Numerical summaries of quantitative data

Numerical summaries of quantitative data

The 95th **percentile** of scores

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"
- "95% of the students had lower scores and 5% had higher"

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"
- "95% of the students had lower scores and 5% had higher"

Three famous percentiles:

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"
- "95% of the students had lower scores and 5% had higher"

Three famous percentiles:

- 25th percentile = **lower quartile**
- 50th percentile = **median**
- 75th percentile = **upper quartile**

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"
- "95% of the students had lower scores and 5% had higher"

Three famous percentiles:

- 25th percentile = **lower quartile**
- 50th percentile = **median**
- 75th percentile = **upper quartile**

In terms of the histogram of scores:

Numerical summaries of quantitative data

The 95th **percentile** of scores is a score, so that:

- "95% of the students had lower scores"
- "95% of the students had lower scores and 5% had higher"

Three famous percentiles:

- 25th percentile = **lower quartile**
- 50th percentile = **median**
- 75th percentile = **upper quartile**

In terms of the histogram of scores:

the total area to the left of this score was 95%.

Cumulative sums

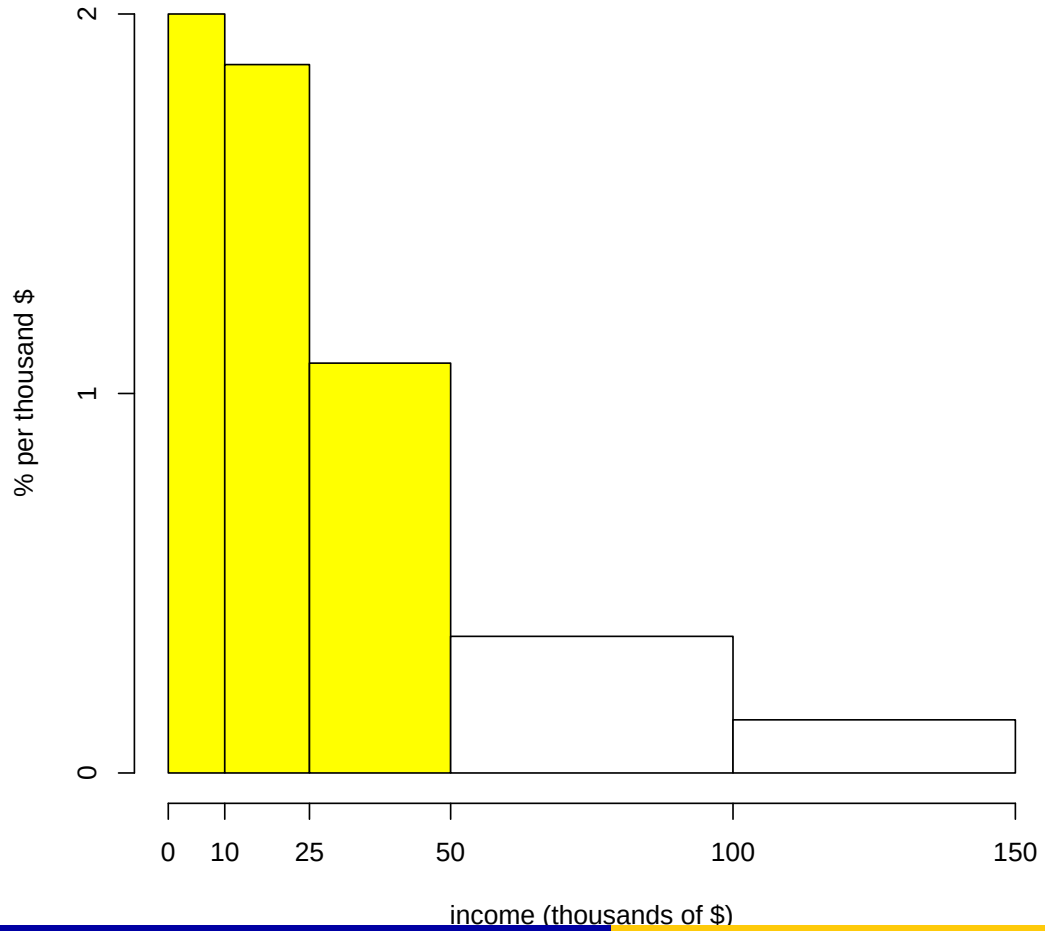
the income distribution, again

income (thousands of dollars)	**percent**	**total percent starting from 0**
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

Percentile from a histogram

Area to the left of \$50,000 $=$ $20\% + 28\% + 27\%$ $=$ 75%

75th percentile of incomes = \$50,000



Estimating percentiles

How to find percentiles that are not at the edges of the bars?

