

Distributions

LLE – Mathematics and Statistics Skills

1 Frequency Table

A bar sells scratch cards, for £3, where the customer can win between £0 and £10, in whole number of £. The winnings record over the past 250 cards sold is given as a frequency table.

1. Complete the relative and cumulative relative frequency columns, to 3 decimal places.

| Winnings, £ | Frequency | Relative Frequency | Cumulative RF |
|-------------|-----------|--------------------|---------------|
| 0 | 48 | 0.192 | 0.192 |
| 1 | 60 | 0.24 | 0.432 |
| 2 | 42 | 0.168 | 0.6 |
| 3 | 31 | 0.124 | 0.724 |
| 4 | 33 | 0.132 | 0.856 |
| 5 | 12 | 0.048 | 0.904 |
| 6 | 8 | 0.032 | 0.936 |
| 7 | 9 | 0.036 | 0.972 |
| 8 | 5 | 0.02 | 0.992 |
| 9 | 1 | 0.004 | 0.996 |
| 10 | 1 | 0.004 | 1 |

Solution

The relative frequencies are the frequency divided by 250 (the total frequency). The cumulative relative frequencies are the running total of the relative frequencies.

2. Identify the mode, by finding the payout with the highest frequency.

Answer: _____ **£1**

Solution

The highest frequency is 60.

3. Use the cumulative relative frequencies to find the median payout (50th percentile).

Answer: _____ **£2**

Solution

There are 43.2% of values up to £1, and then 60% up to £2. The 50% point must have occurred during the £2 group.

4. Find the lower quartile (25th percentile).

Answer: _____ **£1**

Solution

Find the value where the cumulative relative frequency first goes over 25%.

5. Find the upper quartile (75th percentile).

Answer: _____ **£4**

Solution

Find the value where the cumulative relative frequency first goes over 75%.

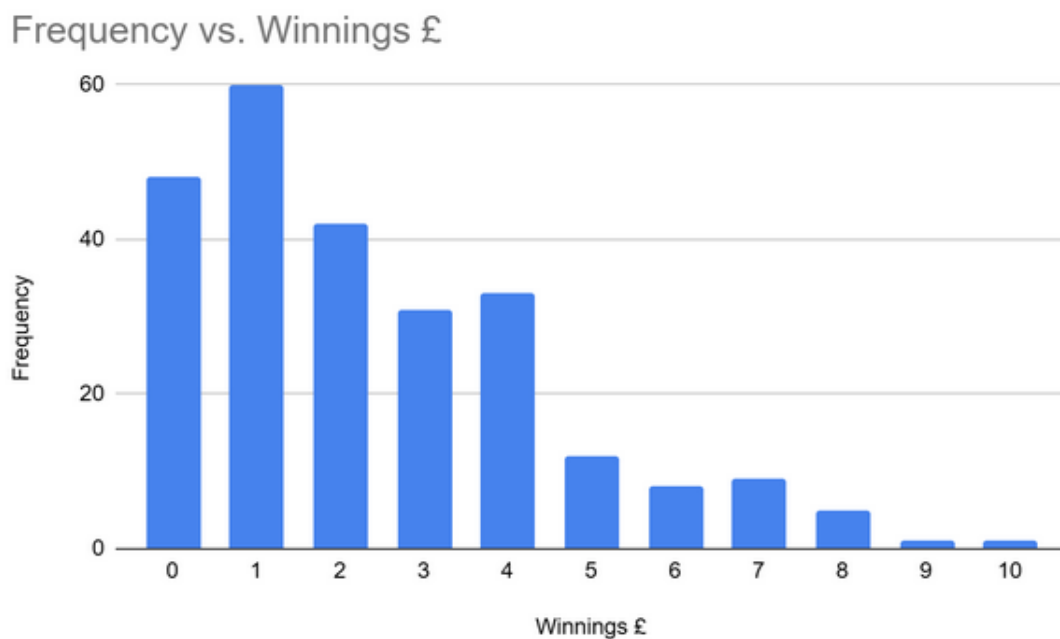
6. Find the 90th percentile.

Answer: _____ £5

Solution

Find the value where the cumulative relative frequency first goes over 90%.

A bar chart of the frequencies of winnings is shown:



7. Describe the skew of the distribution (left-skew, right-skew, or symmetric) based on the shape of the bar chart.

☐ Left-skew ☒ **Right-skew** ☐ Symmetric

Solution

The tail is to the right.

8. Calculate the total revenue on scratchcards (£3 each), total amount paid out in winnings (from table), in order to work out **the profit made**.

Answer: _____ £151

Solution

$$TR = 250 \times 3 = 750$$

$$TC = 48 \times 0 + 60 \times 1 + 42 \times 2 + \dots$$

$$\pi = TR - TC$$

2 Chebyshev's Rule

Chebyshev's Theorem provides an estimate for the minimum proportion of values that lie within a given number of standard deviations from the mean.

For any given distribution, the minimum proportion of observations that lie within k standard deviation of the mean ($k > 1$), is given by:

$$1 - \frac{1}{k^2}$$

1. What is the minimum proportion (giving answer as a percentage) of observations that lie within $\sqrt{2}$ standard deviations of the mean?

Answer: _____ **50%**

Solution

$$1 - \frac{1}{\sqrt{2}^2} = 1 - \frac{1}{2}$$

2. What is the maximum proportion (giving answer as a percentage) of observations that lie beyond 2 standard deviations of the mean?

