Statistics Exam – October 2022

Contents

Problem 1

The Danish government is considering placing a refugee center in Rwanda. People seeking asylum in Denmark would then be sent to Rwanda to apply from there. The probability that a left leaning voter supports this idea is 29%, whereas the probability that a right leaning voter supports it is 59%. The probability that a voter is left leaning is 53%.

USING CALCULATOR:

Subproblem 1

1. Calculate the probability that a voter supports the idea of placing a refugee center in Rwanda.

$$P(voter\ is\ left) = 0.53 \quad P(voter\ is\ right) = 1 - P(voter\ is\ left) = 0.47$$

$$P(support\ given\ left) = 0.1537 \quad P(support\ given\ right) = 0.2773$$

$$P(support) = 0.53 * 0.29 + 0.47 * 0.59 = 0.1537 + 0.2773 = 0.431$$

Subproblem 2

2. Suppose that your uncle supports the idea of placing a refugee center in Rwanda. Find the probability that your uncle is left leaning.

$$P(Left\ supporting) = 0.29$$

$$P(Left) = 0.53$$

$$P(Left\ and\ Left\ supporting) = 0.081461$$

$$P(left\ given\ support) = \frac{P(Left\ and\ Left\ supporting)}{P(Left\ supportimg)} = \frac{0.081461}{0.29} = 0.2809$$

Problem 2

Problem 2

Wash trading –when an investor simultaneously sells and buys the same financial instrument– is illegal as it may be used to manipulate the financial market. The distribution of the number of wash trades per person among persons involved in wash trading in Denmark is given below.

x	1	2	3	4	5
$P(x)$	0.83	0.10	0.04	0.02	0.01

1. Find the expected number of wash trades per person among persons involved in wash trading in Denmark.

1. Using calculator, expected number of wash trades per person:

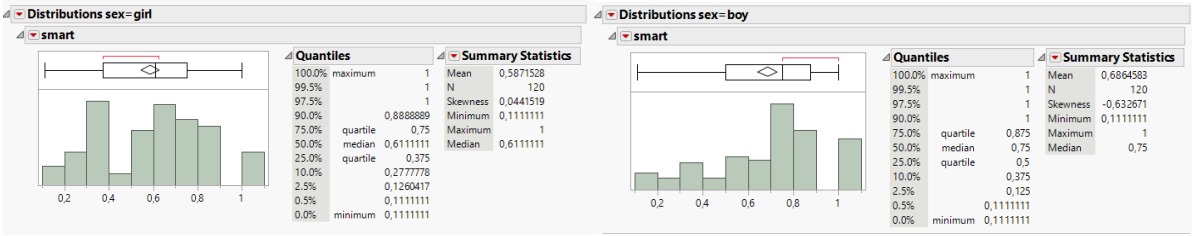
$$Expected\ number\ of\ wash\ trades = \\ 1 * 0.83 + 2 * 0.1 + 3 * 0.04 + 4 * 0.02 + 5 * 0.02 = 1.28$$

Problem 3

Problem 3

The data file contains information about to which degree a child thinks that a person of the same sex as the child is "really, really smart" (`smart`). There is also information about the sex of the child (`sex`).

1. Calculate five-number summaries for the degree to which a girl thinks that the person of the same sex is "really, really smart" as well as for the degree to which a boy thinks that the person of the same sex is "really, really smart". Briefly compare the distributions in the two groups based on the summaries.



The 5 point summary for sex=girl:	The 5 point summary for sex=boy:
$Minimum = 0.1111111$	$Minimum = 0.1111111$
$Lower\ quartile = 0.375$	$Lower\ quartile = 0.5$
$Median = 0.6111111$	$Median = 0.75$
$Upper\ quartile = 0.75$	$Upper\ quartile = 0.875$
$Maximum = 1$	$Maximum = 1$

1: using JMP

The girl's distribution shows that girls are less likely to think that a person of the same sex is "really, really smart" than boys. This can be seen by a lower median of 0.6111111 compared to the boy's 0.75. Furthermore, the lower and upper quartiles are 0.125 lower relatively for girls to boys. The girl's distribution is bimodal with a right skewness of only 0.0441519. The boy's distribution seems to be unimodal with more left skewness (more likely to believe that the same sex is "really, really smart") of -0.632671.

