**SPSS Practical 6:**

**Part A:**

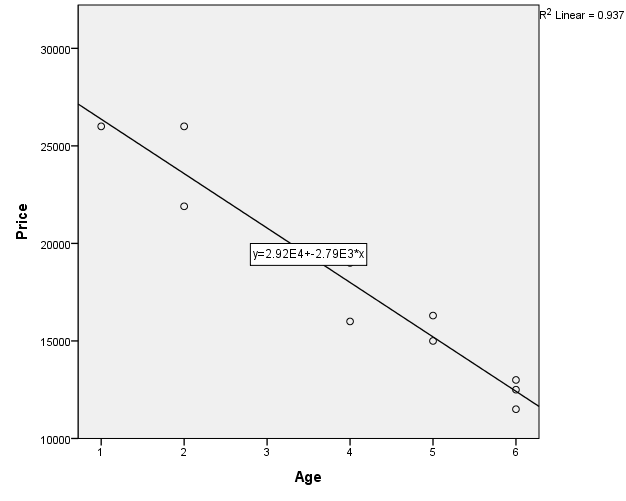
The following data were randomly selected from a sales record, where denote age, in years, and denote sales price, in hundreds of dollars.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 6 | 6 | 6 | 4 | 2 | 5 | 4 | 5 | 1 | 2 |
|  | 125 | 115 | 130 | 160 | 219 | 150 | 190 | 163 | 260 | 260 |

**Answer the following questions:**

1. Use a graph to determine if there is any possible linear relationship between & .

// 어떤것이 predictor 인지 내가 알아 낼 수 있어야함. 논리적으로.



Therer is negative relationship between P & Q

(R = 0.967)

1. Run a Pearson’s test to verify your answer in Part (a).

// 1. 전부 스케일이어야함. 2. Linear 관계가 있어야함. 3. Nomal 추정가능.

여기서는 1번에서 negative임을 확인했으므로 H1 이 r<0 로 해야함.

H1 r!=0 로 할 경우 나머지 다 틀림. 그러므로 pearson 그릴 때에도 onetail 사용해야함.

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Age | Price |
| Age | Pearson Correlation | 1 | -.968\*\* |
| Sig. (1-tailed) |  | .000 |
| N | 10 | 10 |
| Price | Pearson Correlation | -.968\*\* | 1 |
| Sig. (1-tailed) | .000 |  |
| N | 10 | 10 |
| \*\*. Correlation is significant at the 0.01 level (1-tailed). | | | |

: r = 0

: r < 0

The correlation is r(10) = -0.968, sig = 0.000(<0.05)

This test is significant

Reject

Conclusion : There is a negative relationship.

1. Compute the covariance between and .

cov(P,Q) = -9580

|  |  |  |  |
| --- | --- | --- | --- |
| **Correlations** | | | |
|  | | Age | Price |
| Age | Pearson Correlation | 1 | -.968\*\* |
| Sig. (1-tailed) |  | .000 |
| Sum of Squares and Cross-products | 30.900 | -86220.000 |
| Covariance | 3.433 | -9580.000 |
| N | 10 | 10 |
| Price | Pearson Correlation | -.968\*\* | 1 |
| Sig. (1-tailed) | .000 |  |
| Sum of Squares and Cross-products | -86220.000 | 256816000.000 |
| Covariance | -9580.000 | 28535111.111 |
| N | 10 | 10 |
| \*\*. Correlation is significant at the 0.01 level (1-tailed). | | | |

지난번과 똑같음, option 에 cross-product deviation 이용.

1. Determine the regression equation for the data.

Analye -> regression -> linear 여기서 독립, 종속변수 제대로 넣을것

//This is just memo not answer

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVAa** | | | | | | |
| Model | | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 240578912.621 | 1 | 240578912.621 | 118.533 | .000b |
| Residual | 16237087.379 | 8 | 2029635.922 |  |  |
| Total | 256816000.000 | 9 |  |  | ㅇ |
| a. Dependent Variable: Price | | | | | | |
| b. Predictors: (Constant), Age | | | | | | |

여기서 처음부터 SSm, SSr, SSt임을 알 수 있음.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Coefficientsa** | | | | | | |
| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
| B | Std. Error | Beta |
| 1 | (Constant) | 29160.194 | 1143.290 |  | 25.506 | .000 |
| Age | -2790.291 | 256.289 | -.968 | -10.887 | .000 |
| a. Dependent Variable: Price | | | | | | |

Here is the answer.

