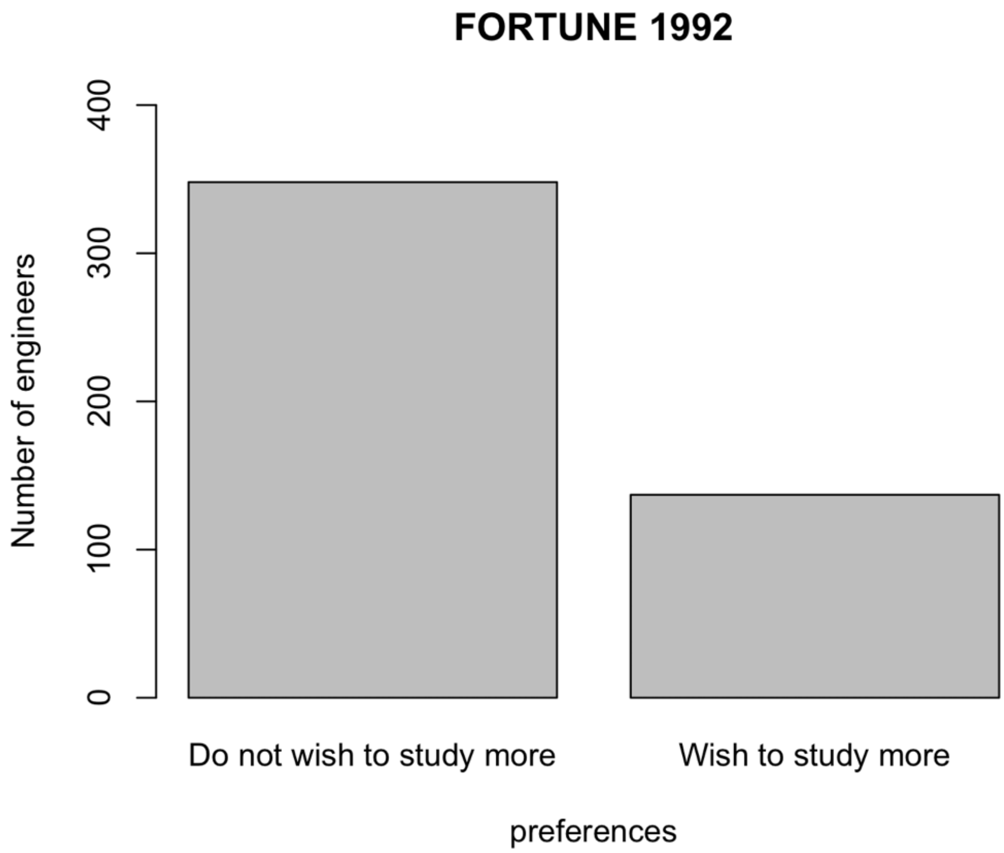
Assignment 2

Domingo Menéndez Rua, Pablo Soëtard García and Jorge Ingelmo Barrada

EXERCISE 1



a)

Looking this diagram, we can deduce that this sample follows a Bernoulli distribution, because as we can see our variable only can takes two values 0 or 1 (do not continue studies or continue studies), this is represented in the graph with two bars, one for the engineers that do not wish to study more(0), and one for the others (1). We can observe that the number of engineers that wish to study more after the B.S degree is around 150, way below the half of the sample that the article claims. There are around 350 engineers in this sample which do not wish to study more.

b) Applicating proportion formula to this sample that follows a Bernoulli distribution, we can conclude that the proportion of engineers that would like to study more is 0,282. This is consistent with the previous diagram.

c) In this step we calculate the standard error of the previous calculus, as we know that the distribution of the sample is a Bernoulli distribution, we can use the formulas based in the Bernoulli distribution for calculate this error, we calculate the error and we obtain that the error is 0.02044.

d) Using the R commands, we have obtained a confidence Interval of $CI_{95}(p) = [0.242, 0.323]$. This holds the results of part b), as the estimator calculated lies in this interval.

e) No, because if we look the confidence interval calculated previously, we can observe that the proportion percentage of engineers that want to study more is between 24,2% and 32,3% this is way less than the 50% said by the Fortune magazine.

