Question 1:

Model 7 with beta estimates:

$\hat{y}$ = 12.192 − .598x_1 − .00598x_2 − .01078x_1x_2 + .086x_1^2+ .00014x_2^2+ .677x4 − .275x_1x_4 − .026x_2x_4 + .013$x_1x_2x_4$ − .782x_3 + .0399x_1x_3 − .021x_2x_3 − .0033$x_1x_2x_3$

difference between the predicted regulated price and predicted deregulated price for any fixed value of mileage.

regulated: x_3 = 0 $\hat{y}$ = 12.134 – 0.760x_1 + 0.086x_1^2

deregulated: x_3 = 1 $\hat{y}$ = 11.037 – 0.7696x_1 + 0.086x_1^2

assuming mileage is x_1: Difference between regulated and deregulated:

12.134 – 0.760x_1 + 0.086x_1^2 – (11.037 – 0.7696x_1 + 0.086x_1^2)

=1.097 + 0.0096x_1

Explanation: To find the difference between the predicted regulated price and the predicted deregulated price, we first created separate for each of these two events. To do so, we plugged in 1 for x_3 for deregulated, and 0 for x_3 for regulated. In doing so, two distinct functions were created to model in each situation. To find the difference between these two, we simply subtracted the deregulated price from the regulated price. The resulting function serves to predict the difference in regulated vs deregulated price based on a particular mileage.

Question 2:

x_2 = 10

x_4 = 1

x_3=1 if deregulated; x3=0 if regulated

$\hat{y}$ = 12.192 − .598x_1 − .00598x_2 − .01078x_1x_2 + .086x_1^2+ .00014x_2^2+ .677x4 − .275x_1x_4 − .026x_2x_4 + .013$x_1x_2x_4$ − .782x_3 + .0399x_1x_3 − .021x_2x_3 − .0033$x_1x_2x_3$

Regulated (x_3 = 0):

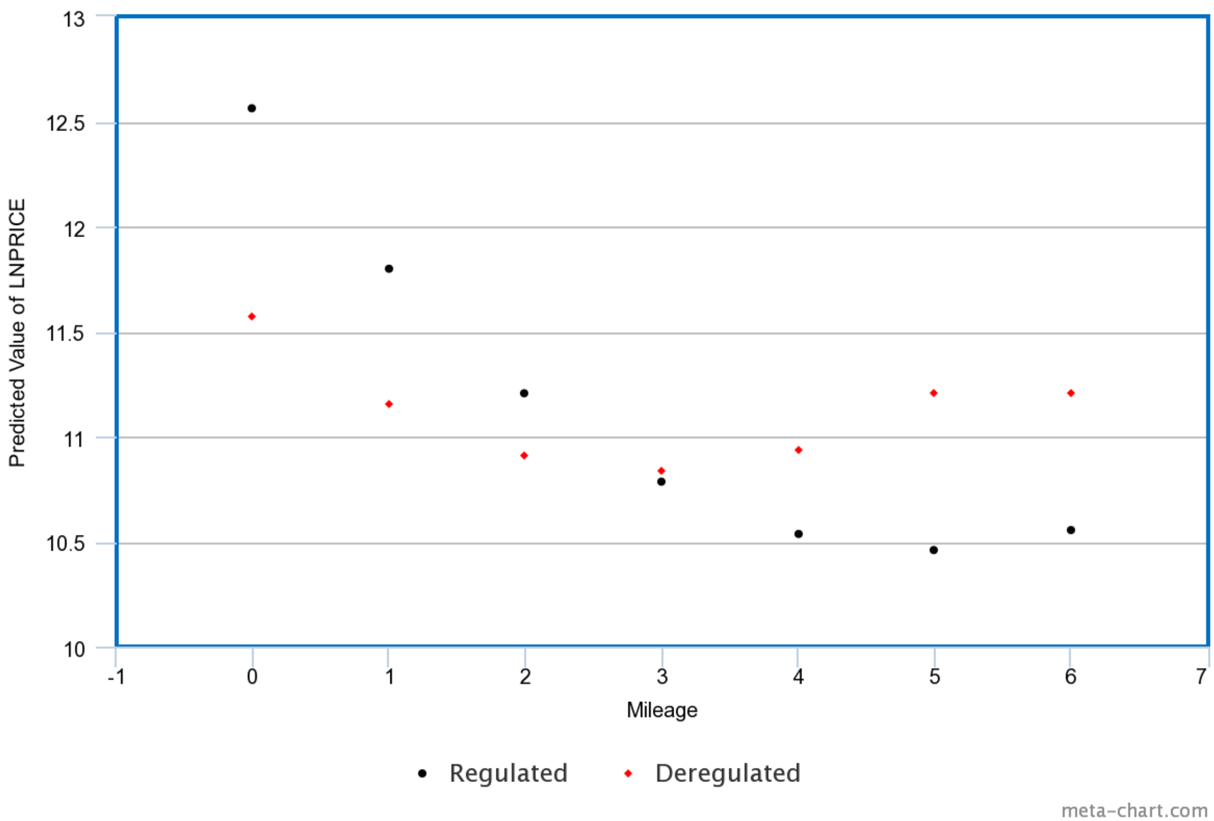
$\hat{y}$ = 12.192 − .598x_1 − .00598(10) − .01078x_1(10) + .086x_1^2+ .00014(10)^2+ .677(1) − .275x_1(1) − .026(10)(1) + .013x_1(10)(1)