Sales = 29160.194 -2790.291(Age)

//여기서도 변수 위치가 바뀌지 않도록 조심할 것.

1. Compute and interpret the coefficient of determination, .

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model Summary** | | | | |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .968a | .937 | .929 | 1424.653 |
| a. Predictors: (Constant), Age | | | | |

//여기에서 r은 pearson 의 r임. R square = coefficient of determination Adjusted R square = population 의 R제곱을 추정해서 보여주는것임. 만약 두 값 차이가 크다면 좋지 않은 data임.

여기서는 r제곱을 이용해서 결론을 내리라고 했으므로

양,음 관계를 쓸 수 없음. 강한지 moderate 인지만 적을 것.

There is strong relationship because = 0.937

여기서 R이 양수라고 하더라도 이건 r제곱을 이용해서 계산한 것 이기 때문에 의미가 없이 무조건 양수가 나옴.

스케터플롯을 확인 해 봐도 이 data의 R은 음수임.

1. Find the predicted sales price when .

Sales = 29160.194 -2790.291(Age)

Sales = 29160.194 -2790.291(7) = 9628.157 9628

1. Find the predicted age when the sales price is $15,000.

15000 = 29160.194 -2790.291(Age) > Age = 5.0748 5

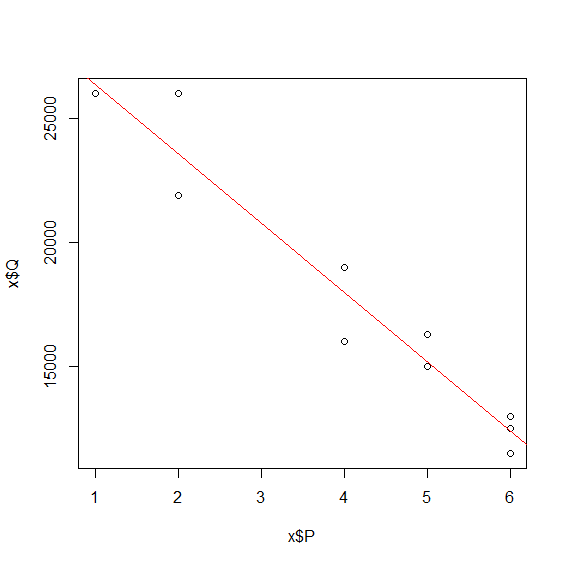
**Part B:**

By using the R Program, compute the following questions: (same dataset as Part A).

1. Graph the data in a scatterplot to determine any possible linear relationship.]

> plot(x$P,x$Q)

> abline(lm(x$Q~x$P),col="red")



1. Compute covariance and Pearson’s correlation coefficient.

Covariance:

>cov() 여기에 인자넣으면됨.

> cov(x$P,x$Q)

[1] -9580

> cor(x$Q,x$P)

[1] -0.9678716

순서를 바꿔서 넣는 것은 그래프를 그릴 것이 아니라면 상관 x

1. Determine the regression equation for the data.

> lm(x$Q~x$P)

Call:

lm(formula = x$Q ~ x$P)

Coefficients:

(Intercept) x$P

29160 -2790

여기서 intercept가 b0다. 뒤에 녀석이 b1

1. Compute and interpret the coefficient of determination, .

>summary(lm(x$Q~x$P))

Multiple R-squared: 0.9368

1. Perform Pearson, Spearman and Kendall test.

각각의 코드

Pearson

>cor.test(x$P,x$Q)

Pearson's product-moment correlation

data: x$P and x$Q

t = -10.887, df = 8, p-value = 4.484e-06

alternative hypothesis: true correlation is not equal to 0

95 percent confidence interval:

-0.9926062 -0.8659576

sample estimates:

cor

-0.9678716

Spearman

>cor.test(x$P,x$Q, method="spearman")

Spearman's rank correlation rho

data: x$P and x$Q

S = 321.32, p-value = 3.149e-05

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

-0.9473695

Kendall’s Tau

>cor.test(x$P,x$Q, method="kendal")

Kendall's rank correlation tau

data: x$P and x$Q

z = -3.3216, p-value = 0.0008951

alternative hypothesis: true tau is not equal to 0

sample estimates:

tau

-0.8690482

Pearson 은 scale 용

나머지는 categorical 이다.

R은 data를 cut 해서 scale을 카테고리로 바꾸지 못함 그래서 warning 이 발생함.

Spss는 cut 해서 바꿔서 넣을 수 있음.