



## Question 2: Income Distribution

Worksheet 2 contains the income / € of 100 professionals.

Column B represents bins that we are going to put the data into (same as below). Column E is how Excel's Data Analysis package receives these bins as input.

Use the data to answer the following questions.

Use Excel to help you establish a frequency distribution with the given bins:

Income / €	Frequency
10 - 20	0
20 - 30	1
30 - 40	12
40-50	33
50 - 60	31
60 - 70	17
70 - 80	5
80 - 90	1

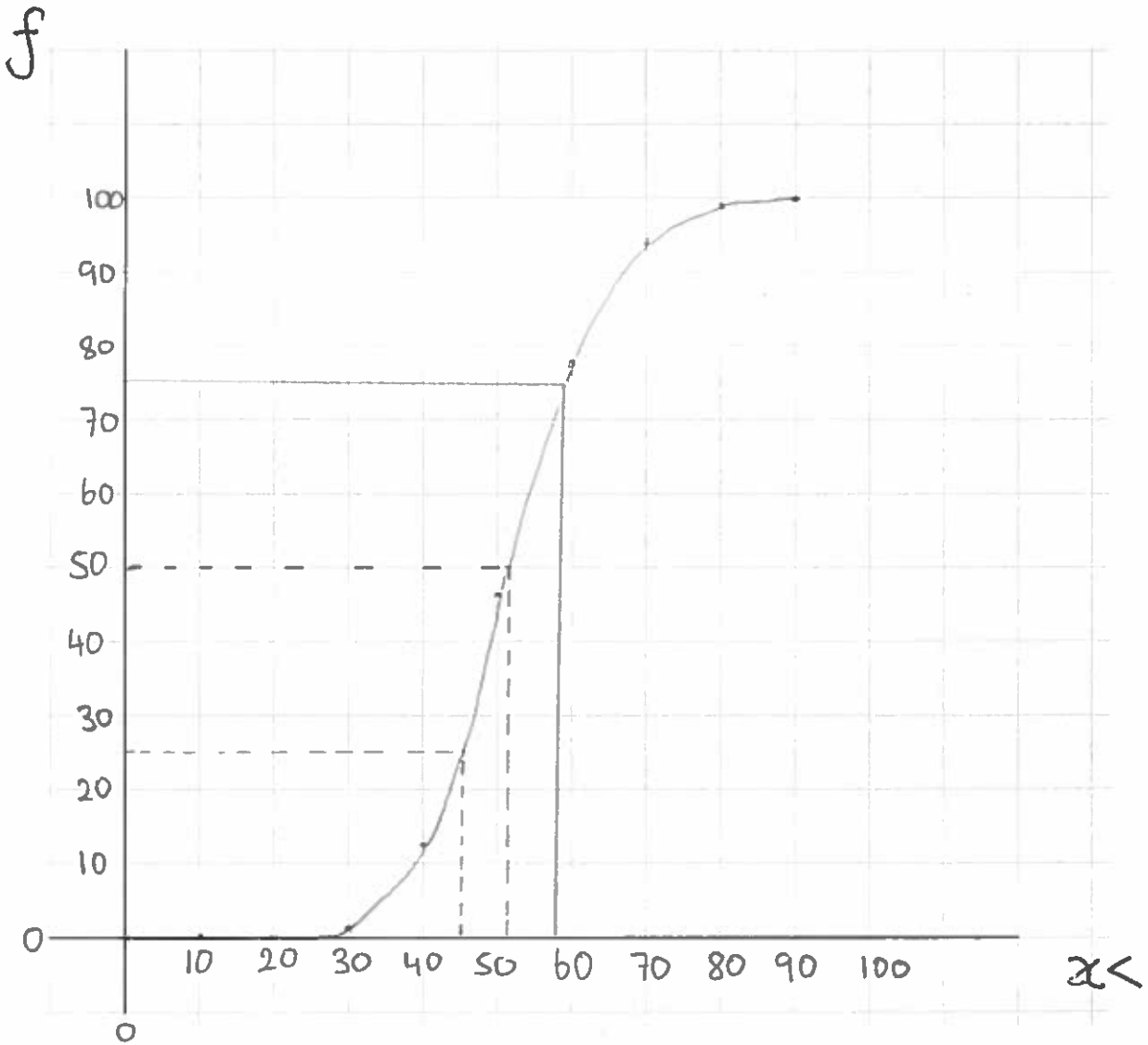
Note: Data  
→ Data Analysis  
→ Histogram