= 12.5632 -0.8508x_1 +0.086x_1^2

Deregulated (x_3 = 1):

$\hat{y}$ = 12.192 − .598x_1 − .00598(10) − .01078x_1(10) + .086x_1^2+ .00014(10)^2+ .677(1) − .275x_1(1) − .026(10)(1) + .013x_1(10)(1) − .782(1) + .0399x_1(1) − .021(10)(1) − .0033x_1(10)(1)

= 11.5712 - 0.5029x_1 + 0.086x_1^2



Explanation: To demonstrate the impact of deregulation on price charged when the origin is Miami and the weight is fixed at 10,000 lbs, we first had to plug those constants into model 7. We replaced x_2 with 10 (10,000 lbs), and we replaced x_4 (origin) with the value 1 to indicate that the origin is Miami (x_4 = 1 if originate in Miami, 0 if originate in Jacksonville). To find the impact of deregulation, we then had to input both 0 and 1 for x_3 in the equation (x_3 =1 if deregulation in effect, 0 if not). We then compared the model for deregulation (x_3=1) with the model for regulation (x_3=0) by creating a scatter plot comparing both models. We were able to plot distance shipped in miles (x_1) vs the predicted value of LNPRICE (y) in order to demonstrate the difference between the regulated and deregulated models because the x variables origin and weight of product were both held fixed.

Question 3:

Analysis:

Variable Assignment:

x_1 = Distance Shipped

x_2 = Weight of Product

$$x_3 = \begin{cases} 1 \; if \; deregulation \; in \; effect \\ 0 \; if \; not \end{cases}$$

$$x_4 = \begin{cases} 1 \; if \; originates \; from \; Miami \\ 0 \; if \; originates \; from \; Jacksonville \end{cases}$$

x_5 = Carrier A

x_6 = Carrier C

x_7 = Carrier D

Carrier B is held as the base level for carriers.

Model 7: Reduced (original) Model

$$E(y) = B_0 - B_1x_1 - B_2x_2 - B_3x_3 + B_4x_4 - B_5x_1^2 + B_6x_2^2 - B_7x_1x_2 + B_8x_1x_3 - B_9x_1x_4 - B_{10}x_2x_3 - B_{11}x_2x_4 - B_{12}x_1x_2x_3 + B_{13}x_1x_2x_4$$

$$\hat{y} = 12.2199 - 0.5670x_1 - 0.0167x_2 - 0.3733x_3 + 0.5999x_4 - 0.0748x_1^2 + 0.0003x_2^2 - 0.0075x_1x_2 + 0.0077x_1x_3 - 0.2239x_1x_4 - 0.0093x_2x_3 - 0.0263x_2x_4 - 0.0011x_1x_2x_3 + 0.0086x_1x_2x_4$$

