Question 1   Term Test 1

a) n = odd

median = $\frac{n+1}{2}$ th value in the ordered list.

-1.86%   -1.35%   -0.8%   1.57%   2.33%   2.71%   3.87%

median = 1.57%

$$\bar{x} = \frac{\sum_{i=1}^{7} x_{i,\,S\&P500}}{7}$$

$$= \frac{1.57 + (-1.35) + 3.87 - 1.86 + 2.33 + 2.71 - 0.8}{7}$$

$= 0.92\%$

(b)

Q1 :   -1.86%   -1.35   -0.8

$n = 3$

$Q_1 = -1.35\%$

2.33%   2.71%   3.87%

$n = 3$

$Q_3 = 2.71\%$

(c) $$S_{x_{S\&P500}} = \sqrt{\frac{\sum_{i=1}^{7}\left(x_i - \bar{x}_{S\&P500}\right)}{6}}$$

$$= \sqrt{\frac{(1.57 - 0.92)^2 + \dots\dots (-0.8 - 0.92)^2}{6}}$$

$$= 2.24$$

(d) From the histogram we note that S&500 monthly returns have one peak and therefore the distribution is <u>unimodal</u>.

Also the left tail is longer (i.e. the overall pattern has higher counts of small values than large values).
Therefore the distribution of S&P500 returns are left skewed or skewed towards small value.

(e) The median is larger then the mean.

This is because the distribubion is skewed towards small values. and there are outliers to the left (extremely small values which are deviations from the overall pattern). The mean is not resistent to outliers and skewness while the median is very resistant.

(f) The scatter points fit more or less on a straight upword slopping line. there fore the form is linear and direction is positive. The strenght is weak or moderate since there are large deviations from the upword slopping line.

Outliers are large deviation from the overall linear form and positive association.

Yes there are outliers. The scatter points with extremely small and

Negative Microsoft returns
with positive S&P 500 returns
are outliers.

Question 2

(a) Denote X^y as the weight of eggs produced by young hens.

$$X^y \sim N(50.9, \sigma^y)$$

$$P(X^y > 54) = 0.28$$

$$P\left(\frac{X^y - 50.9}{\sigma^y} > \frac{54 - 50.9}{\sigma^y}\right) = 0.28$$

$$P\left(Z > \frac{54 - 50.9}{\sigma^y}\right) = 0.28$$

where $Z \sim N(0,1)$

$$P\left(Z > \frac{3.1}{\sigma^y}\right) = 0.28$$

Recall
$P(Z > a) = 1 - P(Z < a)$
$\Rightarrow P(Z<a) = 1 - P(Z<a)$

$$P\left(Z < \frac{3.1}{\sigma^y}\right) = 0.72$$

from Z-table

$$\frac{3.1}{\sigma^y} = 0.58$$

$$\sigma^y = 5.34$$

(b) Denote X as the weight of eggs produced by hens that have reached age of one year.

$$X \sim N(67.1, \sigma)$$

$$P(X > 54) = 0.98$$

$$P(X < 54) = 0.02$$

$$P\left(\frac{X-67.1}{\sigma} < \frac{54-67.1}{\sigma}\right) = 0.02$$

$$P\left(z < -\frac{13.1}{\sigma}\right) = 0.02$$

From z-table

$$-\frac{13.1}{\sigma} = -2.05$$

$$\sigma = 6.39$$

(c) $\sigma^y = 5.34$ (standard deviation of younger hens)
$\sigma = 6.39$ (standard deviation of older hens)

$$\sigma^y > \sigma$$

Since standard deviation is a measure of spread around the mean, younger hens produce more consistent eggs.

