Statistics - October 2021

Problem 1

Question 1

P(Rhesus Positive) = 0,35 + 0,49 = 0,84

The probability of a randomly selected Dane being Rhesus positive is 84%

Question 2

A = Rhesus positive

B = Antigens

P(A) = 0,84

P(B) = 0,59

P(A and B) = 0,49

P(B given A) = 0,5833

Thus, the probability of having antigens in your blood, given that you are Rhesus positive is 58,33%.

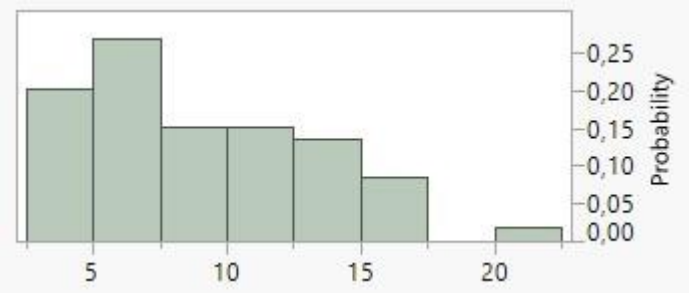
Question 3

We use binomial distribution chance of success is 0,49, which is the probability of P(A and B). Being Rhesus positive and having antigens in your blood. We 5 have trails since 5 people are selected. We want to find the probability of 3 of the 5 trails being a success.

P(3 or more successes) = 0,306.

There is a 30,6% chance of at least 3 out of 5 randomly selected Danes being Rhesus positive and having antigens in their blood.

Problem 2



The distribution is asymmetric with a clear skew to the right. It can be argued that it is unimodal, since there is only one mode that lies between 5 and 7,5. There is one outlier in the distribution to the right.

In the JMP output the mean and standard deviation is given. It is respectively 8,7686667 and 4,4577609. Using this the empirical rule can be applied.

68% of all observations falls within ± 1 standard deviation of the mean:]4,3109 ; 13,2264[

95% of all observations falls within ± 2 standard deviations of the mean:]-0,1469 ; 176842[

99% of all observations falls within ± 3 standard deviations of the mean:]-4,6046 ; 22,1419[

In this distribution the minimum is 2,6 and the maximum is 20,13.

Problem 3

0,53 → 53% of student loan debts are women's

0,42962963 → 42,962963 of the former students in default are women

H_0: $pp = pp_0$

H_a: $pp \neq pp_0$

Assuming normal distribution for the test.

Using Excel the standard error comes out to be 0,001358 withn a test statistic of 73,889980. This gives me a p-value of 0,000. Since the p-value is below the significance level of 0,05 the null hypothesis is being rejected in favor of the alternative hypothesis. We can then conclude that the probability of a former student in default being a women differs from the probability of a former female student having student loan debts.

Problem 4

The proportion of women being hired is 0,261589404

The proportion of men being hired is 0,162754304

The difference between proportions is 0,0988351

Using Excel and assuming normal distribution, the 95% confidence interval comes out to be]0,041597047 ; 0,156073154[

This confidence interval tells us that we can be 95% sure that between 4,1597% and 15,6073% more women are hired as a professor at Danish universities. The confidence interval is also statistically significant, since 0 is not in the interval.

Problem 5

H_0: The effect of corona virus affects the job security equally in the countries

H_a: The effect of corona virus does not affect the job security equally in the countries

Under the null hypothesis the following test is approximately X^2 – distributed with 8 degrees of freedom. The Pearson X^2-test statistic using Excel comes out to be 102,3725. This gives us a p-value of 0,000. Since the p-value is below the significance level of 0,05 we reject the null hypothesis in favor of the alternative hypothesis. We can then conclude that corona virus did not have the same effect on job security in these European countries. It is seen that a lot of people in the UK lost their jobs than in Denmark.

Problem 6

Question 1

Assuming t-distribution for finding the confidence interval.

Using calculator the difference in means is 0,684.

The standard error is $\sqrt{MS\ Error} = \sqrt{20{,}0932} = 4{,}482543921$

The degrees of freedom is 58.

Using the table, the $t_{.05}$ score is 1,671. This is with 60 degrees of freedom, since that is the closest to 58.

$$yy1 - yy2 \pm tt_{.05} ss \sqrt{\frac{1}{nn1} + \frac{1}{nn2}}$$

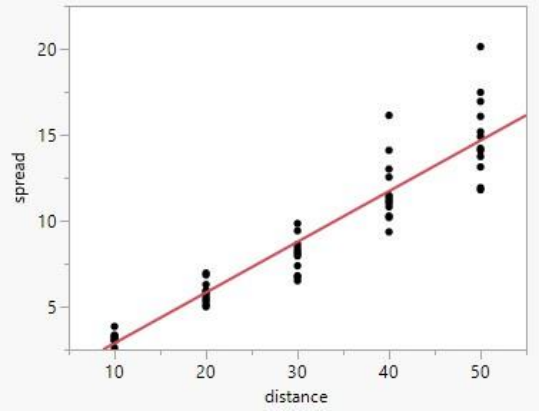
$$0{,}684 \pm 1{,}671(4{,}48254)\sqrt{\frac{1}{30} + \frac{1}{30}}$$

Which is $0{,}684 \pm 1{,}933993428$

The 90% confidence interval ranges from]-1,249993428 ; 2,617993428[

Since 0 falls in the 90% confidence interval, it is statistically insignificant since the difference in means might be 0. This would mean that the expected spread from the two types of cartridges would be the same.

Question 2



The regression equation is $sspprrpppppp = -0{,}078083 + 0{,}2948917 * ppddsstppnnddpp$

H_0: $bb = 0$

H_a: $bb \neq 0$

The JMP output gives me a t-ratio, which is approximately normally t-distributed, of 21,67 and testing with 58 degrees of freedom the p-value comes out to be 0,0001. Thus we reject the null hypothesis in favor of the alternative hypothesis. The conclusion is that the expected spread is proportionate to the distance. The further you are from the target, the greater the spread. This is also indicated by the positive slope.

Question 3

Assuming normal F-distribution and testing with 1 and 58 degrees of freedom.

H_0: The effect on the mean of distance is independent of type of cartridge

H_0: The effect on the mean of distance is dependent of type of cartridge

Using JMP the F-ratio is 0,3894 and the p-value is 0,5351. Since the p-value is above the significance level of 0,05, we accept the null hypothesis and can conclude there is no significant interaction between the type of cartridge and the distance on the spread. In other words, the spread depends on the distance to target and not on the cartridges.