



# SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

## Department of Information Technology

Course Name – 23ADT202 Fundamental of Data  
science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Regression toward the mean





# Introduction to Regression Toward the Mean



Regression toward the mean refers to the tendency of extreme values in a dataset to move toward the average over time. This is common in situations where variability exists due to random factors. Extreme measurements often change to more typical values when observed repeatedly.



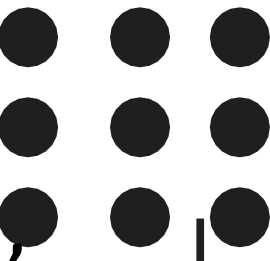
# Key Concept



When a measurement is extreme on the first observation, it is likely to be closer to the average in the second. This shift is not caused by any intervention but by random fluctuations. It's a natural result of inherent variability in measurements, not a predictable trend.



# Example of Regression Toward the Mean



Consider a student who scores unusually high on a test. In the next test, the student's score is likely to regress toward the average. Random factors such as mood or test difficulty affect extreme scores but don't influence subsequent performance consistently.



# Why Does Regression Toward the Mean Happen?



Extreme scores result from both true values and random error. Over time, these random errors tend to average out, so the extreme scores are less likely to appear again. Thus, the extreme observations move closer to the mean when measured again.



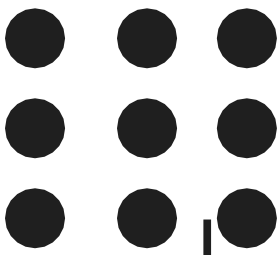
# Illustrative Example - Parents' and Children's Heights?



Consider a dataset where parents' heights are compared to their children's heights. Children of very tall parents tend to be shorter, while children of shorter parents tend to be taller. This demonstrates how both extremes regress toward the population mean.



# Visualization of Regression Toward the Mean?



A visual representation shows how children's heights tend to be closer to the average, even though their parents' heights may be extreme. This shows the natural shift toward the mean.





# Applications in Healthcare, Sports, and Business?



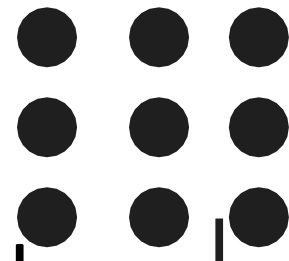
Regression toward the mean applies in various fields:

- **Healthcare:** Patients with extreme symptoms may improve naturally over time.
- **Sports:** Athletes performing exceptionally well may perform closer to average in the next season.
- **Business:** Sales teams with extreme performance may return to average performance in subsequent periods.





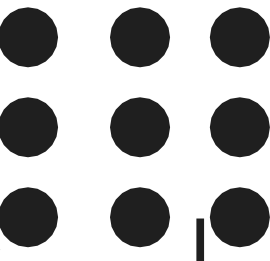
# Common Misinterpretations



It is important not to mistake regression toward the mean for causation. For example, if symptoms improve, it may be due to natural variability rather than the effects of a new treatment. Similarly, ignoring variability can lead to false conclusions about trends.



# Exercise and Conclusion



1. **Analyze the dataset:** Y values in a given dataset will regress toward the mean of X when predicted using linear regression.
2. **Cricket player example:** A player's extreme scores will tend to return to their average over time. Understanding regression toward the mean helps us predict more realistic outcomes.



**THANK YOU**