Question 1

a) The data is categorical, i.e., people who use Android and people who don't.

For categorical data we calculate a proportion.

b) The parameter here is the true proportion of individuals who use an Android device.

It's value is unknown, i.e., $p =$ unknown.

c) The statistic is the proportion calculated in our sample. It provides us with an estimate of p which is unknown.

Thus, $\hat{p} = \frac{359}{500} = 0.718$.

d) 95% confidence interval $\Rightarrow \alpha = 0.05$ remaining $\Rightarrow \alpha/2 = 0.025$ in each tail.

The 95% confidence interval for p is:

$$\hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.718 \pm 1.96 \sqrt{\frac{0.718(0.282)}{500}}$$

$$0.718 \pm 1.96\,(0.0201)$$

$$0.718 \pm 0.0394$$

$$[0.6786,\ 0.7574]$$

Thus, we are 95% confident that the true proportion, p, is between 0.6786 and 0.7574, i.e., that the true market share is between 67.86% and 75.74%.

e) In part (d) note that the margin of error was ± 0.0394. We want to find the value of n that makes this equal to ± 0.02.

$$\Rightarrow 1.96 \sqrt{\frac{0.718(0.282)}{n}} = 0.02$$

$$1.96 \frac{\sqrt{0.718(0.282)}}{\sqrt{n}} = 0.02$$

$$\frac{1.96 \sqrt{0.718(0.282)}}{\sqrt{n}} = 0.02$$

$$\Rightarrow \frac{1}{\sqrt{n}} = \frac{0.02}{1.96 \sqrt{0.718(0.282)}}$$

$$\sqrt{n} = \frac{1.96 \sqrt{0.718(0.282)}}{0.02}$$

$$n = \left(\frac{1.96 \sqrt{0.718(0.282)}}{0.02}\right)^2$$

$$= \frac{(1.96^2)\,[0.718(0.282)]}{(0.02)^2}$$

$$= 1944.57.$$

Thus, we need about 1945 individuals to reduce the margin of error to ± 0.02 (assuming that $\hat{p}$ will still be approximately 0.718 in the next sample).

Question 2

a) The data type is numeric continuous, i.e., money takes decimal values and is generally considered continuous.

In particular, since it is a numeric measurement we calculate the mean.

b) There is no p or $\hat{p}$ in this case since we are not dealing with categorical data.

We do have the following quantities however:

- $\mu =$ unknown.
- $\bar{x} = 42.38$.
- $\sigma =$ unknown.
- $s = 16.80$.

c) 99% confidence interval $\Rightarrow \alpha = 0.01$ remaining $\Rightarrow \alpha/2 = 0.005$ in each tail.

The 99% confidence interval for μ is:

$$\bar{x} \pm z_{0.005} \frac{s}{\sqrt{n}}$$

$$42.38 \pm 2.58 \frac{16.80}{\sqrt{1000}}$$

$$42.38 \pm 2.58\,(0.5313)$$

$$42.38 \pm 1.3707$$

$$[41.01,\ 43.75]$$

We are 99% confident that μ is in the above interval, i.e., that people who click on this Ad go on to spend between \$41.01 and \$43.75.

Question 2 (continued)

d) We wish to be 99% confident that the margin of error is ± 0.50 (rather than ± 1.3707).

$$z_{0.005}\,\frac{s}{\sqrt{n}} = 0.5$$

$$2.58\,\frac{16.80}{\sqrt{n}} = 0.5$$

$$\frac{2.58(16.80)}{\sqrt{n}} = 0.5$$

$$\frac{1}{\sqrt{n}} = \frac{0.5}{2.58(16.80)}$$

$$\sqrt{n} = \frac{2.58(16.80)}{0.5}$$

$$n = \left(\frac{2.58(16.80)}{0.5}\right)^2$$

$$= 7514.81.$$

Thus, we need about 7515 individuals to ensure a margin of error of ± 0.5, i.e., to be within $\pm$ 50c of the true average spend (assuming that s is still approximately equal to 16.80 in the next sample)

Question 3

a) Time taken $\Rightarrow$ numeric continuous. Thus we will be dealing with a mean.

b) $\mu =$ unknown.

c) $\bar{x} = 671.23$ hours.

d) We are sample variance $s^2 = 400$ hours^2. The standard deviation is $s = \sqrt{400} = 20$ hours.

99.9%, i.e., $1-\alpha = 0.999 \Rightarrow \alpha = 0.001$ remaining $\Rightarrow \alpha/2 = 0.0005$ in each tail.

$$\bar{x} \pm z_{0.0005}\,\frac{s}{\sqrt{n}}$$

$$671.23 \pm 3.29\,\frac{20}{\sqrt{45}}$$

$$671.23 \pm 3.29\,(2.981)$$

$$671.23 \pm 9.8089$$

$$[661.42,\ 681.04]$$

We are 99.9% confident that μ is in the above interval. In other words, this particular component is almost certain to last between 661.42 hours and 681.04 hours.

Question 4

a) The true difference in proportions is unknown, i.e., $p_1 - p_2 =$ unknown.

b) We will let index 1 correspond to rural and index 2 correspond to urban:
$\hat{p}_1 = 0.5263$ and $n_1 = 38$;
$\hat{p}_2 = 0.6034$ and $n_2 = 116$.

90% confidence $\Rightarrow \alpha = 0.1$ remaining $\Rightarrow \alpha/2 = 0.05$ in each tail.

$$(\hat{p}_1 - \hat{p}_2) \pm z_{0.05}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$(0.5263 - 0.6034) \pm 1.64\sqrt{\frac{0.5263(0.4737)}{38} + \frac{0.6034(0.3966)}{116}}$$

$$-0.0771 \pm 1.64\,(0.0929)$$

$$-0.0771 \pm 0.1523$$

$$[-0.2294,\ 0.0752]$$

We are 90% confident that the true difference, $p_1 - p_2$, is contained in the calculated interval.

