

Problem 7

/25

This question refers to the `gss` data set, available on Moodle. Use **R** for all calculations. Include all code used to generate answers.

- (a) Which of the following variables are quantitative and which are categorical? What is the level of measurement of each?

- year (explain in just a sentence or two)

Year is quantitative interval data. It is numeric but it doesn't make sense to use ratios and zero is arbitrary.

- marital (no explanation needed)

Marital status is categorical nominal data.

- age (no explanation needed)

Age is quantitative ratio data.

- (b) Find the mean number of hours of television watched per day. Also determine the variance of this variable. Recall that you can force *R* to ignore missing values by adding the argument `na.rm = TRUE` to most functions.

`mean(gss$tvhours, na.rm=TRUE) = 2.98` hours

`var(gss$tvhours, na.rm=TRUE) = 6.71`

- (c) Find the standard deviation of hours of television watched per day. Briefly interpret your answer.

`sd(gss$tvhours, na.rm=TRUE) = 2.59`

The average tv hours in this sample is 2.98 but within that average the actual hours generally vary by up to ± 2.59 hours.

- (d) Determine the five number summary and IQR of the ages of respondents to this survey.

<code>quantile(gss\$age) =</code>	0%	25%	50%	75%	100%
	18	33	46	59	89
IQR = 59 - 33 = 26	Min	Q ₁	M	Q ₃	Max

- (e) What is the 90th percentile of ages in this survey? Briefly but clearly explain what this means.

`quantile(gss$age, 0.9) = 72`

Someone that is 72 is older than 90% of the ages in this sample