Answer: _____ **25%**

Solution

$$1 - \left(1 - \frac{1}{2^2}\right) = \frac{1}{4}$$

3. A distribution has a mean value of 50 and a standard deviation of 5. What is the minimum proportion of observations (giving answer as a

percentage) that are between 40 and 60?

Answer: _____ **75%**

Solution

40 is 2 standard deviations below the mean, 60 is 2 standard deviations above the mean. Chebyshev using $k = 2$.

4. A distribution has a mean value of 60 and a standard deviation of 20. What is the minimum proportion of observations that are between 30 and 90?

☐ 40 % ☐ 44 % ☐ 50 % ☒ **56 %** ☐ Not enough information

Solution

30 to 90 is 1.5 standard deviations, Chebyshev's rule gives 5/9 minimum within.

3 Empirical Rule for Normal Distributions

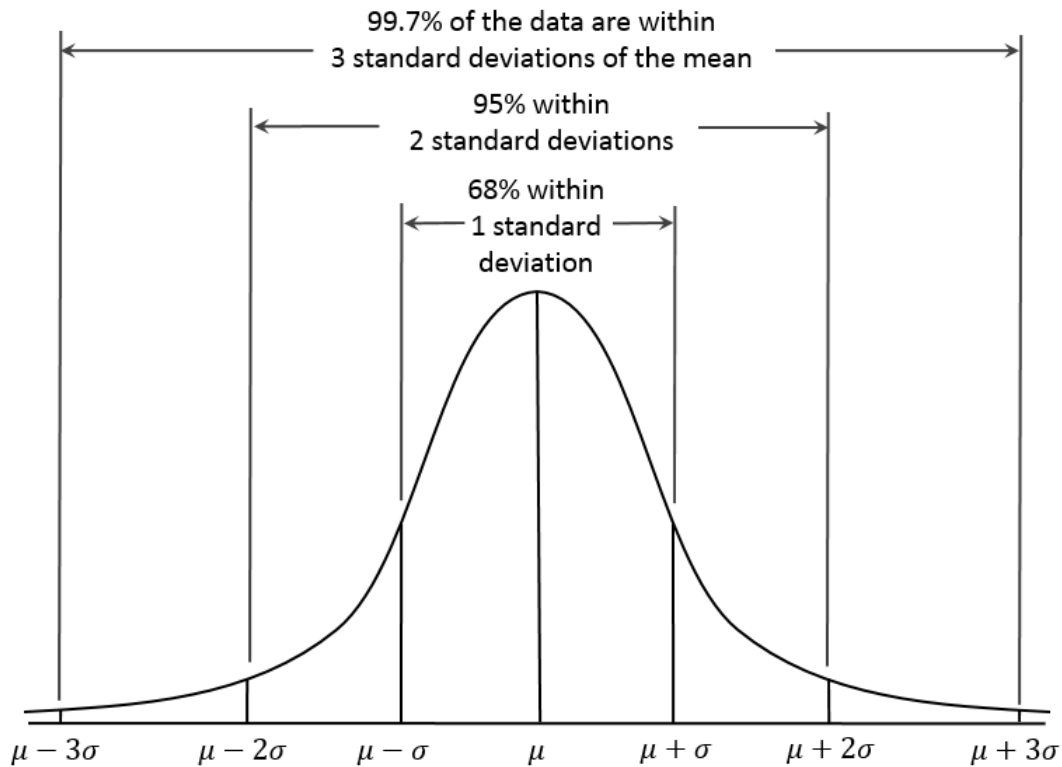
Distributions that are described as normally distributed have a particular shape and have defined characteristics.

The curve is perfectly symmetrical around its centre, meaning the left side is a mirror image of the right.

About 68% of data falls within 1 standard deviation ($\pm\sigma$) of the mean.

About 95% falls within 2 standard deviations ($\pm 2\sigma$).

About 99.7% falls within 3 standard deviations ($\pm 3\sigma$).



1. Approximately what proportion of data lie above 2 standard deviations above the mean, given the data are normally distributed?

☒ **2.5 %** ☐ 5 % ☐ 47.5 % ☐ 95 % ☐ 97.5 %

Solution

95% within 2 standard deviations, 5% outside - shared between above and below 2 standard deviations.

2. Approximately what proportion of data lie above 1 standard deviation below the mean, given the data are normally distributed?

☐ 17 % ☐ 34 % ☐ 68 % ☒ **84 %** ☐ Not possible to tell

Solution

By symmetry 16% lie below 1 standard deviation below the mean, and the whole distribution sums to 100%

3. An office coffee machine distributes liquids, in a normal distribution, that have a mean volume of 250 ml and a standard deviation of 6 ml. What percentage of coffees will be between 238 and 262 ml?

Answer: _____ **95%**

Solution

238 is 2 sd below the mean, and 262 is 2 sd above the mean.

4. An office coffee machine distributes liquids, in a normal distribution, that have a mean volume of 250 ml and a standard deviation of 6 ml. What percentage of coffees will be over 268 ml?

Answer: _____ **0.15%**

Solution

268 ml is 3 sd above the mean, need to use symmetry to split the proportion outside 3 sd into two.

5. The bulb on a photocopier has a mean life of 16000 hours and a variance of 250000 square hours, and the life is normally distributed. What percentage of bulbs last longer than 15000 hours?

Answer: _____ **97.5%**

Solution

Standard deviation is the square root of variance, 15000 hours is then 2 standard deviations below the mean. Use symmetry to work out the probability above this.