

# Homework 1

*Aji John*

Q1. For the scientific question of interest, what are the two primary variables in this study? What is the response variable and what is the predictor variable? What types of variables are they: quantitative or qualitative? Discrete or continuous? Nominal, ordinal, interval, or ratios? Censored (right, left, or both) or uncensored? Explain briefly.

Answer. The two primary variables are 'atrophy', and 'age'. Response variable is 'atrophy', and predictor variable is 'age'. Both the variables are quantitative, 'Age' and 'Atrophy' are discrete.

Q2. What is the population of interest for this study? What is the sample? What is the size of the sample? Are there any individuals in the sample who have missing data on smoking history? If so, provide the participant identification numbers for any study individuals who have missing data on smoking history.

Answer. Adults aged 65 years and older is the population of interest. Sample is 735 individuals randomly selected from medicare rolls. Yes, one individual has missing smoking history, and the id is 545.

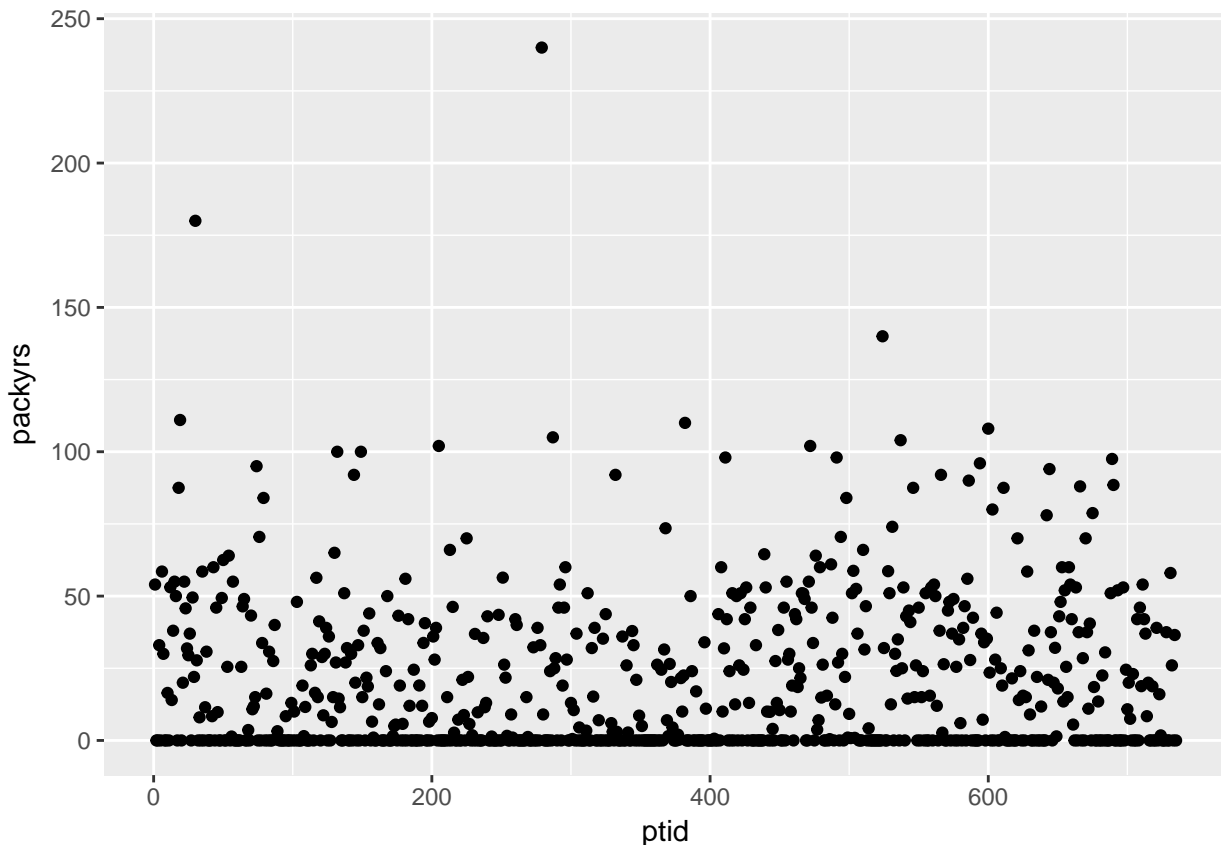
Q3. Problems 4 – 7 ask you to dichotomize the time to death according to death within 5 years of study enrollment or death after 5 years. Why is this valid? Provide descriptive statistics that support your answer.

Answer. Approx. 18% of the individuals who were part of the study were classified as observed to die while on study. Analysis has to be partitioned as 18% of the samples would have incomplete observations.

Death	Groups
0	602
1	133

Q4. Provide a suitable descriptive statistical analysis for selected variables in this dataset as might be presented in a table to be included in a manuscript exploring the association between smoking history in pack years and 5 year all-cause mortality in the medical literature. In addition to the two variables of primary interest, you may restrict attention to age, sex, weight, and prior history of cardiovascular disease, e.g., coronary heart disease (CHD), congestive heart failure (CHF), and stroke.

Answer. We are studying the association between overall mortality and smoking history in pack years.



Min max & Std Dev is required Smoking history.

```

+-----+-----+-----+-----+-----+
| 0 | <= 10 | 11-25 | 26-50 | greater than 50
| (n=321) | (n=73) | (n=99) | (n=148) | (n=93) +=====+=====+=====+=====+
| Age | 75.36(66-99) | - built- r | | | | | | | +-----+-----+-----+-----+-----+
| Sex | $2.10 | - cures | | | | | - tasty | | | +-----+-----+-----+-----+-----+
| Weight | $2.10 | - cures | | | | | - tasty | | | +-----+-----+-----+-----+-----+
| CHF | $2.10 | - cures | | | | | - tasty | | | +-----+-----+-----+-----+-----+
| CHD | $2.10 | - cures | | | | | - tasty | | | +-----+-----+-----+-----+-----+
| Stroke | $2.10 | - cures | | | | | - tasty | | | +-----+-----+-----+-----+-----+
| Death | | - tasty | | |
| percent | | | |
| | - tasty | | | +-----+-----+-----+-----+-----+ * one row excluded for
missing smoking record

```

5.Perform a statistical analysis evaluating an association between smoking history in pack years and 5-year all-cause mortality by comparing mean pack years across groups defined by vital status at 5 years.

Answer. We do a 2 sample t-test, For

Death by 5(0) - get the average pack years Alive by 5(1) - get the average pack years

\*Groups - 0 (thru the study) 1(not thru the study)

As p-value is 0.004 (using critical value of 0.05), We reject the null hypothesis that survivability of individuals based on being a smoker/non-smoker would be the same. 95% Confidence interval, the difference in mean pack years between the survivability groups is 2.89 to 15.666566. The best estimate of the difference, i.e. the point estimate is 9.28245 units, not that unusual to see. In summary, with our analysis we state that pack years has influence on mortality and is found to be significant.

6. Perform a statistical analysis evaluating an association between smoking history and 5-year all-cause mortality by comparing the probability of death within 5 years across groups defined by whether or not study participants have ever smoked. Note that a participant who never smoked has 0 pack years, as is indicated in the documentation file.

Answer. We do a chi-test,

2 by 2 table

Non smokers , and Smokers

7. Perform a statistical analysis evaluating an association between smoking history and 5-year all-cause mortality by comparing the odds of death within 5 years across groups defined by whether or not the study participants ever smoked. #apriori - #fischers exact t-test