

WRITING ASSIGNMENT: HEY BUDDY, WANT TO BUY AN AIRPLANE? (30 POINTS)

Over the past few decades, US. production of small light aircraft has declined dramatically; production of new piston-powered single-engine aircraft dropped 91% from 1979 to 1989. As a result, there is an active market for used aircraft of this category. The purpose of this assignment is to use data on this market to develop a model that can predict and explain the asking price of a particular model of aircraft: the Warrior, manufactured by Piper Aircraft.

The Piper Warrior is a four-passenger airplane with either a 150- or 160-horsepower engine. Introduced in the mid 1970's, it is a modification of the earlier Piper Cherokee.

What factors affect the price of a used airplane? Some factors might include the age of the airframe (i.e., the aircraft body, excluding instruments and engine), the number of hours the aircraft has been flown (a measure of wear on the engine, among other things), and the time since the engine has been overhauled or replaced. Older airplanes may require a new paint job; for an airplane this can cost several thousand dollars.

A Warrior can be equipped with a wide range of features that affects its price. These include radios and electronic navigational and piloting instruments, also called avionics. In new aircraft, the difference between airplanes equipped with basic avionics compared to one fully equipped with the latest high tech gadgets can easily exceed the cost of the aircraft itself. Some particular classes of avionics that could represent considerable investments over the basic package are abbreviated IFR (instruments that permit navigation according to instrument flight rules), DME (distance measuring equipment), LORAN (long-range navigation based on satellite communication), and the ModeC transponder (which sends altitude and position indications to air traffic controllers).

The data here were transcribed from an issue of *Trade-A-Plane*, a national publication advertising used aircraft for sale, in the early 1990's (a typical page is included). The variables recorded include the model YEAR of aircraft, TT (total flight time in hours), SMOH (hours since major overhaul), IFR, DME, LORAN, ModeC, HP (engine horsepower), PAINT (new or recent paint job), and PRICE (these are the abbreviations used by *Trade-A-Plane*). Not all information may be available for all aircraft. The variables IFR, DME, LORAN, ModeC, and PAINT are indicator variables that indicate whether the corresponding item was mentioned as being present in the ad. PRICE is given in thousands of dollars.

Your assignment is to use these data to develop a model that explains the price of used Piper Warriors. In addition to the final model, your paper should describe how arrived at this model, and what strengths and weaknesses the model may have.

The dataset is on my web site in the usual place. As usual, you should integrate plots and tables into your text, and you don't need to hand in every plot or table you make — just the ones that show some interesting feature or support some point you are making. Your written report should be roughly 6–10 *typed* pages, including plots and tables. This is meant to give you an idea of the scope of the assignment, rather than to set minimum or maximum limits on your work. What's important is the completeness and appropriateness of your analysis, rather than its length.

PROBLEMS (PLEASE TYPE AND HAND IN SEPARATELY)

1. RESIDUAL PLOTS (6 POINTS)

- (a) Sketch a picture of a residual plot that shows non-constant variance and linearity.
- (b) Sketch a picture of a residual plot that shows constant variance and nonlinearity.
- (c) Sketch a picture of a residual plot that shows non-constant variance and nonlinearity.

2. SHORT ANSWER (8 POINTS)

For each of the following statements, indicate whether the statement is true or false, and explain your answer briefly.

- (a) A predictor X_3 can be statistically significant in a multiple regression even though the correlation of Y and X_3 (by itself) is 0.
- (b) A predictor X_3 can be statistically *not* significant in a multiple regression even though the correlation of Y and X_3 (by itself) is .8.
- (c) If your dataset has an outlier, transforming Y to $\log(Y)$ can reduce the leverage of this point.
- (d) If your dataset has an outlier, transforming Y to $\log(Y)$ can reduce the Cook's distance of this point.

3. F AND T (6 POINTS)

In a simple linear regression, the F -statistic (F-ratio) tests the null hypothesis $H_0: \beta_1 = 0$ and is equal to $MS_{\text{Regression}} / MS_{\text{Residual}}$. The t -statistic tests the null hypothesis $H_0: \beta_1 = 0$ and is equal to $b_1 / SD(b_1)$.

Because in a simple linear regression these two statistics are testing the same hypothesis, there should be a correspondence between the two. Demonstrate this correspondence by showing that the F -statistic is equal to the t -statistic squared. (Hint: Use $b_1 = r \frac{s_y}{s_x}$ and use the definition of R^2 .)