In particular, note that the interval supports the possibility of no difference: $p_1 - p_2 = 0$. Thus, we conclude that there appears to be no difference in opinions.

c) 52.63% of 38 $\Rightarrow 0.5263(38) = 20$ individuals.

60.34% of 116 $\Rightarrow 0.6034(116) = 70$ individuals.

$\Rightarrow 20 + 90 = 90$ individuals overall from the group of 154 who are in favour of the policy.

d) Overall we have $\hat{p} = \frac{90}{154} = 0.5844$.

$$\hat{p} \pm z_{0.05}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.5844 \pm 1.64\sqrt{\frac{0.5844(0.4156)}{154}}$$

$$0.5844 \pm 0.0651$$

$$[0.5193,\ 0.6495]$$

We are 90% confident that the true overall proportion, p, of individuals in support of the policy is in the above interval.

Question 5

a) The parameter is the true difference between mean temperatures: $\mu_1 - \mu_2$. Of course, its value is unknown.

b) Note that we have:

$n_1 = 50$, $\bar{x}_1 = 40.1$, $s_1 = 2.5$.

$n_2 = 50$, $\bar{x}_2 = 34.8$, $s_2 = 1.1$.

95 % confidence $\Rightarrow \alpha = 0.05$ remaining $= \alpha/2 = 0.025$ in each tail.

$$(\bar{x}_1 - \bar{x}_2) \pm z_{0.025} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(40.1 - 34.8) \pm 1.96 \sqrt{\frac{(2.5)^2}{50} + \frac{(1.1)^2}{50}}$$

$$5.3 \pm 1.96 \sqrt{0.125 + 0.0242}$$

$$5.3 \pm 1.96\,(0.3863)$$

$$5.3 \pm 0.7571$$

$$[4.54,\ 6.06]$$

We are 95% confident that the true difference is between 4.54 and 6.06, i.e., using a Type 1 heat sink leads to a hotter CPU. Thus, we should use Type 2 heat sinks.

c) We want

$$1.96 \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0.4.$$

We assume that $n_1 = n_2 = n$ for simplicity.

$$\Rightarrow 1.96 \sqrt{\frac{(2.5)^2}{n} + \frac{(1.1)2}{n}} = 0.4.$$

$$1.96 \sqrt{\frac{6.25}{n} + \frac{1.21}{n}} = 0.4$$

$$1.96 \sqrt{\frac{6.25 + 1.21}{n}} = 0.4$$

$$1.96 \sqrt{\frac{7.46}{n}} = 0.4$$

$$\frac{1.96\sqrt{7.46}}{\sqrt{n}} = 0.4$$

$$\Rightarrow \frac{1}{\sqrt{n}} = \frac{0.4}{1.96\sqrt{7.46}}$$

$$\sqrt{n} = \frac{1.96\sqrt{7.46}}{0.4}$$

$$n = \left(\frac{1.96\sqrt{7.46}}{0.4}\right)^2$$

$$= 179.11.$$

Thus, we need about 180 CPUs of each type (i.e., 360 altogether) to reduce the margin of error to within ± 0.4.

Question 6

a) Here $n = 8$, $\bar{x} = 6.4$ and $s = 2.2$.

95% confidence $\Rightarrow \alpha = 0.05$ remaining $\Rightarrow \alpha/2 = 0.025$ in each tail.

Since the sample size is small ($n < 30$), we use a t value rather than a z value. Specifically, the degrees of freedom are $\nu = n - 1 = 8 - 1 = 7$.

$$\bar{x} \pm t_{7,\,0.025} \frac{s}{\sqrt{n}}$$

$$6.4 \pm 2.365 \frac{2.2}{\sqrt{8}}$$

$$6.4 \pm 1.8395$$

$$[4.56,\ 8.24]$$

We are 95% confident that computer science students spend, on average, between 2.56 and 8.24 hours gaming in a week.

b) 99% confidence $\Rightarrow \alpha = 0.01$ remaining $\Rightarrow \alpha/2 = 0.005$ in each tail.

$$\bar{x} \pm t_{7,\,0.005} \frac{s}{\sqrt{n}}$$

$$6.4 \pm 3.499 \frac{2.2}{\sqrt{8}}$$

$$6.4 \pm 2.7216$$

$$[3.68,\ 9.12]$$

Note: to increase our confidence in capturing the true μ, the interval width increases.

Question 7

a)

						$\sum$
x	34.1	33.5	32.8	33.1	32.5	166
x^2	1162.81	1122.25	1075.84	1095.61	1056.25	5512.76

$$\bar{x} = \frac{\sum x}{n} = \frac{166}{5} = 33.2.$$

$$s^2 = \frac{\sum x^2 - n\,\bar{x}^2}{n-1} = \frac{5512.76 - 5(33.2^2)}{4} = 0.39.$$

$$s = \sqrt{0.39} = 0.6245.$$

b) 95% confidence $\Rightarrow \alpha = 0.05$ remaining $\Rightarrow \alpha/2 = 0.025$ in each tail.

Since the sample size is small ($n < 30$), we use a t value with $\nu = n - 1 = 5 - 1 = 4$.

$$\bar{x} \pm t_{4,\,0.025}\,\frac{s}{\sqrt{n}}$$

$$33.2 \pm 2.776\,\frac{0.6245}{\sqrt{5}}$$

$$33.2 \pm 2.776\,(0.2793)$$

$$33.2 \pm 0.7753$$

$$[32.42,\ 33.98]$$

c) We are 95% confident that the true value, μ, lies in the above interval. Note that this interval includes the value $\mu = 33$. Thus, it seems that the machine is working as expected.