**LUNG CANCER SURVIVAL ANALYSIS**

**Project Overview**

This project applies Survival Analysis techniques on the veteran lung cancer dataset to study factors that influence patient survival. I used Kaplan-Meier estimators, Cox proportional hazards models, and interaction effects to explore how treatment type, age, performance score, and tumor cell type affect survival outcomes.

**Methods & Tools**

* R Programming (survival package)
* Kaplan-Meier Curves for group comparisons
* Cox Proportional Hazards Model (unadjusted & with interactions)
* Forest plots for visualizing hazard ratios
* Predicted survival curves for example patients

**Insights from the Analysis**

While working on this project, I tried to understand what factors impact the survival of lung cancer patients in the veteran dataset. Here are some key takeaways from my analysis:

1. **Treatment Effect**
   * The “Test” treatment showed a slightly higher hazard ratio (HR 1.35), but it wasn’t statistically significant (p > 0.1).
   * This suggests that the experimental treatment did not clearly improve survival compared to the standard one.
2. **Karnofsky Performance Score (karno)**
   * This turned out to be the strongest predictor.
   * Patients with higher performance scores had significantly better survival outcomes (p < 0.001).
   * The Kaplan-Meier curves by score groups showed a clear separation, meaning functional status at baseline is a very reliable survival indicator.
3. **Tumor Cell Type**
   * Tumor type made a big difference in survival:
     + **Squamous** patients had a median survival of 118 days.
     + **Small Cell** and **Adeno** types both had much poorer survival (median 51 days).
     + **Large cell** patients fell in between, with 156 days median survival.
   * Cox regression confirmed that Adeno and Small Cell cancers had significantly higher hazard ratios (HR > 2–3).
4. **Interaction Between Treatment and Tumor Type**
   * Interestingly, when I included interaction terms, patients with *Small Cell* tumors actually responded differently to the Test treatment, with a hazard ratio of 3.15 (p < 0.05).
   * This suggests that treatment effectiveness may depend on tumor biology, though the sample size was small.
5. **Predicted Example Patient**
   * For a 65-year-old patient with high Karnofsky (80) and Small Cell cancer, the Cox model predicted a median survival of 133 days.
   * The hazard ratio for this patient relative to baseline was about 1.1, indicating slightly worse survival odds.
6. **Overall Model Fit**
   * The Cox model had good discriminatory ability (concordance 0.73), which means it was fairly effective at ranking patients by risk.
   * Forest plots made it easy to visualize which covariates had the largest effect, with Karnofsky and tumor type standing out clearly.

**Takeaway**

This project showed how survival analysis can provide much deeper insights than just comparing means or medians. It helped me practice using **Kaplan-Meier curves, Cox proportional hazards models, and interaction effects** in a real medical dataset. The analysis also highlighted the importance of **clinical factors (performance status and tumor type)** over treatment alone, which is a useful lesson for both data science and healthcare decision-making.