Why Test Data Statistically and Mathematically?

Statistical and mathematical testing provides an objective, data-driven framework for analyzing trends and relationships, especially in research involving large or complex datasets. Here's why these tests are essential for our research:

1. **Objective Evaluation**:
   - Human intuition can be biased, especially when dealing with large datasets. Statistical tests remove subjectivity by providing a quantitative, evidence-based way to evaluate data.
   - In our study, for example, we might **think** that romance films score lower on the Bechdel Test compared to other genres, but statistical tests like the **Chi-Square Test** are able to provide **proof** based on the data.

2. **Detecting Patterns and Trends**:
   - Data may not always reveal obvious patterns. Statistical tests (such as regression analysis) help uncover trends over time or relationships between variables that aren't immediately apparent.
   - For our research, these tools can show whether gender representation has improved in cinema over time (**RQ2**), or whether certain genres or countries perform better on the Bechdel Test (**RQ1**).

3. **Quantifying Relationships**:
   - Relationships between variables often exist, but they need to be quantified. For example, **Spearman's correlation** measures the strength of the relationship between **time** and **Bechdel scores**, and **Cramér's V** quantifies the strength of association between **genres** and **scores**.
   - This provides clarity and numerical insight, avoiding vague conclusions.

4. **Testing Hypotheses**:
   - Statistical testing allows for the rigorous **testing of hypotheses**. We can hypothesize that romance films perform worse on the Bechdel Test or that gender representation has improved over time. Tests like **Chi-Square** and **ANOVA** either support or reject these hypotheses with objective evidence.
   - These tests provide **p-values**, indicating whether our results are significant or likely due to random chance, adding credibility to our conclusions.

5. **Ensuring Accuracy**:
   - Drawing conclusions from descriptive statistics alone (like averages or counts) can lead to inaccuracies. Statistical tests account for variation and randomness, ensuring our conclusions are reliable.
   - For example, when comparing multiple genres or countries, statistical tests ensure that differences in Bechdel scores reflect true differences, not random variations in the sample.

6. **Providing Confidence in Results**:
   - Statistical tests provide measures like **confidence intervals** and **p-values**, which quantify how certain wecan be about our results.
   - This adds reliability to our findings, allowing us to make claims like, “There's a 95% probability that the observed trend in Bechdel scores over time is not random.”

7. **Supporting Deeper Analysis**:
   - Beyond simple group comparisons, advanced techniques like **logistic regression** allow us to model the effects of multiple variables (e.g., **genre**, **year**, or **country**) on Bechdel Test scores.
   - For instance, logistic regression can predict the likelihood of a film scoring higher on the Bechdel Test based on its year of release, providing more nuanced insights.

8. **Identifying Statistical Significance**:
   - Even if we observe differences in our data, it's important to know whether they're **statistically significant**. Tests like **ANOVA** and **Chi-Square** confirm whether the differences or trends are meaningful or just due to random variation.
   - This ensures that our findings, such as the differences in Bechdel Test performance across genres or countries, are real and substantial, not just random fluctuations.

Statistical Tests That Will be Used

Here’s a breakdown of the key statistical tests We are applying in our research, how to perform them, what they receive as input, how they contribute to answering our research questions, and the code snippets to test them:

---

1. Chi-Square Test for Independence (or Ordinal Chi-Square)

- **Purpose**: To determine whether there’s a significant association between two categorical variables (e.g., genre/country and Bechdel Test scores).
- **Receives**:
  - **Contingency table**: Rows representing categories like **genre** (or country) and columns representing the **Bechdel scores** (0-3).
- **Outputs**:
  - **Chi-square statistic** and **p-value**, determining if there's a significant relationship.
- **How it helps**: For **RQ1**, it tests if genres or countries differ significantly in their performance on the Bechdel Test.

---

2. Cramér’s V

- **Purpose**: Measures the strength of association from the Chi-Square test.
- **Receives**:
  - The **Chi-square statistic** and the **contingency table**.
- **Outputs**:

- **Cramér's V** (0-1), with values closer to 1 indicating a stronger association.
- **How it helps**: Quantifies how strong the relationship is between genres/countries and Bechdel Test scores.

---

3. Ordinal Logistic Regression

- **Purpose**: Models the relationship between an ordinal outcome (Bechdel score 0-3) and predictors (e.g., year, genre).
- **Receives**:
  - **Ordinal dependent variable** (Bechdel score) and **independent variables** (year, genre).
- **Outputs**:
  - **Coefficients**, **p-values**, and **predicted probabilities** for the likelihood of achieving a higher Bechdel score.
- **How it helps**: For **RQ2**, shows how time (or other factors) influences the likelihood of films scoring higher on the Bechdel Test.

---

4. Kruskal-Wallis Test

- **Purpose**: Compares distributions of ordinal variables across multiple groups (e.g., genres or countries).
- **Receives**:
  - Groups (genres/countries) and ordinal data (Bechdel scores).
- **Outputs**:
  - **H-statistic** and **p-value**, showing if the distributions are significantly different.
- **How it helps**: For **RQ1**, it shows if genres or countries have significantly different Bechdel scores without assuming normal distribution.

---

5. Spearman's Rank Correlation

- **Purpose**: Measures the monotonic relationship between two variables (e.g., year and Bechdel scores).
- **Receives**:
  - **Two variables**: An ordinal variable (Bechdel scores) and a continuous or ordinal variable (e.g., year).
- **Outputs**:
  - **Correlation coefficient (ρ)** and **p-value** to show the strength and significance of the correlation.
- **How it helps**: For **RQ2**, it shows whether Bechdel Test scores improve over time.

---

6. ANOVA (Analysis of Variance)

- **Purpose**: Compares mean values of the dependent variable across multiple groups.
- **Receives**:
  - The **dependent variable** (Bechdel scores) and **independent groups** (e.g., genres, countries, or decades).
- **Outputs**:
  - **F-statistic** and **p-value** indicating whether group means differ significantly.
- **How it helps**: For **RQ1** or **RQ2**, ANOVA can compare Bechdel Test performance across decades or genres.

---

7. Post-hoc Analysis (Tukey's HSD)

- **Purpose**: Determines which specific groups differ significantly after a significant ANOVA result.
- **Receives**:
  - Dependent and group variables (like ANOVA).
- **Outputs**:
  - Pairwise **p-values**, indicating which groups differ.
- **How it helps**: Provides deeper insights after ANOVA into **which** genres or decades differ significantly.

By applying these statistical and mathematical methods, we add rigor, precision, and credibility to our research findings. These tests allow us to support or refute our hypotheses with clear, objective evidence, ensuring that our conclusions about gender representation in cinema are reliable and scientifically valid.