

Lab 13 *Some of these problems may be more challenging than others. Please feel free to work with others, attend office hours, or post on the course discussion forum if you need help. While collaboration with other students is encouraged, each student is responsible for submitting his or her own work. This assignment should be submitted in one well-commented SAS program. For any questions that require a written answer, do so in the SAS comments. Be sure to re-name the uploaded SAS scripts according to the naming convention `LastnameFirstinitial_Lab#.sas` (e.g., `PileggiS_Lab13.sas`).*

This data set (`mariokart.sas7bdat`) includes all auctions on Ebay for a full week in October, 2009. Auctions were included in the data set if they satisfied a number of conditions. (1) They were included in a search for "wii mario kart" on ebay.com, (2) items were in the Video Games > Games > Nintendo Wii section of Ebay, (3) the listing was an auction and not exclusively a "Buy it Now" listing (sellers sometimes offer an optional higher price for a buyer to end bidding and win the auction immediately, which is an optional Buy it Now auction), (4) the item listed was the actual game, (5) the item was being sold from the US, (6) the item had at least one bidder, (7) there were no other items included in the auction with the exception of racing wheels, either generic or brand-name being acceptable, and (8) the auction did not end with a Buy It Now option. All prices are in US dollars. Our goal for this lab to create models for the total selling price of the Ebay package (`totalPr`).

<code>ID</code>	Auction ID assigned by Ebay.
<code>duration</code>	Auction length, in days.
<code>nBids</code>	Number of bids.
<code>cond</code>	Game condition, either new or used.
<code>startPr</code>	Starting price of the auction.
<code>shipPr</code>	Shipping price.
<code>totalPr</code>	Total price, which equals the auction price plus the shipping price.
<code>shipSp</code>	Shipping speed or method.
<code>sellerRate</code>	The seller's rating on Ebay (number of positive ratings minus the number of negative ratings).
<code>stockPhoto</code>	Whether or not the auction feature photo was a "stock" photo.
<code>wheels</code>	Number of Wii wheels included in the auction.
<code>title</code>	The title of the auctions.

1. Begin by exploring the data set. There are two observations that really don't fit the pattern of the rest with regards to total selling price. Identify these observations and explain why they are outliers in a comment in your SAS code. (*Hint: Examine the title for these observations.*)
2. Create a temporary data set that is a copy of the original marioKart data set. In this temporary data set, remove the outliers that you identified in question 1. Use this data set for the remaining exercises.
3. Continue working with your temporary data set. Create two new indicator variables coded as 0/1:

- (a) `condI` is an indicator for condition of the game (`cond`), where 1=new and 0=used
- (b) `photoI` is an indicator for `stockPhoto`, where 1=yes and 0=no

Write some code to verify that both of these variables were created correctly (as discussed in Lecture 9).

4. Continue working with your temporary data set. Note that marioKart packages can come with 0 through 4 wheels, but only a couple of packages come with 3 or 4 wheels. Create a new variable called `wheelsnew` that is coded as 0 wheels, 1 wheel, or 2+ wheels. Write some code to verify that this variable was created correctly (as discussed in Lecture 9).
5. Examine the relationship between total selling price (`totalPr`) and number of wheels (`wheelsnew`).
 - (a) We want to use analysis of variance to determine if population average total selling price (`totalPr`) differs by the number of wheels a package comes with (`wheelsnew`). Do we have unbalanced data for this analysis? Can we use PROC ANOVA? Answer these questions in a comment in your SAS code.
 - (b) Use analysis of variance to determine if population average total selling price (`totalPr`) differs by the number of wheels a package comes with (`wheelsnew`) at the 0.05 significance level. In a comment in your SAS code, briefly describe your findings.
 - (c) If appropriate, perform pairwise comparisons to determine which population means differ based on the method of your choice. In a comment in your SAS code, briefly describe your findings.
 - (d) Are conditions satisfied for analysis of variance? Perform any SAS procedures necessary to investigate this and comment on your findings in your SAS code.
6. Now let's use regression to examine the relationship between total selling price (`totalPr`) and other variables.
 - (a) Use PROC CORR to examine correlations and produce scatter plots between `totalPr` and the following quantitative variables in the data set. (Quantitative variables include: `duration`, `nBids`, `startPr`, `shipPr`, and `SellerRate`.)
 - (b) From your results in the part (a), which variable has the strongest linear association with `totalPr`? Briefly describe this association in a comment in your SAS code.
 - (c) Estimate the regression model between `totalPr` and the independent variable you selected in part (b). Write the estimated regression equation in a comment in your SAS code.
 - (d) Are conditions satisfied for your regression model in (c)? Perform any SAS procedures necessary to investigate this. In a comment in your SAS code state

what you did to check the conditions and state your findings.

- (e) Use the model statement below in **PROC REG** and substitute in the model selection method of your choice to identify a “best” model for **totalPr**. Identify the final variables in your model in a comment in your SAS code.

_____ SAS Code _____

```
MODEL totalpr = duration nBids startPr shipPr  
                SellerRate wheels condI photoI  
                / SELECTION =    ;
```

_____ SAS Code _____