hw1

Jack

2023-09-25

```
setwd("C:/Users/jackw/Documents/R")
```

QUESTION

NUMBER 1

reading table (q1)

```
library(readr)
indicators <- read_table("datasets/indicators.txt")
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   MetroArea = col_character(),
##   PriceChange = col_double(),
##   LoanPaymentsOverdue = col_double()
## )
```

```
View(indicators)
```

running the linear model function on indictors.csv (q1)

```
library(ISwR)
fit <- lm(PriceChange ~ LoanPaymentsOverdue, indicators)
summary(fit)
```

```
##
## Call:
## lm(formula = PriceChange ~ LoanPaymentsOverdue, data = indicators)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6541 -3.3419 -0.6944  2.5288  6.9163
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           4.5145     3.3240   1.358   0.1933
## LoanPaymentsOverdue  -2.2485     0.9033  -2.489   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.954 on 16 degrees of freedom
## Multiple R-squared:  0.2792, Adjusted R-squared:  0.2341
## F-statistic: 6.196 on 1 and 16 DF,  p-value: 0.02419
```

Q2.2 (A)

finding the inital confidence interval for the lm (q1.1)

```
confint(fit, 'LoanPaymentsOverdue', level=0.95)
```

```
##                         2.5 %     97.5 %
## LoanPaymentsOverdue -4.163454 -0.3335853
```

```
paste("There is evidence of a significant, negative linear association because it's highly pr
oabable that independent variable LoanPaymentsOverdue is negative.")
```

```
## [1] "There is evidence of a significant, negative linear association because it's highly p
roabable that independent variable LoanPaymentsOverdue is negative."
```

Q2.2 (B)

using the lm to estimate E(Y|X=4) and find a predicted 0.95 confidence interval (q1.2)

```
xvalue <- 4
# predict is best used when setting independent variable values.
predict(fit, data.frame(LoanPaymentsOverdue = xvalue), interval = "confidence")
```

```
##         fit       lwr       upr
## 1 -4.479585 -6.648849 -2.310322
```

```
paste("0% is NOT a feasible value for E(Y|X=4) because the upperbound of the confidence inter
val is below the zero value.")
```

```
## [1] "0% is NOT a feasible value for E(Y|X=4) because the upperbound of the confidence inte
rval is below the zero value."
```

QUESTION

NUMBER 2

reading table (q2)

```
library(readr)
invoices <- read_table("datasets/invoices.txt")
```

```
##
## ── Column specification ────────────────────────────────────────────────────────
## cols(
##   Day = col_double(),
##   Invoices = col_double(),
##   Time = col_double()
## )
```

```
View(invoices)
```

running the linear model function on invoices.csv (q1)

```
library(ISwR)
fitq2 <- lm(Time ~ Invoices, invoices)
summary(fitq2)
```

```
##
## Call:
## lm(formula = Time ~ Invoices, data = invoices)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.59516 -0.27851  0.03485  0.19346  0.53083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.6417099  0.1222707   5.248 1.41e-05 ***
## Invoices    0.0112916  0.0008184  13.797 5.17e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3298 on 28 degrees of freedom
## Multiple R-squared:  0.8718, Adjusted R-squared:  0.8672
## F-statistic: 190.4 on 1 and 28 DF,  p-value: 5.175e-14
```

Q2.3 (A)

manually computing confidence interval (q2.1)

confint() only works for independent variables and not for model parameters i.e. b0.

```
confint(fitq2, '(Intercept)', level=0.95)
```

```
##                 2.5 %    97.5 %
## (Intercept) 0.3912496 0.8921701
```

Q2.3 (B)

```r
# t-test
# h0 -> B1 = 0.01
# h0 -> B1 != 0.01
invoice_benchmark <- 0.01
invoice_estimated_value <- 0.0112916
invoice_standard_error <- 0.0008184
# how many standard errors the estimated value is from the hypothesized value
invoice_t_value <- (invoice_estimated_value - invoice_benchmark) / invoice_standard_error
paste("t_value:", invoice_t_value)
```

```
## [1] "t_value: 1.57820136852395"
```

```r
# degrees of freedom = sample size - parameters being estimated
degrees_of_freedom <- 28

# probability of observing a t-value as extreme as the one calculated, assuming the h0=TRUE.
invoice_p_value <- 2 * pt(abs(invoice_t_value), degrees_of_freedom, lower.tail = FALSE)
paste("p_value:", invoice_p_value)
```

```
## [1] "p_value: 0.125751694033898"
```

```r
paste("We fail to reject the null hypothesis because the p-value was found to be ~0.12, great
er than the standard significance level of 0.05. So, we cannot assume that the average proces
sing time is signifcantly different than the benchmark of 0.01")
```

```
## [1] "We fail to reject the null hypothesis because the p-value was found to be ~0.12, grea
ter than the standard significance level of 0.05. So, we cannot assume that the average proce
ssing time is signifcantly different than the benchmark of 0.01"
```

Q2.3 (C)

point estimate (q2.3)

```r
intercept_coef <- coef(fitq2)[1]
invoices_coef <- coef(fitq2)[2]
invoice_num <- 130
# Y = B0 + B1 * Xi
point_estimate <- intercept_coef + invoices_coef * invoice_num
paste("point estimate:", point_estimate)
```

```
## [1] "point estimate: 2.10962361186358"
```

95% prediction interval (q2.3)

```
predict(fitq2, newdata = data.frame(Invoices = c(130)), interval = "prediction", level = 0.9
5)
```

```
##        fit      lwr    upr
## 1 2.109624 1.422947 2.7963
```