Basic Statistics and Probability (2018-19)
Science & Engineering Program Boston University-Faculty of Science UAM

QUIZ 1. Solutions

---

1. a) The shopping mall. Because it is the place where the users are more diverse: everyone has to buy food or clothes. The other places would yield biased samples: for example, in the pool hall we will typically find fans of cue sports.

b) Observational study.

c) The experimental units will be the persons participating in the poll and the population is the people living in that town.

d) The variables measured are:

- the binary variable: smoker/non-smoker (qualitative);
- the age (quantitative);
- how harmful the interviewed person feels second-hand smoke is to adults (qualitative with 4 possible values).

---

2. Let's denote the possible iron deficiency status as D if it is deficient and N if it is not deficient.

a) When choosing two adolescent girls and checking their iron deficiency status, the sample points of the experiment are

$$\text{DD} \qquad \text{DN} \qquad \text{ND} \qquad \text{NN}.$$

The first letter in the pair corresponds to the status of the first girl in the pair.

b) The probability that one of the girls has iron deficiency and the other one does not is

$$P(\text{DN}, \text{ND}) = P(\text{DN}) + P(\text{ND}) = 0.12 \cdot 0.88 + 0.88 \cdot 0.12 = 0.2112$$

c)

$$P(\text{DD}) = P(D)\,P(D) = 0.12\,0.12 = 0.0144$$
$$P(\text{DN}) = P(D)\,P(N) = 0.12\,0.88 = 0.1056$$
$$P(\text{ND}) = P(N)\,P(D) = 0.88\,0.12 = 0.1056$$
$$P(\text{NN}) = P(N)\,P(N) = 0.88\,0.88 = 0.7744$$

---

3. a) The mean age of the trapped miners is $\bar{x} = 1316/33 = 39.88$ years.

b) The interquartile range is IQR $= Q_U - Q_L = 49.00 - 30.00 = 19$, where $Q_U = 49.00$ and $Q_L = 30.00$ are the upper and lower quartiles respectively.
The range is the difference between the maximum and minimum of the sample: range = 63-19 = 44.
The percentage of observations between the two quartiles is 50%.

c)

$$\begin{aligned}\text{Lower inner fence} &= Q_L - 1.5\ \text{IQR} = 30 - 1.5\,19 = 1.51\\ \text{Upper inner fence} &= Q_U + 1.5\ \text{IQR} = 49 + 1.5\ 19 = 77.5\end{aligned}$$

Limits of the whiskers:

$$\begin{aligned}\text{Lower limit} &= \text{Smallest observation inside the inner fences} = 19\\ \text{Upper limit} &= \text{Largest observation inside the inner fences} = 63\end{aligned}$$

There are no outliers, because no observation lies outside the inner fences.

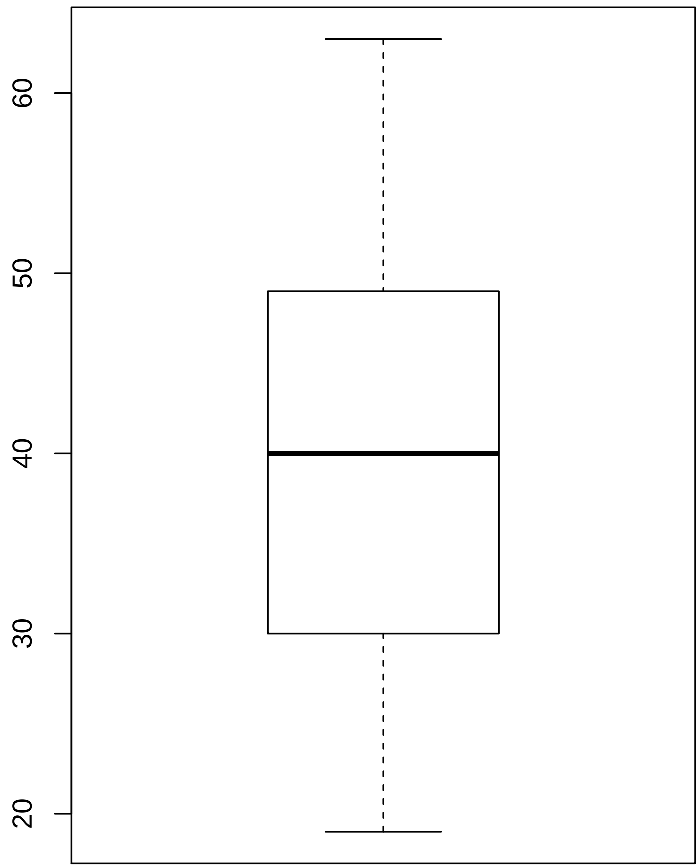

Figure 1

d)

