

# Melanin on Margins: A Study on Skin-Color Bias in the Bollywood Film Industry

Abhik Rana  
(**BS2301**)

May 2025

# Introduction

- Bollywood reaches over 90 countries with a \$2.1B market.
- Despite this reach, there is limited research on colorism in casting.
- This study investigates how skin tone bias manifests in Hindi cinema.

# Research Questions

- ① How pronounced is colorism in Bollywood movies?
- ② How has colorism changed over time?
- ③ What is the distribution of skin tones across roles?
- ④ What are the average luminance trends?

- Top 5 grossing movies each year (2015–2024).
- Extracted  $L^*$  values from CIELAB color space.
- Leading and supporting cast analyzed separately by gender.
- Average luminance used as proxy for skin tone.

# Dataset Cleaning and Preprocessing

- Selected scenes with visible glabella.
- Removed outliers and standardized frame counts.
- Averaged support cast  $L^*$  values per scene.
- Manual filtering of foreign or heavily made-up faces.

# ANOVA: Concept

- Used to compare  $L^*$  across four roles:
  - actor, actress, side\_actor, side\_actress
- Hypothesis:

$$H_0 : \mu_{\text{actor}} = \mu_{\text{actress}} = \mu_{\text{side\_actor}} = \mu_{\text{side\_actress}}$$

- Tested via F-statistic:

$$F = \frac{\text{between-group variance}}{\text{within-group variance}}$$

# ANOVA Results

Source	sum_sq	df	F	PR(>F)
C(role)	26213.77	3	117.25	$1.63 \times 10^{-43}$
Residual	14606.09	196		

Table: ANOVA for Luminance by Role

Null hypothesis rejected: at least one role differs significantly in average luminance.

# ANOVA: Luminance by Year

Source	sum_sq	df	F	PR(>F)
C(year)	1458.47	9	0.78	0.63
Residual	39361.40	190		

Table: ANOVA for Luminance by Year

No significant difference across years.



# Cohen's d Effect Size

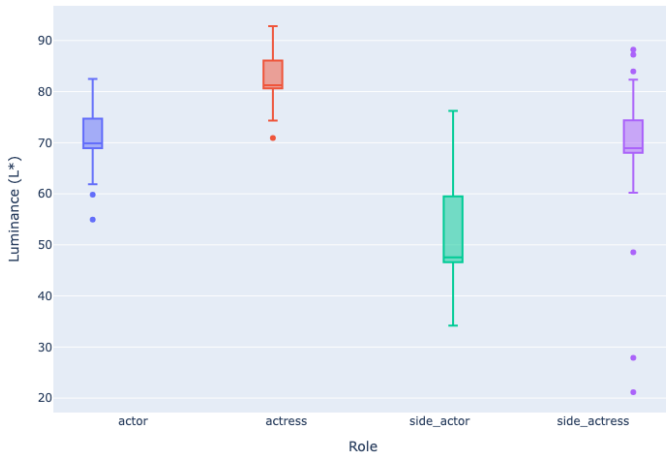
- Formula:  $d = \frac{L_A - L_B}{s_{\text{pooled}}}$
- Actor vs Actress:  $d = -2.319$  (actors significantly lighter)
- Lead vs Side roles:  $d = 1.492$  (leads significantly lighter)

# Distribution of Skin Tones

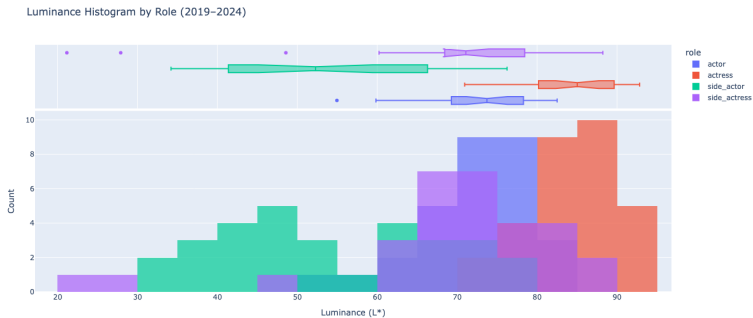
- Multiple visualizations to assess role-based luminance differences.
- Next slides show KDEs, histograms, percentiles, and trends.