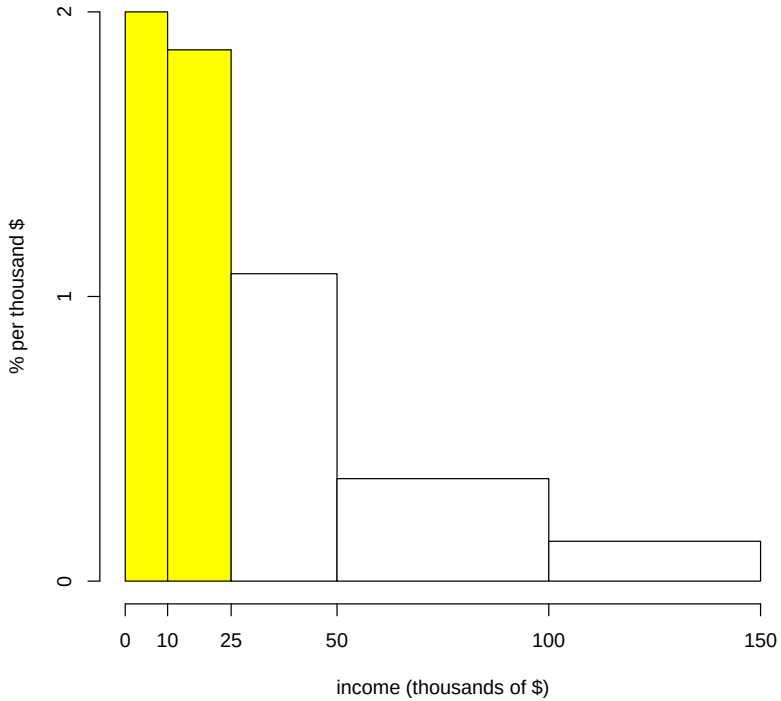
Estimating percentiles

How to find percentiles that are not at the edges of the bars?

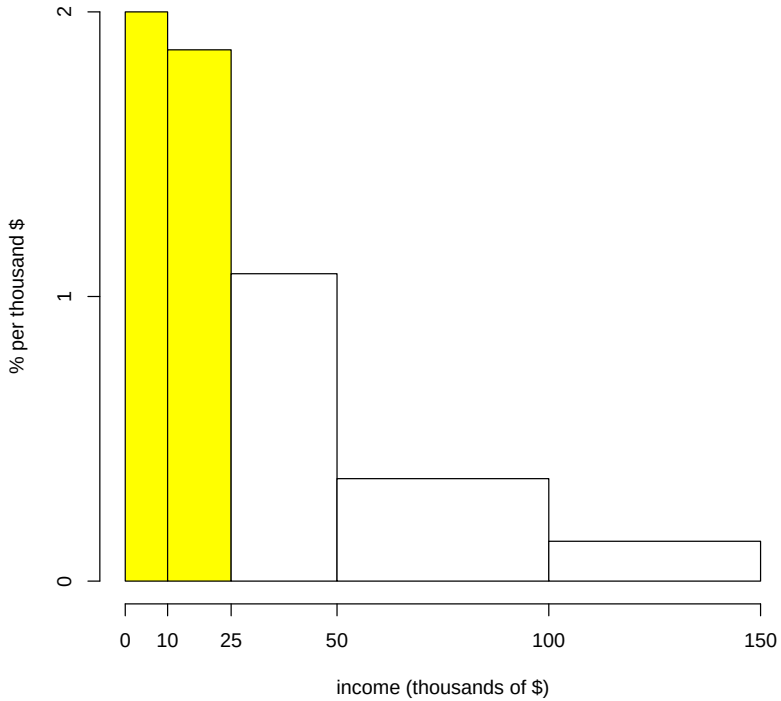
Assuming that the incomes are uniformly distributed within each bar, we can estimate any percentile.