Interval	Absolute frequency
[11,28]	6
(28,38]	9
(38,47]	7
(47,55]	8
(55,64]	3

e) The proportion of miners with an age between 11 and 28 is 6/33= 0.18. The proportion of miners with an age greater than 38 is 18/33=0.55. The absolute and relative frequencies of the interval [11,28] are 6 and 0.18 respectively.

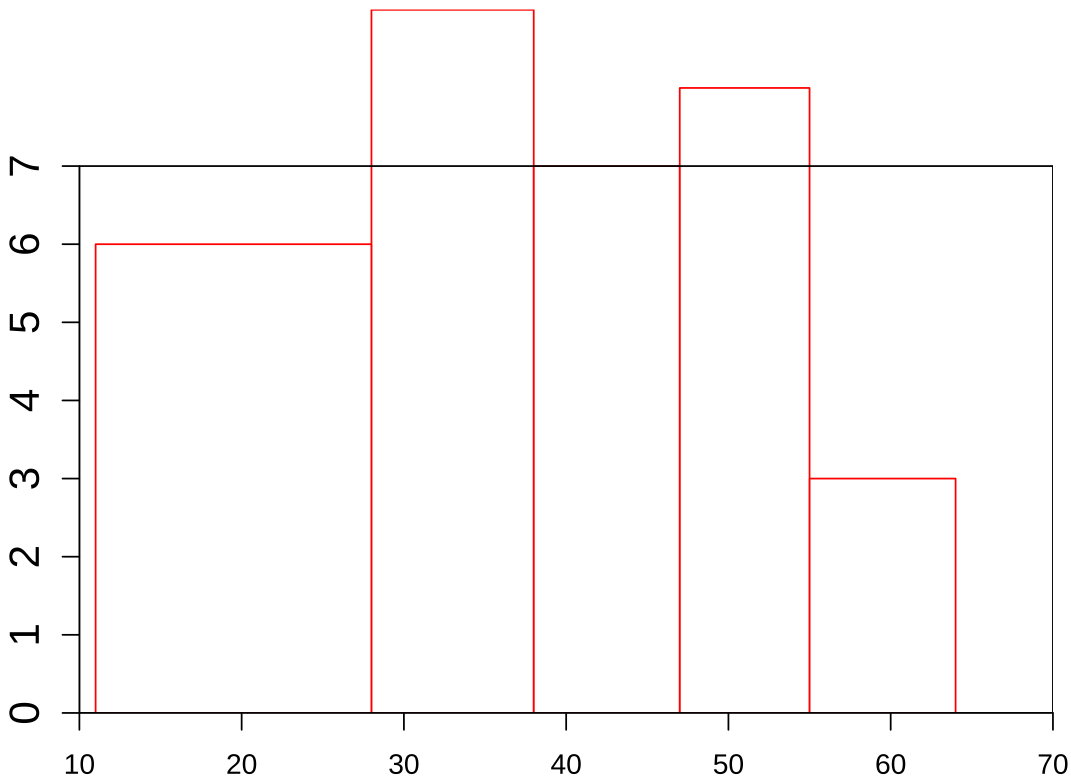

Figure 2: Yes, I drew the frame for the histogram too low, but I promise I did not do it on purpose! It was a slip: sorry!

f) `var(X)` is the sample variance, denoted by s^2. `sd(X)` is the standard deviation, denoted by s. $\texttt{sd(X)} = \sqrt{\texttt{var(X)}}$

g) The z-score corresponding to $x = 20$ is

$$z = \frac{20 - 39.88}{11.62368} = -1.71$$

h) The percentage of measurements in the data set that are above and below the 85th percentile are 85% and 15% respectively.

The 85th percentile of the sample of ages is not lower than 45, because the third quartile (which leaves 75% of the sample to the left) is 49, so the 85th percentile has to be greater than 49.