Student Number:

20-10-2023

Exam name: Statistics (BINBO1139E + + BISHO1005E) - Written sit-in exam (UC)

## Problem 1

**1.**

The probability that a Swede thinks that burning the Koran should be banned is 52,92 %

**2.**

The probability that my Swedish friend is a woman is 0,619047619, ($\approx 61{,}905\ \%$)

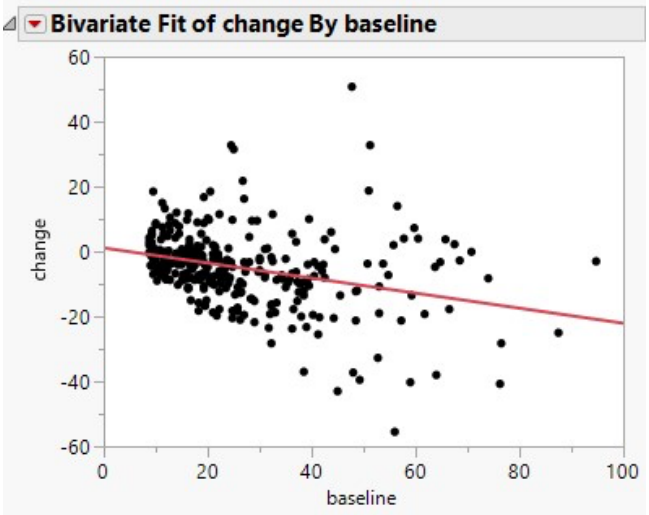
## Problem 2

**1.**

The probability that Danish woman living at the beginning of the 20th century would be too short to be employed at a telephone exchange is 0,115070, ($\approx 11{,}507\ \%$)

**Problem 3 (using jmp)**

**1.**




Looking at this scatterpot, it looks like there is a negative association, when the servility measurement, baseline, goas up the change in serenity tends to go down. Given that negative values indicate improvements, people with a larger severity measurement before the trial seams to have more improvement. The observations for larger values seems to have a larger variability. The observation on 50 could be a potential outliner, which effect the association negatively, it causes a weaker assosiacion.

**The correlation is;** 0,3141735189 between the two variables.

**Problem 4**

**1.**

The 95% confidence interval for the probability that a truck used to transport students has safety flaws, using the standard normal quantile, equals;

$$(0{,}365079) \pm 1{,}96 * (0{,}060657) = ]0{,}246193;\ 0{,}483965\ [$$

Hence with 95% confidence the probability that a truck used to transport students has safety flaws is between 24,6193 % and 48,3965 % , as the proportion is less than 50%, it means that less than 50% of all truks are having safety flaws. A larger proportion of trucks do not have safety flaws.

**Problem 5**

**1.**

To test this I will use the poled estimate t-test.

I do that because I assume equal standard deviations due to;
$62{,}82/33{,}69 = 1{,}864648264 < 2$ ,
Thus the ratio is smaller than two, I assume equal standard deviations.

The t-test statistic equals 2,84952, under the null hypothesis, it is approximately t- distributed with 88 degrees of freedom, I get a p-value 0,005452. As the p- value is below our significance level of 0,05, we have strong evidence against the null hypothesis of no difference and we reject the null hypothesis. Thus we can conclude that the mean spending depends on the type of drink offered, the mean spending is significantly higher for costumers having espresso than for costumers having water. On average it is 30,28 euros higher for costumers having espresso than for costumers having water.

**Problem 6**

The 99% confidence interval for the difference between the probability that an adult female is referred to as a “girl” in university textbooks and the probability that an adult male is referred to as as “boy”, using the standard normal quantile, equals;

$$(0{,}406061) \pm 2{,}57583 * (0{,}066898) = ]0{,}233742;\ 0{,}578379[$$

The difference between the probability that an adult female is referred to as a “girl” in university textbooks and the probability that an adult male is referred to as “boy”, is between 0,233742 and 0,578379. Thus with 99% confidence we can say that there is a larger proportion of adult females that is referred to as a “girl” in university textbooks than that an adult male is referred to as “boy”, there is between 23,3742% and 57,8379% larger proportion of adult females that is referred to as a “girl” in university textbooks than adult male that is referred to as “boy”

Student Number:

20-10-2023

Exam name: Statistics (BINBO1139E + + BISHO1005E) - Written sit-in exam (UC)

**Problem 7 using jmp**

**1.**


Freq: Sold (count)

**Contingency Table**

| Marked with | Count / Row % / Expected | Sort of apple: Elise | Ingrid Marie | Jonagold | Total |
|---|---|---|---|---|---|
| Name & Taste | Count | 200 | 124 | 76 | 400 |
| | Row % | 50,00 | 31,00 | 19,00 | |
| | Expected | 162,376 | 152,97 | 84,6535 | |
| Name Only | Count | 128 | 185 | 95 | 408 |
| | Row % | 31,37 | 45,34 | 23,28 | |
| | Expected | 165,624 | 156,03 | 86,3465 | |
| Total | | 328 | 309 | 171 | 808 |

**Tests**

| N | DF | -LogLike | RSquare (U) |
|---|---|---|---|
| 808 | 2 | 15,046058 | 0,0175 |

| Test | ChiSquare | Prob>ChiSq |
|---|---|---|
| Likelihood Ratio | 30,092 | <,0001* |
| Pearson | 29,882 | <,0001* |

(Using JMP)
The Pearson $\chi\chi^2$-test statistic equals 29,882. Under the null hypothesis, it is approximately $\chi\chi^2$distributed with 2 degrees of freedom, which gives us a P-value <0,0001. We do reject the null-hypothesis of no dependence. We can conclude that including the description of the taste change the distribution of sort sold apple.

