Introduction to Statistics I

Textbook: Elementary Statistics (4th Edition, by Navidi and Monk), and Workshop Statistics (4th Edition by Rossman and Chance).

Previous Lecture

- Measures of Spread: Range, IQR, SD
- Empirical Rule - 1SD: 68%, 2SD: 95%, 3SD: Almost All

§3.3: Measures of Position

When we sample, we might ask ourselves where a particular value x in our sample is relative to our sample mean $\bar{x}$.

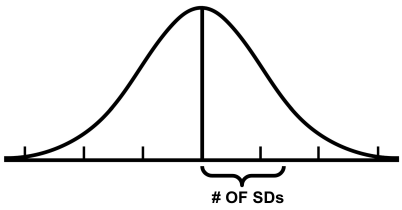
What's the most useful way to describe our value's position?

z - Score

A z-score calculates how many SDs a value x is from the sample mean $\bar{x}$.

(how far is a data pt from the avg value?)

$z = \frac{x-\bar{x}}{s}$, where x is a data pt, $\bar{x}$ is the sample mean, and s is SD.



Applies only to data which is approx. mound shaped!!

Example (z-score): A college accepts two different math placement exams.

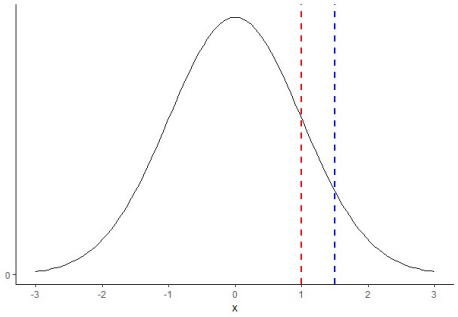
- Exam A scores students from 0 – 20, has $\bar{x} = 10$ and SD of 4.
- Exam B scores students from 0 – 100, has $\bar{x} = 70$ with SD of 10.

Arthur scores 16 on Exam A. Bertha scores 80 on Exam B.

Who did better?

Arthur z-score: $z = \frac{x-\bar{x}}{s} = \frac{16-10}{4} = 1.5$.

Bertha z-score: $z = \frac{x-\bar{x}}{s} = \frac{80-70}{10} = 1.0$.



Bertha (red), Arthur (blue)

Example (neg z-score): A college accepts two different math placement exams.

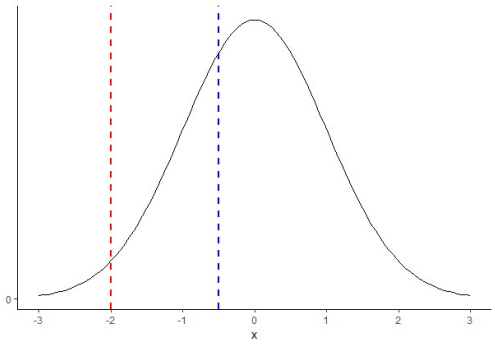
- Exam A scores students from 0 – 20, has mean of 10 and SD of 4.
- Exam B scores students from 0 – 100, has mean of 70 and SD of 10.

Dolly scores 8 on Exam A. Charlie scores 50 on Exam B.

Who did worse?

Dolly z-score: $z = \frac{x-\bar{x}}{s} = \frac{8-10}{4} = -0.5$.

Charlie z-score: $z = \frac{x-\bar{x}}{s} = \frac{50-70}{10} = -2.0$.

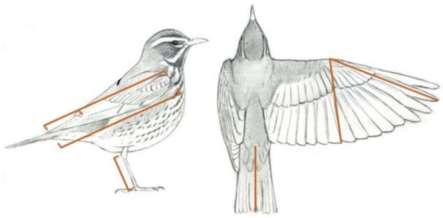


Charlie (red), Dolly (blue)

Example: A study on natural selection in 1898 examined sparrows after a particularly severe winter storm. Some sparrows survived and some didn't. Dr. Bumpus measured the size of the sparrows to see if there was a difference for those who survived and those who didn't.

bit.ly/introstatsdata

Polls: Sparrow Size



Data (length in mm) - **Sparrows who died**:

156	158	160	160	160	161
161	161	161	161	162	162
162	162	162	162	163	163
164	164	165	165	166	166

Min/Max/Median? LQ/UQ?

