1.65$. The following table contains the six observations in Rachel's sample.

y_i:					
8.5	5.0	3.0	2.8	6.0	5.5

a) **(2 marks)** Calculate the sample mean of Rachel's sample.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{8.5+5.0+3.0+2.8+6.0+5.5}{6} = 5.13$$

b) **(2 marks)** Calculate the sample standard deviation of Rachel's sample.

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{(8.5-5.13)^2+(5.0-5.13)^2+(3.0-5.13)^2+(2.8-5.13)^2+(6.0-5.13)^2+(5.5-5.13)^2}{5}}$$

$$= 2.1087$$

c) **(2 marks)** ***Briefly*** explain which sample (ie, Ross's or Rachel's) better estimates the unknown population mean.

$\bar{x} = \bar{y}$ but $S_x < S_y$

Ross's is better because it has smaller S.D.

d) **(2 marks)** ***Briefly*** describe the sampling distribution of the sample mean. Specify any assumptions that you make.

- normal sampling distribution.
- Assumptions: if it is from normal distribution population, the sampling distribution of the mean is always normally distributed.

e) **(2 marks)** Rachel suggests that it would be a serious problem if the distribution of the number of hours that students spend on their smartphones is unknown. Briefly explain (i) whether or not you agree with her concern, (ii) why or why not, and (iii) is a solution to her concern available or not. (Hint: think of the Central Limit Theorem).

(i) & (ii) No, t-distribution or increase sample size

(iii) CLT: $n > 25$, normal sampling distribution of the mean

2. **(12 marks)** The amount of money spent by Tim Horton's customers is assumed to follow a normal distribution with mean μ and a standard deviation $\sigma = \$1.50$. A random sample of 25 customers spent an average of \$8.50.

a) **(2 marks)** Estimate the unknown population mean μ.

$\mu_{\bar{x}} = \mu = 8.50$

b) **(2 marks)** Calculate the standard deviation of the sampling distribution of $\bar{x}$.

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.50}{\sqrt{25}} = 0.3$

c) **(2 marks)** Calculate the probability that the sample mean exceeds \$9.00?

$P(\bar{x} > 9) = P(z > \frac{9-8.5}{0.3}) = P(z > 1.67) = 1 - F(1.67) = 1 - 0.9525$

$= 0.0475.$

d) **(2marks)** Construct a 95% confidence interval for μ.

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 8.5 \pm 1.96 \times \frac{1.5}{\sqrt{25}} = (7.912,\ 9.088)$$

e) **(2 marks)** ***Briefly*** interpret the confidence interval obtained in part (d).

We are 95% confident that the population mean lies between \$7.912 and \$9.088.

f) **(2 marks)** Calculate the required sample size if the sample average is to be within ~~two~~ 0.2 dollars of the unknown population mean. Use a 95% confidence level.

$$n = \frac{z^2_{\frac{\alpha}{2}} \sigma^2}{ME^2} = \frac{1.96^2 \times 1.5^2}{0.2^2} = 216.09, \text{ round up to } 216.$$

3. **(12 marks)** An online shopping website handled an average of 400 customers per day during 30 randomly selected days. Let the daily number of customers be a random variable with a sample variance equal to 225 customers per day.

a) **(2 marks)** Calculate the margin of error for a 99% confidence interval for the true population mean of the daily number of customers.

$$ME = t_{29,0.005} \frac{s}{\sqrt{n}} = 2.756 \left(\frac{\sqrt{225}}{\sqrt{30}} \right) = 7.558$$

OR: $z = 2.58$

$$ME = 2.58 \left(\frac{\sqrt{225}}{\sqrt{30}} \right) = 7.075.$$

b) **(2 marks)** Construct a 99% confidence interval for the true population mean of the daily number of customers.

$CI = \bar{x} \pm ME = 400 \pm 7.558 = (392.442,\ 407.558)$

OR $(392.925,\ 407.075)$

c) **(2 marks)** ***Briefly*** explain which statistical distribution must be used to construct the margin of error and confidence interval in parts (a) and (b).

We need to assume that an independent random sample was selected from a normally distributted population.

d) **(2 marks)** Suppose that the current system can handle [redacted]ustomers per day. If the objective of this website is to handle the daily traffic at least 95% of the time, calculate the minimum increase in customers per day that the website can handle.

Since the current average lies between CI, it doesn't exceed the system, the minimum increase is $410 - 407 = 3$.

e) **(2 marks)** Suppose that the population variance is equal to 100. Construct a 95% margin of error for the true population variance of the daily number of customers.

$-\ \dfrac{(n-1)s^2}{\chi^2_{n-1,\,\alpha/2}} = \dfrac{(30-1)\times 100}{45.722} = 63.427$ as: $\chi^2_{29,\,0.025} = 45.722$

$\chi^2_{29,\,0.975} = 16.047$

$-\ \dfrac{(n-1)s^2}{\chi^2_{n-1,\,1-\alpha/2}} = \dfrac{(30-1)\times 100}{16.047} = 180.719$

f) **(2 marks)** Suppose that the population variance is equal to 100. Construct a 95% confidence interval for the true population variance of the daily number of customers.