Model 7 Complete (New) Model:

$$E(y)=B_0+B_1x_1+B_2x_2+B_3x_1^2+B_4x_2^2+B_5x_3+B_6x_4+B_7x_5+B_8x_6+B_9x_7+B_{10}x_1x_2+B_{11}x_1x_3+B_{12}x_1x_4+B_{13}x_1x_5+B_{14}x_1x_6+B_{15}x_1x_7+B_{16}x_2x_3+B_{17}x_2x_4+B_{18}x_2x_5+B_{19}x_2x_6+B_{20}x_2x_7+B_{21}x_1x_2x_3+B_{22}x_1x_2x_4+B_{23}x_1x_2x_5+B_{24}x_1x_2x_6+B_{25}x_1x_2x_7$$

This model includes the added main effects terms for carrier (A, B, C, and D) and the additional interaction terms for carrier with distance shipped, weight of the product, if deregulation is in effect, and if it originated in Miami or Jacksonville.

Prediction Equation Model 7 With Added Main Effects Terms for Carrier and Interaction Terms for Carrier

$$\hat{y}=12.0709-0.6300x_1-0.0251x_2+0.08516x_1^2+0.0003x_2^2-0.433x_3+0.6114x_4+0.4456x_5+0.2474x_6+0.2161x_7-.0065x_1x_2+0.0436x_1x_3 0.2306x_1x_4-.0512x_1x_5+0.0094x_1x_6+0.0314x_1x_7+0.00304x_2x_3-.0267x_2x_4-.0050x_2x_5+0.0083x_2x_6+0.0305x_2x_7-.0032x_1x_2x_3+0.0090x_1x_2x_4+0.0057x_1x_2x_5-0.0018x_1x_2x_6-0.0053x_1x_2x_7$$

We did not remove any of the added dummy variables for carriers (x5, x6, x7, and interaction terms including those variables) because we did not find significant multicollinearity between any of the three. (see appendix for pairwise correlation)

Hypothesis:

H_0: $B_5 = B_6 = B_7 = B_{13} = B_{14} = B_{15} = B_{18} = B_{19} = B_{20} = B_{23} = B_{24} = B_{25} = 0$
The added terms for carriers (A, B, C, or D) do not cause the curves to differ for each carrier.

H_a: One of B_5, B_6, B_7, B_{13}, B_{14}, B_{15}, B_{18}, B_{19}, B_{20}, B_{23}, B_{24}, B_{25} do not equal 0
The added terms for carrier (A, B, C, or D) do cause the curves to differ for each carrier.

Nested F Test:

The F-stat (10.6206) is greater than the F-critical value (1.77451) so we can determine that the model is statistically significant. There is sufficient evidence (at $\alpha = .05$) that the added terms for carriers cause the models to differ for each carrier.

We reject the null hypothesis that the added terms for carriers (A, B, C, or D) do not cause the curves to differ for each carrier at alpha=.05 (our p-value of 1.5241E-18<.05).

Conclusion:

This report sought to compare if there is a significant difference for trucking prices based on the type of carrier used. In conducting our analysis, we compared the best model for predicting trucking price (Model 7), with that same model including the addition of carrier type and any interaction carrier type might have with other variables.

In conducting our analysis, we looked at multiple models that vary in complexity. Our first model, Model 7, strictly used distance shipped, weight of the product, whether or not deregulation was in effect, and if it was originated in Jacksonville or Miami under the condition that products were all shipped by carrier B. We expanded to see if there would be different results if carriers A, C, and D were used instead. Using statistical tests, we found that there was a significant difference in trucking price for each of the different carriers. Therefore, it would be beneficial to include the information of carrier type when predicting trucking prices.

