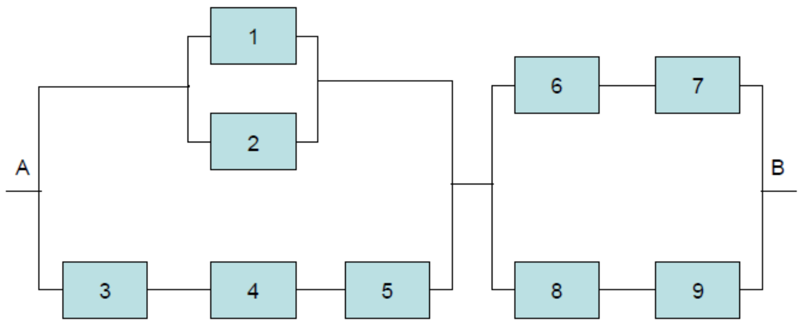
Extraordinary session
Final exam

Time: 120 min.

- **You are not allowed to use any documentation apart from the formula sheet you have received.**
- **Use 4 decimal digits in all calculations and results.**

1. **(2 Points)** In the communication network of 9 connected components according to the following figure, the probability that each component C_i works is $p = 0.9$. The network works if between A and B it is possible to find a component path that works. The operation of each component is assumed to be independent of the rest of the components.



(a) **(1 Point)** Calculate the probability that the block or subnetwork formed by the elements [6,7,8,9] works.

Solution

We denote by F_i that the i-th component works. The subnetwork [6,7,8,9] works if $(F_6 \cap F_7) \cup (F_8 \cap F_9)$: and we know that if $i \neq j$ then $P(F_i \cap F_j) = P(F_i) * P(F_j)$ because the operation of the components is independent.

$$\begin{aligned} P(Subnetwork\ [6,7,8,9]\ works) &= P((F_6 \cap F_7) \cup (F_8 \cap F_9)) \\ &= P(F_6 \cap F_7) + P(F_8 \cap F_9) - P(F_6 \cap F_7 \cap F_8 \cap F_9) \\ &= p^2 + p^2 - p^4 = .9^2 + .9^2 - .9^4 = 0.9639. \end{aligned}$$

(b) **(1 Point)** Find the probability that there is communication between A and B.

Solution

$$P(Networkworks) = P(Subnetwork\ [1,2,3,4,5]\ works) * P(Subnetwork\ [6,7,8,9]\ works),$$

We have that the subnetwork [1,2,3,4,5] works if $(F_1 \cup F_2) \cup (F_3 \cap F_4 \cap F_5)$:
$P(F_1 \cup F_2) = p + p - p^2 = 2p - p^2$,
$P(F_3 \cap F_4 \cap F_5) = p^3$,
$P(Subnetwork\ [1,2,3,4,5]\ works) = (2p - p^2) + p^3 - (2p - p^2)p^3 = 2p - p^2 + p^3 - 2p^4 + p^5$.

Doing the product of both:
$P(Networkworks) = -p^9 + 2p^8 + p^7 - 3p^6 - 2p^4 + 4p^3 \approx 0.9613$.

2. **(3 Points)** The average occupancy (proportion of occupied seats) of the trains of the C1 suburban line was estimated at 59.75% while that of the C2 at 58.28%. For the C1 estimate, 2,400 seats were observed on different trips, while for the C2 estimate, 2,500 were observed.

(a) **(1 Point)** Calculate and interpret a 95% confidence interval for the mean occupancy of line C1. Additional data: $z_{0.025} = 1.96$.

Solution

The confidence interval is

$$\hat{p}_1 \pm z_{0.025}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}} \approx [0.5779; 0.6171],$$

where $\hat{p}_1 = 0.5975$ y $n_1 = 2400$. We can say that the proportion of occupied seats in line C1 is between 57.79% and 61.71% with a confidence level of 95%.

(b) **(1.5 Points)** Evaluates through a unilateral contrast and with a level of significance of 5% if the occupation of the line C1 is greater than that of the C2. Data: $z_{0.05} = 1.645$.

Solution

We have to perform the following contrast: $H_0 : p_1 \leq p_2$ versus $H_1 : p_1 > p_2$.

The test statistic is approximately distributed according to a standard normal and has the following expression:

$$t = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where $p = (0.5975 * 2400 + 0.5828 * 2500)/(2400 + 2500) = 0.59$. Therefore,

$$t = \frac{0.5975 - 0.5828}{\sqrt{0.59 \times 0.41\left(\frac{1}{2500} + \frac{1}{2400}\right)}} \approx 1.0459.$$

On the other hand, since we are working at a 5% level of significance and we have proposed a one-sided test, the critical value will be $z_{0.05}$ =1,645.

