

Week 4 Statistical theory

4. Normal model

Stat 140 - 04

Mount Holyoke College

Dr. Shan Shan

Slides posted at <http://sshanshans.github.io/stat140>

1. Today: normal model

2. Main ideas

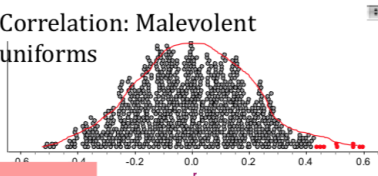
1. Introduction to normal model
2. z -scores

3. Summary

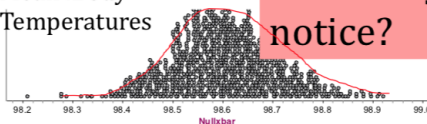
Slope :Restaurant tips



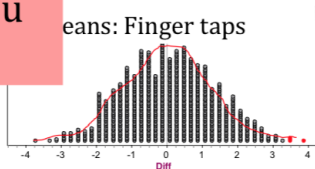
Correlation: Malevolent uniforms



Mean :Body Temperatures

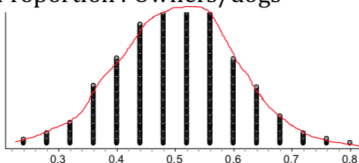


Means: Finger taps

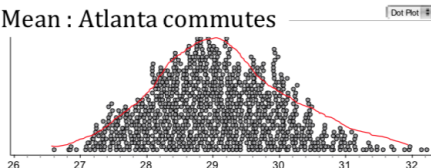


What do you notice?

Proportion : Owners/dogs



Mean : Atlanta commutes



We said

"For most of the statistics we consider, if the sample size is large enough the sampling distribution will be symmetric and bell-shaped."

We will model it as the **normal distribution** (aka the **Gaussian distribution**).

The distribution was named after Carl Friedrich Gauss, the greatest mathematician in history. He proved the fundamental theorem of algebra four ways, inventing a new branch of mathematics each time. He worked in number theory, co-invented the telegraph, and discovered non-Euclidean geometry, but did not publish, fearing controversy.

1. Today: normal model

2. Main ideas

1. Introduction to normal model
2. z -scores

3. Summary

1. Today: normal model

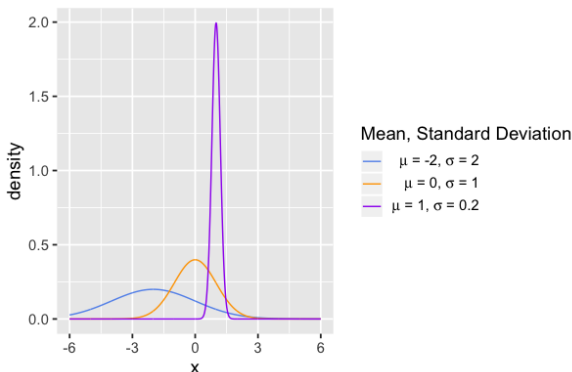
2. Main ideas

1. Introduction to normal model
2. z -scores

3. Summary

The normal distribution is fully characterized by two things:

- ▶ The **mean** of a normal model shows where it is centered.
- ▶ The **standard deviation** of a normal model shows how spread out the normal is.



1. The area under the curve of a normal distribution is equal to the proportion of the distribution falling within that range
2. Knowing just the mean and standard deviation of a normal distribution allows you to calculate areas in the tails and percentiles
3. The probability gives you an idea about how extreme a value

Let's explore together:

https://gallery.shinyapps.io/dist_calc/

The 68–95–99.7 Rule

For any normal distribution,

- ▶ 68% of observations are within $\mu + \sigma$
- ▶ 95% of observations are within $\mu + 2\sigma$
- ▶ 99.7% of observations are within $\mu + 3\sigma$

1. Today: normal model

2. Main ideas

1. Introduction to normal model
2. z -scores

3. Summary

Often, we standardize the data to have mean 0 and standard deviation 1

This is done with z-scores:

- ▶ From x to z :

$$z = (x - \mu) / \sigma$$

- ▶ From z to x :

$$x = \mu + z\sigma$$

Places everything on a common scale.

What does z-score measure in simple English?

Reggie Jackson, the famous baseball player, has an IQ of 132. What percentage of people are smarter than him?

Assume that IQs are normally distributed with mean 100 and standard deviation 16.

We want the area under the normal distribution for IQ that lies to the right of 132. By the z -transformation, this is equivalent to the area under the standard normal distribution that lies to the right of

$$z = \frac{X - \mu}{\sigma} = \frac{132 - 100}{16} = 2$$

The area above 2 is 0.025. Thus about 2.5% of people are smarter than Reggie Jackson.

We find the value of the data (x) that corresponds to a given percentage.

To join Mensa one must be in the top 2.5% of the IQ distribution. What score do you need? That gives the z -value of approximately 2. So 2.5% of the area under the standard normal is above 2.

Now we use the inverse z -transformation:

$$x = \mu + z\sigma = 100 + (2)(16) = 132.$$

One needs an IQ score of at least 132 to join.

1. Today: normal model

2. Main ideas

1. Introduction to normal model
2. z -scores

3. Summary

1. Introduction to normal model
2. z -scores