Looking at the row proportions or the mosaic plot we see that the proportion of sold Elise apples increased when the apples included a description of taste (from 31,37% of all sold apples when name only to 50,00 % of all sold apples when including a description.) While both the proportion of sold Ingrid Marie and Jonagold apples decreased when the apples included a description of taste. (from 45,34% of all sold apples with name only to 31 % of all sold apples when including a description for Ingrid Marie respective from 23,28% of all sold apples with name only to 19 % of all sold apples when including a description for Jona

***Problem 8 using jmp***

**1.**

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 3 | 6403,480 | 2134,49 | 18,0387 |
| Error | 304 | 35971,857 | 118,33 | Prob > F |
| C. Total | 307 | 42375,336 | | <,0001* |

**Parameter Estimates**

**Effect Tests**

| Source | Nparm | DF | Sum of Squares | F Ratio | Prob > F |
|---|---|---|---|---|---|
| baseline | 1 | 1 | 4621,3469 | 39,0552 | <,0001* |
| acupuncture | 1 | 1 | 1591,5803 | 13,4505 | 0,0003* |
| baseline*acupuncture | 1 | 1 | 624,4273 | 5,2771 | 0,0223* |

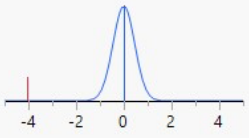
When I fit a model where Change depends on bassline and acupuncture as well as their interaction, I see that there is an interaction as we are rejecting the null hypothesis of no interaction. The F -test statistic is 5,2771, under the null hypothesis of no interaction, it is F-Distributed with 1 and 304 degrees of freedom which gives us a p-value of 0,0223. As the p-value is under our significance level we to reject the null hypothesis and conclude that there is an interaction. Hence I will use this model as my prediction formula for **expected change.**

**t Test**

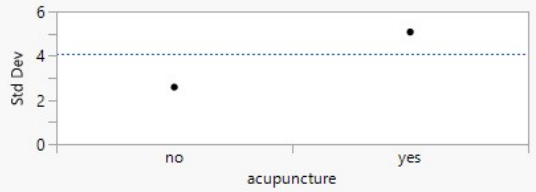
yes-no

Assuming unequal variances

| | | | |
|---|---|---|---|
| Difference | -4,0448 | t Ratio | -8,97213 |
| Std Err Dif | 0,4508 | DF | 248,6115 |
| Upper CL Dif | -3,1569 | Prob > \|t\| | <,0001* |
| Lower CL Dif | -4,9328 | Prob > t | 1,0000 |
| Confidence | 0,95 | Prob < t | <,0001* |




**Tests that the Variances are Equal**




The test statistics equals -8,97213, under the null hypothesis of no difference, it is t-distributed with 248,6115 degrees of freedom, which gives to a p-value < 0,0001. As the p- value is below 0,05 we do reject the null hypothesis of no difference. We can conclude that the expected change in headache severity depends on whether a migraine patient receives the acupuncture treatment or not. As the difference of the two means is negative (-4,0448) it means that the expected improvement is bigger when migraine patients are given acupuncture.

**2.**

**Summary of Fit**

| | |
|---|---|
| RSquare | 0,902486 |
| RSquare Adj | 0,901847 |
| Root Mean Square Error | 1,43084 |
| Mean of Response | -5,02354 |
| Observations (or Sum Wgts) | 308 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 2 | 5779,0524 | 2889,53 | 1411,382 |
| Error | 305 | 624,4273 | 2,05 | Prob > F |
| C. Total | 307 | 6403,4797 | | <,0001* |

**Parameter Estimates**

**Effect Tests**

**Indicator Function Parameterization**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| | Lower 95% | Upper 95% |
|---|---|---|---|---|---|---|
| Intercept | -0,767286 | 0,171997 | -4,46 | <,0001* | -1,105736 | -0,428836 |
| baseline | -0,241795 | 0,005143 | -47,01 | <,0001* | -0,251916 | -0,231674 |
| acupuncture[no] | 4,5737284 | 0,163791 | 27,92 | <,0001* | 4,251425 | 4,8960319 |

The estimated difference in expected change between the two types of treatment corrected for the possible effect of the baseline severity measurement is 4,5737284; a acupuncture has an effect on the expected change that is 4,5737284 lower than a standard treatment when it is corrected for the possible effect of the baseline severity measurement.

A 95% confidence interval, using the t-quantile with 305 degrees of freedom, for the estimated diffrence when a migraine patients is given a acupuncture is; $-4{,}5737284 + 1{,}967772355 \cdot 0{,}163791 = ]-4{,}8960319 ; -4{,}251425[$

**Thus** we are 95% certain that the effect of the acupuncture treatment on the expected change in headache severity corrected for the possible effect of the baseline severity measurement, is between -4,8960319 and -4,251425. Thus when a migraine patients is given a acupuncture treatment the expected change is lower than the expected change for a migraine patients given standard treatment, i.e as it is lower it indicate a larger (expected) improvement when a migraine patients is given acupuncture.