Since $t = 1.045 < z_{0.05} = 1.645$, we will not reject the null hypothesis. In other words, there is not enough evidence against the fact that the mean occupancy of line C1 is equal or less than the mean occupancy of line C2.

(c) **(0.5 Points)** Would the conclusion of the contrast change if the same estimated occupancies had been obtained based on a sample of 10,000 seats from each line?

Solution

The decision would change because increasing the sample sizes increases the precision of the estimate. Specifically, we will have to

$$t = \frac{0.5975 - 0.5828}{\sqrt{0.5902 \times 0.4099\left(\frac{1}{10000} + \frac{1}{10000}\right)}} \approx 2.1133 > 1.645.$$

3. **(3 Points)** In a delivery company (with service 24 hours a day), the number of late deliveries in a day follows a Poisson distribution with a mean of 4 late deliveries per day.

(a) **(1 Point)** What is the probability that the company makes at least 3 late deliveries 2 days in a row?

Solution

Let N_1 be the number of late deliveries in one day, we have that $N_1 \sim Poisson(\lambda = 4)$, therefore the number of late deliveries in two consecutive days satisfies $N_2 \sim Poisson(\lambda = 8)$.

$$P(N_2 \geq 3) = 1 - P(N_2 \leq 2) \approx 1 - 0.0138 = 0.9862.$$

(b) **(1 Point)** Obtain the average time between two consecutive late deliveries. What is the probability that the time between two consecutive late deliveries is less than 8 hours?

Solution

Let X be the time between two consecutive late deliveries. We know that $X \sim Exp(\lambda = 4)$, therefore $E(X) = 1/4$ of a day, that is, 6 hours. The probability that the time between two consecutive late deliveries is less than 8 hours (1/3 day) is:

$$P(X < 1/3) = 1 - e^{-4*1/3} \approx 0.7364.$$

(c) **(1 Point)** What is the probability that on a given day there are no late deliveries? What is the probability that, among 6 randomly chosen days, there are no late deliveries on at least 2 of them?

Solution

The probability of not having delays in one day is: $P(N_1 = 0) = e^(-4) \approx 0.0183$.

If we randomly select 6 days, the number of days without delivery delay $Y \sim Binomial(n = 6; p = 0.0183)$. The probability that at least two days there is no delivery delay is:

$$P(Y \geq 2) = 1 - P(Y \leq 1) = 1 - P(Y = 0) - P(Y = 1) \approx 1 - 0.8951 - 0.1001 = 0.0048.$$

4. **(2 Points)** The following outputs from R correspond to the analysis of a braking system data set of 200 cars. The following variables are measured.

- KM: Kilometers travelled.
- Power: Power of the car.
- ABS: ABS system.
- LEFT_FRONT, RIGHT_FRONT, LEFT_REAR and RIGHT_REAR: four-wheel braking test results.
- Weight: Weight of the car.
- Efficacy: Efficacy of the braking system.

