University of Central Florida
College of Business

QMB 6912
Capstone Project in Business Analytics

Problem Set #4

1 Data Description

This analysis follows the script `Tractor_Tables.R`. The script tabulates various statistics from the variables that may be used to estimate a model for used tractor prices. It uses the data from `TRACTOR7.csv` in the `Data` folder. The dataset includes the following variables.

$$
\begin{aligned}
saleprice_i &= \text{the price paid for tractor } i \text{ in dollars} \\
horsepower_i &= \text{the horsepower of tractor } i \\
age_i &= \text{the number of years since tractor } i \text{ was manufactured} \\
enghours_i &= \text{the number of hours of use recorded for tractor } i \\
diesel_i &= \text{an indicator of whether tractor } i \text{ runs on diesel fuel} \\
fwd_i &= \text{an indicator of whether tractor } i \text{ has four-wheel drive} \\
manual_i &= \text{an indicator of whether tractor } i \text{ has a manual transmission} \\
johndeere_i &= \text{an indicator of whether tractor } i \text{ is manufactured by John Deere} \\
cab_i &= \text{an indicator of whether tractor } i \text{ has an enclosed cab} \\
spring_i &= \text{an indicator of whether tractor } i \text{ was sold in April or May} \\
summer_i &= \text{an indicator of whether tractor } i \text{ was sold between June and September} \\
winter_i &= \text{an indicator of whether tractor } i \text{ was sold between December and March}
\end{aligned}
$$

I have downloaded the file `TRACTOR7.csv`, loaded the data described above into R, calculated the summary statistics for these data, and finally, presented these statistics in LaTeX tables. These operations are all performed by the script `Tractor_Tables.R` in the `Code` folder. The script uses an R package called `xtable` to automate the production of the tables from a data frames in R.)

I analyze the data in subsets, according to the makes of tractors, calculating the summary statistics for each subset and present these statistics in the LaTeX tables that follow.

2 Summary by Make of Tractors

Table 1 lists summary statistics for numeric variables in separate rows for subsamples defined by the make of tractors, according to whether they were manufactured by John Deere.

	John Deere	Other
Min. saleprice	2800	1500
Mean saleprice	27973	19557
Max. saleprice	200000	135000
Min. horsepower	16.00	18.0
Mean horsepower	99.13	101.4
Max. horsepower	491.00	535.0
Min. age	2.00	2.00
Mean age	18.03	15.56
Max. age	33.00	33.00
Min. enghours	22	1
Mean enghours	4023	3449
Max. enghours	10734	18744

Table 1: Summary by Make of Tractor

3 Indicator Variables by Make of Tractors

Table 2 lists the frequencies of observations of each brand of tractor across a series of indicator variables, which includes whether a tractor has a diesel engine, four-wheel-drive transmission, a manual transmission, or an enclosed cab to protect the operator.

	John Deere	Other	Totals
Total	39	237	276
Gasoline	9	16	25
Diesel	30	221	251
2WD	19	101	120
4WD	20	136	156
Automatic	9	73	82
Manual	30	164	194
No Cab	25	101	126
Has Cab	14	136	150

Table 2: Indicator Variables by Make of Tractor

4   Season Sold by Make of Tractor

Table 3 lists the frequencies of observations of sales of the brand categories of tractors over the four seasons.

	John Deere	Other	Totals
Fall	14	89	103
Spring	9	53	62
Summer	6	58	64
Winter	10	37	47
Totals	39	237	276

Table 3: Season Sold by Make of Tractor

5   Average Price of Tractors by Season Sold

Table 4 lists the average price of tractors sold over the four seasons.

	John Deere	Other
Fall	35875	16574
Spring	37483	23958
Summer	11908	16062
Winter	17990	25909

Table 4: Average Price of Tractors by Season Sold

6 Correlation Matrices

To plot the correlation between variables, we use the logarithm of prices, which had a better-behaved distribution, according to our results in Problem Set #3. We separate the analysis into numerical and indicator variables

Table 5 shows the correlation between the log. of tractor prices and the numeric variables horsepower, age and engine hours.

	Log. of Price	Horsepower	Age	Engine Hours
Log. of Price	1.000	0.649	-0.441	-0.046
Horsepower	0.649	1.000	0.039	0.378
Age	-0.441	0.039	1.000	0.559
Engine Hours	-0.046	0.378	0.559	1.000

Table 5: Correlation Matrix of Log. Prices and Numeric Variables

Table 6 shows the correlation between the log. of tractor prices and the indicators for whether a tractor has a diesel engine, four-wheel-drive transmission, a manual transmission, or an enclosed cab to protect the operator.

	Log. of Price	Diesel	FWD	Manual	Cab
Log. of Price	1.000	0.331	0.522	0.039	0.588
Diesel	0.331	1.000	0.207	0.347	0.344
FWD	0.522	0.207	1.000	0.006	0.223
Manual	0.039	0.347	0.006	1.000	0.280
Cab	0.588	0.344	0.223	0.280	1.000

Table 6: Correlation Matrix of Log. Prices and Indicator Variables