Homework 9

Code

6210422036: Tanat Iempreedee

Hide

```
library(ggplot2)
library(fpp2)
library(forecast)
library(GGally)

data <- read.table('c5p9.csv',sep=',',header=TRUE)
data
```

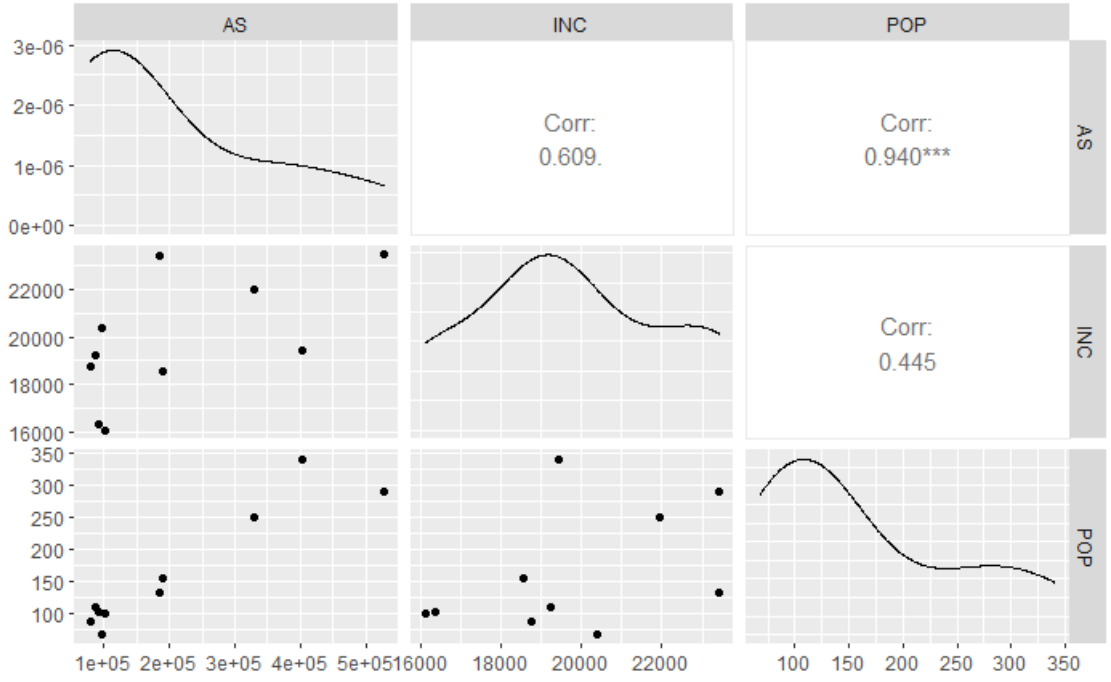
Area <int>	AS <int>	INC <int>	POP <dbl>
1	185792	23409	133.17
2	85643	19215	110.86
3	97101	20374	68.04
4	100249	16107	99.59
5	527817	23423	289.52
6	403916	19426	339.98
7	78283	18742	89.53
8	188756	18553	155.78
9	329531	21953	248.95
10	91944	16358	102.13

1-10 of 10 rows

Correlations

Hide

```
ggpairs(data, columns=2:4)
```



Fit linear regression

Hide

```
reg <- lm(AS~INC+POP, data=data)
summary(reg)
```

```

Call:
lm(formula = AS ~ INC + POP, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-46425 -32764  -1817  17198  89799

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.075e+05  1.263e+05  -2.435 0.045076 *
INC          1.456e+01  6.884e+00   2.115 0.072276 .
POP          1.397e+03  1.875e+02   7.452 0.000143 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 47640 on 7 degrees of freedom
Multiple R-squared:  0.9295,    Adjusted R-squared:  0.9094
F-statistic: 46.17 on 2 and 7 DF,  p-value: 9.288e-05
```

1) Estimate values for b_0, b_1, b_2 for model

- $\beta_0 = -307500$
- $\beta_1 = 14.56$
- $\beta_2 = 1397$

$AS = -307500 + 14.56(INC) + 1397(POP)$

2) Are the signs you find for the coefficients consistent with your expectations?

Both coefficients for INC and POP are positive
- Area with higher income should have a higher sales
- Also, greater population should increase the sales

3) Are the coefficients for the two explanatory variables significantly different from zero?

- **INC**: β_1 is significantly different from zero but at significance level of 0.9
- **POP**: β_2 is significantly different from zero

4) What percentage of the variation in AS is explained by this model

R-squared = 0.9295

This model can explain the variation in AS by 92.95%

5) What point estimate of AS would you make for a city where INC=23175 and POP=128.07? Also find the 95% prediction interval

Hide

```
newX <- data.frame(INC=23175, POP=128.07)
forecast(reg, newdata = newX)
```

	Point Forecast <dbl>	Lo 80 <dbl>	Hi 80 <dbl>	Lo 95 <dbl>	Hi 95 <dbl>
1	208811.3	128328.6	289294	74308.43	343314.2

1 row

- Point estimate = 208,811.3
- Prediction interval = [74308.43, 343314.2]