



BITS Pilani
DIGITAL
Excellence made yours








BITS Pilani
DIGITAL
Excellence made yours

Probability and Statistics

Week 2 Live Session

Dr. Pushpendra Gupta

Today's Focus

-  Hands-on Visualization using Google Colab
-  Recap
-  Chebyshev's Inequality in Practice
-  Choosing the Right 'Average'
-  Skewness: What is your data is really saying

Week 1 Recap



Central Tendency



Variability Measures



Data Visualization



Distribution Analysis

Week 2 Recap



Quartile Deviation and IQR



5 Point Summary



Box Plot



Infer from Statistical Summary



Chebyshev's Inequality in Practice



BITS Pilani
DIGITAL
Excellence made yours



The Universal Guarantee:

For ANY dataset, at least $\left(1 - \frac{1}{k^2}\right)$ of data lies within k standard deviations



No assumptions about distribution shape needed!

$k = 2$: At least 75% of data within 2σ

$k = 3$: At least 89% of data within 3σ



Perfect for AI because:

Real-world data is often non-normal

New data distributions are unknown

Normal Distribution



Chebyshev's Inequality in Practice



BITS Pilani
DIGITAL
Excellence made yours



The Universal Guarantee:

For ANY dataset, at least $\left(1 - \frac{1}{k^2}\right)$ of data lies within k standard deviations

at most $\frac{1}{k^2}$ of the data will be outside k standard deviations from the mean.

k	At most outside $k\sigma$	At least within $k\sigma$	Normal comparison
1.5	44.4%	55.6%	86.6% within
2	25%	75%	95% within
2.5	16%	84%	98.8% within
3	11.1%	88.9%	99.7% within



Chebyshev's Inequality in Practice

A payment system processes transactions with mean Rs. 500 and standard deviation Rs. 150. Without knowing the distribution, what can we say about transactions outside normal ranges?

using Chebyshev with $k = 2$:

- Range: $\mu \pm 2\sigma = 500 \pm 300 = [200, 800]$
- Chebyshev guarantees: At most 25% of transactions fall outside $[200, 800]$
- Equivalently: At least 75% of transactions are between 200 and 800

This 25% is an upper bound on outliers, not the probability that any specific flagged transaction is fraudulent!

$$P(\text{transaction outside } [200, 800]) \leq 0.25$$

A manufacturing process produces widgets with mean weight 100g and standard deviation 5g. What percentage of widgets must weigh between 85g and 115g?

This range is $\mu \pm 3\sigma$ ($k = 3$) By Chebyshev's inequality:

At least $1 - 1/9 = 88.9\%$ of widgets fall in this range.



Chebyshev in Action - Anomaly Detection



BITS Pilani
DIGITAL
Excellence made yours

Fraud Detection System:

Monitor transaction amounts for each user

No assumptions about spending patterns

Use Chebyshev bounds to flag unusual transactions

Implementation:

Calculate mean and σ for user's historical data

Flag transactions beyond 2.5σ as "suspicious"

Guaranteed to catch extreme anomalies

Business Impact:

Reduce false positives (incorrectly flagged as fraudulent)

Catch 84%+ of extreme cases automatically

Thank You

The content provided in this digital training programme has been developed by professors / external vendors. All intellectual property rights in the content, including but not limited to video, text, slides, graphics, and audio materials, are owned exclusively by BITS Pilani ("BITS"), unless otherwise expressly stated. This programme may also include licensed or contributed works from third-party experts, faculty, and collaborating institutions, all of whom are duly acknowledged where appropriate.

All product names, logos, brands, and entity names referenced within this content are the property of their respective trademark owners. The inclusion of such names, logos, or brands does not imply any affiliation with, or endorsement by, these entities.

This material is intended solely for personal learning and academic purposes. Except for content that is explicitly confirmed to be in the public domain, no part of this content may be reproduced, copied, modified, distributed, republished, uploaded, posted, publicly displayed, or transmitted in any form or by any means – electronic, mechanical, photocopying, recording, or otherwise – without the prior express written consent of BITS Pilani or the applicable rights holder.

Any unauthorized use, commercial exploitation, resale, or public sharing of this content in any manner constitutes a violation of applicable intellectual property laws. BITS Pilani reserves the right to pursue all appropriate legal remedies, including but not limited to civil action and criminal prosecution under the Copyright Act, 1957, and other relevant statutes in force in India.

By accessing or using this digital training programme, you agree to comply with the terms set forth in this disclaimer.