Appendix

Codes Used and Outcome:

```
data cs2;
infile '/folders/myfolders/TRUCKING4.txt' dlm='09'x firstobs=2;
input PRICPTM        DISTANCE    WEIGHT       PCTLOAD     ORIGIN $ MARKET
$ DEREG $     CARRIER $    PRODUCT     LNPRICE;
run;

data sad;
set cs2;
mia = 0;
if ORIGIN = 'MIA' then mia = 1;
yes=0;
if dereg = 'YES' then yes = 1;
A=0;
C=0;
D=0;
if carrier = 'A' then A =1;
if carrier = 'C' then C =1;
if carrier = 'D' then D =1;
x1 = DISTANCE;
x2 = WEIGHT;
x3 = yes;
x4 = MIA;
x5 = A;
x6 = C;
x7 = D;
x1sq = x1**2;
x2sq = x2**2;
x1x2 = x1*x2;
x1x3 = x1*x3;
x1x4 = x1*x4;
x1x5 = x1*x5;
x1x6 = x1*x6;
x1x7 = x1*x7;
x2x3 = x2*x3;
x2x4 = x2*x4;
x2x5 = x2*x5;
x2x6 = x2*x6;
x2x7 = x2*x7;
x1x2x3 = x1*x2*x3;
x1x2x4 = x1*x2*x4;
x1x2x5 = x1*x2*x5;
x1x2x6 = x1*x2*x6;
x1x2x7 = x1*x2*x7;
```

```
run;

*checks for multicollinearlity;
proc corr data=sad nosimple;
var x1 x2 x3 x4 x5 x6 x7;
run;
```

The CORR Procedure

7 Variables:	x1 x2 x3 x4 x5 x6 x7

Pearson Correlation Coefficients, N = 448 Prob > |r| under H0: Rho=0							
	x1	x2	x3	x4	x5	x6	x7
x1	1.00000	0.03484 0.4620	-0.06832 0.1488	0.10616 0.0246	-0.05417 0.2526	0.01907 0.6872	0.05098 0.2816
x2	0.03484 0.4620	1.00000	-0.03532 0.4558	0.01328 0.7792	0.02231 0.6376	-0.08523 0.0715	0.03421 0.4702
x3	-0.06832 0.1488	-0.03532 0.4558	1.00000	-0.02498 0.5979	0.13870 0.0033	-0.19205 <.0001	-0.12324 0.0090
x4	0.10616 0.0246	0.01328 0.7792	-0.02498 0.5979	1.00000	-0.07777 0.1002	0.04355 0.3578	-0.02474 0.6015
x5	-0.05417 0.2526	0.02231 0.6376	0.13870 0.0033	-0.07777 0.1002	1.00000	-0.33976 <.0001	-0.31275 <.0001
x6	0.01907 0.6872	-0.08523 0.0715	-0.19205 <.0001	0.04355 0.3578	-0.33976 <.0001	1.00000	-0.25438 <.0001
x7	0.05098 0.2816	0.03421 0.4702	-0.12324 0.0090	-0.02474 0.6015	-0.31275 <.0001	-0.25438 <.0001	1.00000

```
proc reg data=sad plots = none;
model lnprice = x1 x2 x3 x4 x5 x6 x7 /vif;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: LNPRICE

Number of Observations Read	448
Number of Observations Used	448

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	141.13647	20.16235	117.10	<.0001
Error	440	75.75806	0.17218		
Corrected Total	447	216.89453			

Root MSE	0.41494	R-Square	0.6507
Dependent Mean	10.85452	Adj R-Sq	0.6452
Coeff Var	3.82276		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	11.83871	0.06503	182.06	<.0001	0
x1	1	-0.29744	0.01425	-20.87	<.0001	1.01968
x2	1	-0.03325	0.00227	-14.66	<.0001	1.01143
x3	1	-0.38640	0.04167	-9.27	<.0001	1.07986
x4	1	0.19735	0.03962	4.98	<.0001	1.02032
x5	1	0.36548	0.05106	7.16	<.0001	1.41011
x6	1	0.27046	0.05674	4.77	<.0001	1.42119
x7	1	0.38854	0.05843	6.65	<.0001	1.36587

*Beta parameters;
*For Model 7:

```
proc reg data=sad plots = none;
model lnprice = x1 x2 x1sq x2sq x3 x4 x1x2 x1x3 x1x4 x2x3 x2x4 x1x2x3 x1x2x4;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: LNPRICE