Hence establish a cumulative frequency distribution:

$x <$	f
20	0
30	1
40	13
50	46
60	77
70	94
80	99
96	100

Done "by hand"

Hence, produce a rough sketch of the ogive on the graph below. Fill in all the missing elements including axes labels and ~~some~~ tick marks.



Hence give a rough estimate of the median.

$$\text{Median} \approx \text{€}52.$$

Also use the ogive to give a rough estimate of the interquartile range.

$$Q_1 \approx \text{€}45 \quad Q_3 \approx \text{€}58$$

$$\Rightarrow Q_3 - Q_1 \approx \text{IQR} \approx \text{€}13.$$

### Question 3: Age of Software

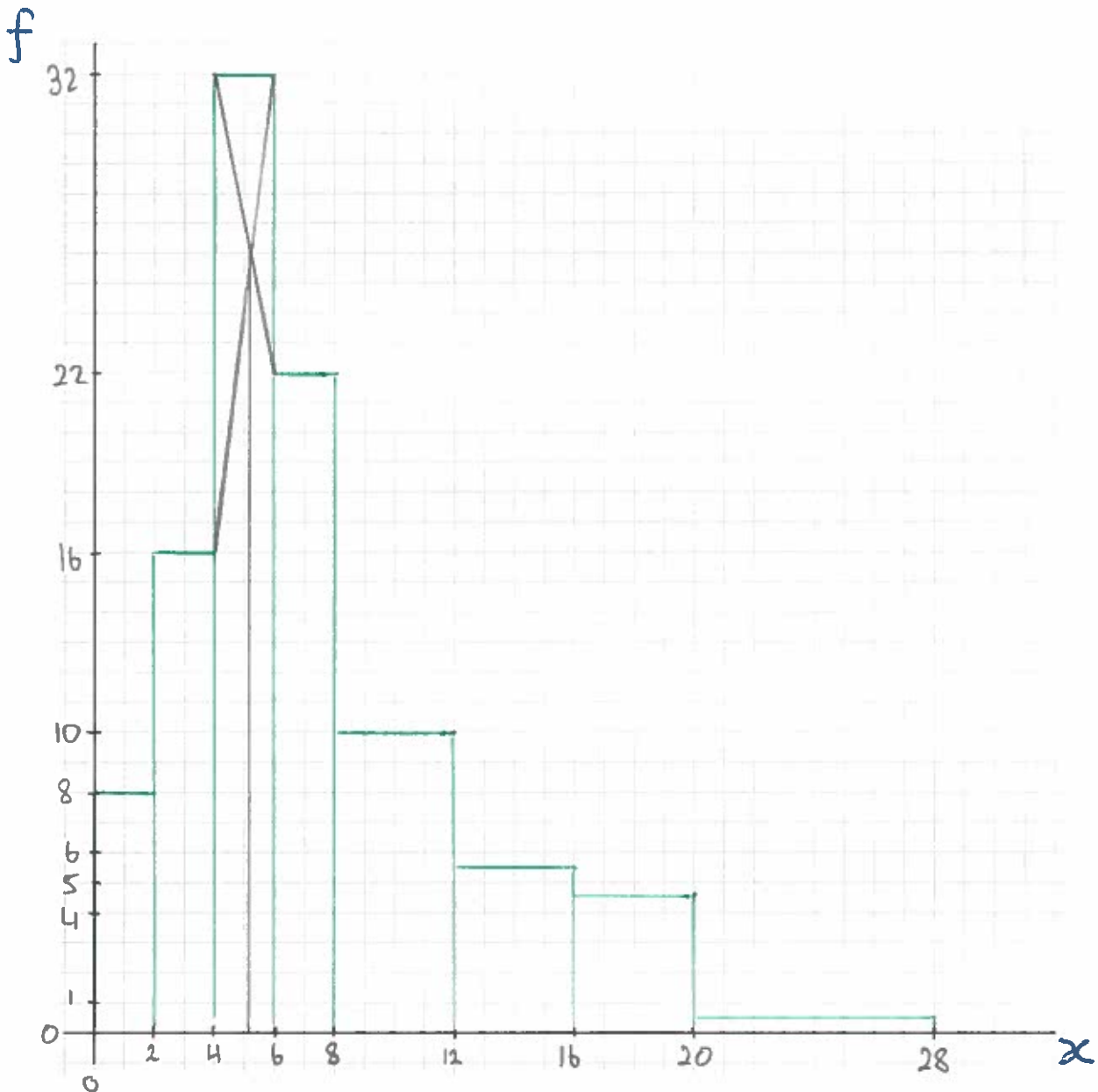
Below,  $x$ , represents the age / weeks of virus definitions on a sample of machines used by a large firm and  $f$  represents of how many machines are in the given bin/range.

