Ben Goodwin

HW9

DS6371

Question 1)

Beta\_hat\_1 = (Sum (X\_i-X\_bar) (Y\_i-Y\_bar) )/ (sum X\_i-X\_bar)^2

(Sum (X\_i-X\_bar) (Y\_i-Y\_bar) )=4461

(sum X\_i-X\_bar)^2=42247.37

Beta\_hat\_1 = (Sum (X\_i-X\_bar) (Y\_i-Y\_bar) )/ (sum X\_i-X\_bar)^2=0.10559232

Beta\_hat\_1 = 4461/42247.37^2=0.1056

Beta\_0\_hat= Y\_bar-Beta\_1\_hat\*X\_bar

Y\_bar=81

X\_bar=90.233

Beta\_0\_hat=81-.1056\*90.2333=71.4713635

SE(beta\_1\_hat)=sigma\_hat\*(sqrt(1/(n-1)s\_x^2))

SE(beta\_0\_hat)=sigma\_hat\*(sqrt(1/n + xbar^2/(n-1)s\_x^2))

Df=30-2

Df=28

SE(beta\_0\_hat)=4.9549

SE(beta\_1\_hat)=0.0507

Y\_hat =Beta\_0\_hat+beta\_1\_hat\*x\_i

1. Least squares regression equation

y=71.4713635+0.1056\*x\_i

1. Interpretation of slope and intecept

Slope interpretation: Each 1 unit increase in payroll results in a 0.1056 increase in wins.

Intercept interpretation: The predicted number of wins with a 0 payroll is 71.4713645

B) Six-step hypothesis test for slope

**Critical Values: = \_2.048\_**

**Test Statistic: 2.083**

**Pvalue: 0.0465**

**Decision: Reject Ho!**

Conclusion: We will reject the null hypothesis and conclude that the slope is different from zero.

C) CI for Beta\_1

95% CI: = 0.1056+/-2.083\*0.0507= [0.00173383,0.2112081]

D) Verification of results:

SAS input:

data MLB;

input TEAM $ PAYROLL WINS;

datalines;

NYY 206 95

BOS 162 89

CHC 146 75

PHI 142 97

NYM 134 79

DET 123 81

CHW 106 88

LAA 105 80

SF 99 92

MIN 98 94

LAD 95 80

HOU 92 76

SEA 86 61

STL 86 86

ATL 84 91

COL 84 83

BAL 82 66

MIL 81 77

TB 72 96

CIN 71 91

KC 71 67

TOR 62 85

ARZ 61 65

CLE 61 69

WAS 61 69

FA 57 80

TEX 55 90

OAK 52 81

SD 38 90

PIT 35 57

;

run;

proc reg data=MLB;

model WINS = PAYROLL;

run;

SAS output:

Table

Description automatically generated

Chart

Description automatically generated

Six-step hypothesis test for slope

**Critical Values: = \_2.048\_**

**Test Statistic: 2.083**

**Pvalue: 0.0465**

**Decision: Reject Ho!**

Conclusion: We will reject the null hypothesis and conclude that the slope is different from zero.

95% CI: (0.00173383 0.2094509)

Question 2)

1. I) y=21.70019+0.59681xi

II) Slope interpretation: Each 1 unit increase in science results in a 0.59681 increase in math.

Intercept interpretation: The predicted science score with a 0 math score is 21.70019.

R Input:

A picture containing graphical user interface

Description automatically generated

R output:

Text, letter

Description automatically generated

B)

Six-step hypothesis test for slope

**Critical Values: = \_2.326\_**

**Test Statistic: 11.437**

**Pvalue: 2.e^-16**

**Decision: Reject Ho!**

Conclusion: We will reject the null hypothesis and conclude that the slope is different from zero.

Six-step hypothesis test for slope

**Critical Values: = \_2.326\_**

**Test Statistic: 7.879**

**Pvalue: 2.15e^-13**

**Decision: Reject Ho!**

Conclusion: We will reject the null hypothesis and conclude that the slope is different from zero.

R output:

Text, letter

Description automatically generated

C) 99% CI for slope and intercept:

99% CI for slope: 0.59681-2.326x0.05217, 0. 59681+2.326x0.05217

=[0.461094,0.7325341]

99% CI for intercept: 21.70019-2.326x2.75429, 21.70019+2.326x2.75429

= [14.536591,28.8637921]

II) Interpretation

We are 99% confident that the population parameter for the slope is between 0.461094 and 0.732534.

We are 99% confident that the population parameter for the intercept is 14.536591,28.8637921.

D) Verification in R

R input:

A picture containing logo

Description automatically generated

R output:

A picture containing text

Description automatically generated

Bonus:

3 1D in R

R input:

A picture containing text

Description automatically generated

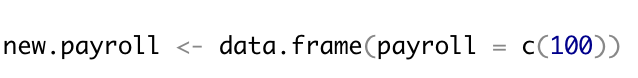
R output:

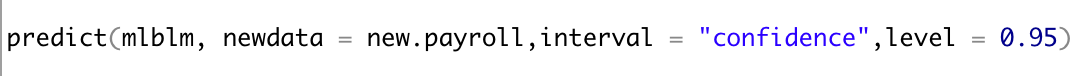
Text

Description automatically generated

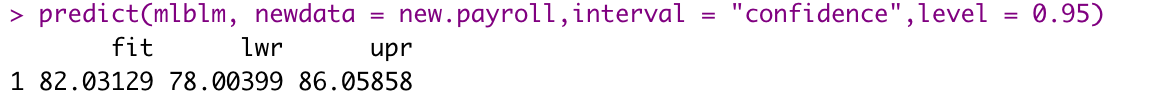
5)

A) R input:

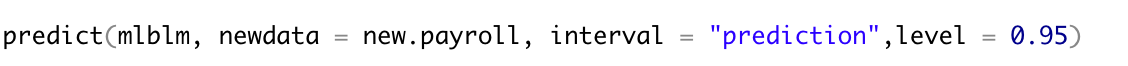
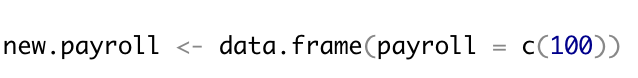




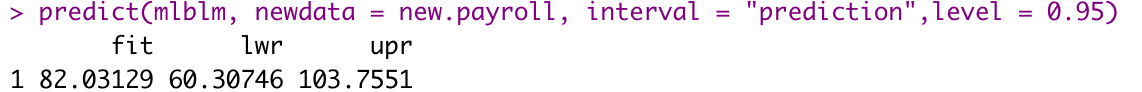
R output:



B) R input:



R output:



C) A prediction interval shows the uncertainty around a single value, while a confidence interval reflects the uncertainty around the mean prediction values. Thus, a prediction interval will be generally much wider than a confidence interval for the same value.