Assignment4_JianghongMan

Jianghong Man

9/28/2020

Chapter 14

Question 2

a) Y' = 2 * 6 + 9 = 21
b) 14 = 2X + 9, then X = 2.5

Question 4

The standard error of the estimate is a measure of the accuracy of predictions. The formula is as follows:

For population:

$\sigma_{est} = \sqrt{\frac{\Sigma(Y-Y')^2}{N}}$

For sample:

$\sigma_{est} = \sqrt{\frac{\Sigma(Y-Y')^2}{N-2}}$

Question 6 chongxinsuan

```
x <- c(2,4,4,5,6)
y <- c(5,6,7,11,12)
res <- cor.test(x, y, method = "pearson")
res
```

```
##
##  Pearson's product-moment correlation
##
## data:  x and y
## t = 3.7819, df = 3, p-value = 0.0324
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1360488 0.9940671
## sample estimates:
##       cor
## 0.9091846
```

```
r <- lm(y ~ x)
summary(r)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##       1       2       3       4       5
##  1.0000 -1.8182 -0.8182  1.2727  0.3636
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.1818     2.2234   0.082   0.9400
## x             1.9091     0.5048   3.782   0.0324 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.497 on 3 degrees of freedom
## Multiple R-squared:  0.8266, Adjusted R-squared:  0.7688
## F-statistic:  14.3 on 1 and 3 DF,  p-value: 0.0324
```

a) r = 0.9092; p = 0.0324. Therefore, r is significantly different from 0 and we can reject the null hypothesis in this case.

b) Slope = 1.9091, intercept = 0.1818. p-value is 0.0324 for the slope, which indicates that it differs significantly from zero.

c) t for confidence interval is 3.7819, and 95 percent confidence interval:

Lower limit = 0.9940671

Upper limit = 0.1360488

Question 8

t = 1.8517, df = 18, p = 0.0805. Thus, it is not significant at 0.05 level.

Question 10

```r
pre <- c(59,52,44,51,42,42,41,45,27,63,54,44,50,47,55,49,45,57,46,60,65,64,50,74,59)
post <- c(56,63,55,50,66,48,58,36,13,50,81,56,64,50,63,57,73,63,46,60,47,73,58,85,44)
re <- lm(post ~ pre)
summary(re)
```

```
##
## Call:
## lm(formula = post ~ pre)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -24.401  -6.351   2.288   6.486  22.354
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  16.1552    13.5774   1.190  0.24624
```

```
## pre           0.7869     0.2596   3.032  0.00593 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.61 on 23 degrees of freedom
## Multiple R-squared:  0.2855, Adjusted R-squared:  0.2544
## F-statistic: 9.191 on 1 and 23 DF,  p-value: 0.005933
```

Post = 16.1552 + 0.7869 x Pre

Y' = 0.7869 * 43 + 16.1552 = 49.9919

Question 12

b = r * (Sy/Sx) = -0.6 * (3 / 2.5) = -0.72

A = My - b * Mx = 12 - (-0.72) * 10 = 19.2

Thus, Y' = 19.2 - 0.72X

Question 14

True

Question 16

False

Question 18

```r
AM_data <- read_excel("angry_moods.xls")
colnames(AM_data) <- c("Gender", "Sports", "Anger_Out", "Anger_In",
                       "Control_Out", "Control_In", "Anger_Expression")

r2 <- lm(AM_data$Anger_Out ~ AM_data$Control_Out)
summary(r2)
```

```
##
## Call:
## lm(formula = AM_data$Anger_Out ~ AM_data$Control_Out)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.488 -2.440 -0.295  2.193 10.560
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         28.49482    2.02477   14.07  < 2e-16 ***
## AM_data$Control_Out -0.52413    0.08386   -6.25 2.18e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.45 on 76 degrees of freedom
## Multiple R-squared:  0.3395, Adjusted R-squared:  0.3308
## F-statistic: 39.07 on 1 and 76 DF,  p-value: 2.183e-08
```

a) slpoe = -0.52413

b) intercept = 28.49482

c) Yes

d) p = 2.18e-08 < .001, significant

e) standard error = 3.45

Question 19

```r
sat_data <- read_excel("sat.xls")
colnames(AM_data) <- c("high_GPA", "math_SAT", "verb_SAT", "comp_GPA",
                       "univ_GPA")

regression_sat <- lm(sat_data$univ_GPA ~ sat_data$high_GPA)
summary(regression_sat)
```

```
##
## Call:
## lm(formula = sat_data$univ_GPA ~ sat_data$high_GPA)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.69040 -0.11922  0.03274  0.17397  0.91278
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        1.09682    0.16663   6.583 1.98e-09 ***
## sat_data$high_GPA  0.67483    0.05342  12.632  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2814 on 103 degrees of freedom
## Multiple R-squared:  0.6077, Adjusted R-squared:  0.6039
## F-statistic: 159.6 on 1 and 103 DF,  p-value: < 2.2e-16
```

a) slope = 0.67483

b) y-intercept = 1.09682

c) Y' = 2.2 * 0.67483 + 1.09682 = 2.581446

d) Y' = 4.0 * 0.67483 + 1.09682 = 3.79614

Chapter 16

Question 1

The log transformation can be used to make highly skewed distributions less skewed. It can be valuable both for making patterns in the data more interpretable and for making the relationship between variables clearer.

Question 2

1000

Question 3

3,4,5

Question 4

-2

Question 5

2