Week 6

ASSIGNMENT 6

For the following set of questions 1, 2, 3, 4, 5 use the dataset bonds.txt. This dataset contains 2 variables, Coupon rate and Bid price.

1) What is the relationship between the variables, Coupon rate and Bid price?

- Coupon rate = 99.95 + 0.24 * Bid price
- Bid price = 99.95 + 0.24 * Coupon rate
- Bid price = 74.7865 + 3.066 * Coupon rate
- Coupon rate = 74.7865 + 3.066 * Bid price

Accepted Answers:
*Bid price = 74.7865 + 3.066 * Coupon rate*

2) Choose the correct option that best describes the relation between the variables Coupon rate and Bid price in the given data.

- Strong positive correlation
- Weak positive correlation
- Strong negative correlation
- Weak negative correlation

Accepted Answers:
Strong positive correlation

3) What is the R-Squared value of the model obtained in $\mathrm{Q1}$?

- 0.2413
- 0.12
- 0.7516
- 0.5

Accepted Answers:
0.7516

4) What is the adjusted R-Squared value of the model obtained in $\mathrm{Q1}$?

- 0.22
- 0.7441
- 0.088
- 0.5

Accepted Answers:
0.7441

5) Based on the model relationship obtained from Q1, what is the residual error obtained while calculating the bid price of a bond with coupon rate of 3?

- 10.5155
- -10.5155
- 6.17
- 0

Accepted Answers:
10.5155

$$Bid price = 74.7866 + 3.0661 (Coupon Price)$$

$$\downarrow 3$$

$$Bid Price = 74.7866 + 3.066 \times 3$$

$$= 83.9845$$

$$y|_{x=3} = 94.50$$

$$y - \hat{y} = 94.50 - 83.7845$$

$$=$$

6) State whether the following statement is True or False.
Covariance is a better metric to analyze the association between two numerical variables than correlation.

- True
- False

Answer: b

Solution

Correlation is a better metric to determine the association than covariance as correlation does not have any units and is also not dependent on the range of values of the variables.

Accepted Answers:
False

7) If R^2 is 0.6, SSR=200 and SST=500, then SSE is

- 500
- 200
- 300
- None of the above

Answer: c

Solution:

SSR = SST - SSE

SST=SSR+SSE
SSE=SST-SSR
SSE=500-200
SSE=300

Accepted Answers:
300

8) Linear Regression is an optimization problem where we attempt to minimize

- SSR (residual sum-of-squares)
- SST (total sum-of-squares)
- SSE (sum-squared error)
- Slope

Answer: c

Linear Regression is a minimization problem where we tend to minimize the loss function or the sum-squared error (SSE) function, (i.e.,)

SSE

$$min \sum_{i=1}^{n} (y_i - \underbrace{\beta_0 - \beta_1 x_i)}_{\hat{y}}^2$$

Accepted Answers:
SSE (sum-squared error)

The model built from the data given below is $\underbrace{Y = 1.39x + 6.09}$. Find the values for R^2 and Adjusted R^2.

β_1 β_0

X	8.55	0.32	8.4	6.96	3.23	8.18
Y	17.03	6.19	16.53	16.06	10.92	19.34

$n=6$

a. R^2 is 0.95 and Adjusted R^2 is 0.93
b. R^2 is 0.69 and Adjusted R^2 is 0.67
c. R^2 is 0.93 and Adjusted R^2 is 0.95
d. None of the above

S_{xx}

S_{yy}

S_{xy}

n

R^2

R_a^2

Answer: a

Solution:

```
> x=c(8.55,0.32,8.4,6.96,3.23,8.18)
> y=c(17.03,6.19,16.53,16.06,10.92,19.34)
> n=6
> sxx=sum(x^2)-sum(x)^2 /n
> syy=sum(y^2)-sum(y)^2 / n
> sxy=sum(x * y)-(sum(x)*sum(y))/n
> # Coefficient of determination R-squared
> sse=syy - sxy^2 / sxx
> r2=(syy - sse) /syy
> round(print(r2),2)
[1] 0.9468673
[1] 0.95
> # R-squared adjusted
> r2_adj=r2 - (1 - r2) * ((2 - 1) / (length(y) - 2))
> round(print(r2_adj),2)
[1] 0.9335841
[1] 0.93
```

10) Identify the parameters β_0 and β_1 that fits the linear model $\beta_0 + \beta_1 x$ using the following information: total sum of squares of X, $SS_{XX} = 52.53$, $SS_{XY} = 52.01$, mean of X, $\bar{X}$=4.46, and mean of Y, $\hat{Y}$=6.32.