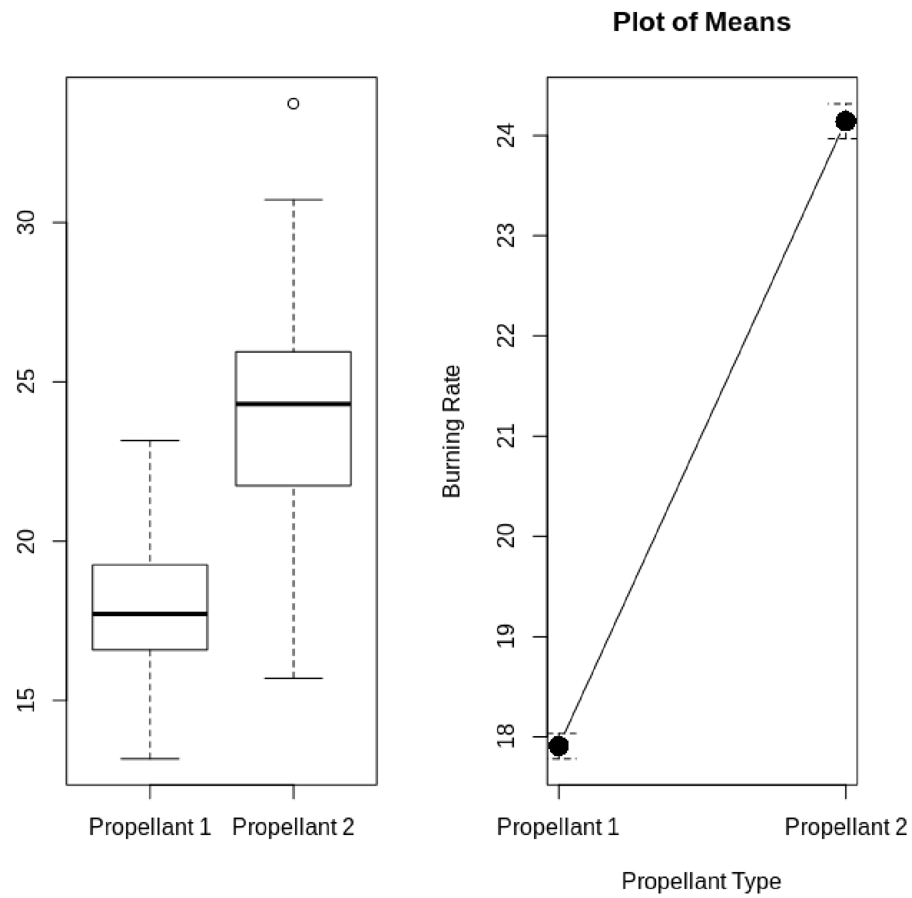
f) Let's suppose H_0: p >= ½, and H_1: p < ½, we have performed a hypothesis test with R and we have obtained that we can reject the null hypothesis. This holds with the calculated P-value = 4.804725e-22 < 0.05, so we can conclude that there is some evidence that the proportion of the sample is less than ½.

EXERCISE 2

a) The number of specimens of propellant 1 tested are 250 and those of propellant 2 are 300.

b) With the data and plots below, we can see that the mean of the two samples is very different, both in the summary of indexes result and in the plot of means. The sample of propellant 2 seems to be more spread than those of propellant 1, these can be seen in the difference of the standard deviation as well as in the boxplot, being the box of propellant 2 "higher" than the one of propellant 1.

```
                  mean       sd      IQR        cv      0%      25%      50%
Propellant 1 17.90745 2.008253 2.667775 0.1121462 13.1651 16.58272 17.71375
Propellant 2 24.14054 3.041428 4.190775 0.1259884 15.6983 21.75290 24.30790
                   75%     100% BurnRate:n
Propellant 1 19.25050 23.1515        250
Propellant 2 25.94367 33.7341        300
```