Min: 156, Median: 162, Max: 166, LQ: 161, UQ: 163.5

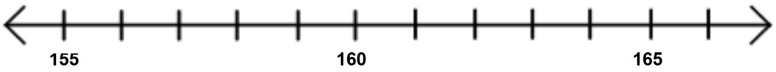
Five Number Summary (FNS) summarizes data: Describes where the four quarters of data fall.

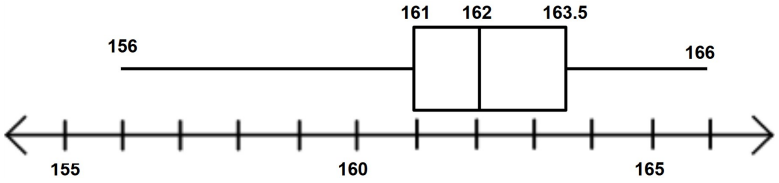
Min, Lower Quartile, Median, Upper Quartile, Max.

Back to Example - FNS:

Min: 156, LQ: 161, Median: 162, UQ: 163.5, Max: 166

How to graph?





Box Plot

Min:156, LQ: 161, Med: 162, UQ: 163.5, Max: 166

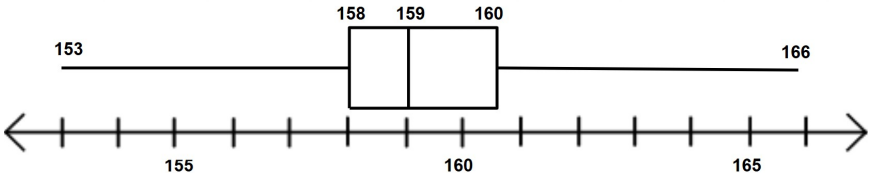
Sparrows who Survived

153	154	154	155	156	156	157	157	158
158	158	158	159	159	159	159	159	159
160	160	160	160	160	160	160	160	160
161	161	161	161	162	163	165	166	

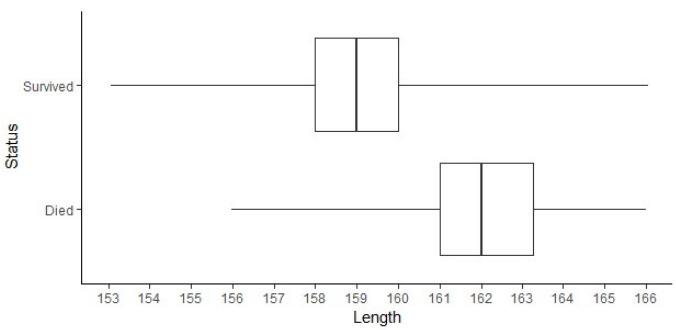
Min/LQ/Median/UQ/Max?

Min: 153, LQ: 158, Median: 159, UQ: 160, Max: 166.

Graph? Compare w/other graph?

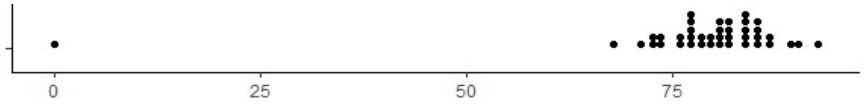


Min:153, LQ:158, Med:159, UQ:160, Max:166



Compare

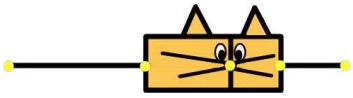
Modified Box Plots & Outliers



Outliers: Unusually large and/or small values may be outliers. We want to describe distrs excluding the outliers.

We need to define what outliers are, so we can exclude them.

