Solutions **Part I**

1. a) [1] No. They differ from sample to sample..

   b) [3] This is to be expected since Minitab is selecting random samples and therefore each sample will have different values. Thus the values of the sample means and the sample variances will also differ.

   c) [4] No. Because random samples are being selected, different samples will be obtained each time you repeat the experiment..

2. a) [3] See Histograms

   b) [1] Does the histogram of the sample means based on samples of size 5 look approximately bell-shaped? __**No it is very skewed** __

   c) [3] Why is this to be expected? __**the Poisson distribution with $\mu = 0.1$ is very skewed and n = 5 is too small for the CLT to apply to the distribution of $\overline{X}$.**

   d) [1] Does the histogram of the sample means based on samples of size 30 look approximately bell-shaped? _ **No**__ most will still be fairly skewed.

   e) [3] Why might this be expected? _**The Poisson distribution with $\mu = 0.1$ is a very skewed distribution so n has to be quite large before $\overline{X}$ becomes approximately normal by the CLT.**

   f) [1] Does the histogram of the Poisson sample means based on samples of size 100 look approximately bell-shaped? ___**somewhat**___

3. a) [1] The value of $z_{\alpha/2}$ used to compute the interval is: __**$z_{.05} = 1.645$**__

   b) [1] The length of the interval is: _ **$2z_{.05}\ \sigma/\sqrt{n} = 2(1.645)(4/\sqrt{9}) = 4.3867$** __

   c) [1] The number of intervals over the 30 replications that actually contain μ is: _**will vary slightly**_

   d) [3] Compare the percentage of the intervals that contain μ to the stated confidence level of 90%: __ **(answer to (c))/30 X100% compared to 90%** ___

Part II

1. Let X= height of a North American woman. X is normal , $\mu = 64$, $\sigma^2 = 4$

   a) $P(X > 66) = P(Z > (66 - 64)/2) = P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 =$ **0.1587**

   b) $\overline{X}$ = average height of a random sample of n = 4 women. Therefore, $\overline{X}$ is normal, $\mu = 64$, standard error $\sigma/\sqrt{n} = 2/2 = 1$. Therefore,

   $P(\overline{X} > 66) = P(Z > (66 - 64)/1) = P(Z > 2) = 1 - P(Z < 2) = 1 - 0.9772 =$ **0.0228**

   c) n = 100, so standard error of $\overline{X}$ is $\sigma/\sqrt{n} = 2/10 = 0.2$

   $P(\overline{X} > 66) = P(Z > (66 - 64)/0.2 = P(Z > 10) =$ **0** to 4 decimal places

   d) If X is not normally distributed, then the only part that can be done is (c) since n = 100 is large enough for the CLT to apply. Cannot do (a) as the distribution of X is not known. Also cannot do (b) as n = 4 is not large enough to apply the CLT so the distribution of $\overline{X}$ is unknown.

2. X = # of customers/week at a store. Distribution of X has $\mu = 5000$, $\sigma = 500$

   $\overline{X}$ = average # of customers of a random sample of n = 64 stores. $E(\overline{X}) = \mu = 5000$,

   s.e.($\overline{X}$) = $\sigma/\sqrt{n}$ = 500/8 = 62.5 The CLT should apply for n = 64 (unless distribution very skewed)

   a) Approximate $P(4980 < \overline{X} < 5075) = P\{(4980 - 5000)/62.5 < Z < (5075 - 5000)/62.5\} =$
   $P(-0.32 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.32) = 0.8849 - 0.3745 =$ **0.5104**

   b) $P(\overline{X} < 4075) = P\{Z < (4075 - 5000)/62.5\} = P(Z < -14.8) =$ **0** to many decimal places.

c) Want $P(-75 < \overline{X} - \mu < 75) = P(-75/62.5 < Z < 75/62.5) = P(-1.2 < Z < 1.2) =$
$P(Z < 1.2) - P(Z < 1.2) = 0.8849 - 0.1151 =$ **0.7698**

3. $n = 144$ is large enough that $(\overline{X} - \mu)/\ s/\sqrt{n})$ will be approximately standard normal.

