Percentiles and Quantiles

Reading

Section 1.6.5

Practice Problems

1.9.6 (Page 65) 1.46, 1.48, 1.49a, 1.50, 1.51, 1.53, 1.54

Notes

- **Percentiles** measure how many of the values are below a given value.
- So the 90th percentile is the x value such that 90% of the values are below it.
- Lower percentiles correspond to lower values.
- Important percentiles:
  - 25th percentile = first quartile = **Q1**
  - 50th percentile = median = **M**
  - 75th percentile = third quartile = **Q3**
- These three percentiles, together with the min and max, split the data into 4 quarters:
  - from min to Q1
  - from Q1 to M
  - from M to Q3
  - from Q3 to max
- Each of these quarters has 25% of the data (the same number of values!)
- These numbers are called the **five number summary**.
- **Interquartile Range** is the distance between Q1 and Q3:

$$IQR = Q3 - Q1$$

- Boxplots are a graphical representation of the 5 number summary.
  - A "box" from Q1 to Q3
  - A thick line marking the median

- "whiskers" extend to min/max
- These boxplots can be misleading in the presence of outliers.

- **Outlier test**. Values that are more than $1.5 \times IQR$ from the quartiles are suspected outliers.
- **Modified boxplot**:
  - Whiskers extend to last values before the suspected outliers
  - suspected outliers marked with a symbol

Practice

The following table contains the populations for each state, in millions. Determine the 5 number summary and draw the boxplot.

State	Pop(mil)	State	Pop(mil)	State	Pop(mil)
WY	0.56	MS	2.97	NJ	8.79
DC	0.60	IA	3.05	NC	9.54
VT	0.63	CT	3.57	GA	9.69
ND	0.67	OK	3.75	MI	9.88
AK	0.71	OR	3.83	OH	11.54
SD	0.81	KY	4.34	PA	12.70
DE	0.90	LA	4.53	IL	12.83
MT	0.99	SC	4.63	FL	18.80
RI	1.05	AL	4.78	NY	19.38
NH	1.32	CO	5.03	TX	25.15
ME	1.33	MN	5.30	CA	37.25
HI	1.36	WI	5.69		
ID	1.57	MD	5.77		
NE	1.83	MO	5.99		
WV	1.85	TN	6.35		
NM	2.06	AZ	6.39		
NV	2.70	IN	6.48		
UT	2.76	MA	6.55		
KS	2.85	WA	6.72		
AR	2.92	VA	8.00		