Statistics for Data Analytics

Continuous Assessment

Lecturer: Dr Shahram Azizi Sazi

Student' full name:

Please provide your solution in word/pdf,… and submit it on Moodle before deadline. Notably, submissions after deadline will not be graded.

Questions

1- In a financial network, an agent works properly with p=0.8. Let us assume that 5 agents work in this network.
   a) Define X to be the possible numbers of agents who properly work, compute the probability table for X.
   b) What is P(X>4)?
   c) Find the expectation and variance X.

2- A manufacturing process produces ball bearings with diameters that have a normal distribution with known standard deviation of .04 centimeters. Ball bearings with diameters that are too small or too large are undesirable. In order to test the claim that μ=0.50 centimeters, perform a two-tailed hypothesis test at the 5% level of significance. Assume that a random sample of 25 gave a mean diameter of 0.51 centimeters. Perform a hypothesis test (step procedure outlined in class) and state your decision.

3- A specific price dataset is analysed and, the summary of ANOVA table is given as follows:

Oneway

Descriptives

PRICE

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	5	81.2000	29.8781	13.3619	44.1015	118.2985	40.00	120.00
2.00	5	77.0000	20.5061	9.1706	51.5383	102.4617	59.00	110.00
3.00	5	55.4000	13.5019	6.0382	38.6352	72.1648	40.00	73.00
Total	15	71.2000	23.7523	6.1328	58.0464	84.3536	40.00	120.00

ANOVA

PRICE

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1916.400	2	958.200		.189
Within Groups	5982.000	12	498.500		
Total	7898.400	14			

Find the F-statistic and express your decision.

4- An opinion poll surveyed a simple random sample of 1000 students. Respondents were classified by gender (male or female) and by opinion (Reservation for women, No Reservation, or No Opinion). Results are shown in the observed contingency table below.

	Opinion on Women Reservation			Row total
	Yes	No	Can't Say	
Male	200	150	50	400
Female	250	300	50	600
Column total	450	450	100	1000

Does the gender and opinion on women reservation are independent? Use a 0.05 level of significance. To do so,

a. State the hypotheses.
b. Find the statistic value.
c. Find the critical value.
d. Explain your decision and Interpret results.

5- The **delivery dataset** is analysed in R and the output of Regression analysis is as follows:

```
> fit <- lm(Time ~ Cases + Distance , data = delivery)
> summary(fit)

Call:
lm(formula = Time ~ Cases + Distance, data = delivery)

Residuals:
    Min      1Q  Median      3Q     Max
-5.7880 -0.6629  0.4364  1.1566  7.4197

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.341231   1.096730   2.135 0.044170 *
Cases       1.615907   0.170735   9.464 3.25e-09 ***
Distance    0.014385   0.003613   3.981 0.000631 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.259 on 22 degrees of freedom
Multiple R-squared: 0.9596, Adjusted R-squared: 0.9559
F-statistic: 261.2 on 2 and 22 DF,  p-value: 4.687e-16
```

(a) List the assumptions for the linear regression.
(b) Using the above output, specify the response and independent variables. Find the coefficients' estimates for independent variables.
(c) Identify the significant independent variables at level $\alpha = 0.05$.
(d) Provide the predictive model and find the predicted value of **time** where **cases are two** and **Distance is three**.