Therefore, we conclude that boys are more likely to think that a person of the same sex is "really, really smart" than a girl is to think that.

Problem 4

Problem 4

Two hundred and twenty (220) of 282 Danish homeless older adults were vaccinated against COVID-19.

1. **Test whether the probability that a homeless older adult was vaccinated differs from 94.6%, the overall probability of being vaccinated in Denmark. If it does, briefly explain how.**

USING A CALCULATOR

$$\hat{P}_{homeless} = \frac{220}{282} = 0.780141844 \quad \hat{P}_{vaccinationDenmark} = 0.946$$

$$Difference\ in\ proportions = -0.165858156$$

1. Test statistic for the hypothesis that the probability that a homeless older adult was vaccinated equals 0.946 is

$$\frac{\frac{220}{282} - 0.946}{\sqrt{0.946 * (1 - 0.946)/282}} = -12.32307066$$

Under the null, it's approximately normally distributed, thus resulting in a p-value indistinguishable from 0. Therefore, we reject the null hypothesis and can conclude that the probability that a homeless older adult was vaccinated significantly differs from the overall probability of being vaccinated in Denmark. A smaller proportion of homeless older adults are vaccinated compared to the overall population which can be seen by the difference in proportions (estimate: -0.165858156)

Problem 5

Problem 5

A study found that 17 of 40 women lied about their height in their online dating profile. For men, it was 22 of 40, who lied.

1. Calculate a 95% confidence interval for the difference between the probabilities that a woman lies about her height and that a man lies about his height. What can you conclude based on this confidence interval?

1. Using calculator

$$\hat{P}_{women} = 0.425$$

$$\hat{P}_{men} = 0.55$$

$$Estimated\ difference = -0.125$$

$$Standard\ error = 0.110891276$$

A 95% confidence interval is calculated:

$$]-0.342342906; 0.092342906[$$

As the predicted difference in proportion includes 0, difference cannot be assumed between a woman lying about height and a man lying about height. There is not a significant difference at a level of 5% for any of the proportions.

Problem 6

Problem 6

Dust exposure is a health hazard. To investigate the risk, dust exposure was measured for a group of drill & blast tunnel workers and a group of outdoor concrete workers. The table below contains summary statistics for the measured dust exposure for the two groups.

Group	Mean	StdDev	N
Drill & blast	18.0	7.8	115
Outdoor concrete	6.5	3.4	220

1. Calculate a 99% confidence interval for the difference between the expected dust exposure for drill & blast tunnel workers and the expected dust exposure for outdoor concrete workers. What can you conclude based on this confidence interval?

1. As $\frac{7.8}{3.4} = 2{,}294118$, we do not assume same standard deviations as they are not within a factor of 2 of each other.

$$Difference\ of\ means = 11.5$$

$$Standard\ error = 0.762620$$

$$Degrees\ of\ freedom = 137.066$$

Assuming unequal standard deviations, using the calculated (using calculator) standard error and degrees of freedom, we get a 99% confidence interval of:

$$]9.507891; 13.492109[$$

Since the 99% confidence interval does not include 0, we may assume that there is a significant difference in risk of dust exposure in the two groups. The risk of dust exposure is higher in the Drill&Blast group by looking at the difference of means.

Problem 7

Problem 7

The table below contains the number of male and female COVID-19 patients in a Chinese hospital and the amount of oxygen they required: No oxygen, supplemental oxygen, or ventilation.

	Amount of oxygen required		
Sex	None	Supplemental	Ventilation
Female	161	403	31
Male	157	461	63

1. Test whether the amount of oxygen required depends on the sex of the patient. If it does, briefly explain how.

Using JMP:

Tests

N	DF	-LogLike	RSquare (U)
1276	2	4,6302812	0,0045

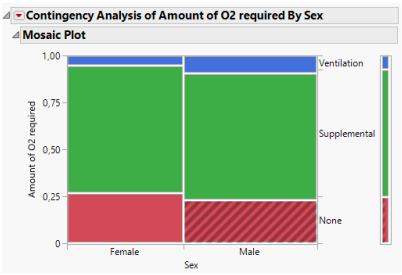
Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	9,261	0,0098*
Pearson	9,082	0,0107*

Pearson's χ^2-test statistic is equal to 9.082. Under the null, it is χ^2-distributed with 2 degrees of freedom. We are given a p-value of 0.0107, and therefore, we reject the null hypothesis of no dependence. We conclude that the amount of oxygen required depends on sex.

