7COM1079-2024 Student Group No: A169

Names of Student Attendees:

1. Ghouse Mohiddin Shaik
2. Muqtadir Siddiqui Mohammed Abdul
3. Kashif Uddin Mohammed
4. Shahzad Hussain
5. Hassan Qureshi

Group Name: A169 Name of Student Presenting: Ghouse Mohiddin Shaik

Research question: Is there a correlation between a movie's duration and its average votes on IMDb?

Tutorial Presentation for Feedback

Date: 10/11/2024

University of Hertfordshire UH

❖ **Sample data:** This dataset consists of movie details from various years, focusing on their duration and average vote"

	imdb_title_	title	original_tit	year	date_publi	genre	duration	country	language	director	writer	productior	actors	descriptior	avg_vote	votes	budget	usa_gross_	worlwide_g	metascore	reviews_fr	reviews_from_critics
2	tt0000574	The Story o	The Story o	1906	########	Biography,	70	Australia		Charles Ta	Charles Ta	J. and N. Ta	Elizabeth T	True story	6.1	537	$ 2250				7	7
3	tt0001892	Den sorte (	Den sorte (	1911	########	Drama	53	Germany,	Denmark	Urban Gad	Urban Gad	Fotorama	Asta Nielse	Two men o	5.9	171					4	2
4	tt0002101	Cleopatra	Cleopatra	1912	########	Drama, His	100	USA	English	Charles L.	Victorien S	Helen Gar	Helen Gar	The fabled	5.2	420	$ 45000				24	3
5	tt0002130	L'Inferno	L'Inferno	1911	########	Adventure,	68	Italy	Italian	Francesco	Dante Aligl	Milano Filn	Salvatore F	Loosely ad	7	2019					28	14
6	tt0002199	From the M	From the M	1912	1913	Biography,	60	USA	English	Sidney Olc	Gene Gaur	Kalem Con	R. Henders	An accoun	5.7	438					12	5
7	tt0002423	Madame D	Madame D	1919	########	Biography,	85	Germany	German	Ernst Lubit	Norbert Fa	Projektions	Pola Negri	The story o	6.8	709					11	9
8	tt0002445	Quo Vadis?	Quo Vadis?	1913	########	Drama, His	120	Italy	Italian	Enrico Gua	Henryk Sie	SocietÃ Ita	Amleto No	An epic Ital	6.2	241	ITL 45000				6	4
9	tt0002452	Independe	Independe	1912	########	History, W	120	Romania		Aristide De	Aristide De	Societatea	Aristide De	The movie	6.7	187	ROL 400000				3	1
10	tt0002461	Richard III	Richard III	1912	########	Drama	55	France, US	English	AndrÃ© Ca	James Kea	Le Film d'A	Robert Ger	Richard of	5.5	211	$ 30000				7	1
11	tt0002646	Atlantis	Atlantis	1913	########	Drama	121	Denmark	Danish	August Blo	Axel Garde	Nordisk Fil	Olaf FÃ¸ns	After Dr. Fr	6.7	310					9	9
12	tt0002844	FantÃ´ma	FantÃ´ma	1913	########	Crime, Dra	54	France	French	Louis Feuil	Marcel Alla	SociÃ©tÃ©	RenÃ© Na	Inspector J	7	1853					9	29
13	tt0003014	Ingeborg H	Ingeborg H	1913	########	Drama	96	Sweden		Victor SjÃ¶	Nils Krok	Svenska Bi	Hilda Borg	Single mot	7.1	888					16	7

❖ **Columns used**: 1. duration (minutes)
2. average_vote (out of 10)

❖ The dataset contains **81,273 rows** and **22 columns** in total.

University of Hertfordshire UH

Dataset ID: DS231 (IMDb movies.csv)

This dataset offers valuable insights into how movie duration impacts audience ratings.

Our *Independent variable* duration.

This Independent variable datatype : Interval/measurement data.

Our *Dependent variable* is average_vote.

This Dependent variable datatype: Interval/measurement data.

University of Hertfordshire UH

Our Research Question:

Is there a correlation between average votes (dependent interval variable) and movie duration (independent interval variable)?

