

## STAT 500 - Homework 1

Due in class September 17

The dataset `teengamb` concerns a study of teenage gambling in Britain. More details about the study can be found in Ide-Smith & Lea (1988) *Journal of Gambling Behavior*, 4, 110-118. It contains five variables:

- **sex**: 0 = male, 1 = female
- **status**: socioeconomic status score based on parents' occupation
- **income**: in pounds per week
- **verbal**: verbal score in words out of 12 correctly defined
- **gamble**: expenditure on gambling in pounds per year

Make a numerical and graphical summary of the data, commenting on any features that you find interesting. Limit the output you present to a quantity that a busy reader would find sufficient to get a basic understanding of the data. For the variable “sex”, assign labels “male” and “female” and make sure R is treating it as a categorical variable before you compute the summary.

Also answer the following questions, **incorporating the answers into your report**:

1. In your summary, report the means and medians of variables “income” and “gamble”. Explain why the means are larger than the medians.
2. How many different values are there for the variable “verbal”? Based on the boxplot and/or other summaries of the variable “verbal”, what values of the verbal score would you consider to be outliers? Give the row numbers of the outlier observations.

**Hints:** After you install R (if it is not already installed on the system you are using) and the `faraway` package, you can type `library(faraway)` to load the package and `data(teengamb)` to load the data. Useful R functions for this homework: `data()`, `summary()`, `hist()`, `plot()`, `boxplot()`. You can always type `help(subject)` to get detailed help on the `subject`, e.g. `help(plot)`. Or you can type `help.start()` to get interactive help with a search engine.

**Solutions to this homework should be no longer than 3 pages.**