# Luminance Distribution by Role

Luminance Distribution by Role (2015–2024)

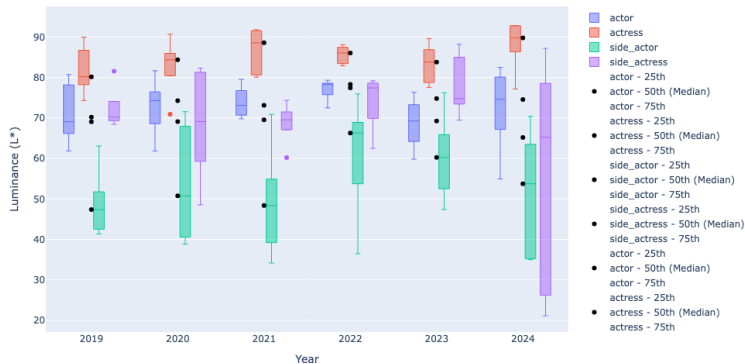


# Luminance Histogram by Role

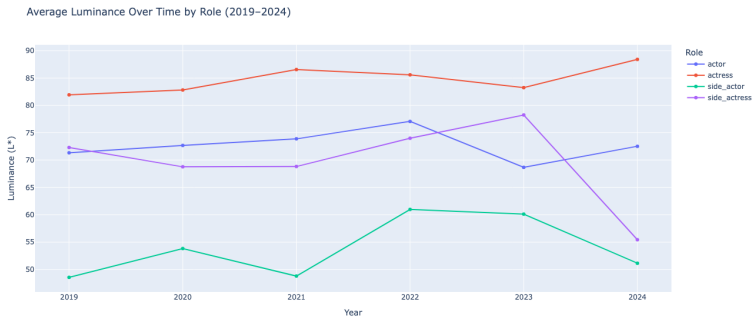


# Luminance Distribution with Percentiles

Luminance Distribution (2019–2024) by Role with Percentile Overlays



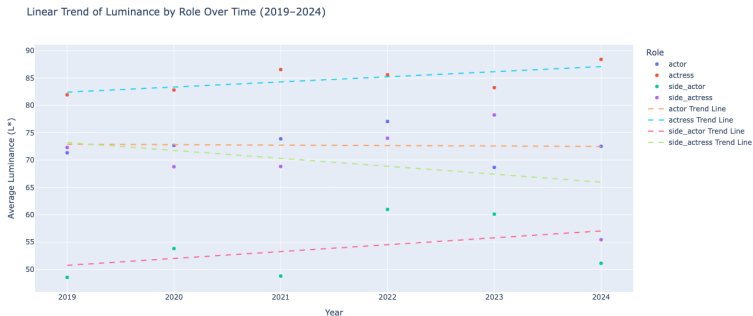
# Average Luminance Over Time by Role



# Observed Trends

- Persistent preference for lighter skin in lead roles.
- Clustering and trendline analyses support earlier findings.

# Linear Trend of Luminance by Role





# K-Means Clustering of Luminance



# Clustering Interpretation

- Cluster 0: Darker tones → More side roles.
- Cluster 2: Lighter tones → Mostly lead roles.
- Supports presence of colorism through unsupervised learning.

# Conclusion

- Strong empirical evidence of colorism.
- Leads have significantly lighter skin than supporting characters.
- Temporal trends show little progress.
- Statistical and ML methods converge on similar findings.

# Acknowledgments

I would like to express my sincere gratitude to  
**Professor Garga Chatterjee** and **Professor Kuntal Ghosh**  
for their invaluable guidance and mentorship throughout this study.

Thank You

Questions?