- 1.9 and 0.99
- 10.74 and 1.01
- 4.42 and 1.01
- None of the above

Answer: a

Solution:

$$\beta_1 = \frac{SS_{XY}}{SS_{XX}} = \frac{52.01}{52.53} = 0.99$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 6.32 - (0.99 * 4.46) = 1.9$$

Accepted Answers:
1.9 and 0.99

Extra Questions

1) Choose the correct option that best describes the relation between the variables x and y in the given data

X	-2	-1	5	-5	2
Y	-3	-4	-10	0	-7

- Randomly sampled
- Negatively correlated
- Positively correlated
- None of the above

Answer: b

Solution:

```
> x=c(-2,-1,5,-5,2)
> y=c(-3,-4,-10,0,-7)
> cor(x,y)
[1] -1
```

The model built from the data given below is $Y = 0.2x + 60$. Find the values for R^2 and Adjusted R^2 .

X	80	75	85	70	65
Y	85	70	80	95	70

- R^2 is 0.022 and Adjusted R^2 is −0.303
- R^2 is 0.022 and Adjusted R^2 is −0.0303
- R^2 is 0.022 and Adjusted R^2 is 0.303
- None of the above

Accepted Answers:
R^2 is 0.022 and Adjusted R^2 is −0.303

For the following set of questions 5,6,7 use the dataset ***women.csv***. This dataset contains 2 variables ***height*** **(in cms)** and ***weight*** **(in kgs)**

What is the relationship between the variables height and weight?

a. weight = -87.51667 + 3.45 * height
b. height = 3.45 * weight - 87.51667
c. weight = 0.28 * height + 25.723456
d. None of the above

Answer: a

Solution:

```
> data=read.csv("women.csv")
> model_1=lm(weight~height,data=data)
> print(model_1)

Call:
lm(formula = weight ~ height, data = data)

Coefficients:
(Intercept)       height
     -87.52         3.45
```

What is the R-Squared value of the model obtained in Q5?

a. 0.7417
b. 0.991
c. 0.583
d. None of the above

Answer: b

Solution

summary(model_1)

```
Residual standard error: 1.525 on 13 degrees of freedom
Multiple R-squared:  0.991,	Adjusted R-squared:  0.9903
F-statistic:  1433 on 1 and 13 DF,  p-value: 1.091e-14
```

The R-squared value for the model fitted between the variables ***weight*** (as y, in kg) and ***height*** (as x, in cm) is found to be 0.991

Based on the model relationship obtained from Q5, what is the residual error obtained while calculating the weight of a woman with height 69 cms?

a. -361.08333
b. 0.63333
c. 0.0345
d. 0.53333

Answer: d

Solution:

Residual Error = Predicted Value - Actual Value

Actual Value = 150 kgs; Predicted Value = (3.45 * 69 - 87.51667) = 150.53333 kgs

Residual Error = 150.53333 - 150 = 0.53333 kgs

Practice Questions

1) The higher the value of R^2 for a model, the observations are more closely grouped around:

- the origin
- the best fit line
- average values of the predicted variable
- the intercept

2) The standard assumption of ordinary least squares regression is that:

- ○ there are no errors in measurements of independent and dependent variables
- ○ there are errors only in measurement of independent variable
- ✓ there are errors only in measurement of dependent variable
- ○ there are errors both in measurements of independent and dependent variables

3) The relationship between the dependent and independent variables in a simple linear regression is described by

- ○ F-statistic
- ○ predicted value and error
- ○ standardised residuals
- ○ coefficient and intercept

4) The Pearson's correlation coefficient between two parameters is calculated to be 0.10. What can be inferred from the correlation coefficient egarding
the relationship between the two parameters?

- ○ There exists a weak negative relationship between two variables
- ○ There exists a strong negative relationship between two variables
- ○ There exists a weak positive relationship between two variables
- ○ Correlation coefficient cannot possess this value

5) Standardised residuals have:-

- binomial distribution with n degrees of freedom
- t distribution with n-2 degrees of freedom
- log-normal distribution with n-2 degrees of freedom
- chi-square distribution with n degrees of freedom

Independent Dependent

Predictor Response