Freq: Count

Contingency Table

	Amount of O2 required		
Sex	Row %	None	Supplemental	Ventilation
	Female	27,06	67,73	5,21
	Male	23,05	67,69	9,25



Based on the row proportions or the mosaic plot, we are able to see that more men require ventilation than women (9.25% men and 5.21% women). Furthermore, more women require none than men (27.06% and 23.05%, respectively). We see almost no difference in supplemental.

Problem 8

Problem 8

The data file contains information about to which degree a child thinks that a person of the same sex as the child is "really, really smart" (`smart`). There is also information about the sex of the child (`sex`) as well as information about the age of the child (`age`; categorical). Furthermore a variable, `sexage`, with the combination of sex and age is included in the data file; this variable may or may not be useful.

1. Test whether the degree to which a child thinks that the person of the same sex as the child is "really, really smart" depends on the age of the child. If it does, briefly explain how.

Subproblem 1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
age	2	0,820086	0,410043	7,3571	0,0008*
Error	237	13,208965	0,055734		
C. Total	239	14,029051			

USING JMP

1. The test statistic for the null hypothesis of no difference between "smart" and age equals 7.3571. It is F-distributed under the null hypothesis with (2, 237) degrees of freedom, yielding a p-value of 0.0008. Thus, we reject the null and can assume that there is a significant difference between at least one of the age groups.

Connecting Letters Report

Level		Mean
five	A	0,71892361
seven	B	0,60399306
six	B	0,58750000

Levels not connected by same letter are significantly different.

Conducting a Tukey HSD connecting letters report using JMP, we can see that there is a significant difference between age group five and six. However, we cannot assume difference between age group six and seven.
Therefore, five-year-olds are significantly more likely to believe that a person of the same sex is "really, really smart" compared to six- and seven-year-olds.

2 IS ON NEXT PAGE:

Subproblem 2

2. Test whether the effect of the child's age on the degree to which a child thinks that the person of the same sex as the child is "really, really smart" depends on the sex of the child. If it does, briefly explain how.

   It may be useful to use five year old boys as the reference level in the analysis.

USING JMP

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
sex	1	1	0,59169560	11,2586	0,0009*
age	2	2	0,82008584	7,8022	0,0005*
sex*age	2	2	0,31939140	3,0386	0,0498*

2. The test statistic of no interaction between sex and age equals 3.0386. It is F-distributed with (2, 234) degrees of freedom, and this gives us a p-value of 0.0498. At a significance level of 5%, we reject the hypothesis of no interaction and can assume that the degree to which a child of a certain age thinks that the person of the same sex is "really, really smart" depends on the sex of the child.

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0,5434028	0,036247	14,99	<,0001*
sexage[boy&five]	0,1760417	0,051262	3,43	0,0007*
sexage[girl&five]	0,175	0,051262	3,41	0,0008*
sexage[boy&six]	0,1319444	0,051262	2,57	0,0107*
sexage[girl&six]	-0,04375	0,051262	-0,85	0,3943
sexage[boy&seven]	0,1211806	0,051262	2,36	0,0189*

Looking at the parameter estimates, we see that as children grow up, they are less likely to think that a person of the same sex is "really, really smart". Furthermore, it proves that boys are more likely to think that a person of the same sex is "really, really smart". If we look at sexage[boy&five], we see that they are 0.1760417 more likely than a sexage[girl&five] to think that a person of same sex is "really, really smart". This stays true for sexage[boy&six] and sexage[boy&seven] at 0.1319444 and 0.1211806, respectively, when compared to girls the same age. We even see a special case, where girls aged 6 are 0.04375 less likely to think that a person of same sex is "really, really smart" when compared to a boy aged seven, further proving the point of sex playing a significant role.