Jacob Igel ISA 291 - A

1. State the first order multiple linear regression model.
   1. 𝑌=𝛽0+𝛽1𝑥1+𝛽2𝑥2+𝜖
2. Interpret the meaning of the population slope coefficient for the Door.
   1. When looking at the population of the doors, the cars will only have either two or four car doors since there are no cars with an odd number of doors.
3. Read the data into R and create scatter plots between the y=Price and x 1=Liter. Repeat with y=Price and x 2=Doors. For each plot, does the relationship appear linear/nonlinear?  Strong/Weak? Positive/Negative?

![Chart, scatter chart

Description automatically generated]()

Liter vs Price: With this scatter plot, it appears that as the number of liters in a car increases, the price also increases – this appears to be a strong positive linear relationship.

![Chart, line chart, scatter chart

Description automatically generated]()

Doors vs Price: With this scatter plot, it appears that as the number of doors increases, the price drops/a negative relationship. The scatter plot seems to have an extremely week relationship with a lot of variation.

1. Regress y=Price on x 1=Liter and x 2=Doors.  Print a summary of the regression object.

Residuals:

Min 1Q Median 3Q Max

-11283 -5438 -3005 2602 40030

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 10291.1 1516.2 6.787 0.0000000000222 \*\*\*

Liter 4923.0 261.2 18.845 < 0.0000000000000002 \*\*\*

Doors -1105.8 339.7 -3.255 0.00118 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8158 on 801 degrees of freedom

Multiple R-squared: 0.3205, Adjusted R-squared: 0.3188

F-statistic: 188.9 on 2 and 801 DF, p-value: < 0.00000000000000022

1. State the regression equation.
   1. Y = 10291.1 + 4923.0x1 - 1105.8x2
2. Find the estimated Price of a used 4-door car with a 3.1 Liter engine.
   1. avg\_yhat = 10291.1 + (4923.0\*3.1) - (1105.8\*4)
   2. avg\_yhat = 21129.2
3. Conduct a test of overall model significance.

confint(reg1, level = .95)

2.5 % 97.5 %

(Intercept) 7314.882 13267.3525

Liter 4410.211 5435.7770

Doors -1772.650 -439.0008

1. Is Liter a significant linear predictor of Price in this model?
   1. > cor(cars$Liter, cars$Price)

[1] 0.5581458

* 1. We can say that liters is somewhat significant in being a predictor to price in this model due to the correlation being somewhat close to 1.

1. Construct a 95% confidence interval to estimate the population slope coefficient for Liter.  Interpret the interval.
   1. Liter 4410.211 5435.7770
   2. We are 95% confident that the population slope of a liter, given the amount of liters the car can hold, will increase the range of price between 4410.2 and 5435.7.
2. Is Doors a significant linear predictor of Price in this model?
   1. > cor(cars$Doors, cars$Price)

[1] -0.1387497

* 1. We can say that doors is not a predictor of price in this model due to the low correlation between Price and Doors.

1. Construct a 95% confidence interval to estimate the population slope coefficient for Doors. Interpret the interval.
   1. Doors -1772.650 -439.0008
      1. We can say with 95% confidence that the price of a car can drop between 1772.6 and 439 depending on the amount of doors.
2. What is the Coefficient of Determination?  Adjusted Coefficient of Determination?  What do these values measure?
   1. Coefficient of Determination - 0.3205156
   2. Adjusted Coefficient of Determination - 0.3188191
   3. These values help the viewer measure the statistical relationship between the intercept and the slopes within the regression model.
3. Compute a 95% confidence interval to estimate the mean price of all used cars with Liter = 3.1 and Doors = 2.

> predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="confidence",level = 0.95)

fit lwr upr

1 23340.75 22177.94 24503.56

1. Compute a 95% prediction interval to estimate the price of a car with Liter = 3.1 and Doors = 2.

> predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="prediction",level = 0.95)

fit lwr upr

1 23340.75 7284.348 39397.15

1. Explain why the interval in #14 is wider than the interval in #13.
   1. With prediction intervals, they are taking every single variation in the regression into account therefore developing a larger interval. Confidence intervals only look at a specific value and goes off that.
2. Plot the residuals vs. the fitted values. Do the residuals appear randomly scattered?  Do the residuals appear centered at zero? Does the variance of the residuals appear constant?
   1. The residuals look scattered but in certain blobs around zero but the majority of points are not centered at zero. The variance appears to be inconsistent looking across the graph.

![Chart, box and whisker chart

Description automatically generated]()

1. Construct a QQ plot of the residuals and a histogram of the residuals. Do the residuals appear normally distributed?
   1. Looking at the graphs below, it appears that the residuals are skewed to the right. This does not conclude that they are normally distributed.