(d) Denote $X \sim N(\mu, \sigma)$

$$P(X < 54) = 0.08$$

$$P\left(\frac{X-\mu}{\sigma} < \frac{54-\mu}{\sigma}\right) = 0.08$$

$$P\left(z < \frac{54-\mu}{\sigma}\right) = 0.08$$

From z-Table

$$\frac{54-\mu}{\sigma} = -1.41 \quad \text{Equation 1}$$

$$P(X > 70) = 0.13$$

$$P(X < 70) = 0.87$$

$$P\left(\frac{X-\mu}{\sigma} < \frac{70-\mu}{\sigma}\right) = 0.87$$

$$P\left(z < \frac{70-\mu}{\sigma}\right) = 0.87$$

From z-table $\rightarrow \frac{70-\mu}{\sigma} = 1.13$ Equation 2

$$\frac{54-\mu}{\sigma} = -1.41 \Rightarrow \sigma = \left.\frac{54-\mu}{-1.41}\right\} 1$$

$$\left.\frac{70-\mu}{\sigma} = 1.13\right\} 2$$

Substituting 1 into 2

$$70-\mu\left(\frac{-1.41}{54-\mu}\right) = 1.13$$

$$-98.7 + 1.41\mu = 1.13(54-\mu)$$

$$-98.7 + 1.41\mu = 61.02 - 1.13\mu$$

$$2.54\mu = 159.72$$

$$\boxed{\mu = 62.88} \quad 3$$

Substituting 3 into 1

$$\sigma = \frac{54 - 62.88}{-1.41} = 6.297 \approx 6.3$$

$$\boxed{\sigma = 6.3}$$

(e) (i) Judgemental Sampling.

The sampling procedure is likely to be biased since larger hens may produce heavier eggs produced by ten largest hens is not a good representation of the population.

(ii) Convenience Sampling

The sampling procedure is likely to be biased since heavier eggs may slide the furthest and therefore land closest to the entrance of the coup.

Question 3

(a)

Row Varicble : accidents

Column Varicble : cellphone

		accidents Yes	accidents No	Margin Row
Cellphone	Yes	20	58	78
Cellphone	No	10	92	102
	Margin column	30	150	Total 180

(b)

		accidents Yes	accidents No	Margin row
Cellphone	Yes	$\frac{1}{9}$	$\frac{29}{90}$	$\frac{13}{30}$
Cellphone	No	$\frac{1}{18}$	$\frac{23}{45}$	$\frac{17}{30}$
	Margin column	$\frac{1}{6}$	$\frac{5}{6}$	1

(c) Distribution of accidents condition on cellphone = yes

		accidents	
		Yes	No
Cellphone	Yes	$\frac{10}{39}$	$\frac{29}{39}$

Distribution of accidents conditional on cellphone = no

		accidents	
		Yes	No
Cellphone	No	$\frac{5}{51}$	$\frac{46}{51}$

(d) From part (b) (the joint distribution) we note that the relative frequency for accident = yes and cellphone = yes combination / interaction is larger than the relative frequency for accident = yes and cellphone = no combination / interaction

Similarly from part (c) the relative frequency of accidents = yes is larger for the cellphone = yes group than for cellphone = no group (i.e. $\frac{10}{39} > \frac{5}{51}$)

Therefore there appears to be a positive association between cellphone use and car accidents. However since this is an observation study a positive association does not necessarily imply a cause-and-effect relationship.

Question 4

(a) $S = \{HHH, THH, HTH, HHT, THT, TTH, HTT, TTT\}$

There a 8 possible outcomes

(b) $A = \{TTH, HTT, TTT\}$

$B^c = \{TTT\}^c$

$A \cap B^c = \{TTH, HTT\}$

$$P(A \cap B^c) = P(\{TTH\}) + P(\{HTT\})$$
$$= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

(c) $A = \{TTH, HTT, TTT\}$

$C = \{HHH, HTH, HHT, HTT\}$

$A \cup C = \{TTH, HTT, TTT, HHH, HTH, HHT\}$

$$P(A \cup C) = P(\{TTH\}) + P(\{HTT\}) + P(\{TTT\})$$
$$+ P(\{HHH\}) + P(\{HTH\}) + P(\{HHT\})$$
$$= \frac{6}{8}$$

(d)

$A = \{TTH, HTT, TTT\}$

$A^c = \{HHH, THH, HTH, HHT, THT\}$

To show that A and A^c partition the sample space we need to explicitly show that

(1) A and A^c are disjoint

and

(2) $A \cup A^c = S$

(1) A and A^c have no common outcomes therefore A and A^c are disjoint.

(2) $\{TTH, HTT, TTT\} \cup \{HHH, THH, HTH, HHT, THT\}$

$= \{TTH, HTT, TTT, HHH, THH, HTH, HHT, THT\}$

$= S$: sample space