Statistics II

Problem sheet 6 - Solutions

1. In each case we express the event involving X in terms of an event involving $Z = \frac{X-100}{10}$.

Note that $Z \sim N(0,1)$ so that tables can be used.

i) $P(X>115) = P(Z > \frac{115-100}{10}) = P(Z>1.5)$

$= 1 - \Phi(1.5)$

$= 1 - 0.9332 = 0.0668$

ii) $P(X<92) = P(Z< \frac{92-100}{10}) = P(Z<-0.8)$

$= \Phi(-0.8) = 1 - \Phi(0.8) = 1 - 0.7881 = 0.2119$

iii) $P(95<X<12) = P(\frac{95-100}{10} < Z < \frac{120-100}{10})$

$= P(-0.5 < Z < 2)$

$= \Phi(2) - \Phi(-0.5)$

$= \Phi(2) - (1 - \Phi(0.5))$

$= \Phi(2) + \Phi(0.5) - 1 = 0.9772 + 0.6915 - 1$

$= 0.6687$

2. Let X be the gestation period for a human birth. Then $X \sim N(266, 16^2)$

i) $P(X > 290) = P\left(Z > \frac{290-266}{16}\right)$

$= P\left(Z > \frac{24}{16}\right) = P(Z > 1.5)$

$= 0.0668$ (from 3 i))

ii) $P(250 < X < 282) = P\left(\frac{250-266}{16} < Z < \frac{282-266}{16}\right)$

$= P(-1 < Z < 1)$

$= \Phi(1) - \Phi(-1) = 2\Phi(1) - 1$

$= 1.6826 - 1 = 0.6826$

This gives the rule-of-thumb that the probability of a normal random variable taking a value within 1 standard deviation of the mean (i.e. between $\mu - \sigma$ and $\mu + \sigma$) is approximately $2/3$. Also $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$.

3 (i) $P(\mu-\sigma \leq X \leq \mu+\sigma) = P(X \leq \mu+\sigma) - P(X \leq \mu-\sigma)$

$$= P\left(\frac{X-\mu}{\sigma} \leq \frac{\mu+\sigma-\mu}{\sigma}\right) - P\left(\frac{X-\mu}{\sigma} \leq \frac{\mu-\sigma-\mu}{\sigma}\right)$$

$= P(Z \leq 1) - P(Z \leq -1)$, where $Z \sim N(0,1)$

$= \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1))$

$= 2\,\Phi(1) - 1$

From statistical tables $\Phi(1) = 0.8413$, and

$P(\mu-\sigma \leq X \leq \mu+\sigma) = 2 \times 0.8413 - 1$

$= \underline{0.6826}$

(ii) Working as above, we have:

$P(\mu-k\sigma \leq X \leq \mu+k\sigma) = P(X \leq \mu+k\sigma) - P(X \leq \mu-k\sigma)$

$= P(Z \leq k) - P(Z \leq -k)$, $Z \sim N(0,1)$.

$= \Phi(k) - \Phi(-k) = \underline{2\,\Phi(k) - 1}$

Then (using the tables):

$P(\mu-2\sigma \leq X \leq \mu+2\sigma) = 2\,\Phi(2) - 1 = 2 \times 0.9772 - 1$

$= \underline{0.9544}$

and

$P(\mu-3\sigma \leq X \leq \mu+3\sigma) = 2\,\Phi(3) - 1 = 2 \times 0.99865 - 1$

$= \underline{0.9973}$.

4\. (i) In the general case when $X \sim N(\mu, \sigma^2)$ we have

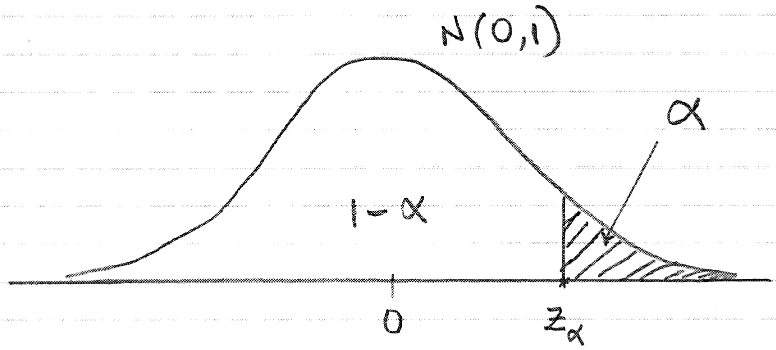
$$P(X \geq x_\alpha) = \alpha \iff P\left(\frac{X-\mu}{\sigma} \geq \frac{x_\alpha - \mu}{\sigma}\right) = \alpha$$

$$\iff P\left(Z \geq \frac{x_\alpha - \mu}{\sigma}\right) = \alpha \quad , \text{ where } Z \sim N(0,1)$$

or $$P(Z \geq z_\alpha) = \alpha \quad \text{with } z_\alpha = \frac{x_\alpha - \mu}{\sigma} \qquad (*)$$

We therefore want to find a value z_α from the $N(0,1)$ distribution, such that $P(Z \geq z_\alpha) = \alpha$

i.e.