![Chart, line chart

Description automatically generated]()

![Chart, histogram

Description automatically generated]()

**CODE INPUT**

setwd("~/Desktop/School/ISA 291")

cars = read.csv("Cars.csv", header = TRUE, stringsAsFactors = TRUE)

cars

options(scipen = 999)

reg1 = lm(Price ~ Liter + Doors, data = cars)

reg1

summary(reg1)

plot(cars$Liter, cars$Price)

abline(lm(Price ~ Liter, data = cars), col = 'red')

cor(cars$Liter, cars$Price)

plot(cars$Doors, cars$Price)

abline(lm(Price ~ Doors, data = cars), col = 'red')

cor(cars$Doors, cars$Price)

avg\_yhat = 10291.1 + (4923.0\*3.1) - (1105.8\*4)

avg\_yhat

confint(reg1,newdata = data.frame(Liter), level = .95)

summary(reg1)$adj.r.squared

anova(reg1)

predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="confidence",level = 0.95)

predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="prediction",level = 0.95)

plot(reg1$fitted.values, reg1$residuals)

abline(0,0, col = 'red')

hist(reg1$residuals, breaks = 30)

qqnorm(reg1$residuals)

qqline(reg1$residuals)

**CODE OUTPUT**

> setwd("~/Desktop/School/ISA 291")

>

> cars = read.csv("Cars.csv", header = TRUE, stringsAsFactors = TRUE)

