Descriptive Statistics Review

Measurement Scales and their Properties

Scales	Properties	Examples
Categorical/Nominal	Identity	gender, political affiliation
Ordinal	+Magnitude	rank ordering, e.g. placement in a race
Interval	+Equal Unit Size	Fahrenheit or Celsius
Ratio	+Absolute Zero	time, weight, height, Kelvin

Frequency Distributions

Frequency	f	how many times a score occurs
Proportion	f/n	where n is the sample size
Cumulative Frequency	cf	number of scores $\leq$ a given value
Cumulative Proportion	$cp = \frac{cf}{n}$	

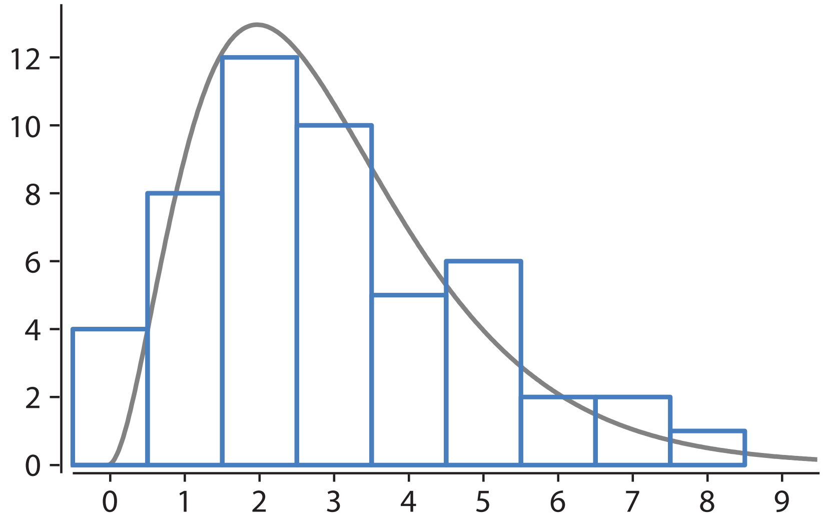

Figure 1: A histogram for a unimodal distribution with positive skew. Bars denote data from 50 observations, curve denotes an underlying smooth distribution.

x	f	cf
0	4	4
1	8	12
2	12	24
3	10	34
4	5	39
5	6	45
6	2	47
7	2	49
8	1	50
9	0	50

Describing the Shape of a Distribution

1. A distribution is *symmetric* if there is an axis about which the tails are the same
2. *Skewness* quantifies asymmetry: positive skew (long right tail) vs negative skew (long left tail)
3. *Modality*: how many peaks are there (e,g, unimodal, bimodal, multimodal)

Descriptive Statistics Summary

Measures of Central Tendency:

1. Mean: Average, sensitive to outliers
2. Median: 50th percentile, insensitive to outliers, but not as useful in statistical inference
3. Mode: Most frequent value, peak(s) of a smooth distribution

Sample Mean	$\bar{X} = \frac{\sum X}{n}$
Sum of Squares	$SS = \sum (X - \bar{X})^2$
Sample Variance	$s^2 = \frac{\sum (X - \bar{X})^2}{n-1} = \frac{SS}{n-1}$
Sample Standard Deviation	$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}} = \sqrt{s^2}$
Population Mean	$\mu = \frac{\sum X}{N}$
Population Variance	$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$

Notes: s^2 is calculated using $n - 1$. Using n yields a biased estimator. N is the population size.

Percentile Rank	$\frac{cf - 0.5f}{n} \cdot 100\%$
Range	$max(X) - min(X)$
Inter-quartile Range	$75^{th} - 25^{th}$percentile (or Q3-Q1)

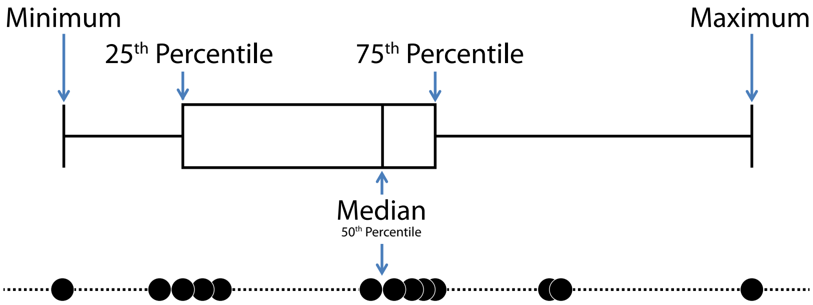

Figure 2: Box Plot. Whiskers often use different standards, such as 1.5xIQR, and attempt to remove outliers.