Egypt University of Informatics - Data Analysis Course

Project Title

Genes or the Diagnosis Stage: Who Controls the Cancer Clock?

Submitted by

Mira Micheal (23-101038), Mennatallah Amr (23-101022), Maya Shalash (23-101004), Zeina Shaalan (23-101012)

Introduction

Every cancer story is a race against time. But what matters more in that race - your genetic risk, or how early the cancer is caught?

This project investigates that exact question. Using a dataset of global cancer patients from 2015-2024, we stripped away distractions and focused on two factors:

- Genetic Risk Score (0-10): A person's hereditary susceptibility to cancer.
- Cancer Stage at Diagnosis (0 to IV): How advanced the disease was when first discovered.

By analyzing their impact on survival time and cancer severity, we aim to answer:
What truly shapes a cancer patient's outcome: their DNA, or the stage at diagnosis?

Research Question

Which has a stronger influence on survival and cancer severity - your genetic predisposition, or the stage at which you're diagnosed?

Data Used

Feature Descriptions:

- Genetic_Risk: Score from 0 to 10 estimating hereditary cancer risk
- Cancer_Stage: Stage of cancer at diagnosis (0 to IV)
- Survival_Years: Years survived post-diagnosis (0 to 10)
- Target_Severity_Score: Composite score from 0.9 to 9.16 reflecting cancer aggressiveness

Egypt University of Informatics - Data Analysis Course

Total patients analyzed: [insert number]

Years: 2015-2024

Visual Analysis

Plot A: Genetic Risk vs. Survival Years

- Finding: Patients with high genetic risk show somewhat lower survival, but the trend is weak.
- Conclusion: Genetics play a role, but do not consistently predict survival.

Plot B: Cancer Stage vs. Survival Years

- Finding: Survival drops sharply from Stage I to Stage IV.
- Conclusion: Stage at diagnosis is a much stronger determinant of survival.

Plot C: Cancer Stage vs. Severity Score

- Finding: Severity increases almost linearly with stage.
- Conclusion: Stage at diagnosis is directly correlated with cancer aggressiveness.

Hypothesis Testing

Null Hypothesis (H0): Genetic risk explains severity and survival as well as diagnosis stage.

Alternative Hypothesis (H): Cancer stage has a stronger effect than genetic risk.

Tests Used:

- Linear Regression
- Pearson Correlation
- ANOVA

Results:

- Genetic risk has weak correlation (r 0.2)
- Cancer stage has strong correlation (r 0.7)
- ANOVA: highly significant ($p < 0.001$)

Conclusion: Cancer stage is a far stronger predictor than genetic risk.

Limitations

- Genetic risk is modeled, not sequenced
- Staging may vary slightly between systems
- Environmental factors were excluded intentionally

Final Conclusion

Genes might whisper your risk, but the cancer stage shouts your future.

Our analysis confirms:

The earlier the diagnosis, the better the outcome - even in patients with high genetic risk.

Cancer stage at diagnosis is the strongest and most reliable predictor of both survival and severity.