Elevators

- You're setting regulations for an elevator for 8 people

Elevators

- You're setting regulations for an elevator for 8 people
- Total weight of 8 randomly chosen people is normally distributed.
  - mean is 1200 *lb*
  - standard deviation is 200 *lb*

Elevators

- You're setting regulations for an elevator for 8 people
- Total weight of 8 randomly chosen people is normally distributed.
  - mean is 1200 *lb*
  - standard deviation is 200 *lb*
- How often will weight be above 1600 *lb*?

Elevators

- You're setting regulations for an elevator for 8 people
- Total weight of 8 randomly chosen people is normally distributed.
  - mean is 1200 *lb*
  - standard deviation is 200 *lb*
- How often will weight be above 1600 *lb*?
  - (68-95-99.7 rule)

Elevators

- You're setting regulations for an elevator for 8 people
- Total weight of 8 randomly chosen people is normally distributed.
  - mean is 1200 *lb*
  - standard deviation is 200 *lb*
- How often will weight be above 1600 *lb*?
  - (68-95-99.7 rule)
- How often will weight be above 1750 *lb*?
  - Need to compute the z-score

z-scores

- The z-score computes how many standard deviations a point is above the mean.

z-scores

- The z-score computes how many standard deviations a point is above the mean.
  - $z(x) = \dfrac{x - \mu}{\sigma}$

z-scores

- The z-score computes how many standard deviations a point is above the mean.
  - $z(x) = \dfrac{x - \mu}{\sigma}$
- The corresponding number on the z-score table five what the percent of the area to the left of x.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979

z-scores

- The z-score computes how many standard deviations a point is above the mean.
  - $z(x) = \dfrac{x - \mu}{\sigma}$
- The corresponding number on the z-score table five what the percent of the area to the left of x.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979

Warranties

- You're a car manufacturer

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years
- Willing to replace 4% of failed engines

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years
- Willing to replace 4% of failed engines
- How long of a warranty can you give?

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years
- Willing to replace 4% of failed engines
- How long of a warranty can you give?
  - Find the z-score representing .04 area, and then find x

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years
- Willing to replace 4% of failed engines
- How long of a warranty can you give?
  - Find the z-score representing .04 area, and then find x

z	0.09	0.08	0.07	0.06	0.05	0.04	0.03
-1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336
-1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418
-1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516
-1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630

Warranties

- You're a car manufacturer
- Lifetime of engine is normally distributed
  - mean is 10 years
  - standard deviation is 2 years
- Willing to replace 4% of failed engines
- How long of a warranty can you give?
  - Find the z-score representing .04 area, and then find x

z	0.09	0.08	0.07	0.06	0.05	0.04	0.03
-1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336
-1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418
-1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516
-1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630

Sample Data

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$
  - With inflation, that's $\$124,000$

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$
  - With inflation, that's $\$124,000$
  - That included NBA star Ralph Sampson's $\$1,165,500$ salary

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$
  - With inflation, that's $\$124,000$
  - That included NBA star Ralph Sampson's $\$1,165,500$ salary
- Mean is sensitive to outliers

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$
  - With inflation, that's $\$124,000$
  - That included NBA star Ralph Sampson's $\$1,165,500$ salary
- Mean is sensitive to outliers
  - A data point is an **outlier** if its value is extreme, and not typical of most of the data

Sample Data

- In 1984, U.Va. announced that the mean salary of a graduate from the Department of Rhetoric was $\$55,000$
  - With inflation, that's $\$124,000$
  - That included NBA star Ralph Sampson's $\$1,165,500$ salary
- Mean is sensitive to outliers
  - A data point is an **outlier** if its value is extreme, and not typical of most of the data
- Want a type of average that is not sensitive to outliers

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.
  - Responses: 5,7,3,38,7

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.
  - Responses: 5,7,3,38,7
- What is the mean?

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.
  - Responses: 5,7,3,38,7
- What is the mean?
- The median is another kind of average:
  - List the data in order, and take the middle number

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.
  - Responses: 5,7,3,38,7
- What is the mean?
- The median is another kind of average:
  - List the data in order, and take the middle number
  - If there are two middle numbers, take their mean

Mean and Median

- Suppose we ask 5 people how many hours of TV they watch.
  - Responses: 5,7,3,38,7
- What is the mean?
- The median is another kind of average:
  - List the data in order, and take the middle number
  - If there are two middle numbers, take their mean
- What is the median number of hours watched?

Quartiles

- Another type of average is the interquartile range:

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
    - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
  - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles
    - These groups are called the first through fourth quartiles

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
  - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles
    - These groups are called the first through fourth quartiles
    - Which quartiles will contain outliers?