Median of income distribution

Median of income distribution

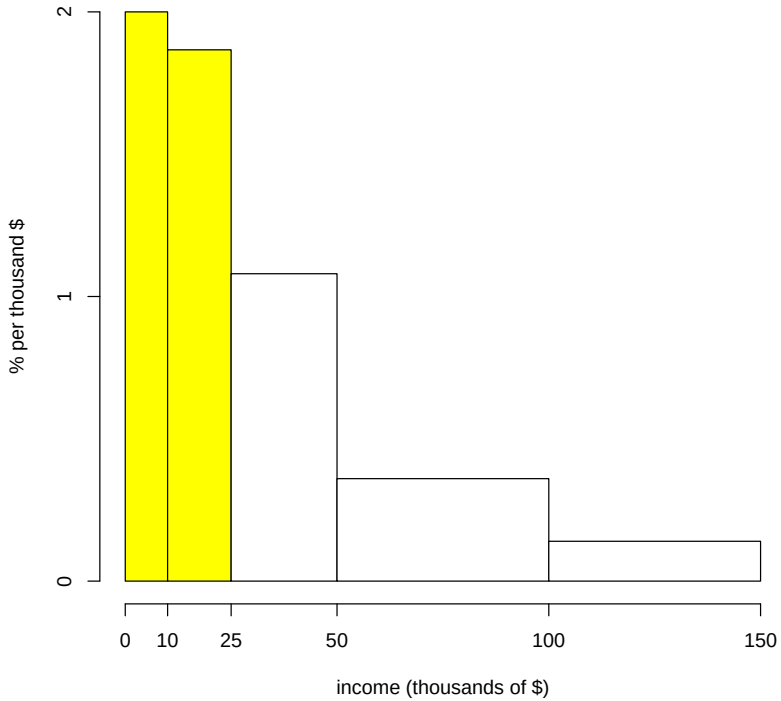


Median of income distribution



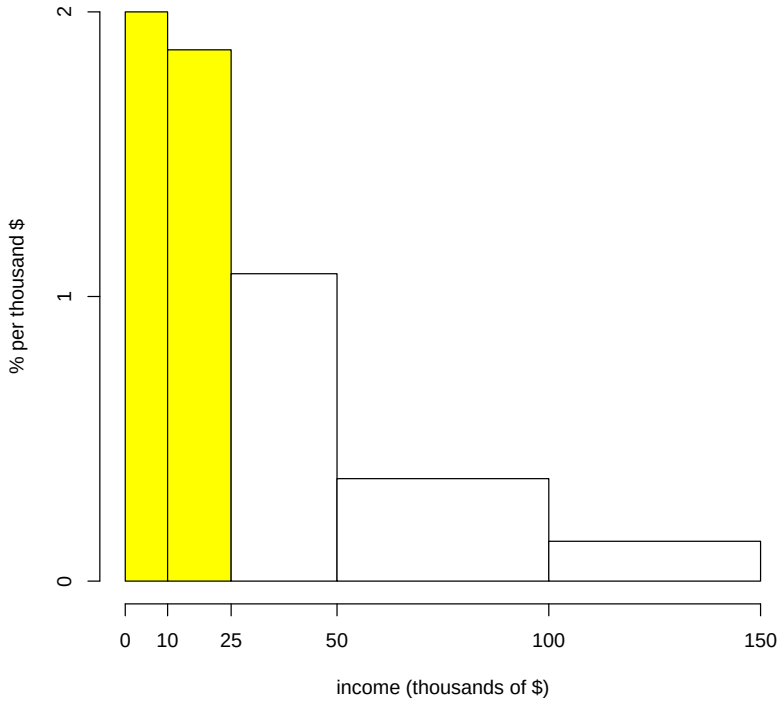
Area between 0 and 25 is 20% + 28% = 48%; need another 2%.