Two models of the efficacy of the braking system have been considered depending on the rest of the variables.

```
> BrakingITV$ABS = as.factor(BrakingITV$ABS)
> summary(BrakingITV)
       KM              Power           ABS      LEFT_FRONT       RIGHT_FRONT
 Min.   :   889   Min.   : 24.40   0:157   Min.   :1.200   Min.   :1.100
 1st Qu.: 78940   1st Qu.: 51.00   1: 43   1st Qu.:1.800   1st Qu.:1.800
 Median :116089   Median : 65.00           Median :2.000   Median :2.000
 Mean   :125131   Mean   : 65.18           Mean   :2.058   Mean   :2.042
 3rd Qu.:163750   3rd Qu.: 79.00           3rd Qu.:2.300   3rd Qu.:2.300
 Max.   :399000   Max.   :141.00           Max.   :3.200   Max.   :3.000

 LEFT_REAR        Weight          Efficiency
 Min.   :0.100   Min.   :0.400   Min.   : 640.0   Min.   :33.72
 1st Qu.:0.800   1st Qu.:0.800   1st Qu.: 907.2   1st Qu.:54.26
 Median :1.000   Median :1.000   Median :1020.5   Median :60.33
 Mean   :1.046   Mean   :1.049   Mean   :1031.0   Mean   :59.57
 3rd Qu.:1.200   3rd Qu.:1.200   3rd Qu.:1145.0   3rd Qu.:63.85
 Max.   :2.100   Max.   :2.500   Max.   :1603.0   Max.   :79.93

> modelo1 = lm(EFFICACY ~ KM + Power + ABS + WEIGHT + LEFT_FRONT + RIGHT_FRONT
               + LEFT_REAR + RIGHT_REAR, data = BrakingITV)
> summary(modelo1)

Call:
lm(formula = EFFICACY ~ KM + Power + ABS + WEIGHT + LEFT_FRONT +
    RIGHT_FRONT + LEFT_REAR + RIGHT_REAR, data = BrakingITV)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.307e+01  2.911e+00  21.661  < 2e-16 ***
KM          -5.242e-06  5.612e-06  -0.934  0.35142
Power       -2.285e-03  2.813e-02  -0.081  0.93536
ABS1        -1.100e+00  1.171e+00  -0.939  0.34872
WEIGHT      -3.799e-02  4.420e-03  -8.596 2.93e-15 ***
LEFT_FRONT   4.287e+00  1.897e+00   2.260  0.02496 *
RIGHT_FRONT  6.619e+00  2.126e+00   3.114  0.00213 **
```

```
LEFT_REAR    7.312e+00  2.577e+00   2.838  0.00503 **
RIGHT_REAR   6.414e+00  2.400e+00   2.673  0.00817 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.065 on 191 degrees of freedom
Multiple R-squared:  0.4568,Adjusted R-squared:  0.434
F-statistic: 20.08 on 8 and 191 DF,  p-value: < 2.2e-16
```

```
> modelo2 = lm(EFFICACY ~ WEIGHT + LEFT_FRONT + RIGHT_FRONT + LEFT_REAR
               + RIGHT_REAR, data = BrakingITV)
> summary(modelo2)

Call:
lm(formula = EFFICACY ~ WEIGHT + LEFT_FRONT + RIGHT_FRONT + LEFT_REAR +
    RIGHT_REAR, data = BrakingITV)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 63.881056   2.397147  26.649  < 2e-16 ***
WEIGHT      -0.038817   0.003905  -9.939  < 2e-16 ***
LEFT_FRONT   4.143495   1.869605   2.216  0.02784 *
RIGHT_FRONT  6.470138   2.103471   3.076  0.00240 **
LEFT_REAR    6.652091   2.477723   2.685  0.00789 **
RIGHT_REAR   6.680863   2.364479   2.826  0.00521 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.044 on 194 degrees of freedom
Multiple R-squared:  0.4528,Adjusted R-squared:  0.4387
F-statistic: 32.11 on 5 and 194 DF,  p-value: < 2.2e-16
```

(a) **(0.5 Points)** What is the best model? Justify your answer.

Solution

The second model is more suitable because all variables are significant and the adjusted R^2 is slightly larger.

(b) **(0.5 Points)** Write the equation of the model chosen in the previous question and interpret its coefficient of determination, R^2.

Solution

$EFFICACY = 63.8811 - 0.0388 * WEIGHT + 4.1435 * LEFT\_FRONT + 6.4701 * RIGHT\_FRONT + 6.6521 * LEFT\_REAR + 6.6809 * RIGHT\_REAR.$

The model explains 45.28% of the variability observed in the Efficacy of the braking system.

(c) **(1 Point)** A car with a Weight of 1200kg and with four-wheel braking test results equal to 2.0, how likely is it to have an Efficacy greater than 60 in its braking system?

Solution

We will use the second model. First of all, we obtain the average prediction for a car with these characteristics:

$$EFFICACY = 63.8811 - 0.0388 * 1200 + 4.1435 * 2 + 6.4701 * 2 + 6.6521 * 2 + 6.6809 * 2 = 65.2143,$$

On the other hand, we have $\widehat{\sigma} = 5,044$, so we can assume that $EFFICACY|Weight = 1200\ AND\ Test\ Wheel = 2) \sim \mathcal{N}(\mu = 65.2143, \sigma = 5.044)$:

$$\begin{aligned} P(EFFICACY > 60|Weight = 1200\ and\ Test\ Wheel = 2) &= P\left(\frac{EFFICACY - \mu}{\sigma} > \frac{60 - 65.2143}{5.044}\right) \\ &\approx P\left(Z > -1.0338\right) \approx 0.8494. \end{aligned}$$