Midterm

STAT-UB.0003 – Regression and Forecasting Models

The exam is closed book and notes, with the following exception: you are allowed to bring one letter-sized page of notes into the exam (front and back). You are also permitted use of a calculator. Each part of each problem is worth 5 points. There are 75 points total. There is no penalty for guessing incorrectly on a multiple choice problem. Partial credit may be awarded for all problems, as long as you show work.

For the problems involving calculations, you must show all work to get full credit. For short-answer problems, there should not be any symbols in your final answer (p, n, λ, etc.), but you do not need to fully simplify your answer. It is ok to have quantities like $\binom{5}{2}$, $e^{-3.1}$, etc. in your final answers on these problems.

NYU Stern Honor Code:

I will not lie, cheat or steal to gain an academic advantage, or tolerate those who do.

Signature: _____ Date: _____

Name: _____

Short Answer (Pickup Truck Analysis)

Use the Pickup Truck Data Analysis provided at the end of the exam to answer the problems in this section.

1. (5 points) In the context of the fitted regression model, interpret the coefficient of "miles." Your answer should be a single sentence.

2. (5 points) In the fitted regression model, does the sign of the coefficient of "miles" make sense? Why or why not? Answer in 1–2 sentences.

3. (5 points) Identify a reasonable sample and population for the pickup truck data.

4. (5 points) There is clearly a strong association between "price" and "miles." Give an argument for why there might not be a causal relationship between these two variables. Answer in 1–2 sentences.

5. (5 points) For the pickup truck regression model, is there evidence of nonconstant variance in the errors? Explain why or why not in 1–2 sentences.

Multiple Choice (Pickup Truck Analysis)

Use the Pickup Truck Data Analysis provided at the end of the exam to answer the problems in this section.

6. (5 points) Using the pickup truck data, which of the following is an approximate 95% confidence interval for the true (population) value of β_1, the coefficient of "miles"?

   A. $(-1.4, 1.2)$

   B. $(-0.14 - 0.07)$

   C. $(67.4, 94.6)$

   D. $(-0.110, -0.099)$

   E. Not enough information to determine.

7. (5 points) In the context of the fitted regression model, which of the following is approximately equal to the price range (in thousands of dollars) of 95% of all cars listed on Craigslist Chicago with 25K miles?

   A. $(11.5, 17.4)$

   B. $(1.1, 17.2)$

   C. $(6.4, 22.5)$

   D. $(6.2, 12.1)$

   E. Not enough information to determine.

8. (5 points) Approximately what is the mean price of all pickup trucks listed on Craigslist Chicago that have 100,000 miles?

    A. $6,600
    B. $8,566
    C. $66,000
    D. $17,100
    E. Not enough information to determine.

9. (5 points) What is the ID of the observation with the largest positive residual?

    A. 2
    B. 7
    C. 3
    D. 18
    E. Not enough information to determine.

10. (5 points) The t statistic for testing the null hypothesis that there is no linear relationship between miles and expected price is:

    A. 11.58

    B. 8.87

    C. -6.46

    D. 11.91

    E. Not enough information to determine.

11. (5 points) Using the pickup truck data, which of the following is an approximate 95% confidence interval for the mean price (in thousands of dollars) of all pickup trucks listed on Craigslist Chicago?

    A. $(-3.2, 20.3)$

    B. $(14.1, 20.0)$

    C. $(6.6, 10.5)$

    D. $(15.7, 18.4)$

    E. Not enough information to determine.

12. (5 points) Which of the following statements are supported by the data analysis?

    A. There is no significant evidence that the linear regression model is useful.

    B. The 95% confidence interval for β_0, the regression intercept, contains 0.

    C. The true (population) value of the β_0, the regression intercept, is equal to 17.054.

    D. There is strong evidence that the linear regression model is useful.

    E. Not enough information to determine.

13. (5 points) Why might it be a bad idea to use the fitted regression model to predict the listed price of a car with 160K miles?

    A. 160K miles is outside the range of the data.

    B. The predicted price is negative.

    C. The regression standard error is too high.

    D. It doesn't make sense to have a car with 160K miles.

    E. None of the above. There is nothing wrong with using the model in this way.

Additional Problems

The following two problems do not relate to the Pickup Truck data.