$x$	0-2	2-4	4-6	6-8	8-12	12-16	16-20	20-28
$f$	8	16	32	22	20	11	9	2

Produce a histogram on the graph below. Label the axes and include **some** tick marks.

The following table may prove useful

$x$	0-2	2-4	4-6	6-8	8-12	12-16	16-20	20-28
$f$	8	16	32	22	20	11	9	2
# s.w.	1	1	1	1	2	2	2	4
$h$	8	16	32	22	10	5.5	4.5	0.5



Comment on the Skewness.

Skewed to the right

Estimate the mode using the histogram.

Mode  $\approx 5.2$  weeks

Confirm the accuracy of your estimate by using this formula from the tables.

$$\begin{aligned}\text{Mode} &= L_M + C_M \cdot \frac{d_M - d_{M-1}}{2d_M - (d_{M-1} + d_{M+1})} \\ &= 4 + 2 \cdot \frac{32 - 16}{2(32) - (16 + 22)} = 5.231 \text{ weeks}\end{aligned}$$

With the help of Excel, calculate the mean and (sample) standard deviation.

[HINT: Worksheet 3 contains a useful template. The below formulae are useful:

$$\bar{x} = \frac{\sum fx}{\sum f} \quad s = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f - 1}}$$

$$\bar{x} = \frac{934}{120} \approx 7.783 \text{ weeks}$$

$$s = \sqrt{\frac{2984}{119}} \approx 5.008 \text{ weeks}$$

**Question 4: File Back-Ups**

200 workers at a call centre are supposed to back up their call log every day to a central database. The IT Manager notices that on average only 1% of workers fail to do this.

On Monday morning, the IT Manager checks the database. Calculate the following probabilities using Excel's BINOM.DIST command:

k=	0	1	2	3	At most 2	More than 3
Probability of k workers failing to back-up	0.1340	0.2707	0.2720	0.1814	0.6767	0.1420

||  
 $P[0 \text{ or } 1 \text{ or } 2]$       ||  
 $P[\text{not } - (0,1,2,3)]$

**Question 5: Spam Email**

Despite a recent review of email security, a firm's central email system is still experiencing a high volume of spam email. During an 8 hour work day, on average 4 spam emails are being distributed to all staff.

What is the probability of exactly 4 emails being sent to all staff during an 8 hour day? Use Excel's POISSON.DIST command to help.

0.1954

In the period of four hours before lunch, calculate the following probabilities. Use Excel's POISSON.DIST command to help.

$X \sim \text{Po}_i[2]$

k=	0	1	2	3	More than 3
Probability of k spam emails before lunch	0.1353	0.2707	0.2707	0.1804	0.1429

||  
 $P[\text{not } - (0,1,2,3)]$