**Thyroid Cancer Risk Analysis Summary:**

**Statistical/Hypothetical Question:**

The aimed of this study is to determine whether radiation exposure and family history significantly increase the risk of thyroid cancer.

* **Null Hypothesis (H₀)**: Radiation exposure and family history do not significantly impact thyroid cancer risk.
* **Alternative Hypothesis (H₁)**: Radiation exposure and family history increase the risk of thyroid cancer.

**Outcome of Exploratory Data Analysis (EDA)**

1. **Descriptive Statistics & Outliers**
   * Most patients were between 30-70 years old, with outliers in younger (15-20) and older (80-90) age groups.
   * Nodule size between 1-4 cm was common, with larger nodules (>5 cm) being potential risk factors.
   * Radiation Exposure was rare (~15%), while 30% had a family history of thyroid cancer.
   * 77% of cases were benign, while 23% were malignant.
2. **Correlation & Distribution Analysis**
   * Pearson Correlation Analysis showed that family history (r = 0.14) and radiation exposure (r = 0.089) were weakly correlated with diagnosis.
   * Probability Mass Function (PMF) indicated that larger nodules (3-5 cm) were more common in malignant cases.
   * Cumulative Distribution Function (CDF) revealed 75% of nodules were smaller than 3.76 cm, with larger ones being higher risk.
3. **Hypothesis Testing (Chi-Square Test)**
   * Chi-square test confirmed a significant association (p < 0.05) between Radiation Exposure and Diagnosis, leading to the rejection of the null hypothesis.
4. **Regression Analysis (Logistic Regression)**
   * Radiation Exposure and Family History were the strongest predictors (p < 0.05).
   * Age and Nodule Size were NOT significant predictors.
   * Pseudo R² = 0.0246, indicating more variables are needed for better prediction.

**What Was Missed in the Analysis?**

* There is Lack of additional environmental & lifestyle factors like smoking, obesity, diet, and hormone levels (TSH, T3, T4).
* There is no consideration for ethnicity or geographic factors, which could influence risk.
* Interaction effects between multiple risk factors (e.g., radiation exposure & genetic predisposition) were not analyzed.

**Were There Any Missing Variables?**

Yes. Variables that could have helped are:

* Iodine Deficiency (Known risk factor for thyroid disorders)
* TSH, T3, T4 Levels (Hormonal influence on thyroid function)

**Incorrect Assumptions & Challenges Faced**

* Assumption: Nodule Size would be a strong predictor, but it was not statistically significant in regression analysis.
* Challenge: Understanding why Pseudo R² was very low (0.0246), suggesting missing predictive factors.
* Complexity: Interpreting interactions between radiation exposure, genetics, and hormonal levels was difficult without additional variables.

**Final Conclusion**

1. Radiation Exposure significantly increases thyroid cancer risk.
2. Family History is a strong genetic predictor.
3. Age and Nodule Size do not significantly impact cancer risk.
4. Future studies should explore additional genetic, lifestyle, and environmental factors for improved prediction models.

By: Roshan GC

DSC530 Term Project

Bellevue University