We can use statistical tables.

(i) If $\alpha = 0.05$, tables of the $N(0,1)$ distribution (e.g. NCST p35) give us

$$P(Z \geq 1.6449) = 0.05 \quad , \text{ i.e. } z_{0.05} = 1.6449.$$

Then from $(*)$ we have

$$x_{0.05} = \mu + z_{0.05} \times \sigma \Rightarrow x_{0.05} = 60 + 1.6449 \times 10$$

$$\Rightarrow \underline{x_{0.05} = 76.449}$$

(ii) Similarly we find

$$P(Z \geq 1.96) = 0.025 \Rightarrow z_{0.025} = 1.96$$

and

$$x_{0.025} = \mu + z_{0.025} \cdot \sigma \Rightarrow x_{0.025} = 60 + 1.96 \times 10$$

$$\Rightarrow \underline{x_{0.025} = 79.6}$$

5 i) The median of the data is the observation in position $\frac{38+1}{2} = 19\frac{1}{2}$. Therefore we take

$$\text{Median} = \frac{x_{19} + x_{20}}{2} = \frac{29+36}{2} = \underline{32.5}$$

$$Q_1 = \text{Observation in position } \left(\frac{38+1}{4}\right) = x_{(9.75)}$$

$$= x_{(9)} + 0.75\{x_{(10)} - x_{(9)}\} = 18 + 0.75 \times 1 = \underline{18.75}$$

[Notice that we have used linear interpolation between the 9^{th} and 10^{th} observation.]

$$Q_3 = \text{Observation in position } \left(\frac{3(38+1)}{4}\right) = x_{(29.25)} = \underline{44.25}$$

ii) Stem-and-leaf diagram

Stem	Leaf
0	1159
1	0347899
2	12555679
3	6899
4	0113445669
5	00449

OR

(Lengthened plot)

Stem	
0	11
0	59
1	034
1	7899
2	12
2	555679
3	
3	6899
4	011344
4	5669
5	0044
5	9

Perhaps there is some evidence that the distribution is not symmetric. The evidence is not particularly strong.

6. i) <u>5-number summary:</u>

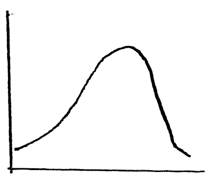
Min: 5, Max: 58, Median $= \frac{1}{2}(39+39) = 39$.

$Q_1 =$ Observation in position $\left(\frac{54+1}{4}\right) = x_{(13.75)} = 27$

$Q_3 =$ —— " —— $\left(\frac{3(54+1)}{4}\right) = x_{(41.25)} = 47$

ii) <u>Stem-and-leaf plot</u>

Stems	Leaves
0	
0	5889
1	4
1	89
2	224
2	67779
3	001133
3	55666999
4	11122444
4	555777799
5	1123
5	5688

No, the plot looks asymmetric i.e. .

The distribution <u>does not</u> seem to be consistent with a Normal distribution.

iii) According to my calculations, for this data set

$$\sum x_i = 1962, \qquad \sum x_i^2 = 80\,860$$

$$\Rightarrow \bar{x} = \frac{1962}{54} = 36.33, \quad s^2 = \frac{1}{53}\left(\sum x_i^2 - \frac{(\sum x_i)^2}{54}\right)$$

$$= 180.64$$

$$\Rightarrow \text{Sample S.D. is } s = \sqrt{180.64} = 13.44$$

Observations within 1 SD of mean lie in range:

$$[36.33 - 13.4,\ 36.33 + 13.4] = [22.93,\ 49.73]$$

There are 37 observations in this range (as a %, $\frac{37}{54} = 68.5\%$)

Observations within 2 SD of mean lie in range:

$$[36.33 - 26.9,\ 36.33 + 26.9] = [9.43,\ 63.23]$$

There are 50 observations in this range i.e. 92.6% of data.

Even though the data don't look normal, the (68 - 95) rule is very approximately correct.