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
  - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles
    - These groups are called the first through fourth quartiles
    - Which quartiles will contain outliers?
  - Define Q_1 to be the median of the first and second quartile

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
  - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles
    - These groups are called the first through fourth quartiles
    - Which quartiles will contain outliers?
  - Define Q_1 to be the median of the first and second quartile
  - Define Q_3 to be the median of the third and fourth quartiles

Quartiles

- Another type of average is the interquartile range:
- Take: $1, 3, 11, 20, 50, 16, 9, 2, 1, 9, 16, 24, 1, 5, 15, 22$
  - Sort the data into numerical order, and divide them into four equal (consecutive) groups, called quartiles
    - These groups are called the first through fourth quartiles
    - Which quartiles will contain outliers?
  - Define Q_1 to be the median of the first and second quartile
  - Define Q_3 to be the median of the third and fourth quartiles
  - Define the interquartile range, IQR, to be $IQR = Q_3 - Q_1$

Outliers

- Rule of thumb: an outlier is a data point that is:
  - less than $Q_1 - 1.5 \cdot IQR$, or

Outliers

- Rule of thumb: an outlier is a data point that is:
  - less than $Q_1 - 1.5 \cdot IQR$, or
  - more than $Q_2 + 1.5 \cdot IQR$

Outliers

- Rule of thumb: an outlier is a data point that is:
  - less than $Q_1 - 1.5 \cdot IQR$, or
  - more than $Q_2 + 1.5 \cdot IQR$
- What are the outliers in the previous data set?

Batting Average

- In baseball, the batting average is

$$\text{batting average} = \frac{\text{number of hits}}{\text{number of at-bats}}$$

Batting Average

- In baseball, the batting average is

$$\text{batting average} = \frac{\text{number of hits}}{\text{number of at-bats}}$$

- Consider the following data for two players:

	Hits (2012)	Attempts (2012)	Hits (2013)	Attempts (2013)
Player A	55	100	298	1000
Player B	372	1000	25	100

Batting Average

- In baseball, the batting average is

$$\text{batting average} = \frac{\text{number of hits}}{\text{number of at-bats}}$$

- Consider the following data for two players:

	Hits (2012)	Attempts (2012)	Hits (2013)	Attempts (2013)
Player A	55	100	298	1000
Player B	372	1000	25	100

- What are the batting averages for the two players in each year?

Batting Average

- In baseball, the batting average is

$$\text{batting average} = \frac{\text{number of hits}}{\text{number of at-bats}}$$

- Consider the following data for two players:

	Hits (2012)	Attempts (2012)	Hits (2013)	Attempts (2013)
Player A	55	100	298	1000
Player B	372	1000	25	100

- What are the batting averages for the two players in each year?
- Player *A* appears to be a better hitter

Simpson's Paradox

- Now compute the overall batting average for each player

Simpson's Paradox

- Now compute the overall batting average for each player
- Overall, it seems that Player B is the better hitter!

Simpson's Paradox

- Now compute the overall batting average for each player
- Overall, it seems that Player B is the better hitter!
- This is an example of **Simpson's Paradox**:

Simpson's Paradox

- Now compute the overall batting average for each player
- Overall, it seems that Player B is the better hitter!
- This is an example of **Simpson's Paradox**:
  - A trend that appears in different groups of data may disappear when these groups are combined. Using aggregate data, the trend may reverse itself.

Berkeley Graduate School

- 1973 admissions data for UC Berkeley graduate school:

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

Berkeley Graduate School

- 1973 admissions data for UC Berkeley graduate school:

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

- Case was made that admissions was biased against women.

Berkeley Graduate School

- 1973 admissions data for UC Berkeley graduate school:

	Applicants	Admitted
Men	8442	**44%**
Women	4321	35%

- Case was made that admissions was biased against women.
- Alternative explanations?

Berkeley Graduate School

- Breakdown among six largest departments:

Dept.	Men Applicants	Men Admitted	Women Applicants	Women Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

Berkeley Graduate School

- Breakdown among six largest departments:

Dept.	Men Applicants	Men Admitted	Women Applicants	Women Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

- Most departments individually favored women

Berkeley Graduate School

- Breakdown among six largest departments:

Dept.	Men Applicants	Men Admitted	Women Applicants	Women Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

- Most departments individually favored women
- Explanations?

Berkeley Graduate School

- Breakdown among six largest departments:

Dept.	Men Applicants	Men Admitted	Women Applicants	Women Admitted
A	825	62%	108	**82%**
B	560	63%	25	**68%**
C	325	**37%**	593	34%
D	417	33%	375	**35%**
E	191	**28%**	393	24%
F	373	6%	341	**7%**

- Most departments individually favored women
- Explanations?
- It was concluded that women were more likely to apply to more competitive departments with low rates of admission.

Kidney Stone Treatment

- A medical study compared kidney stone treatments:

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy
- Treatment A was successful 78% of the time (273/350)

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy
- Treatment A was successful 78% of the time (273/350)
- Treatment B was successful **83**% of the time (289/350)

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy
- Treatment A was successful 78% of the time (273/350)
- Treatment B was successful **83**% of the time (289/350)
- Break treatment among stone size:

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy
- Treatment A was successful 78% of the time (273/350)
- Treatment B was successful **83**% of the time (289/350)
- Break treatment among stone size:

	Treatment A	Treatment B
Small Stones	**93% (81/87)**	87% (234/270)
Large Stones	**73% (192/263)**	69% (55/80)

Kidney Stone Treatment

- A medical study compared kidney stone treatments:
  - Treatment A: all open surgical procedures
  - Treatment B: percutaneous nephrolithotomy
- Treatment A was successful 78% of the time (273/350)
- Treatment B was successful **83**% of the time (289/350)
- Break treatment among stone size:

	Treatment A	Treatment B
Small Stones	**93% (81/87)**	87% (234/270)
Large Stones	**73% (192/263)**	69% (55/80)

- Doctors were performing the better treatment to the more serious stones.