Number of Observations Read	448
Number of Observations Used	448

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	13	144.65825	11.12756	66.86	<.0001
Error	434	72.23628	0.16644		
Corrected Total	447	216.89453			

Root MSE	0.40797	R-Square	0.6670
Dependent Mean	10.85452	Adj R-Sq	0.6570
Coeff Var	3.75857		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.21986	0.18789	65.04	<.0001
x1	1	-0.56696	0.07840	-7.23	<.0001
x2	1	-0.01673	0.01688	-0.99	0.3222
x1sq	1	0.07484	0.00860	8.70	<.0001
x2sq	1	0.00034869	0.00035207	0.99	0.3225
x3	1	-0.37331	0.13660	-2.73	0.0065
x4	1	0.59991	0.18372	3.27	0.0012
x1x2	1	-0.00754	0.00498	-1.51	0.1307
x1x3	1	0.00766	0.04416	0.17	0.8623
x1x4	1	-0.22387	0.06201	-3.61	0.0003
x2x3	1	-0.00930	0.01138	-0.82	0.4142
x2x4	1	-0.02632	0.01529	-1.72	0.0859
x1x2x3	1	0.00111	0.00345	0.32	0.7483
x1x2x4	1	0.00864	0.00514	1.68	0.0935

*Model 7 + added terms

```
proc reg data=sad plots=none;
model lnprice = x1 x2 x1sq x2sq x3 x4 x5 x6 x7 x1x2 x1x3 x1x4 x1x5 x1x6 x1x7 x2x3 x2x4
x2x5 x2x6 x2x7 x1x2x3
x1x2x4 x1x2x5 x1x2x6 x1x2x7;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: LNPRICE

Number of Observations Read	448
Number of Observations Used	448

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	161.05565	6.44223	48.69	<.0001
Error	422	55.83887	0.13232		
Corrected Total	447	216.89453			

Root MSE	0.36376	R-Square	0.7426
Dependent Mean	10.85452	Adj R-Sq	0.7273
Coeff Var	3.35121		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	12.07092	0.19788	61.00	<.0001
x1	1	-0.63004	0.07961	-7.91	<.0001
x2	1	-0.02511	0.01727	-1.45	0.1467
x1sq	1	0.08516	0.00783	10.88	<.0001
x2sq	1	0.00026832	0.00032024	0.84	0.4026
x3	1	-0.43328	0.12726	-3.40	0.0007
x4	1	0.61138	0.16541	3.70	0.0002
x5	1	0.44755	0.14523	3.08	0.0022
x6	1	0.24741	0.15586	1.59	0.1132
x7	1	0.21606	0.19297	1.12	0.2635
x1x2	1	-0.00650	0.00508	-1.28	0.2016
x1x3	1	0.04372	0.04160	1.05	0.2939
x1x4	1	-0.23062	0.05570	-4.14	<.0001
x1x5	1	-0.05121	0.04593	-1.11	0.2655
x1x6	1	0.00943	0.04755	0.20	0.8429
x1x7	1	0.03138	0.05729	0.55	0.5842
x2x3	1	0.00304	0.01090	0.28	0.7808
x2x4	1	-0.02665	0.01376	-1.94	0.0534
x2x5	1	-0.00501	0.01211	-0.41	0.6793
x2x6	1	0.00834	0.01520	0.55	0.5837
x2x7	1	0.03049	0.01483	2.06	0.0403
x1x2x3	1	-0.00315	0.00341	-0.92	0.3555
x1x2x4	1	0.00900	0.00462	1.95	0.0520
x1x2x5	1	0.00571	0.00370	1.54	0.1240
x1x2x6	1	-0.00175	0.00447	-0.39	0.6951
x1x2x7	1	-0.00529	0.00435	-1.22	0.2250

```
*nested F-test;
data nestedFtest;
Fstat = ((72.23628-55.83887)/(12))/(55.83887/(448 -13 -1));
pvalue = SDF('F',Fstat,12,448 -13 -1);
Fcritical = quantile('F',.95,12,448 -13 -1);
proc print data=nestedFtest;
run;
```

Obs	Fstat	pvalue	Fcritical
1	10.6206	1.5241E-18	1.77451