**Lab 3**  *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_Lab3.sas`*).*

The `mariokart` data set 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`).

`mariokart.sas7bdat`:

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

1. Identify the `mariokart.sas7bdat` file from either PolyLearn or the shared drive. Save the file to a location on your computer or your flash drive.

2. Create a library reference called `mylib` to access the `mariokart.sas7bdat`.

3. Explore the *total selling price* variable with any of PROCS that we have learned so far (`PROC CONTENTS`, `PROC PRINT`, `PROC MEANS`, `PROC FREQ`, `PROC UNIVARIATE`). There are a 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. (In future classes we will learn how to "clean" these observations, but for now we will leave them as is.)

4. An eBay official claims that the population mean number of bids received exceeds 10. Test the hypotheses $H_0$: $\mu = 12$ vs $H_a$: $\mu \neq 12$ and compute a 95% confidence interval for $\mu$ using PROC UNIVARIATE. Fill in the blanks in the paragraph below to interpret the results.

Based on (insert #) packages for sale on eBay, the average number of bids a package received is _____ with a standard deviation of _____. We used a one sample t-test to test $H_0$: $\mu = 12$ vs $H_a$: $\mu \neq 12$, where $\mu$ represents the population average number of bids. The test statistic is $t = $ _____ and the p-value is _____. We (do / do not) have evidence that the population mean number of bids differs from 12. Furthermore, a 95% confidence interval for the population mean number of bids is (_____, _____). This (does / does not) provide evidence that the population mean number of bids exceeds 12.

5. Replicate your summary statistics and confidence interval findings (not the hypothesis test) from the previous question using PROC MEANS. Below is the output that you are trying to achieve. *Hint: you need to use statistics key words.*

**Question 5 Output**

**The MEANS Procedure**

| | | | | |
|---|---|---|---|---|
| | | | **Lower 95%** | **Upper 95%** |
| **N** | **Mean** | **Std Dev** | **CL for Mean** | **CL for Mean** |
| 143 | 13.5384615 | 5.8787864 | 12.5666438 | 14.5102793 |

Analysis Variable : nBids

6. Use a SAS procedure to identify the percent of packages that were sold with 2 wheels. Note your findings as a comment in your SAS code.

7. Examine the relationship between condition of the package and number of wheels by creating a contingency table with wheels on the rows and condition on the columns. In a comment in your SAS code, report the percent of new packages that came with zero wheels and the percent of used packages that came with zero wheels. Modify your table so that only frequencies and these percents are presented. *(Your table will still be 4x2, with total rows and columns, but each cell should only have two numbers:*

*a frequency and the relevant percent to compare new versus used packages.)*

8. Perform a chi-squared test to determine if there is an association between condition of the package and whether or not the package has a stock photo; in addition, print the expected cell counts for the table. (*Hint: Use the help file, this is an option on the TABLES statement.*) The contingency table should only display frequencies and expected cell counts, as shown in the output below. Fill in the blanks in the following paragraph to interpret the results.

We used a chi-squared test to assess $H_0$: there (is / is not) an association be-

tween condition and stock photo versus $H_a$: there (is / is not) an association

between condition and stock photo. The chi-squared test statistic is $\chi^2 =$

_____and the $p$-value is _____. We (do / do not) have evidence of an associa-

tion between condition of the package and stock photo. Furthermore, *(insert #)*

expected cell counts exceed 5, so the expected cell count condition for the chi-

squared test (is / is not) satisfied.

### Question 8 Output

**The FREQ Procedure**

| Frequency Expected | Table of cond by stockPhoto | | |
|---|---|---|---|
| | | stockPhoto | |
| | **cond** | **no** | **yes** | **Total** |
| | **new** | 4<br>15.678 | 55<br>43.322 | 59 |
| | **used** | 34<br>22.322 | 50<br>61.678 | 84 |
| | **Total** | 38 | 105 | 143 |

**Statistics for Table of cond by stockPhoto**

| Statistic | DF | Value | Prob |
|---|---|---|---|
| Chi-Square | 1 | 20.1681 | <.0001 |
| Likelihood Ratio Chi-Square | 1 | 22.9505 | <.0001 |
| Continuity Adj. Chi-Square | 1 | 18.4781 | <.0001 |
| Mantel-Haenszel Chi-Square | 1 | 20.0271 | <.0001 |
| Phi Coefficient | | -0.3755 | |
| Contingency Coefficient | | 0.3516 | |
| Cramer's V | | -0.3755 | |

| Fisher's Exact Test | |
|---|---|
| Cell (1,1) Frequency (F) | 4 |
| Left-sided Pr <= F | <.0001 |
| Right-sided Pr >= F | 1.0000 |
| | |
| Table Probability (P) | <.0001 |
| Two-sided Pr <= P | <.0001 |

**Sample Size = 143**