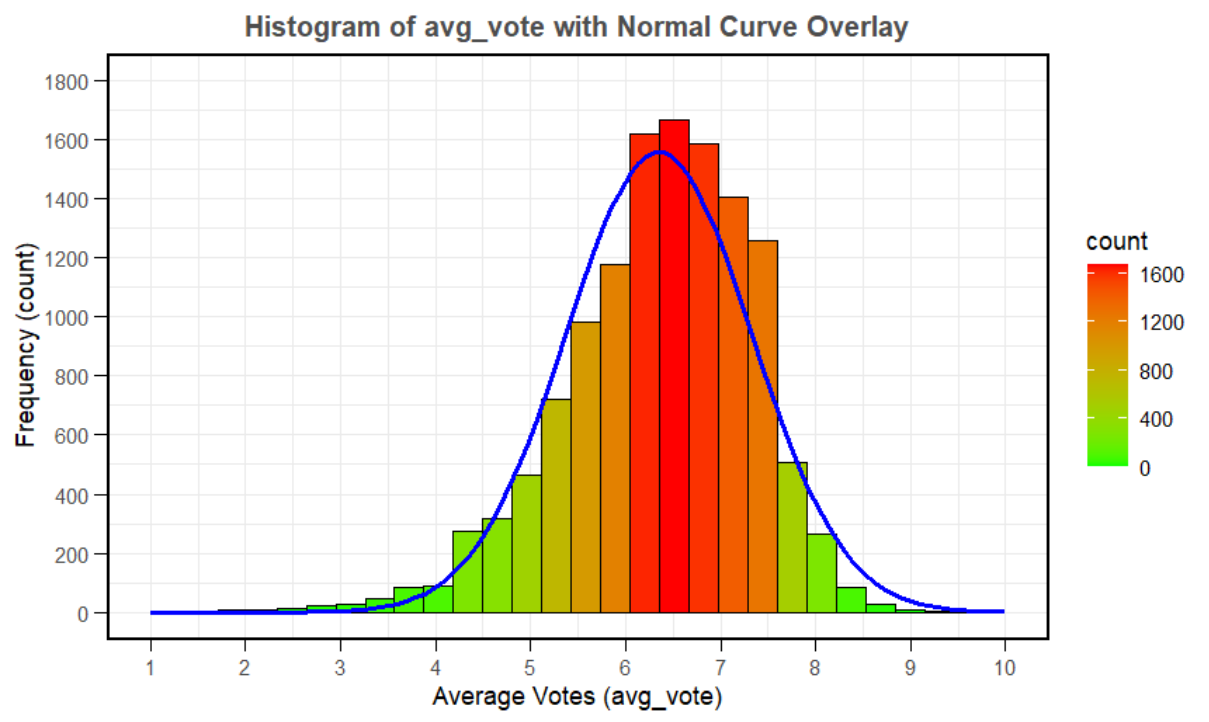
Hypotheses:

- *Null Hypothesis (H_0):* There is no correlation between average votes and movie duration.
- *Alternative Hypothesis (H_1):* There is a correlation between average votes and movie duration.

University of Hertfordshire UH

Here is a **Histogram** showing the frequencies of our dependent variable(average_vote) to include the normal curve overlay.

Our RQ asks about Correlation



Choose one:

1. The blue normal curve **overlay follows** the contours of the underlying data, so for our analysis we will use a parametric test for correlation: Pearson's r

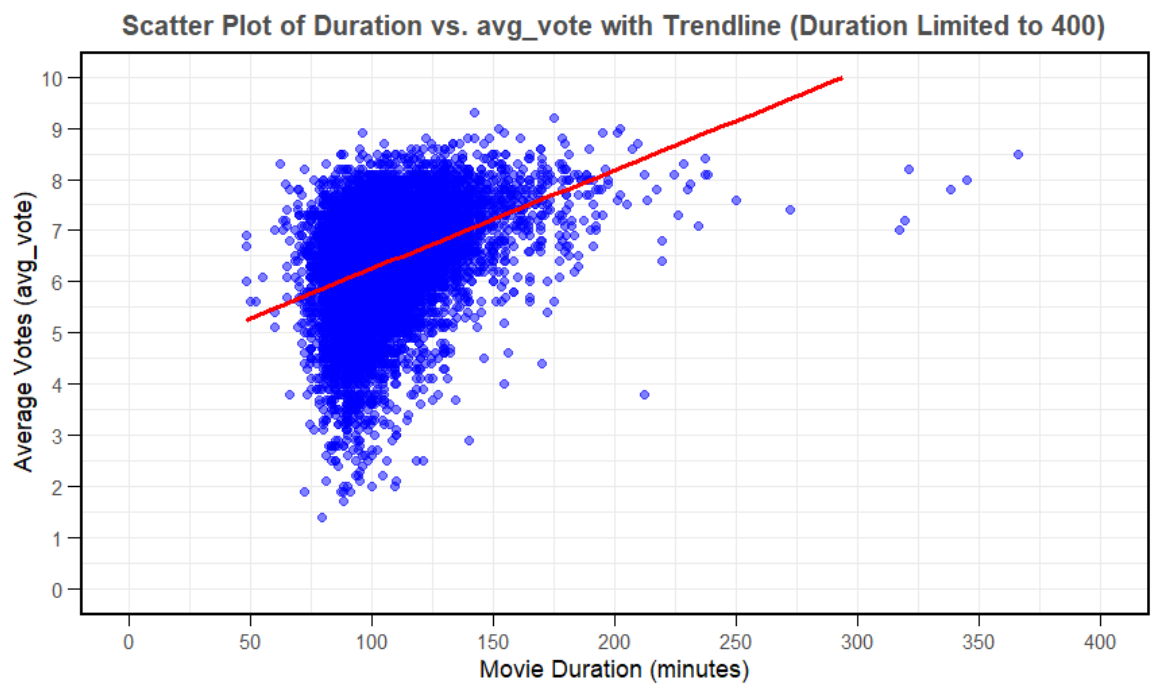
OR

The normal curve overlay **does not follow** the shape of the underlying data, so for our analysis we use the non-parametric test for correlation that does not assume normality: Spearman's Rho or Kendal's Tau

The example here is borderline, in terms of shape, so when in doubt choose the non-parametric equivalent.

Here is the **Scatter** Plot of our independent (duration) vs. dependent (average_vote) variables with trendline.

Our RQ asks about Correlation



Choose one:

1. The blue normal curve **overlay follows** the contours of the underlying data, so for our analysis we will use a parametric test for correlation: Pearson's r

OR

The normal curve overlay **does not follow** the shape of the underlying data, so for our analysis we use the non-parametric test for correlation that does not assume normality: Spearman's Rho or Kendal's Tau

The example here is borderline, in terms of shape, so when in doubt choose the non-parametric equivalent.

University of Hertfordshire UH