Final
COR1-GB.1305 – Statistics and Data Analysis

The exam is open book and notes, and you are 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, but only if 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: ——

Multiple Choice

1. (5 points) Suppose that in testing the null hypothesis $H_0 : \mu = 0$ against the alternative $H_a : \mu \neq 0$, you observe the test statistic $t = 0.80$. 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.

2. (5 points) An analyst tells you that using a sample size of $n = 100$, she estimates that your company profits for next year will be 4.8 million dollars. She tells you that the standard error of her estimate is 0.8 million dollars. Give an approximate 95% confidence interval for next year's company profits.

    A. (4.72, 4.88)

    B. (3.20, 6.40)

    C. (4.64, 4.96)

    D. (4.00, 5.60)

    E. (2.80, 6.80)

3. (5 points) To compare the effectiveness of Atkins to conventional dieting, we collect two samples: the weight losses of 33 people who went on the Atkins diet, and the weight losses of 30 people who used a conventional diet. For the people on the Atkins diet, the mean weight loss was 15.42 pounds, and the standard deviation was 14.37. For the people on the conventional diet, the mean loss was 7.01 pounds, and the standard deviation was 12.36 pounds. Find a 99% confidence interval for the difference in expected weight loss between all Atkins and conventional dieters.

   A. $(7.8, 9.0)$

   B. $(-0.3, 17.1)$

   C. $(1.8, 15.0)$

   D. $(9.0, 21.9)$

   E. Not enough information to determine.

Short Answer

4. (10 points) A random sample of 50 adults were asked how much they spend on lottery tickets, and were interviewed about various socioeconomic variables. The variables are

PercLott = Percentage of total household income spent on the lottery. (This is Y).
YrsEdu = Number of years of education,
Age = The person's Age,
Kids = Number of Children,
Income = Personal income (Thousands of Dollars).

Here is the Minitab regression output:

```
Analysis of Variance

Source      DF  Adj SS  Adj MS  F-Value  P-Value
Regression   4  404.42  101.10    17.72    0.000
  YrsEdu     1   60.68   60.68    10.63    0.002
  Age        1    0.21    0.21     0.04    0.850
  Kids       1    0.55    0.55     0.10    0.761
  Income     1   23.30   23.30     4.08    0.050
Error       45  256.80    5.71
Total       49  661.22


Model Summary

    S    R-Sq  R-Sq(adj)  R-sq(pred)
2.389  61.16%     57.71%      52.16%


Coefficients

Term         Coef    SE Coef  T-Value  P-Value   VIF
Constant   15.070      2.444     6.17    0.000
YrsEdu    -0.5911     0.1813    -3.26    0.002  1.47
Age       0.00647    0.03395     0.19    0.850  2.81
Kids       0.0816     0.2665     0.31    0.761  1.93
Income   -0.06663    0.03305    -2.02    0.050  1.58


Regression Equation

PercLott = 15.1 - 0.591 YrsEdu + 0.0065 Age + 0.082 Kids - 0.0666 Income
```

Use the regression analysis from the lottery data to answer parts (a) and (b).

(a) In the context of the estimated multiple regression model, what is the interpretation of coefficient of `Income`?

(b) For the lottery data, what can we conclude from the regression F test?

Pickup Truck Data Analysis

The next 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

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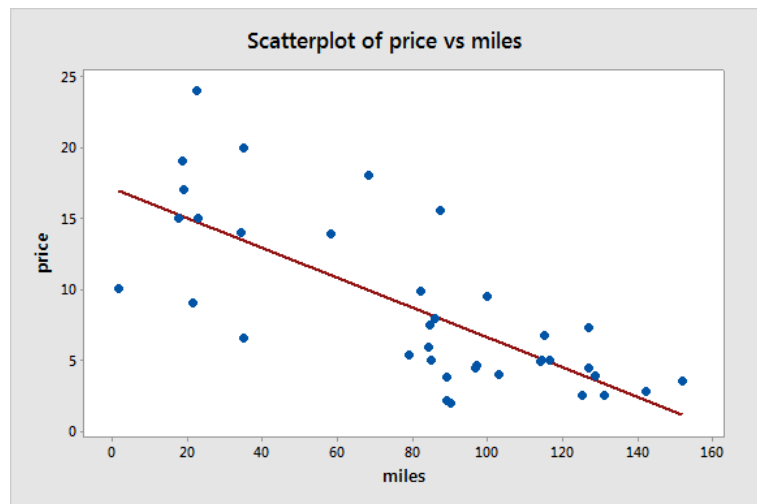
Here are descriptive statistics for the two variables. For the sake of the exam, some values have been replaced by question marks.

```
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
```

