1. A local ice cream shop keeps track of how much ice cream they sell versus the temperature on that day. The manager is trying to figure out how the temperature affects ice cream sales. Below are the figures for the last 10 days in May.

|  |  |
| --- | --- |
| Temperature (oC) | 14 16 12 15 18 22 19 25 23 21 |
| Ice Cream Sales ($) | 210 320 190 330 400 520 410 610 540 420 |

(

a) Determine a linear regression model equation.

b) On a particular day the temperature reaches 21 oC, how much ice cream does the manager predict will be sold?

c) Compute the correlation coefficient between temperature and statistics ice cream sales.

2. The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below.

|  |  |  |  |
| --- | --- | --- | --- |
| Result | Drug | Placebo | Total |
| Nausea | 36 (25.15) | 13 ( ) | 49 |
| No nausea | 254 ( ) | 262 ( ) | 516 |
| Total | 290 | 275 | 565 |

a) Compute the expected number of patients of each cell when the treatment is independent of the side effect of nausea. Fill in the parenthesis of the above table.

b) What is the chi-squared value for the testing of independence between treatment and side effect of nausea?

c) Test if the relationship between treatment and side effect of nausea is independent using significance level of 0.05.

3. Five students from a certain high school are sampled, and their SAT Verbal scores are 560, 500, 470, 660, and 640. ( )

a) Compute the mean and the sample standard deviation of the SAT scores.

b) Compute 95% confidence interval for the population mean assuming that the scores follows normal distribution.

c) We want to construct a 95% confidence interval with width of 60 (i.e., ). How many scores should be samples?

d) Perform the hypothesis testing for the mean SAT Verbal score is larger 500 with significance level 0.10. versus

4. The amount of coffee sold (in 1000 kg) in two locations are as follows

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | mean | Sample stdev |
| Location 1 | 26.3, 26.5, 16.2, 16.2, 26.4, 16.8, 16.1, 16.9 | 20.175 | 5.163 |
| Location 2 | 44.4, 44.7, 34.9, 54.1, 44.8, 22.6, 36.4 | 40.271 | 10.038 |

a) Test if the variance of the two groups are equal or not with significance level 0.05.

versus

b) Compute the confidence interval of the difference of the means assuming that the variances are different (use .

c) Perform the hypothesis test if the true average differs for the two locations assuming that the variances are different (use . versus

5. A random sample of 50 bulbs was selected, the lifetime of each bulb determined, resulting the following data.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample size | Sample mean | Sample standard dev. | Sample variance |
| 50 | 738.44 | 38.20 | 1459.24 |

a) Compute the confidence interval of the mean lifetime of the bulb (use .

b) Perform the hypothesis test if the true mean lifetime is less than 750. (use .

versus

c) When a level 0.05 test is used for versus , we want to compute , the probability of a type II error (the probability that the null hypothesis is not rejected when the alternative hypothesis is true).

c-1) Compute the rejection region of the test.

c-2) Compute . You may use R statement pnorm( ) to compute the probability

6. (11 pt) Suppose we want to examine the safety of compact cars, midsize cars, and full-size cars. We collected a sample of three for each of the treatments (cars types). Using the data provided below, test whether the mean pressure applied to the driver’s head during a crash test is equal for each types of car. Use α =0.05.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| Compact cars | 643, 655, 702 | 2000 | 666.67 |
| Midsize cars | 469, 427, 525 | 1421 | 473.67 |
| Full-size cars | 484, 456, 402 | 1342 | 447.33 |
|  |  | 4763 | =529.22 |

1. State the null and alternative hypothesis
2. Compute the sum of squares due to type of cars.
3. Fill in the blank space of the ANOVA table. If you could not compute the sum of squares due to type of cars in b), assume that it is 80000.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Source of  Variation | df | Sum of  Squares | Mean  Squares | f |
| Treatments  Error  Toral | \_\_\_\_\_(1)\_\_\_\_\_  \_\_\_\_\_(2)\_\_\_\_  8 | \_\_\_\_\_(3)\_\_\_\_\_  \_\_\_\_\_(4)\_\_\_\_\_  96304 | \_\_\_\_\_(5)\_\_\_\_\_  \_\_\_\_\_(6)\_\_\_\_\_ | \_\_\_\_\_(7)\_\_\_\_\_ |

1. Perform the hypothesis test based on the above table and the critical value of the F distribution.

(use .

e) Construct the 95% confidence interval of the difference between the mean pressure of compact cars and midsize cars.

7. The age of the cars, x years, and the mileage of the cars, y thousand miles of 10 used cars are shown as follows:

|  |  |
| --- | --- |
| X (age) | 2 2.5 3 4 4.5 4.5 5 3 6 6.5 |
| Y (mileage) | 22 34 33 37 40 45 49 30 58 58 |

(cf: 19.9, =1268.4,

=153.9, )

The above data was fitted using a linear regression model by R.

1. What percentage of the total variation of sales amount can be explained by the linear regression line?
2. Compute the residual standard error.

The following R output is the result of the execution of the regression analysis of the above data.

> x <- c(2, 2.5, 3, 4, 4.5, 4.5, 5, 3, 6, 6.5)

> y <- c(22, 34, 33, 37, 40, 45, 49, 30, 58, 58)

> reg <- \_\_(1)\_\_ ( \_\_(2)\_\_~ \_\_(3)\_\_)

> summary(\_\_(4)\_\_)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.907 3.0386 \_\_(5)\_\_ 0.0191 \*

x 7.73 \_\_(6)\_\_ 11.035 4.05e-06 \*\*\*---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: \_\_\_\_\_\_\_ on 8 degrees of freedom

Multiple R-squared: \_\_\_\_\_\_\_\_, Adjusted R-squared: \_\_\_\_\_\_\_\_

F-statistic: 121.8 on 1 and 8 DF, p-value: 4.05e-06

1. What is the R statements or values that fit to the above parenthesis of (1)-(6).
2. Determine a linear regression model equation from the above R output.

e) Obtain the fitted value of y when x=5

f) Compute the residual when x=5

g) Construct 95% confidence interval of .

h) Construct the 95% confidence interval of the mean free-flow when mileage is 5.5()

(The standard error of when x=5.5 can be obtained as 1.393.)

> data <- data.frame(x=c(5.5))

> predict(reg, data, se.fit = TRUE)

$se.fit

[1] 1.392831

Cf)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| z | 0.5 | 1 | 1.28 | 1.645 | 2 | 2.33 | 3 |
| ) | 0.6915 | 0.8413 | 0.9 | 0.95 | 0.9772 | 0.99 | 0.9984 |

, , ,

=11.143, =7.378, =5.024, =9.49, =5.99, =3.841,

, ,

, ,

, , , ,

, , , ,

, ,

, , ,

, , ,

4.578 3.948 5.158 4.339