Exercise 1:

(a) Do all students have the same chance of entering the sample? Explain your answer.

Solution: No. The probability of students entering the sample depends on whether the student is enrolled into the class or not. In cluster sampling, the individual units (sub sampling units) of the population are only included in the sample if they are in a cluster (primary sampling unit) that is in the sample.

(b) Describe three significant differences.

Solution: ① The Registrars's sample can contain repeated individuals. In this way, cluster samples are less precise than simple radom samples.

② The Registrars's sample doesn't have the same probability of the individual units. But for a simple random sample from among all the students, all of them have the same chance being selected to the sample.

③ The Registrars's sample divides the population into several groups. It takes classes into groups while the simple random sample from among all the students takes students in to a whole selected group.

④ The Registrars's sample obtains data on every sampling unit in each of the randomly selected clusters.

Exercise 2:

Solution: P(a woman over 50 has breast cancer) = 0.01 = P(A)

P(test positive | have cancer) = 0.9 = P(B|A)

P(test positive | no cancer) = 0.08 = P(B|A^c)

Interested in P(a woman over 50 has breast cancer | test positive)

= P(A|B)

$P(A|B) = P(B|A)P(A)/P(B)$

$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c)$

$= 0.9 \times 0.01 + 0.08 \times (1-0.01) = 0.0882$

$\Rightarrow P(A|B) = (0.9 \times 0.01) / 0.0882 = \frac{5}{49} = 0.1020408 \approx 10.2\%$

Exercise 3:

(a) Solution: Use Markov's inequality.

$P(X \geq t) \leq \frac{\mu}{t}$ $\quad E(X) = \mu = 50$

$\therefore P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$

the probability that this week's production will exceed 75 is $\frac{2}{3}$.

(b) Solution: $\sigma^2 = 25$ $\quad \mu = 50$

use Chebyshev's theorem:

$P\left(\frac{|X-\mu|}{\sigma} \leq k\right) \geq 1 - \frac{1}{k^2} \Rightarrow P\left(\frac{|X-50|}{5} \leq k\right) \geq 1 - \frac{1}{k^2}$

we want to know $P(40 \leq X \leq 60) \Rightarrow k = 2$

$\therefore 1 \geq P(|X-50| \leq 10) = P(40 \leq X \leq 60) \geq 1 - \frac{1}{4} = \frac{3}{4}$

Exercise 4:

1. $E(X_i) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$

$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{100} X_i}{100}\right) = \frac{\sum_{i=1}^{100} E(X_i)}{100} = \frac{100\mu}{100} = \mu = \frac{7}{2}$

2. $Var(X_i) = E(X_i^2) - E(X_i)^2$

$E(X_i^2) = \frac{1}{6}\times 1^2 + \frac{1}{6}\times 2^2 + \frac{1}{6}\times 3^2 + \frac{1}{6}\times 4^2 + \frac{1}{6}\times 5^2 + \frac{1}{6}\times 6^2 = \frac{91}{6}$

$E(X_i)^2 = (\frac{7}{2})^2 = \frac{49}{4}$

$Var(X_i) = \frac{91}{6} - \frac{49}{4} = \frac{182-147}{12} = \frac{35}{12}$

$Var(\bar{x}) = \frac{1}{100^2}\sum_{i=1}^{100} Var(X_i) = \frac{35}{12}\times\frac{1}{100} = \frac{7}{240}$

3. approximately normally distributed

4. $SD = \sqrt{Var(\bar{x})} = 0.1707825128$

$P(3.25 \le \bar{x} \le 3.75) = 85.68\%$