The University of British Columbian —Okanagan

DATA 310

Assignment 2

Due at 11:59 pm on Oct 27, 2024. Submit it on Canvas.

Question 1

The rate of spread R (in m/s) of a wildfire is related to wind speed W (in km/h). Suppose data have been collected and the output from the summary of function `lm` is shown below

```
R    W
30   35
32   40
18   20
35   50
12   15
17   22
20   27
22   36
24   35
```

```
Fire.lm <- lm(R ~ W, data=Fire_Spread)
summary(Fire.lm)

##
## Call:
## lm(formula = R ~ W, data = Fire_Spread)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5126 -0.8563 -0.6167  1.8922  4.1377
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.10187    2.93655   1.056 0.325918
## W            0.65030    0.08952   7.264 0.000168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.792 on 7 degrees of freedom
## Multiple R-squared:  0.8829,Adjusted R-squared:  0.8661
## F-statistic: 52.77 on 1 and 7 DF,  p-value: 0.0001679
```

(a) (1 point) Find the F-statistic and MSE from the summary output.

(b) (5 points) Build an ANOVA table as below using F score, MSE and degrees of freedom from the output of `summary`. Please write the detailed calculation and process.

Source	df	SS	MS	F
Reg.				
Error				
Total				

(c) (1 point) Check the result from previous question using `anova` function

(d) (1 point) Calculate $\sqrt{F}$ and compare the t-value for $\hat{\beta}_1$ in the output

Question 2

A machine sequentially performs two tasks. The first task requires **an unobservable exponentially distributed amount of time ε, with unknown mean** $1/\lambda$. The second task requires an amount of time, proportional to an observable positive-valued variable x which is independent of ε; the proportionality constant is positive, but otherwise unknown, and can be represented by β. The total time y required to process a single job is the sum of the times it takes to complete the two tasks.

(a) (1 point) Write down a linear statistical model relating y to x.

(b) (6 points) Derive maximum likelihood estimators for β and λ.

(c) (4 points) Fit the linear model, obtained in (a), to the following data.

```
> x
 [1] 1.8 0.1 0.2 1.4 2.4 0.9 0.9 0.6 0.2 3.7
> y
 [1] 2.0 0.4 1.9 2.7 3.2 1.1 1.5 1.1 0.8 4.1
```

(d) (3 points) Obtain a scatterplot of the data, and overlay the fitted line.

(e) (2 points) Calculate residuals. Comment.

(f) (4 points) Suppose ordinary least-squares had been used to fit the model in (a). Would this have resulted in a biased estimate for β? If so, estimate the bias that would have arisen if least-squares had been used in part (c).

(g) (2 points) What condition is required so that the least-squares estimate of β is unbiased, assuming a regression through the origin model?

Question 3

Use R to complete the following problem, but do not use built-in functions such as `lm()` and `predict()`. You need write the codes to calculate the question. A toy car has been released from ramps having nine different angles. Distance travelled (in m) has been measured in each case.

angle	1.3	4.0	2.7	2.2	3.6	4.9	0.9	1.1	3.1
distance	0.43	0.84	0.58	0.58	0.70	1.00	0.27	0.29	0.63

(a) (2 points) Plot the data. What is the predictor and what is the response? Is a linear model reasonable? (On physical grounds?, On statistical grounds?)

(b) (3 points) Compute estimates of the slope and intercept for a linear regression model. (Do all intermediate calculations, i.e. compute $S_{xy}, S_{xx}, \bar{y}, \bar{x}$, etc.)

(c) (5 points) Provide 95% confidence intervals for the slope and intercept parameters. (Again, show all intermediate calculations.)

(d) (5 points) Complete the analysis of variance table and test whether or not `distance` depends upon `angle`.

(e) (2 points) Find a 95% confidence interval for the expected distance at an angle of 2.5 degrees. Compute the corresponding 95% prediction interval.

Question 4

Use R for this problem. (You may use built-in functions such as `lm` and `aov`, etc..) The data frame `fossum` (in *Ch2*) contains several types of measurements on a sample of female possums. Among the variables measured are `totlngth` (total length, in cm) and `hdlength` (head length, in mm). Data `fossum` in the package `DAAG`

(a) (1 point) Plot `hdlngth` against `totlngth`.

(b) (3 points) Estimate the line obtained by regressing `hdlngth` on `totlngth`.

(c) (2 points) Obtain the ANOVA table for these data.

(d) (2 points) Test the hypothesis that `hdlngth` is unrelated to `totlngth`.

(e) (2 points) Provide a 95% prediction interval for the head length of a female possum with a total length of 85 cm.

(f) (3 points) Plot the residuals against the fitted values. Do they look randomly distributed?