Median of income distribution



Area between 0 and 25 is 20% + 28% = 48%; need another 2%.
The area of the 25-50 bar is 27%.

Median of income distribution



Area between 0 and 25 is 20% + 28% = 48%; need another 2%.

The area of the 25-50 bar is 27%.

So the **median** income must be somewhere in the 25-50 interval.

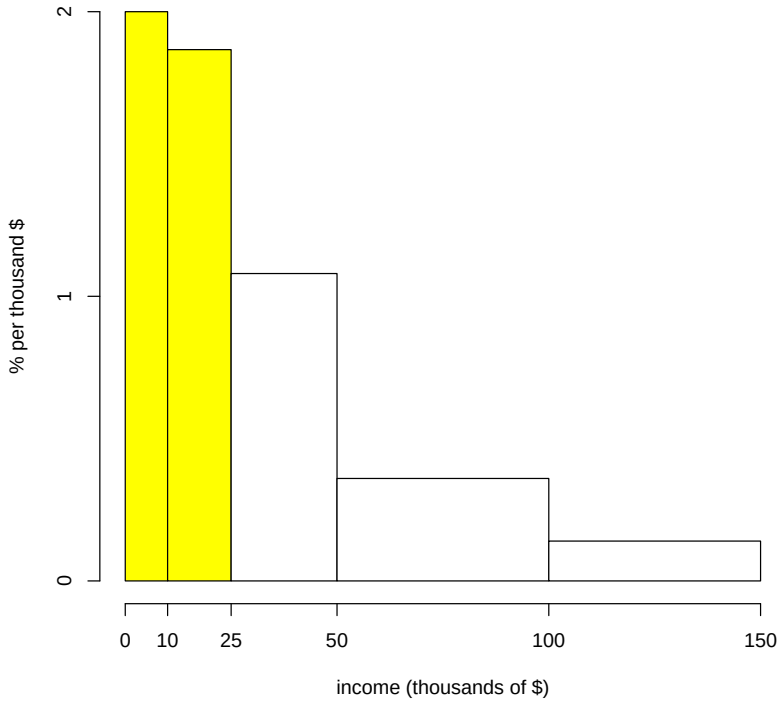
Estimating the median

Yellow bars have a total area of 48%.

Area of 25-50 bar = 27%, and we want 2% of it to get to the median.

estimated median $= 25 + \frac{2}{27} \times (50\text{-}25)$

$=$ 26.85 thousand dollars $=$ \$26,850, roughly



Estimating the 80th percentile

income (thousands of dollars)	percent	total percent starting from 0
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

- The **80th percentile** must be in the 50-100 interval.

Estimating the 80th percentile

income (thousands of dollars)	percent	total percent starting from 0
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

- The **80th percentile** must be in the 50-100 interval.
- The bars from 0 to 50 contain 75% of the area; we need another 5%.

Estimating the 80th percentile

income (thousands of dollars)	percent	total percent starting from 0
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

- The **80th percentile** must be in the 50-100 interval.
- The bars from 0 to 50 contain 75% of the area; we need another 5%.
- The area of the 50-100 bar is 18%.

Estimating the 80th percentile

income (thousands of dollars)	percent	total percent starting from 0
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

- The **80th percentile** must be in the 50-100 interval.
- The bars from 0 to 50 contain 75% of the area; we need another 5%.
- The area of the 50-100 bar is 18%.

estimated 80th percentile $= 50 + \frac{5}{18} \times$ (100-50)
$=$ 63.89 thousand dollars $=$ \$63,890, roughly

Estimating the 80th percentile

income (thousands of dollars)	percent	total percent starting from 0
$0 - 10$	20	20
$10 - 25$	28	48
$25 - 50$	27	75
$50 - 100$	18	93
$100 - 150$	7	100

- The **80th percentile** must be in the 50-100 interval.
- The bars from 0 to 50 contain 75% of the area; we need another 5%.
- The area of the 50-100 bar is 18%.

estimated 80th percentile $= 50 + \frac{5}{18} \times$ (100-50)
$=$ 63.89 thousand dollars $=$ \$63,890, roughly

assuming a uniform distribution within bars