

Second midterm exam
December 15th, 2021

Time:
75 minutes

- **You are not allowed to use any documentation apart from the formula sheet you have received, and the Z(0,1) table.**
- **Use 4 decimal digits in all calculations and results.**

1. (3 points) The following estimator, to be used on a simple random sample of size 4, is proposed to estimate the average of failures in a batch of computer equipment:

$$\hat{\mu} = \frac{X_1 + X_2 - 2X_3 - 2X_4}{12}$$

- (0,5 points) Check if the estimator is not unbiased
 - (1 point) Calculate its bias. Does the estimator overestimate or underestimate the parameter value?
 - (0,5 points) Calculate its variance
 - (1 point) For which sample size its variance is smaller than that of the sample mean distribution.
2. (4 points) Randomly selected performance programs are periodically executed on a network to check its performance. The programs are selected from a collection which have essentially the same type of instructions and input-output calls. 51 measures are taken, obtaining an average response time of 2,7 seconds with a sample variance of 0,3721 s². Quality standards establish that average response times should be smaller than 3 seconds.
- (2 points) Can it be stated that the network fulfills this condition at a significance level of 1,5%?
 - (2 points) When reviewing the work, data collection errors were detected in 2% of the samples. After threatening the responsible of the disaster with the dismissal, a sample of size 60 was taken, obtaining an error percentage of 1,8%. Has the percentage of errors been reduced to the same level of significance?

3. (3 points) Data are available from a sample of students from the Polytechnic School. The information available is:

- **Average:** Final grade in Advanced Maths
- **IA:** Academic Index before following the subject
- **Assistance:** Binary variable that shows whether the student attends class frequently. 1 = less than 5 absences, 0 = 5 or more absences.

The following R outputs show the summary statistics of the three variables:

```
summary(IA)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4.558   6.580   7.170   7.138   7.716   9.788

summary(Promedio)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
5.281   6.856   7.355   7.344   7.815   9.511

table(Asistencia)
Asistencia
 0  1
21 79
```

A multiple regression model is built with these variables in order to explain variable Average. The result is summarized in the following R output:

```
> summary(lm(Promedio ~ IA + Asistencia))

Call:
lm(formula = Promedio ~ IA + Asistencia)

Residuals:
    Min       1Q   Median       3Q      Max
-1.25934 -0.38176 -0.03849  0.39748  1.41526

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.38034     0.40747   8.296 6.29e-13 ***
IA             0.47299     0.05399   8.761 6.33e-14 ***
Asistencia    0.74319     0.13387   5.552 2.46e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5448 on 97 degrees of freedom
Multiple R-squared:  0.5168, Adjusted R-squared:  0.5068
F-statistic: 51.87 on 2 and 97 DF,  p-value: 4.801e-16
```

- (1 point) Write the equation for the model and interpret its determination coefficient.
- (1,5 points) Interpret all model coefficients and analyze whether they are significant or not to explain variable Average.
- (0,5 points) From the model results can you tell if it is interesting for students to attend class frequently? Justify your answer.