**Is the cancer type related to sex? Does it depend on the type of cancer?**

1. A graph of male and female

   Description automatically generatedA table with numbers and labels

   Description automatically generatedPerformed a chi-squared test on the entire data set, this uncovered a strong association between the type of cancer and sex (p-value = 0.01317). However, this is fairly obvious given some cancer types within the data set are exclusive to one sex - prostate, testicle, ovary, uterus, breast (whilst breast cancer can occur in men, 1% of total USA breast cancer patients are men - [Male breast cancer - Symptoms and causes - Mayo Clinic](https://www.mayoclinic.org/diseases-conditions/male-breast-cancer/symptoms-causes/syc-20374740)).
2. Filtered out those data types that are known to be exclusive to one sex, then conducted another chi-squared test on the remaining cancer types. The association disappears (p-value = 0.6277).
3. A table with numbers and a number of labels

   Description automatically generated with medium confidenceAs the chi-squared test shows no significant associations, calculated the proportions of each type of cancer for each sex. Whilst this seems to show there is association between sex and some types of cancer, notice how all the types showing a extreme proportion have very small totals – hence their proportions are skewed.
4. Used a logistic regression to investigation whether the filtered types of cancer can be used as a categorical predictor to predict gender. The logistic regression model helps determine whether the type of cancer significantly influence the probability of being male or female.

***Notes on logistic regression:*** *a type of regression analysis used when the dependent variable (outcome) is in binary form (can be used in this context as only two possible outcomes – male or female). This helps us to investigate the relationship between these variables. The model estimates the probability that a given observation belongs to a category based on the value of the predictors.*

However, model provides no useful insights.

1. Ultimately, the dataset is flawed and not useful. Once you filter out the known single-sex exclusive cancer types, there are 16 different types of cancer and 80 total data points. Meaning that it is very difficult to make conclusions for each specific cancer type, given the data set for each is so small and easily skewed by one data point.
2. So, as a result, cancer type is related for sex for some types of cancer. However this is fairly trivial. And, given this data set alone, there is not enough evidence to say that there is a significant association between the type of cancer and sex.