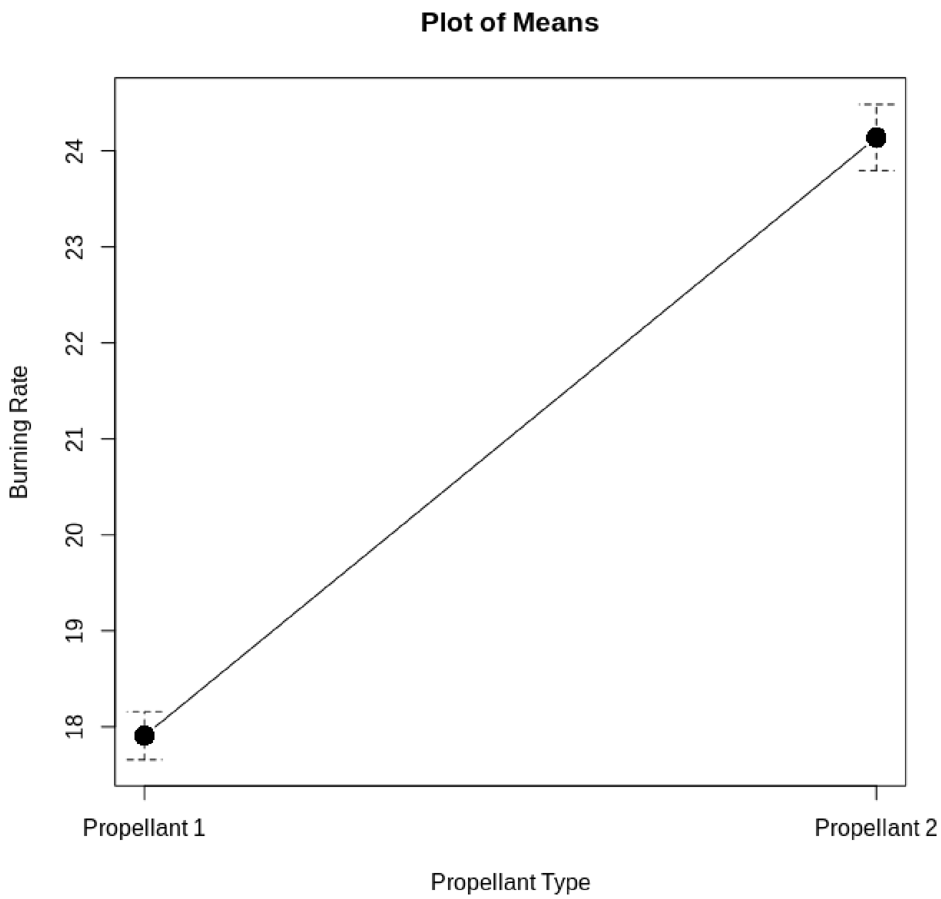


c) Let's suppose H_0: $\mu_{P1} = \mu_{P2}$, and H_1: $\mu_{P1} \neq \mu_{P2}$, we have performed a hypothesis test with R and we have obtained that we can reject the null hypothesis. This holds with the calculated P-value = 1.021955e-109 < 0.05, so we can conclude that there is some evidence that mean burning rate is different when using propellant 1 from when using propellant 2.

d) With the result above, we can conclude that there is no relationship in the burning rates and the type of propellant in the population considered, as we have rejected the null hypothesis.

e) $CI_{95}(\mu_{P1} - \mu_{P2}) = [-6.658841, -5.807349]$, as we can see, the interval limits are both negative, these means that, with a 95% confidence, Propellant 2 has a greater burning rate than Propellant 1. As 0 is not contained in the interval this also holds the results provided in c) and d).

