Statistic exam 2022 October

Problem 1

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

$$P(S) = P(S|L) * P(L) + P(S|Not\ L) * P(Not\ L) = 0.29 * 0.53 + 0.59 * 0.47$$
$$= 0.1537 + 0.2773 = 0.431$$

The probability of a voter supports the idea of placing a refugee center in Rwanda is 0.431

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

$$P(L|S) = \frac{P(L) * P(S|L)}{P(L) * P(S|L) + P(Not\ L) * P(S|Not\ L)} = \frac{0.53 * 0.29}{0.53 * 0.29 + 0.47 * 0.59} = \frac{0.1537}{0.431}$$
$$= 0.356612529$$

The probability that your uncle is left given that he supports the idea is 0.356612529

Problem 2

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

$$1 * 0.83 + 2 * 0.1 + 3 * 0.04 + 4 * 0.02 + 5 * 0.01 = 0.83 + 0.2 + 0.12 + 0.08 + 0.05$$
$$= 1.28$$

The expected number of wash trades per person involved in wash trading is 1.28

Problem 3

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 five number summaries for the degree to which a girl thinks that the person of same sex (girl) is smart is:

$$Minimum = 0.111$$
$$1st\ quantile = 0.375$$
$$Mediam = 0.6111$$
$$3rd\ quantile = 0.75$$
$$Maximum = 1$$

The five number summaries for the degree to which a boy thinks that the person of same sex (boy) is smart is:

$$Minimum = 0.111$$
$$1st\ quantile = 0.5$$
$$Median = 0.75$$
$$3rd\ quantile = 0.875$$
$$Maximum = 1$$

The minimum and maximum is the same for both boys and girls. However, the median is higher for boys (0.75) than girls (0.611), which mean that boys to a higher degree thinks that same sex (boy) is smart than the degree of girls thinking the same sex (girls) is smart. The 1^{st}

and 3rd quantile for boys is 0.5 and 0.875, which for girls is 0.375 and 0.75. This again results in boys having a higher degree of thinking that same sex (boys) is smart than girls do.

Problem 4

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.

Hypothesis Test (Proportion)			
Significance Level		0,05	
H_o:	p =	0,946	
$1-P_o$		0,054	
P# of population X		220	
n (sample size)		282	
p̂ (population proportion)		0,780142	
Standard Error		0,013459	
Z Statistic		-12,323071	
P-Value		0,000000	Reject the null
P-Value (two-tail)		0,000000	Reject the null

Proportion:

$$z = \frac{P - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}}$$

The test statistic for the hypothesis that the probability of an homeless older have been vaccinated is equals 0.946 is -12.323071. This gives a p-value way lower than 0.0001. Therefore, we reject the null hypothesis of no difference. We may conclude that a smaller proportion of homeless people (estimate is 0.780142 =78.0142%) have been vaccinated than 94.6% which is the overall probability of been vaccinated in Denmark.

Problem 5

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?

Hypothesis Testing (Diff of Prop)		
Confidence Interval	0,95	
α	0,025	
Null Hypothesis (prob 0)	0	
n_1 (sample size)	40	
population of #1	17	
Proportion ($\hat{p}_1$)	0,425	
n_2 (sample size)	40	
population of #2	22	
Proportion ($\hat{p}_2$)	0,55	
Estimate for Difference	-0,125	
Pooled Proportions	0,4875	
Standard Error (CI)	0,110891276	
Standard Error (test)	0,111768455	
Z Statistic	-1,959963985	For calculation
Z Statistic (see formula ttr)	-1,118383538	Report
p̂	0,4875	
Lower Bound	-0,342342906	
Upper Bound	0,092342906	
Reference for conclusion (Null, p1 or p2)	0	
As our predicted difference in proportion includes 0 we can not assume difference in the proportion between p_1 and p_2.		

Eg. Proble

Hypothesis Test (Diff of Mean)		
Significance Level	0,05	
Null Hypothesis (prob 0)	0	
Z Statistic	-1,118383538	Report
Two Tailed P-Value p > |t|	0,2634032	Don't reject the null
P-Value p > t	0,1317016	Don't reject the null
P-Value p < t	0,8682984	Don't reject the null

The test statistic to test the hypothesis of no difference between the two probabilities equals -1.118383538. The test statistic is approximately standard normal distributed which gives us a p-value of 0.2634032. The confidence interval is $]-0.34234291; 0.09234291[$, which 0 is included in the interval. Because 0 is included and the p-value is larger than 5%, we can therefore not assume difference in the probabilities of a woman lies more about her height than a man do.

Problem 6

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?