> cars

Price Mileage Make Model Trim Type

1 17314.10 8221 Buick Century Sedan 4D Sedan

2 17542.04 9135 Buick Century Sedan 4D Sedan

3 16218.85 13196 Buick Century Sedan 4D Sedan

4 16336.91 16342 Buick Century Sedan 4D Sedan

5 16339.17 19832 Buick Century Sedan 4D Sedan

6 15709.05 22236 Buick Century Sedan 4D Sedan

7 15230.00 22576 Buick Century Sedan 4D Sedan

8 15048.04 22964 Buick Century Sedan 4D Sedan

9 14862.09 24021 Buick Century Sedan 4D Sedan

10 15295.02 27325 Buick Century Sedan 4D Sedan

11 21335.85 10237 Buick Lacrosse CX Sedan 4D Sedan

12 20538.09 15066 Buick Lacrosse CX Sedan 4D Sedan

13 20512.09 16633 Buick Lacrosse CX Sedan 4D Sedan

14 19924.16 19800 Buick Lacrosse CX Sedan 4D Sedan

15 19774.25 23359 Buick Lacrosse CX Sedan 4D Sedan

16 19344.17 23765 Buick Lacrosse CX Sedan 4D Sedan

17 19105.13 24008 Buick Lacrosse CX Sedan 4D Sedan

18 18543.43 26034 Buick Lacrosse CX Sedan 4D Sedan

19 17808.20 32896 Buick Lacrosse CX Sedan 4D Sedan

20 17968.84 34665 Buick Lacrosse CX Sedan 4D Sedan

21 22358.88 8970 Buick Lacrosse CXL Sedan 4D Sedan

22 23785.92 10577 Buick Lacrosse CXL Sedan 4D Sedan

23 22926.09 14363 Buick Lacrosse CXL Sedan 4D Sedan

24 21895.76 16508 Buick Lacrosse CXL Sedan 4D Sedan

25 21273.06 18908 Buick Lacrosse CXL Sedan 4D Sedan

26 21460.01 19467 Buick Lacrosse CXL Sedan 4D Sedan

27 21183.12 21394 Buick Lacrosse CXL Sedan 4D Sedan

28 20406.10 22596 Buick Lacrosse CXL Sedan 4D Sedan

29 21058.14 24469 Buick Lacrosse CXL Sedan 4D Sedan

30 19556.90 25245 Buick Lacrosse CXL Sedan 4D Sedan

31 23447.69 15755 Buick Lacrosse CXS Sedan 4D Sedan

32 23547.24 16235 Buick Lacrosse CXS Sedan 4D Sedan

33 23016.01 18147 Buick Lacrosse CXS Sedan 4D Sedan

34 22230.03 22102 Buick Lacrosse CXS Sedan 4D Sedan

35 22625.07 23612 Buick Lacrosse CXS Sedan 4D Sedan

36 21799.17 24439 Buick Lacrosse CXS Sedan 4D Sedan

37 21341.26 25212 Buick Lacrosse CXS Sedan 4D Sedan

38 21683.03 26779 Buick Lacrosse CXS Sedan 4D Sedan

39 20986.02 27096 Buick Lacrosse CXS Sedan 4D Sedan

40 20902.10 29649 Buick Lacrosse CXS Sedan 4D Sedan

41 20698.08 2992 Buick Lesabre Custom Sedan 4D Sedan

42 20099.26 10036 Buick Lesabre Custom Sedan 4D Sedan

43 18145.13 18339 Buick Lesabre Custom Sedan 4D Sedan

44 17944.86 19592 Buick Lesabre Custom Sedan 4D Sedan

45 19027.86 21797 Buick Lesabre Custom Sedan 4D Sedan

46 18348.90 23852 Buick Lesabre Custom Sedan 4D Sedan

47 17750.88 25040 Buick Lesabre Custom Sedan 4D Sedan

48 17772.97 25052 Buick Lesabre Custom Sedan 4D Sedan

49 17394.02 25464 Buick Lesabre Custom Sedan 4D Sedan

50 17645.75 27830 Buick Lesabre Custom Sedan 4D Sedan

51 21908.37 17353 Buick Lesabre Limited Sedan 4D Sedan

52 21956.34 17787 Buick Lesabre Limited Sedan 4D Sedan

53 21646.12 19562 Buick Lesabre Limited Sedan 4D Sedan

54 21575.46 20137 Buick Lesabre Limited Sedan 4D Sedan

55 20952.22 20158 Buick Lesabre Limited Sedan 4D Sedan

56 21562.05 23767 Buick Lesabre Limited Sedan 4D Sedan

57 19981.13 24323 Buick Lesabre Limited Sedan 4D Sedan

58 19425.85 27839 Buick Lesabre Limited Sedan 4D Sedan

59 19191.99 29187 Buick Lesabre Limited Sedan 4D Sedan

60 19641.74 31324 Buick Lesabre Limited Sedan 4D Sedan

61 25589.98 2308 Buick Park Avenue Sedan 4D Sedan

62 25098.63 10014 Buick Park Avenue Sedan 4D Sedan

63 23420.71 18910 Buick Park Avenue Sedan 4D Sedan

64 22661.05 20105 Buick Park Avenue Sedan 4D Sedan

65 23493.08 20453 Buick Park Avenue Sedan 4D Sedan

66 22435.20 22287 Buick Park Avenue Sedan 4D Sedan

67 21878.12 23237 Buick Park Avenue Sedan 4D Sedan

68 23077.57 23798 Buick Park Avenue Sedan 4D Sedan

69 21698.01 25489 Buick Park Avenue Sedan 4D Sedan

70 21831.82 25564 Buick Park Avenue Sedan 4D Sedan

71 26831.19 4695 Buick Park Avenue Special Ed Ultra 4D Sedan

72 26060.34 9795 Buick Park Avenue Special Ed Ultra 4D Sedan

73 26781.81 12052 Buick Park Avenue Special Ed Ultra 4D Sedan

74 26302.07 13050 Buick Park Avenue Special Ed Ultra 4D Sedan

75 26190.27 17335 Buick Park Avenue Special Ed Ultra 4D Sedan