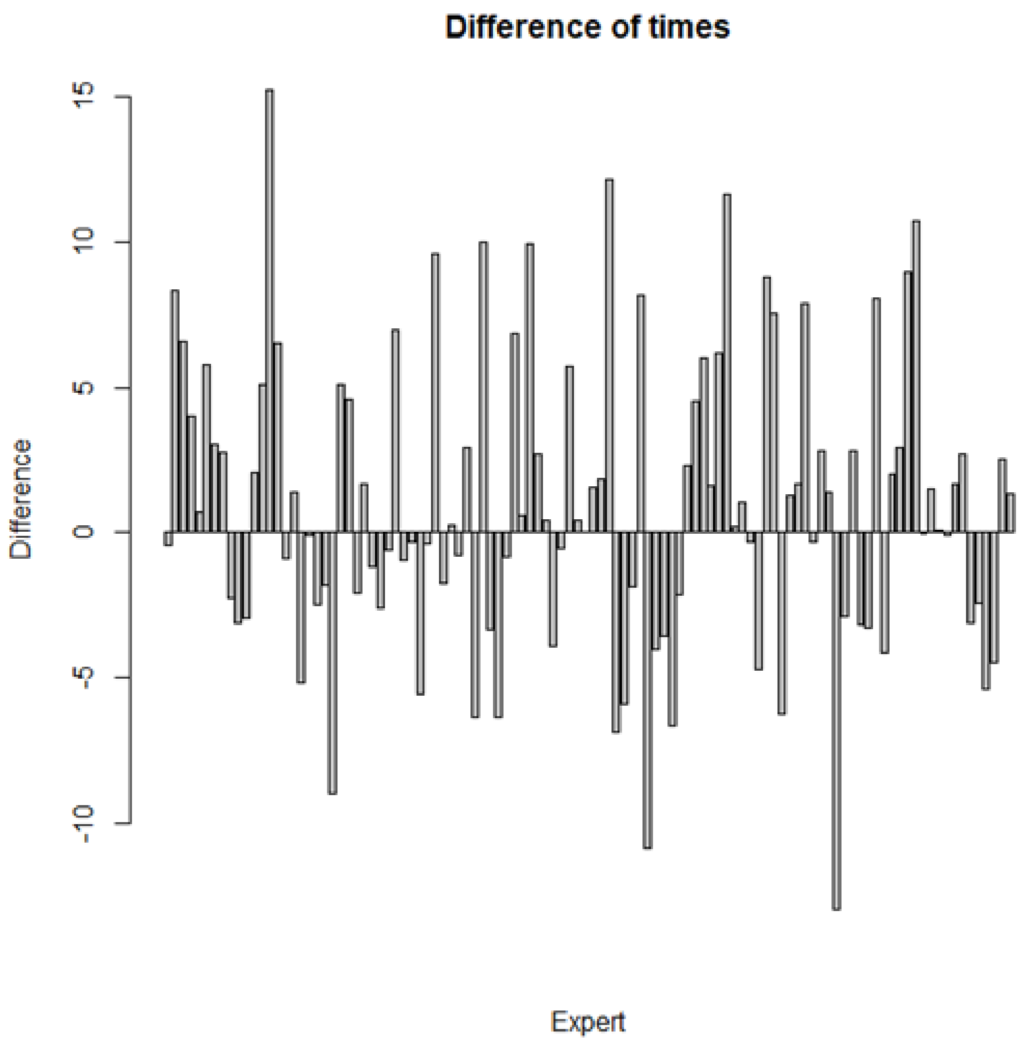
f) $CI_{95}(\mu_{P1}) = [17.65729, 18.1576]$ and $CI_{95}(\mu_{P2}) = [23.79498, 24.48611]$



EXERCISE 3

a) This part is done in the .Rhistory file.

b)



What we can see in this dataset is that the difference between both languages is in general a high negative number or a high positive number, this means that 0 is also part of that ”interval”, hence we can predict a possible relation between the means of both sample sets.

c) Solution: $CI_{95}(\mu_{L1} - \mu_{L2}) = [-0.0251642, 1.924453]$ what we can see here is that the value 0 is part of the interval, so we can’t conclude that the first language is faster than the second or vice versa. This holds with what we have observed analyzing the sample set.

d) We calculate the maximum error for a confidence level of 95% (procedure on R.history file) and the result is 0.9748085.

e) Let's suppose H_0: $\mu_{P1} = \mu_{P2}$, and H_1: $\mu_{P1} \neq \mu_{P2}$, we have performed a hypothesis test and we obtained that we have failed to reject the null hypothesis. This holds with the calculated P-value = 0.05616 > 0.05, so we can conclude that there is no evidence that mean time to code the function is different when using language 1 from when using language 2. This also holds with the confidence interval calculated before, as 0 belongs to that interval.