Confidence Interval (Diff of Mean)		
Confidence Interval	0,99	
α	0,01	
Mean	18,000	
Standard Deviation	7,8000000	
n (sample size)	115	
Mean	6,50	
Standard Deviation	3,4000000	
n (sample size)	220	
Difference of Means	11,500000	Report
Standard Error	0,762620	Report
Degrees of Freedom	137,066067	Report
T Statistic	2,612192	
Lower Bound	9,507891	
Upper Bound	13,492109	

	Hypothesis Test (Diff of Mean)		
1	Significance Level	0,05	
2	1- or 2-tailed	2	
	Null Hypothesis (prob 0)	0	
	T Statistic	15,07959833	Report
	P-Value	0,00000	Reject the null

As the two standard deviations are highly different, we use the "confidence means unequal" test. The 99% confidence interval is]9.507891; 13.492109[using a t-quantile with 137.066067 degrees of freedom. The confidence interval tells us that on average significantly the drill & blast tunnel workers is exposed to more dust exposure than outdoor concrete workers. This means that blast tunnel workers are exposed from 9.5 to 13.5 percentage points more dust than outdoor concrete workers. The p-value is far below 0.0001 with an test statistic equal 15.07959833 which also signals a difference in exposure.

Problem 7

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

Number of Rows (R)	2
Number of Columns (C)	3
Level of Significance	5%

RxC Table	None	Supplemental	Ventilation		
Female	161	403	31		
Male	157	461	63		
Column Totals	318	864	94	0	0

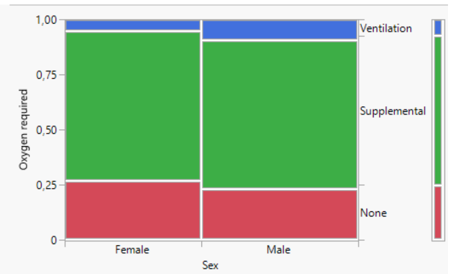
RESULTS*			
Chi-square Statistic	9,0825	Report	Large values are evidence against the null hyp
P-value	0,010660	Report	Reject the null hypothesis
Degrees of freedom	2	Report	Depends on r and c
Standard deviation:	2,0000	$s = \sqrt{2df}$	
# of std.devs. From expe	3,5412		

The Pearson test statistic of no difference between sex equals 9,0825. Under the null it is χ^2 distributed with 2 degrees of freedom, giving a p-value of 0.010660. Therefore, the null hypothesis is rejected, and oxygen required depends on the sex of the patient. Looking at a mosaic plot, the proportion of male requiring ventilation (9.25) is almost the double of female requiring ventilation (5.21). The proportion of supplemental oxygen requiring is almost the same: female (67.73) is a bit higher than male (67.69). Female (27.06) is also higher in the proportion of none oxygen requiring than male (23.05).

[redacted]

21-10-2022

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



Weight: Column 3

Contingency Table

Oxygen required

Sex	Count / Total % / Col % / Row %	None	Supplemental	Ventilation	Total
Female	Count	161	403	31	595
	Total %	12,62	31,58	2,43	46,63
	Col %	50,63	46,64	32,98	
	Row %	27,06	67,73	5,21	
Male	Count	157	461	63	681
	Total %	12,30	36,13	4,94	53,37
	Col %	49,37	53,36	67,02	
	Row %	23,05	67,69	9,25	
Total	Count	318	864	94	1276
	Total %	24,92	67,71	7,37	

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*

Problem 8:

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.

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			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
five	80	0,718924	0,02639	0,66693	0,77092
seven	80	0,603993	0,02639	0,55200	0,65599
six	80	0,587500	0,02639	0,53550	0,63950

Std Error uses a pooled estimate of error variance

The test statistic for testing of no difference between the three ages equal 7.3571. Under the null it's F-distributed with 2 and 237 degrees of freedom, which gives a p-value of 0.0008. Therefore, we reject the null hypothesis of no difference between the three ages, and we may conclude that at least one of the ages differs from the other two.

Tukey's procedure shows that age 5 (letter A) differs from age six and seven (both have letter B). The degree to which a child thinks that the person of same sex as the child is "really smart" is highest for the 5-year-old (0.71892361), and falls when a child is getting older. Age 6 and 7 has almost the same level of degree, however seven (0.60399306) is a little higher than six (0.5875).

Connecting Letters Report

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

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.

Effect Tests

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

The test statistic for testing the hypothesis that 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 smart" does not depend on the sex of the child is 3.0386. Under the null it's F-distributed with 2 and 234 degrees of freedom, giving a p-value of 0.0498. The null is therefore rejected, and it seems the effect of the child's age and the degree of thinking same sex is smart depends on the sex of the child.

In the picture below, we can clearly see how the effect of the child's age and the degree of thinking same sex is smart depends on the sex of the child. Girls are the red cross and boys are the blue circles. As we can see below, girls have a lower degree of thinking that same sex is smart in every age. In age 5 the degree is almost equal between the two sex (0.7184 for girls and 0.7194 for boys), but girls degree is a little bit less. In age 6 girls have a degree of 0.49965278 and boys have 0.67534722, which clearly is a big difference. In age 7 for girls have a degree of 0.54340278 and for boys the degree is 0.66458333, again a big difference.

[redacted] 21-10-2022

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