Quiz 6

JIraphan

1 December 2563

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

2. Physicians depend on laboratory test results when managing medical problems such as diabetes or epilepsy. In a uniformity test for glucose tolerance, three different laboratories were each sent nt = 5 identical blood samples from a person who had drunk 50 milligrams (mg) of glucose dissolved in water. The laboratory results (in mg/dl) are listed here:

For each $i = 1,2,3$, let μ_i be the mean of test results (in mg/dl) from Lab i.
The null hypothesis is H_0: $\mu_1 = \mu_2 = \mu_3$.
The alternative hypothesis is H_1: At least one μ_i is different from others.

```
Lab1 <- c(121.3,111.9,110.1,105.4,101.6)
Lab2 <- c(99.5,113.2,108.9,109.1,100.4)
Lab3 <- c(104.2,109.7,102.3,111.2,106.6)
DAT2 <- cbind.data.frame("Lab1"=Lab1, "Lab2"=Lab2,"Lab3"=Lab3)
DAT2
```

```
##    Lab1  Lab2  Lab3
## 1 121.3  99.5 104.2
## 2 111.9 113.2 109.7
## 3 110.1 108.9 102.3
## 4 105.4 109.1 111.2
## 5 101.6 100.4 106.6
```

```
DAT2stacked=stack(DAT2)
DAT2stacked
```

```
##    values  ind
## 1   121.3 Lab1
## 2   111.9 Lab1
## 3   110.1 Lab1
## 4   105.4 Lab1
## 5   101.6 Lab1
## 6    99.5 Lab2
## 7   113.2 Lab2
## 8   108.9 Lab2
## 9   109.1 Lab2
```

```
## 10  100.4 Lab2
## 11  104.2 Lab3
## 12  109.7 Lab3
## 13  102.3 Lab3
## 14  111.2 Lab3
## 15  106.6 Lab3
```

```
mod <- aov(values ~ as.factor(ind), data=DAT2stacked)
summary(mod)
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## as.factor(ind)  2   42.8   21.42   0.611  0.559
## Residuals      12  421.1   35.09
```

From the anova table, p-value is 0.559 which is greater than 0.05. Then we do not reject the null hypothesis H_0. Then there is NO differenc e in average readings for the three labatories.

3. A professor of economics wants to study the relationship between income (y in $1000s) and education (x in years). A random sample eight individuals is taken and the results are shown below.

```
Education <- c(16,11,15,8,12,10,13,14)
Income <- c(58,40,55,35,43,41,52,49)
mod3 <- lm(Income ~ Education)
summary(mod3)
```

```
##
## Call:
## lm(formula = Income ~ Education)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.624 -2.398  0.782  1.150  3.556
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.6165     4.3539   2.438 0.050583 .
## Education     2.9098     0.3449   8.437 0.000151 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.436 on 6 degrees of freedom
## Multiple R-squared:  0.9223, Adjusted R-squared:  0.9093
## F-statistic: 71.18 on 1 and 6 DF,  p-value: 0.0001513
```

Determine the coefficient of determination. $R^2 = 0.9223$.

Discuss what its value tells you about the two variables.

92.03 % of the variation in income is explained by the variation in years of education.

Calculate the Pearson correlation coefficient.

```r
sqrt(0.9223)

## [1] 0.9603645
```

$r = 0.9603645$.

Why does it have the sign it has?

It has a positive sign since the slope of the regression line (= 2.9098) is positive.

What is the equation of the estimated regression line?

$Income = 10.6165 + 2.9098 \times Education$

Is there a linear relationship between education and income? Explain?

Yes, since R^2 close to 1.

4. A study was conduced to determine the effect of extra help sessions attended on students ability to avoid mistakes on a 20- multiple choice test. The data shown below represent the number of extra help sessions attended (x) and the average number of mistakes (y) recorded.

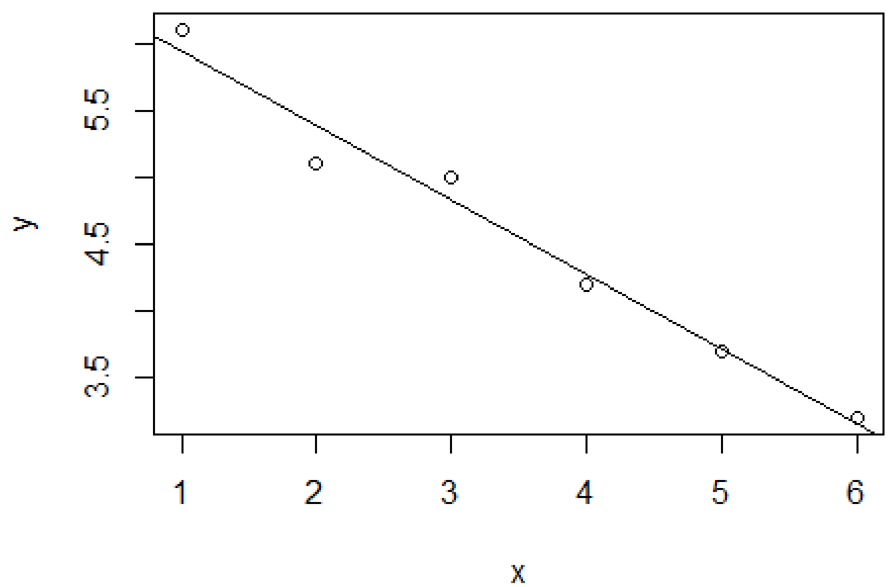
```r
x <- 1:6
y <- c(6.1,5.1,5.0,4.2,3.7,3.2)
mod4 <-lm(y ~ x)
summary(mod4)

##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##        1        2        3        4        5        6
##  0.15714 -0.28571  0.17143 -0.07143 -0.01429  0.04286
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.50000    0.17593   36.95  3.2e-06 ***
## x           -0.55714    0.04518  -12.33 0.000248 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.189 on 4 degrees of freedom
## Multiple R-squared:  0.9744, Adjusted R-squared:  0.968
## F-statistic: 152.1 on 1 and 4 DF,  p-value: 0.0002484
```

Use the regression formulas to find the least-squares line for the data.

$\hat{y} = 6.50000 - 0.55714 \times x$

Plot the six points and graph the line.

```
plot( y ~ x)
abline( a=6.50000,b=-0.55714)
```



Does the line appear to provide a good fit to the data points?

Yes.

Use the least-Squares line to predict the value of y when x = 3.5.

```
6.50000-0.55714*3.5

## [1] 4.55001
```

Do the data provide sufficient evidence to indicate that y and x are linearly related at the 1% level of significance?

Yes, since $p - value = 0.000248 ***$.

The Multiple R-squared is 0.9744. Then 97.44 % of the variation of y can be explained by the variation in the $x -$ value.