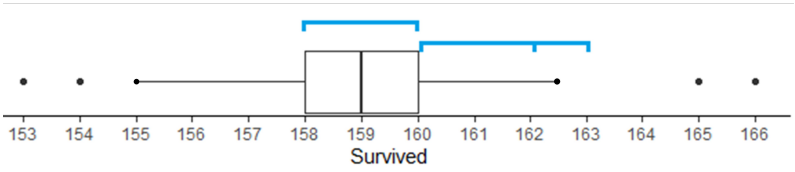
Pts in the box are not outliers. What about pts in the whiskers?



Boxplot w/whiskers

"Far away" pts should be outliers. What is "far away" enough?

Let's define pts more than 1.5 times the width of the box away as "too far away."



Defining what an outlier is

Outlier Calculation

Recall: width of the box is IQR.

Outliers Definition: pts more than 1.5 IQRs from nearest quartile.

- **Lower Outliers**: any pt below $LQ - 1.5(IQR)$.
- **Upper Outliers**: any pt above $UQ + 1.5(IQR)$.

Continuing our example:

Recall sparrows who survived: $LQ = 158,\ \ UQ = 160.$

Lower & upper outlier cut-offs?

$IQR = 160 - 158 = 2.$

$1.5(IQR) = 1.5 \times 2 = 3.$

Lower outlier is any pt below: $LQ - 1.5(IQR) : 158 - 3 = 155.$

Upper outlier is any pt above: $UQ + 1.5(IQR) : 160 + 3 = 163.$

153	154	154	155	156	156	157	157	158
158	158	158	159	159	159	159	159	159
160	160	160	160	160	160	160	160	160
161	161	161	161	162	163	165	166	

How many outliers are there?

Recall sparrows who died: $LQ = 161,\ UQ = 163.5.$

Lower & upper outlier cut-offs?

$IQR = 163.5 - 161 = 2.5.$

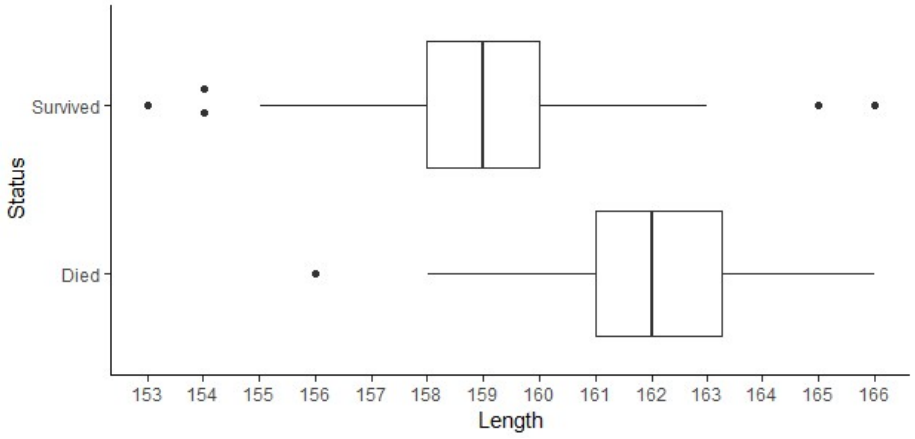
$1.5(IQR) = 2.5 \times 1.5 = 3.75.$

Lower outlier is any pt below: $LQ - 1.5(IQR) : 161 - 3.75 = 157.25.$

Upper outlier is any pt above: $UQ + 1.5(IQR) : 163.5 + 3.75 = 167.25.$

156	158	160	160	160	161
161	161	161	161	162	162
162	162	162	162	163	163
164	164	165	165	166	166

How many outliers?



Modified Box Plots (shows outliers)

Outliers are marked with a • .

Whiskers extend to last non-outlier.

Activity: 3.3a

Activity: 3.3b.

bit.ly/introstatsdata

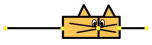
Data: IceCream

Applets: Dotplots

Convert grams to cups for icecream:

Inchcalculator.com/convert/gram-to-cup

What did we learn?

- z-score: $\frac{x-\bar{x}}{s}$
- Five Number Summary (FNS)
- Box plot 
- Outliers ●●● ●● ○
- Modified Box Plot (shows outliers)