

Stat 401 B – Lecture 2

Results from Lab 0

- Guessed values are biased towards the high side.
- Judgment sample means are biased toward the high side and are more variable.

1

Random Sampling

- When random sampling from a population with mean, μ , and standard deviation, σ , the distribution of the sample mean
 - is centered at μ
 - with standard deviation $\frac{\sigma}{\sqrt{n}}$

2

Central Limit Theorem

- When random sampling from a population whose shape is not normal, the shape of the distribution of the sample mean will approach the normal as the sample size increases.

3

Stat 401 B – Lecture 2

Statistical Thinking

- Starts with a question –
- Did cars and trucks made in 2004 meet the Corporate Average Fuel Economy (CAFE) standard?

4

Statistical Thinking

- What data would help us answer this question?
- Data on average fuel economy for a random sample of cars and trucks.

5

Fuel Economy Data (MPG)

26.5 24.5 25.0 25.0 24.5 23.0
21.0 21.5 24.0 21.0 23.5 20.5
17.0 22.0 22.0 22.0 22.0 21.0
20.0 25.0 22.5 19.5 19.5 16.0
22.0 29.0 18.0 18.5 17.5 23.5
24.0 23.0 22.5 25.5 18.0 22.5

6

Stat 401 B – Lecture 2

Average Fuel Economy

$$\bar{Y} = 22.0139 \text{ mpg}$$

- Why is this a meaningful summary?
- How do you interpret the sample average (mean) value?

7

What other summaries?

- Stem and leaf
- Five number summary
- Box plot
- Histogram
- Sample standard deviation

8

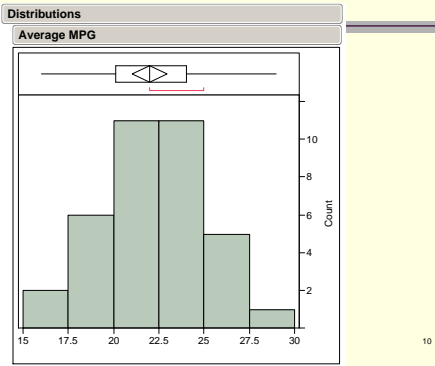
Stem and Leaf

29	0	
28		
27		
26	5	
25	0005	Maximum = 29.0 mpg
24	0055	Upper Quartile = 24.0 mpg
23	0055	
22	00000555	Median = 22.0 mpg
21	0005	Lower Quartile = 20.25 mpg
20	05	
19	55	Minimum = 16.0 mpg
18	005	
17	05	
16	0	

9

Stat 401 B – Lecture 2

Box plot, Histogram



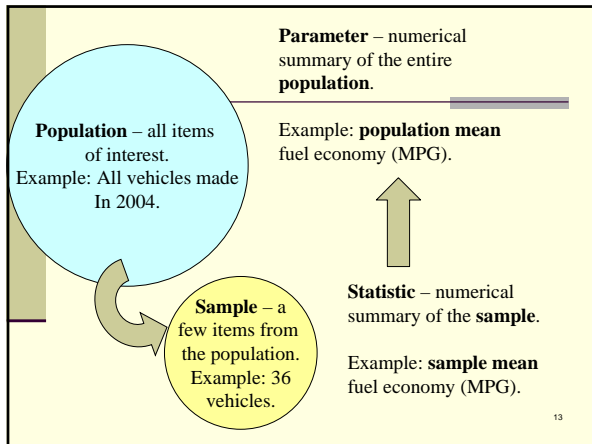
Mean and standard deviation

Distributions	
Average MPG	
Moments	
Mean	22.013889
Std Dev	2.8472696
Std Err Mean	0.4745449
upper 95% Mean	22.977266
lower 95% Mean	21.050511
N	36

Statistical Inference

- So far we have looked as displaying and summarizing sample data.
- What do these data tell us about all cars and trucks in 2004.

Stat 401 B – Lecture 2



One-sample model

$$Y = \mu + \varepsilon$$

- Y represents a value of the variable of interest
- μ represents the population mean
- ε represents the random error associated with an observation

14

Conditions

- The random error term, ε , is
 - Independent
 - Identically distributed
 - Normally distributed with standard deviation, σ

15

Stat 401 B – Lecture 2

Confidence Intervals

- A set of reasonable or plausible values for the population mean.
- We can state how confident we are that intervals based on random samples actually capture the population mean.

16

95% Confidence Interval

$$\bar{Y} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

t^* from a t - table with
95% confidence and
 $n - 1$ degrees of freedom

17

95% Confidence Interval

$$\bar{Y} \pm t^* \left(\frac{s}{\sqrt{n}} \right)$$

$$22.0139 \pm 2.030 \left(\frac{2.8473}{\sqrt{36}} \right)$$

$$22.0139 \pm 0.9633$$

21.051 mpg to 22.977 mpg

18
