**HOMEWORK 1**

Use the weights of freshmen males in September to construct a frequency distribution. Begin with a lower class limit of 50 kg and use a class width of 10 kg.

|  |  |  |  |
| --- | --- | --- | --- |
| 52 | 69 | 72 | 78 |
| 56 | 69 | 72 | 78 |
| 62 | 69 | 73 | 81 |
| 63 | 69 | 75 | 88 |
| 64 | 69 | 75 | 92 |
| 66 | 70 | 75 | 94 |
| 66 | 70 | 76 | 96 |
| 67 | 72 | 78 | 97 |

1. Build the frequency table with relative and cumulative frequencies

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Weight class | Frequency | Relative frequency | Cumulative frequency | Density Frequency per kg |
| 50- | 2 | 0.063 | 2 | 0.063/10=0.0063 |
| 60- | 11 | 0.344 | 13 | 0.0344 |
| 70- | 13 | 0.406 | 26 | 0.0406 |
| 80- | 2 | 0.063 | 28 | 0.0063 |
| 90- | 4 | 0.125 | 32 | 0.0125 |
| Total | 32 | 1 | 32 |  |

(quiz 1, 2,3)

1. Draw a histogram to show the distribution of weights



quiz 4: 3 histograms are correct, on the Y axis you can have: frequency OR relative frequency OR density (frequency per kg), the last one is better as it is the only one that accounts for intervals with different length!

1. Compute cumulative relative frequencies

|  |  |  |
| --- | --- | --- |
| Weight class | Relative frequency | Relative cumulative frequency |
| 50- | 0.063 | 0.063 |
| 60- | 0.344 | 0.407 |
| 70- | 0.406 | 0.813 |
| 80- | 0.063 | 0.876 |
| 90- | 0.125 | 1 |
| Total | 1 | 1 |

(quiz 5)

Quiz 5: Please note that the relative cumulative frequency at 90 kg is 0.876 (or 87.6%) as this is the frequency cumulated up to the end of the interval (80 to 90 excluded). This means that 87.6% of the sample weights less than 90 kg. You might also find it by individual data (28 out of 32 freshmen males weights less than 90 kg).

1. Draw the cumulative distribution function



Black line: cumulative relative frequency using individual data

Red line: cumulative relative frequency using aggregated data

quiz 6: yellow arrow-> 80 kg corresponds to a cumulative relative frequency of 81.3% (the exact value I got from individual data 100\*26/32). This can also be identified using aggregated data.

quiz 7: blu arrow-> the 90th percentile has, by definition, a cumulative relative frequency of 0.9. In the example it corresponds to 92/94 kg approximately (a visual approx. is ok). Having individual data I can also derive the exact value of the 90th percentile as 0.9\*(32+1)=29.7

29th obs=92

30th obs=94

92+(94-92)\*(29.7-29)=93.4

1. Calculate the sample mean

$$\overbar{x}=\frac{1}{n} \sum\_{i=1}^{n}x\_{i}= \frac{52+56+…+96+97}{32}=\frac{2353}{32}=73.5$$

(quiz 8)

1. Calculate the standard deviation

$n=32$

$\sum\_{}^{}x\_{i}=2353$

$\sum\_{}^{}x\_{i}^{2}=52^{2}+56^{2}+…+96^{2}+97^{2}=2704+3136+…+9216+9409=176569$

$s=\sqrt{\frac{1}{n-1}\sum\_{}^{}(x\_{i}-\overbar{x})^{2}}=\sqrt{\frac{1}{n-1}\left(\sum\_{}^{}x\_{i}^{2}-\frac{(\sum\_{}^{}x\_{i})^{2}}{n}\right)}=\sqrt{\frac{1}{32-1}\left(176569-\frac{2353^{2}}{32}\right)}=\sqrt{\frac{3550}{31}}=10.7$

(quiz 9 e 10)

1. Calculate the mode

$$mode=69 $$

(quiz 11)

1. Calculate quartiles and indicate median

$$first quartile: i=q\*\left(n+1\right)=0.25\*\left(32+1\right)=8.25\rightarrow 8th=67; 9th=69 \rightarrow q1= 67+\left(69-67\right)\*\left(8.25-8\right)=67.5$$

$$ Alternative simpler method 2:\frac{67+69}{2}=68$$

$$median:i=0.5\*33=16.5\rightarrow 16th=72;17th=72\rightarrow q2=72$$

$$third quartile:i=0.75\*33=24.75\rightarrow 24th=78;25th=78\rightarrow q3=78$$

(quiz 12,13,14)

1. Identify minimum and maximum value and calculate range

$$min=52$$

$$max=97$$

$$range=97-52=45$$

(quiz 15)

1. Draw a box and whisker plot (also called box-plot)



(quiz 16)

*Outliers below 53 and 93 kg :*

*First quartile=68*

*Third quartile=78*

*Interquartile range = 78-68=10*

*68-10\*1.5=53*

*78+10\*1.5=93*

1. Answer to question in the course web page