YuHW02ST430

Haozhe (Jerry) Yu

2023-09-15

Question 1

For this problem, use the grade point average data described in KNNL Problem #1.19

```
actgpa <- as_tibble(read.delim2("https://users.stat.ufl.edu/~rrandles/sta4210/Rclassnotes/data/textdatas
            sep = "",
            strip.white=TRUE,
            col.name = c("GPA","ACT")
            ))
actgpa$GPA <- as.numeric(actgpa$GPA)
actgpa
```

```
## # A tibble: 119 x 2
##      GPA   ACT
##    <dbl> <int>
##  1  3.88    14
##  2  3.78    28
##  3  2.54    22
##  4  3.03    21
##  5  3.86    31
##  6  2.96    32
##  7  3.96    27
##  8  0.5     29
##  9  3.18    26
## 10  3.31    24
## # i 109 more rows
```

```
actgpam <- lm(GPA~ ACT,data=actgpa)

sactgpam <- summary(actgpam)

# Questions Answered will be with in-line code. I will provide an example of the inline code I use to d

str(sactgpam)
```

```
## List of 11
##  $ call         : language lm(formula = GPA ~ ACT, data = actgpa)
##  $ terms        :Classes 'terms', 'formula'  language GPA ~ ACT
##   .. ..- attr(*, "variables")= language list(GPA, ACT)
```

```
##   .. ..- attr(*, "factors")= int [1:2, 1] 0 1
##   .. .. ..- attr(*, "dimnames")=List of 2
##   .. .. .. ..$ : chr [1:2] "GPA" "ACT"
##   .. .. .. ..$ : chr "ACT"
##   .. ..- attr(*, "term.labels")= chr "ACT"
##   .. ..- attr(*, "order")= int 1
##   .. ..- attr(*, "intercept")= int 1
##   .. ..- attr(*, "response")= int 1
##   .. ..- attr(*, ".Environment")=<environment: R_GlobalEnv>
##   .. ..- attr(*, "predvars")= language list(GPA, ACT)
##   .. ..- attr(*, "dataClasses")= Named chr [1:2] "numeric" "numeric"
##   .. .. ..- attr(*, "names")= chr [1:2] "GPA" "ACT"
##  $ residuals    : Named num [1:119] 1.252 0.58 -0.416 0.112 0.546 ...
##   ..- attr(*, "names")= chr [1:119] "1" "2" "3" "4" ...
##  $ coefficients : num [1:2, 1:4] 2.0679 0.0404 0.3203 0.0127 6.4571 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "(Intercept)" "ACT"
##   .. ..$ : chr [1:4] "Estimate" "Std. Error" "t value" "Pr(>|t|)"
##  $ aliased      : Named logi [1:2] FALSE FALSE
##   ..- attr(*, "names")= chr [1:2] "(Intercept)" "ACT"
##  $ sigma        : num 0.619
##  $ df           : int [1:3] 2 117 2
##  $ r.squared    : num 0.0791
##  $ adj.r.squared: num 0.0712
##  $ fstatistic   : Named num [1:3] 10.1 1 117
##   ..- attr(*, "names")= chr [1:3] "value" "numdf" "dendf"
##  $ cov.unscaled : num [1:2, 1:2] 0.267445 -0.010464 -0.010464 0.000423
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "(Intercept)" "ACT"
##   .. ..$ : chr [1:2] "(Intercept)" "ACT"
##  - attr(*, "class")= chr "summary.lm"
```

```
sactgpam
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = actgpa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73842 -0.32556  0.04421  0.44644  1.25203
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.06789    0.32025   6.457 2.53e-09 ***
## ACT          0.04036    0.01273   3.170  0.00194 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6193 on 117 degrees of freedom
## Multiple R-squared:  0.07911,    Adjusted R-squared:  0.07124
## F-statistic: 10.05 on 1 and 117 DF,  p-value: 0.001944
```

```
sactgpam$sigma
```

```
## [1] 0.619263
```

```
sactgpam$df[2]
```

```
## [1] 117
```

```
help("tibble")
```

```
## starting httpd help server ... done
```

```
str(confint(actgpam,level = 0.95))
```

```
##  num [1:2, 1:2] 1.4336 0.0151 2.7021 0.0656
##  - attr(*, "dimnames")=List of 2
##   ..$ : chr [1:2] "(Intercept)" "ACT"
##   ..$ : chr [1:2] "2.5 %" "97.5 %"
```

```
confint(actgpam,level = 0.95)[2,2]
```

```
## [1] 0.06557706
```

```
confint(actgpam,level = 0.95)
```

```
##                   2.5 %     97.5 %
## (Intercept) 1.43364511 2.70213072
## ACT         0.01514958 0.06557706
```

a. What is the estimate of σ from this analysis?

The estimate for σ is 0.619263 with 117 degrees of freedom.

Example inline code: `sactgpam$sigma`

b. Give a point estimate and 95% confidence interval for the slope and interpret each of these in words.

The 95% confidence interval for the slope is 0.0151496 to 0.0655771. This indicates that 95% of intervals from samples of the same size will contain β_1.

c. Give a point estimate and 95% confidence interval for the y-intercept.

The 95% confidence interval for the y-intercept is 1.4336451 to 2.7021307. This indicates that 95% of intervals from samples of the same size will contain β_0.

d. Obtain a 95% interval estimate of the mean GPA for students whose ACT test score is 28. Interpret your confidence interval.

```r
oned <- tibble(ACT=25)
predict(actgpam,oned,interval="confidence",level=0.95) #%>% str()
```

```
##        fit      lwr      upr
## 1 3.076971 2.964378 3.189564
```

```r
predict(actgpam,oned,interval="predict",level=0.95)
```

```
##        fit      lwr      upr
## 1 3.076971 1.845395 4.308546
```

```r
predict(actgpam,oned,interval="confidence",level=0.95)[2]
```

```
## [1] 2.964378
```

The 95% confidence interval for students whose ACT score is 28 is 2.9643776 to 3.1895642. This indicates that 95% of intervals from samples of the same size will contain the true mean GPA of students whose ACT is 28.

e. Predict GPA using a 95% prediction interval for students whose ACT test score is 28.

The 95% prediction interval for students whose ACT score is 28 is 1.8453953 to 4.3085464. This indicates that 95% of students in the population who scored a 28 on their ACT will have a GPA in this range.

f. Would it be reasonable to consider inference on the intercept for this problem? Please provide justification for your answer.

No. It would not be reasonable to consider inference on the intercept for this problem. The ACT is scored on a scale of 1 to 36, so it is impossible for a student to score a 0 on the intercept, making inference on the intercept unreasonable.

g. For each of the following hypothesis tests, give the value of the test statistic, the degrees of freedom, the p-value (if you cannot obtain the pvalue, give the critical value for the test statistic), and clearly state your conclusion.

```r
help("pt")
summary(actgpam)
```

```
##
## Call:
## lm(formula = GPA ~ ACT, data = actgpa)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.73842 -0.32556  0.04421  0.44644  1.25203
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.06789    0.32025   6.457 2.53e-09 ***
## ACT          0.04036    0.01273   3.170  0.00194 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6193 on 117 degrees of freedom
## Multiple R-squared:  0.07911,    Adjusted R-squared:  0.07124
## F-statistic: 10.05 on 1 and 117 DF,  p-value: 0.001944
```

Including Plots

You can also embed plots, for example:

```
## [1] 21
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.