76 25508.21 17480 Buick Park Avenue Special Ed Ultra 4D Sedan

77 23348.02 24027 Buick Park Avenue Special Ed Ultra 4D Sedan

78 23406.69 25387 Buick Park Avenue Special Ed Ultra 4D Sedan

79 23159.54 25869 Buick Park Avenue Special Ed Ultra 4D Sedan

80 21536.74 37128 Buick Park Avenue Special Ed Ultra 4D Sedan

81 51154.05 2202 Cadillac CST-V Sedan 4D Sedan

82 49248.16 6685 Cadillac CST-V Sedan 4D Sedan

83 46747.67 15343 Cadillac CST-V Sedan 4D Sedan

Cylinder Liter Doors Cruise Sound Leather

1 6 3.1 4 1 1 1

2 6 3.1 4 1 1 0

3 6 3.1 4 1 1 0

4 6 3.1 4 1 0 0

5 6 3.1 4 1 0 1

6 6 3.1 4 1 1 0

7 6 3.1 4 1 1 0

8 6 3.1 4 1 1 0

9 6 3.1 4 1 0 1

10 6 3.1 4 1 1 1

11 6 3.6 4 1 0 0

12 6 3.6 4 1 1 0

13 6 3.6 4 1 1 0

14 6 3.6 4 1 1 1

15 6 3.6 4 1 1 1

16 6 3.6 4 1 1 0

17 6 3.6 4 1 0 0

18 6 3.6 4 1 1 1

19 6 3.6 4 1 1 0

20 6 3.6 4 1 1 1

21 6 3.6 4 1 1 0

22 6 3.6 4 1 1 1

23 6 3.6 4 1 1 1

24 6 3.6 4 1 0 1

25 6 3.6 4 1 0 0

26 6 3.6 4 1 0 1

27 6 3.6 4 1 0 0

28 6 3.6 4 1 0 0

29 6 3.6 4 1 1 1

30 6 3.6 4 1 0 0

31 6 3.8 4 1 1 0

32 6 3.8 4 1 1 0

33 6 3.8 4 1 1 1

34 6 3.8 4 1 0 1

35 6 3.8 4 1 0 1

36 6 3.8 4 1 0 0

37 6 3.8 4 1 1 1

38 6 3.8 4 1 1 0

39 6 3.8 4 1 0 0

40 6 3.8 4 1 1 1

41 6 3.8 4 1 0 1

42 6 3.8 4 1 1 1

43 6 3.8 4 1 1 0

44 6 3.8 4 1 0 0

45 6 3.8 4 1 0 1

46 6 3.8 4 1 1 0

47 6 3.8 4 1 1 0

48 6 3.8 4 1 1 0

49 6 3.8 4 1 1 0

50 6 3.8 4 1 1 1

51 6 3.8 4 1 0 0

52 6 3.8 4 1 1 0

53 6 3.8 4 1 1 1

54 6 3.8 4 1 1 0

55 6 3.8 4 1 1 0

56 6 3.8 4 1 0 1

57 6 3.8 4 1 1 0

58 6 3.8 4 1 1 0

59 6 3.8 4 1 1 1

60 6 3.8 4 1 1 1

61 6 3.8 4 1 1 0

62 6 3.8 4 1 1 0

63 6 3.8 4 1 0 1

64 6 3.8 4 1 0 0

65 6 3.8 4 1 1 0

66 6 3.8 4 1 1 1

67 6 3.8 4 1 1 0

68 6 3.8 4 1 0 1

69 6 3.8 4 1 1 1

70 6 3.8 4 1 1 1

71 6 3.8 4 1 1 0

72 6 3.8 4 1 1 0

73 6 3.8 4 1 1 0

74 6 3.8 4 1 1 0

75 6 3.8 4 1 0 1

76 6 3.8 4 1 1 0

77 6 3.8 4 1 0 1

78 6 3.8 4 1 0 1

79 6 3.8 4 1 0 0

80 6 3.8 4 1 1 1

81 8 5.7 4 1 1 1

82 8 5.7 4 1 0 1

83 8 5.7 4 1 1 1

[ reached 'max' / getOption("max.print") -- omitted 721 rows ]

>

> options(scipen = 999)

> reg1 = lm(Price ~ Liter + Doors, data = cars)

> reg1

Call:

lm(formula = Price ~ Liter + Doors, data = cars)

Coefficients:

(Intercept) Liter Doors

10291 4923 -1106

> summary(reg1)

Call:

lm(formula = Price ~ Liter + Doors, data = cars)

Residuals:

Min 1Q Median 3Q Max

-11283 -5438 -3005 2602 40030

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 10291.1 1516.2 6.787 0.0000000000222 \*\*\*

Liter 4923.0 261.2 18.845 < 0.0000000000000002 \*\*\*

Doors -1105.8 339.7 -3.255 0.00118 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8158 on 801 degrees of freedom

Multiple R-squared: 0.3205, Adjusted R-squared: 0.3188

F-statistic: 188.9 on 2 and 801 DF, p-value: < 0.00000000000000022

>

> plot(cars$Liter, cars$Price)

> abline(lm(Price ~ Liter, data = cars), col = 'red')

> cor(cars$Liter, cars$Price)

[1] 0.5581458

>

> plot(cars$Doors, cars$Price)

> abline(lm(Price ~ Doors, data = cars), col = 'red')

> cor(cars$Doors, cars$Price)

[1] -0.1387497

>

> avg\_yhat = 10291.1 + (4923.0\*3.1) - (1105.8\*4)

> avg\_yhat

[1] 21129.2

>

> confint(reg1,newdata = data.frame(Liter), level = .95)

2.5 % 97.5 %

(Intercept) 7314.882 13267.3525

Liter 4410.211 5435.7770

Doors -1772.650 -439.0008

>

> summary(reg1)$adj.r.squared

[1] 0.3188191

>

> anova(reg1)

Analysis of Variance Table

Response: Price

Df Sum Sq Mean Sq F value Pr(>F)

Liter 1 24442819155 24442819155 367.239 < 0.00000000000000022 \*\*\*

Doors 1 705281249 705281249 10.596 0.001181 \*\*

Residuals 801 53313282459 66558405

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="confidence",level = 0.95)

fit lwr upr

1 23340.75 22177.94 24503.56

> predict(reg1,newdata=data.frame(Liter = 3.1,Doors=2),interval="prediction",level = 0.95)

fit lwr upr

1 23340.75 7284.348 39397.15

>

> plot(reg1$fitted.values, reg1$residuals)

> abline(0,0, col = 'red')

>

>

> hist(reg1$residuals, breaks = 30)

>

> qqnorm(reg1$residuals)

> qqline(reg1$residuals)