a) An approximate 99% confidence interval estimate for the true average vacation cost is
$(\bar{x} - z_{0.005}\ s/\sqrt{n},\ \bar{x} + z_{0.005}\ s/\sqrt{n}) = (2386 - 2.575[400/\sqrt{144}],\ 2386 + 2.575[400/\sqrt{144}])$
$= (2386 - 85.83,\ 2386 + 85.83) =$ **($2300.17, $2471.83)**

b) An approximate confidence interval is $(\bar{x} - 2.575\ s/\sqrt{n},\ \ \bar{x} + 2.575\ s/\sqrt{n})$ as in (a)
Thus want $2.575\ s/\sqrt{n} \leq 60$, i.e. $n \geq [\,2.575(400)/60]^2 = 294.69$ I.e. sample size **n = 295**

4. Want to find $\bar{x}$. A confidence interval is of the form $(\bar{x} - a,\ \bar{x} + a)$, therefore $\bar{x}$ is in the middle of the interval. For interval (5.6, 6.4), **$\bar{x}$ = (5.6 + 6.4)/2 = 6.0**

5. Let p_1 = population proportion aware of product in Toronto p_2 = proportion for Vancouver

a) An approximate 95% confidence estimate for $p_1 - p_2$ is

$$\left(\hat{p}_1 - \hat{p}_2 \pm z_{.025}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}\right) = (0.631 - 0.798\ \pm\ 1.96\ \sqrt{.631(.369)/1000 + .798(.202)/1000}\,) =$$

$(-\,0.167 - 0.0389,\ -\,0.167 + 0.0389) =$ **(- 0.2059. – 0.1281)**

b) We are approximately 95% confident that the true difference between the proportion of Toronto consumers that were aware of the product and the proportion of Vancouver consumers who were aware is between – 0.2050 and – 0.1281. That is, the proportion of Vancouver consumers is greater than that of Toronto consumers by 0.1281 to 0.2050 with approx. 95% confidence. (Or proportion of Toronto consumers < proportion of Vancouver ..)

6. Sample sizes are large enough for the CLT to apply and to use s_1^2 and s_2^2 to estimate σ_1^2 and σ_2^2.

a) An approximate 90% confidence interval estimate for $\mu_1 - \mu_2$ is $\left(\bar{x}_1 - \bar{x}_2 \pm z_{0.05}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}\right) =$

$(12 - 9 - 1.645\sqrt{6/100 + 4/100},\ 12 - 9 + 1.645\sqrt{6/100 + 4/100}\,) =$
(3 – 0.52, 3 + 0.52) = (2.48, 3.52)

b) We are 90% confident that the true mean downtime difference between machine I and machine II is between 2.48 and 3.52 minutes. I.e. we are 90% confident that the average downtime for machine I is greater than that for machine II by between 2.48 & 3.52 min.

7. X = # of voters out of sample of 250 who would support incumbent. n = 250 is large enough to apply the normal approximation to the binomial.

a) An approximate 90% confidence interval for the population proportion of voters who support the incumbent is $\left(\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}\,\hat{q}}{n}}\right) =$

$(0.4 - 1.645\sqrt{(0.4)(0.6)/250},\ \ 0.4 +\ 1.645\sqrt{(0.4)(0.6)/250}\,) =\ (0.4 - 0.05,\ 0.4 + 0.05) =$
(0.35, 0.45)

b) Want $z_{0.05}\sqrt{\hat{p}\hat{q}/n} \leq 0.04$ where use $\hat{p} = 0.4$ from (a). So $1.645\sqrt{(0.4)(0.6)}/\sqrt{n} \leq 0.04$.
So $n \geq (1.645/0.04)^2\,(.4)(.6) = 405.9$ Therefore would need a sample of at least **406**

c) If have no previous estimate for p use $\hat{p} = 0.5$. Therefore, $1.645\sqrt{(0.5)(0.5)}/\sqrt{n} \leq 0.04$
$n \geq (1.645/0.04)^2\,(.5)(.5) = 422.8$ Need a sample of at least **n = 423**