We used Minitab to fit a linear regression model to the pickup truck data. For the sake of the exam, some values have been replaced by question marks.

```
Analysis of Variance

Source          DF    Adj SS    Adj MS  F-Value  P-Value
Regression       1    675.97   675.973    ?????    0.000
  miles          1    675.97   675.973    ?????    0.000
Error           35    567.76    16.222
  Lack-of-Fit   34    566.31    16.656    11.53    0.230
  Pure Error     1      1.44     1.445
Total           36   1243.73
```

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
4.02761  54.35%     53.05%      48.38%
```
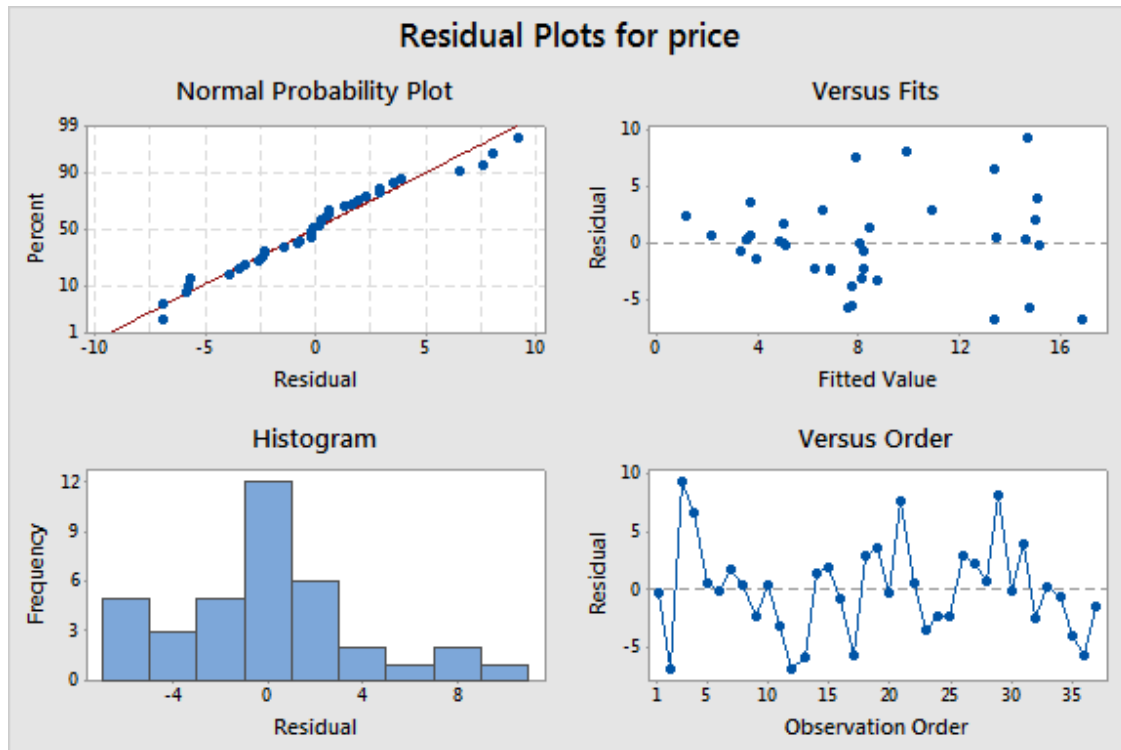
```
Coefficients

Term          Coef  SE Coef  T-Value  P-Value   VIF
Constant     17.05     1.47    ?????    0.000
miles      -0.1048   0.0162    ?????    0.000  1.00
```

```
Regression Equation

price = 17.05 - 0.1048 miles
```

Here are residual diagnostic plots for the regression fit:

Residual Plots for price

Short Answer (Pickup Truck Data Analysis)

Use the Pickup Truck Data Analysis to answer the problems in this section.

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

6. (5 points) Identify a reasonable sample and population for the pickup truck data.

7. (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 Data Analysis)

Use the Pickup Truck Data Analysis to answer the problems in this section.

8. (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. $(-1.43, 1.22)$

      B. $(-0.137, -0.072)$

      C. $(67.4, 94.6)$

      D. $(-0.110, -0.099)$

      E. Not enough information to determine.

9. (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.

10. (5 points) Approximately what is the mean price of all pickup trucks listed on Craigslist Chicago that have 100,000 miles?

 A. $6,570

 B. $8,566

 C. $65,700

 D. $17,050

 E. Not enough information to determine.

11. (5 points) What is the observation ID ("Obs.") of the data point with the largest positive residual?

 A. 2

 B. 7

 C. 3

 D. 18

 E. Not enough information to determine.

12. (5 points) The t statistic for testing the null hypothesis that there is no linear relationship between miles and expected price is:

    A. 11.59

    B. 8.86

    C. -6.46

    D. 11.91

    E. Not enough information to determine.

13. (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.

14. (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 17.05.

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

    E. Not enough information to determine.