$CI: (63.427,\ 180.719)$

4. **(10 marks)** The following table gives the weight of six randomly selected users before and after implementing a diet programme.

	(X) Weight (lbs.)	(Y)
User	Before implementation	After implementation
1	136	125
2	205	195
3	157	150
4	138	140
5	175	165
6	166	160

Difference
d_i
11
10
7
-2
10
6
Σ 42

a) **(2 marks)** Calculate the difference in weight loss for each user.

$d_i = y_i - x_i$

b) **(2 marks)** Calculate the sample mean difference in weight loss.

$$\bar{d} = \frac{\Sigma d_i}{n} = \frac{42}{6} = 7.0$$

c) **(2 marks)** Calculate the sample standard deviation of the difference in weight loss.

$$S_d = \sqrt{\frac{\Sigma (d_i - \bar{d})^2}{n-1}} = 4.816$$

d) **(2 marks)** Calculate the margin of error for a 95% confidence interval for the true mean difference in weight loss.

$$ME: \ t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}} = (2.571) \times \frac{4.816}{\sqrt{6}} = 5.056$$

$$OR: \ z_{\frac{\alpha}{2}} = 1.96 \rightarrow ME = 3.854$$

e) **(2 marks)** Construct a 95% confidence interval for the true mean difference in weight loss.

$$C.I. = \ \bar{d} \pm t_{n-1,\alpha/2} \frac{S_d}{\sqrt{n}} = 7 \pm (2.571) \times \frac{4.816}{\sqrt{6}} = (1.94, \ 12.06)$$

$$OR, \ Z_{\frac{\alpha}{2}} = 1.96 \rightarrow C.I. = (3.146, \ 10.854)$$

5. **(14 marks)** A student who misses more than three classes per semester is considered a "non-attending" student. A random sample of 300 students (180 females and 120 males) were asked if they were "non-attending". Thirty females and 40 males said that they were. Let p_1 and p_2 be the respective population proportions of female and male students that are non-attending.

a) **(2 marks)** Calculate the sample proportions of non-attending female and male students.

$$\hat{p}_1 = \frac{30}{180} \doteq \frac{1}{6}$$

$$\hat{p}_2 = \frac{40}{120} = \frac{1}{3}$$

b) **(2 marks)** Calculate the margin of error for a 95% confidence level for the true population proportion of non-attending males.

$$ME = z \sqrt{\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 1.96 \sqrt{\frac{0.3(1-0.3)}{120}} = 0.082.$$

c) **(2 marks)** Construct a 95% confidence interval for the true population proportion of non-attending males.

$$C.I. = \hat{p}_2 \pm ME = 0.3 \pm 0.082 = (0.218, 0.382)$$

d) **(2 marks)** ***Briefly*** interpret the confidence interval in part (c).

We are 95% confident that the proportion of "non-attending" males lies between 29.7% and 30.3%.

e) **(2 marks)** Assume that we want the estimator $\hat{p}_2$ to be within three percentage points of the true population proportion with 95% confidence. Calculate how many male students need to be sampled.

$$n = \frac{Z^2 (0.25)}{ME^2} = \frac{1.96^2 \times 0.25}{0.03^2} = 1067.1 \text{ round up to } 1067$$

f) **(2 marks)** Calculate the margin of error for a 95% confidence level for the true population difference between p_1 and p_2.

Recall: $\hat{p}_1 = \frac{30}{180} = 0.167, \quad \hat{p}_2 = \frac{40}{120} = 0.3$

$$ME = Z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} = 1.96\sqrt{\frac{0.167(1-0.167)}{180} + \frac{0.3(1-0.3)}{120}}$$

$$= 0.0049.$$

g) **(2 marks)** Construct a 95% confidence interval for the true population difference between p_1 and p_2.

$$C.I. = (\hat{p}_1 - \hat{p}_2) \pm ME = (0.167 - 0.3) \pm 0.0049$$

$$= (-0.1379, -0.1281)$$

6. **(6 marks)** The number of daily parking violations in Montreal is assumed to be normally distributed with mean μ and variance σ^2. Over a period of 25 days, the sample variance was found to be $s^2 = 100$.

a) **(2 marks)** Calculate the lower confidence limit for a 95% confidence interval for the population variance.

with $\chi^2_{n-1, 1-\alpha/2} = \chi^2_{24, 0.975} = 12.401$

$\chi^2_{n-1, \alpha/2} = \chi^2_{24, 0.025} = 39.364$

$$LCL = \frac{(n-1)s^2}{\chi^2_{n-1, \alpha/2}} = \frac{24 \times 100}{39.364} = 60.97$$

b) **(2 marks)** Calculate the upper confidence limit for a 95% confidence interval for the population variance.

$$UCL = \frac{(n-1)s^2}{\chi^2_{n-1, 1-\alpha/2}} = \frac{24 \times 100}{12.401} = 193.53$$

c) **(2 marks)** ***Briefly*** interpret the results in parts (a) and (b).

We are 95% confident that the variance lies between 60.97 and 193.53.