14. (5 points) Which of the following is the most appropriate tool for checking the assumption that the mean of the regression errors is equal to zero at every value of the predictor?

    A. Mean and standard deviation of the residuals.

    B. Normal probability plot of residuals.

    C. Plot of residuals versus fits.

    D. Plot of residuals versus order.

    E. Histogram of residuals.

15. (5 points) Suppose that in testing the null hypothesis $H_0 : \mu = 0$ against the alternative $H_a : \mu \neq 0$, with known population variance, you observe the test statistic $z = 0.80$. Which of the following is approximately equal to the p-value corresponding this test statistic?

    A. 0.58

    B. 0.21

    C. 0.29

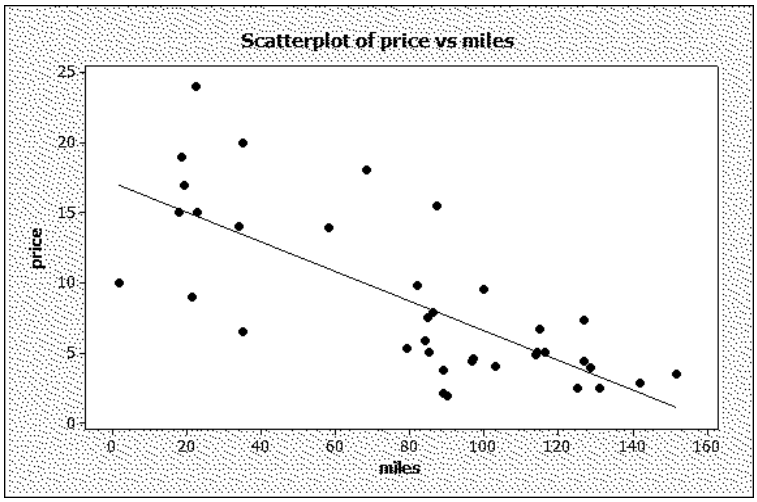
    D. 0.42

    E. None of the above.

Pickup Truck Data Analysis for Problems 1–13

The first part of the exam involves the following dataset: a random sample of 37 pickup trucks listed for sale on Craigslist Chicago. The table gives the observation ID ("Obs."), the asking price, in thousands of dollars ("Price"), and the number of miles showing on the truck's odometer, in thousands of miles ("Miles"). The adjacent graph shows a scatterplot of Price versus Miles, with the fitted least squares regression line.

Obs.	Price	Miles
1	14.995	17.638
2	9.998	1.500
3	23.950	22.422
4	19.980	34.815
5	2.800	142.000
6	7.900	86.000
7	6.700	115.000
8	14.980	22.702
9	5.900	84.000
10	3.900	128.500
11	5.000	85.000
12	6.500	35.000
13	8.995	21.195
14	9.800	82.000
15	16.985	19.000
16	2.500	131.000
17	2.100	89.000
18	13.900	58.170
19	7.295	127.025
20	4.900	114.000
21	15.495	87.226
22	13.995	34.000
23	5.350	79.000
24	4.595	97.000
25	4.000	103.000
26	9.490	99.760
27	3.450	151.923
28	4.395	127.000
29	17.999	68.277
30	4.999	114.400
31	18.995	18.490
32	4.400	96.600
33	5.000	116.455
34	7.485	84.652
35	3.800	89.000
36	1.900	90.000
37	2.500	125.000



Here are descriptive statistics for the two variables. For the sake of the exam, some values have been replaced by question marks.

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
price	37	8.566	?????	5.878	1.900	4.197	6.500	13.948	23.950
miles	37	81.02	?????	41.36	1.50	34.91	87.23	114.70	151.92

Fitting a linear regression model to the pickup truck data gives the following output. For the sake of the exam, some values have been replaced by question marks.

```
The regression equation is
price = 17.1 - 0.105 miles


Predictor      Coef  SE Coef      T      P
Constant     17.054    1.472  ?????  0.000
miles      -0.10477  0.01623  ?????  0.000


S = 4.02761   R-Sq = 54.4%   R-Sq(adj) = 53.0%
```

Here are residual diagnostic plots for the regression fit: