

# Análisis Estadístico de Películas Blockbuster

1977 – 2019

Ronald Provance Valladares  
Agustín Alberto Carbajal Romero  
Dylan Ramsés Cabrera Morales

Universidad de La Habana — MATCOM

# Objetivos y Motivación

- La industria cinematográfica mueve miles de millones anualmente
- Comprender qué impulsa el éxito es clave para productores e inversores

## Preguntas de investigación:

- ¿Qué factores influyen en la recaudación mundial?
- ¿Existen clusters de películas similares?
- ¿Ha cambiado la recaudación promedio con el tiempo?

# Descripción del Dataset

- Fuente: Kaggle
- Período: 1977–2019
- Top 10 películas por año
- Total: 430 películas

## Variables principales:

- Presupuesto: \$15M – \$365M
- Recaudación
- Rating IMDb (1–10)
- Género
- Clasificación MPAA

# Estadísticos Descriptivos

Variable	Media	Mediana	Desv. Est
Presupuesto	\$108.5M	\$100.0M	\$67.8M
Recaudación	\$698.0M	\$654.0M	\$425.0M
Rating IMDb	6.87	6.90	0.95
ROI	6.43	5.52	6.21

## Hallazgos:

- Alta variabilidad en presupuestos y recaudación
- Ratings generalmente altos
- ROI muy variable

# Patrones Exploratorios

## Géneros más comunes:

- Acción (32 %)
- Aventura (28 %)
- Sci-Fi (18 %)
- Drama (12 %)

## Clasificación MPAA:

- PG-13: 45 %
- R: 38 %
- PG: 12 %
- G: 5 %

**Conclusión clave:** Películas PG-13 tienden a recaudar más.

# Pregunta 1: Regresión Lineal Múltiple

Se estimó un modelo con:

- Presupuesto
- Rating IMDb
- Género (dummies)
- Clasificación MPAA (dummies)

## Tratamiento:

- Variables continuas estandarizadas
- Variables categóricas codificadas
- Variables no significativas eliminadas

# Validación del Modelo

- Media residuos  $\approx 0$
- Durbin-Watson  $\approx 1,44$
- Heterocedasticidad corregida (HC3)
- No multicolinealidad (Cond.# = 4.56)
- Normalidad confirmada

# Resultados de Regresión

## Rendimiento:

- $R^2 = 0,697$
- F-statistic < 0,001

## Factores importantes:

- Presupuesto: +0.63 % por 1 %
- Rating IMDb: +13 % por punto
- Animación: +31 %
- PG-13: +32 %

# Conclusiones de Regresión

- Presupuesto es el factor más determinante
- Rating influye significativamente
- Clasificación PG-13 amplía audiencia
- Animación destaca en recaudación

**Éxito = interacción de múltiples factores**

## Pregunta 2: K-Means Clustering

- Algoritmo no supervisado
- Variables:
  - Presupuesto
  - Rating IMDb
  - Género
- Estandarización + One-hot encoding

# Selección de K

Métrica	Valor
Silhouette	0.52
Davies-Bouldin	0.89
Calinski-Harabasz	185.3

**K óptimo: 4 clusters**

# Clusters Identificados

## **Cluster 0 – Blockbusters Probados**

- Presupuesto alto, rating alto

## **Cluster 1 – Apuestas Riesgosas**

- Presupuesto muy alto, rating medio

## **Cluster 2 – Joyas Ocultas**

- Presupuesto moderado, rating alto

## **Cluster 3 – Estándar Masivo**

- Presupuesto y rating promedio

# Hallazgos Clustering

- Presupuesto no garantiza éxito crítico
- Múltiples modelos de negocio coexisten
- Acción y Sci-Fi requieren mayores presupuestos
- Drama logra éxito con menos inversión

# Pregunta 3: ANOVA

## Hipótesis:

- $H_0$ : Recaudación igual en todas las décadas
- $H_1$ : Alguna década difiere

## Supuestos verificados:

- Normalidad
- Homogeneidad de varianzas

# Resultados ANOVA

Métrica	Valor
F	5.234
p-value	0.0012
$\eta^2$	0.08

**Decisión:** Rechaza  $H_0$

# Recaudación por Década

- 1970s: \$550.2M
- 1980s: \$618.5M
- 1990s: \$701.3M
- 2000s: \$742.8M
- 2010s: \$821.5M

Incremento del 49 % desde 1970s

# Interpretación ANOVA

- Diferencias estadísticamente significativas
- Década explica 8 % de la variación
- Otros factores también influyen fuertemente

**Tendencia creciente clara**

# Limitaciones

- Solo Top-10 por año (sesgo)
- Variables limitadas
- Período específico

# Conclusiones Finales

- Presupuesto importa, pero no lo es todo
- Múltiples caminos al éxito
- Calidad predice recaudación
- Industria usa estrategias diferenciadas

**El éxito no es una fórmula única**

# Cierre

Datos: Kaggle

Técnicas: Regresión, Clustering, ANOVA

Período: 1977–2019

430 películas analizadas