Netshock
Burn Bright

INFERENTIAL STATS

WHAT ARE INFERENTIAL STATISTICS?

Descriptive statistics describe the data, examples being:

- 40% of company sales were on Pizza
- An average of xx customers entered the store per day
- Where does a particular value stand with respect to other values in the dataset
- Generally graphical representation of data

Inferential statistics use our descriptive statistics to draw conclusions. For example, if we did a phone survey of 1,000 people (a sample) and 150 said that they disliked pizza, we could infer that 15% of all men dislike pizza. We can use descriptive statistics to predict or estimate from a sample for the whole population.

DISTRIBUTIONS

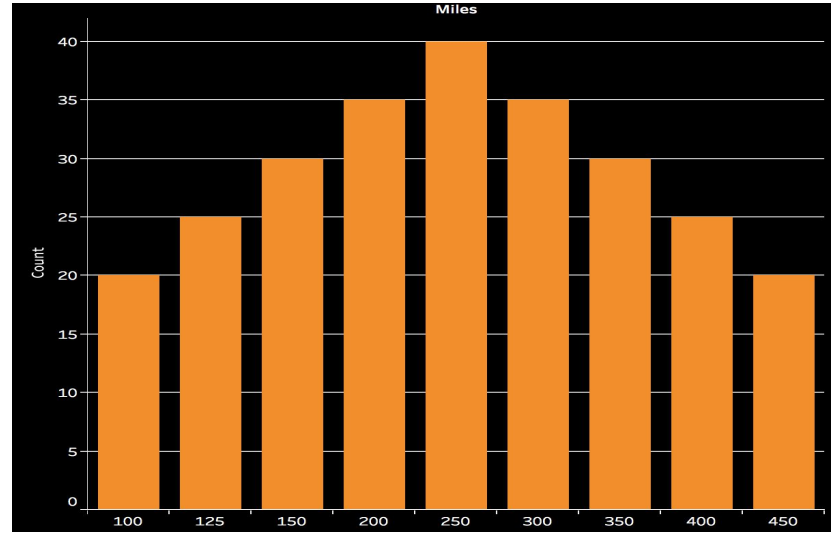
A distribution lists all possible values and how often they occur.

For example, miles driven without needing to refuel using the data below, gives us this distribution.

This is a probability

distribution.

100	20 people
125	25
150	30
200	35
250	40
300	35
350	30
400	25
450	20



Netshock
Burn Bright

STANDARD NORMAL DISTRIBUTIONS

With a normal distribution:

1. Mean, median and mode are equal
2. Symmetrical
3. No skew (50% of values are less than the mean & 50% greater)
4. Highest point is the mean (because its the same as the mode)

A standard normal distribution has a mean of 0 and a standard deviation of 1.

Why do we standardize?

- By standardizing, we move from saying ‘my height is 10.1cm from the mean’ and move to saying ‘my height is 2 standard deviations from the mean’
- Generally we say that in a standard normal distribution:
  - 68% of values are within 1 standard deviation of the mean
  - 95% of values are within 2 standard deviations of the mean
  - 99.7% of values are within 3 standard deviations of the mean

STANDARD NORMAL DISTRIBUTIONS

With a normal distribution:

1. Mean, median and mode are equal
2. Symmetrical
3. No skew (50% of values are less than the mean & 50% greater)
4. Highest point is the mean (because its the same as the mode)

A standard normal distribution has a mean of 0 and a standard deviation of 1.

Why do we standardize?

- By standardizing, we move from saying ‘my height is 10.1cm from the mean’ and move to saying ‘my height is 2 standard deviations from the mean’
- Generally we say that in a standard normal distribution:
  - 68% of values are within 1 standard deviation of the mean
  - 95% of values are within 2 standard deviations of the mean
  - 99.7% of values are within 3 standard deviations of the mean

STANDARD NORMAL DISTRIBUTIONS

Let's look at an example:

If we have 2 students and we want to compare performance.

- One has done a maths exam, the results of which are in a normal distribution with a mean of 44 and standard deviation of 6. The student scored 49.
- One has done an English exam for which all results had a mean of 45 and a standard deviation of 9. The student got a score of 55.

So now, which student did better with respect to all other students that took the exam? The English exam taker has a higher score, but relative to her class, she may have underperformed.

At the end of the following section, you see that the English student did slightly better than the Math student - but not by much!

EXAMPLE: ENGLISH EXAM

Take our original points and calculate the mean and standard deviation.

English Scores
33
39
41
41
43
44
49
55
66

Mean: 45

Std Dev: 9

A standardized distribution has a mean of 0 and std dev of 1

Netshock
Burn Bright

EXAMPLE: ENGLISH EXAM

Now, let's subtract the mean from each datapoint.

ORIGINAL DATA
33
39
41
41
43
44
49
55
66

Mean: 45

Std Dev: 9

-12.666667
-6.6666667
-4.6666667
-4.6666667
-2.6666667
-1.6666667
3.33333333
9.33333333
20.3333333

New Mean: 0

Std Dev: 9

The mean is now 0, yay!

Now for the standard deviation

Netshock
Burn Bright

EXAMPLE: ENGLISH EXAM

Finally, let's divide the new values by the standard deviation

Subtracted Vals		
-12.666667		-1.291106218
-6.6666667		-0.679529589
-4.6666667		-0.475670712
-4.6666667		-0.475670712
-2.6666667		-0.271811835
-1.6666667		-0.169882397
3.33333333		0.339764794
9.33333333		0.951341424
20.3333333		2.072565245

Mean: 0

Std Dev: 9

New Mean: 0

New Std Dev: 1

The score is within 1 standard deviation above the mean

Netshock
Burn Bright

EXAMPLE: MATHS EXAM

Take our original points and calculate the mean and standard deviation.

English Scores
33
39
41
41
43
44
49
55
51

Mean: 44

Std Dev: 6

A standardized distribution has a mean of 0 and std dev of 1

Netshock
Burn Bright

EXAMPLE: MATHS EXAM

Now, let's subtract the mean from each datapoint.

ORIGINAL DATA
33
39
41
41
43
44
49
55
51

Mean: 44

Std Dev: 6

-11
-5
-3
-3
-1
0
5
11
7

New Mean: 0

New Std Dev: 6

The mean is now 0, yay!

Now for the standard deviation

Netshock
Burn Bright

EXAMPLE: MATHS EXAM

Finally, let's divide the new values by the standard deviation

Subtracted Vals		
-11		-1.639783183
-5		-0.745355992
-3		-0.447213595
-3		-0.447213595
-1		-0.149071198
0		0
5		0.745355992
11		1.639783183
7		1.043498389

Mean: 0

Std Dev: 6

New Mean: 0

New Std Dev: 1

The score is within 1 deviation of the mean

Netshock
Burn Bright