Copenhagen Business School

Exam in Statistics

HA-IB 2nd year
BSc International Shipping and Trade 1st year

Tuesday 16 January 2018 9:00-13:00

Solutions

Problem 1

Put

$$N = \textit{uses Nutella}$$
$$B = \textit{uses butter}$$

1.

$$P(N \textbf{ and } B) = P(B|N) \cdot P(N) = P(B|N) \cdot (1 - P(\text{not } N)) = 0.43 \cdot (1 - 0.31) = 0.2967$$

Problem 2

1. Probability of at most two

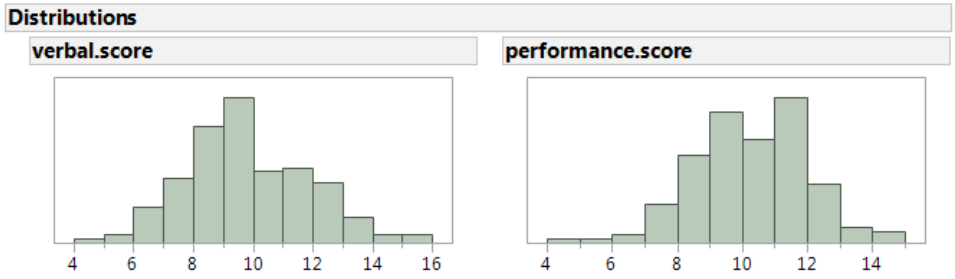
$$P(1) + P(2) = 0.606$$

2. Expected household size is

$$\sum_x x \cdot P(x) = 2.471$$

Problem 3

1.



Both distributions appear approximately symmetric. Both distributions may be bi-modal and slightly skewed (verbal sccores to the right, performance scores to the left).

2. As the observations are paired, the appropriate confidence interval is the confidence interval for the mean of the differences:]0.1778;0.6777[(using a t-quantile with 173 degrees of freedom).

Confidence Intervals

Parameter	Estimate	Lower CI	Upper CI	1-Alpha
Mean	0,427778	0,17783	0,677725	0,900
Std Dev	1,993722	1,832901	2,188452	0,900

The conclusion is that we are 90% confident that the performance score is between 0.1778 and 0.6777 higher than the verbal score.

Using tables rather than JMP the appropriate quantile is 1.660 corresponding to degrees of freedom equal to 100 which yields a confidence interval of]0.1769;0.6787[.

Problem 4

1. The difference between the two proportions equal 0.02662 and the standard error equals 0.02593; hence the confidence interval equals $]-0.0774;0.0242[$ (using the standard normal quantile of 1.96).

   Based on the confidence interval we cannot rule out that the "Coconut Eclair" is equally disliked by men and women; as the confidence interval includes 0, the difference is not significant.

   JMP answers $]-0.0773;0.0243[$ using a different formula.

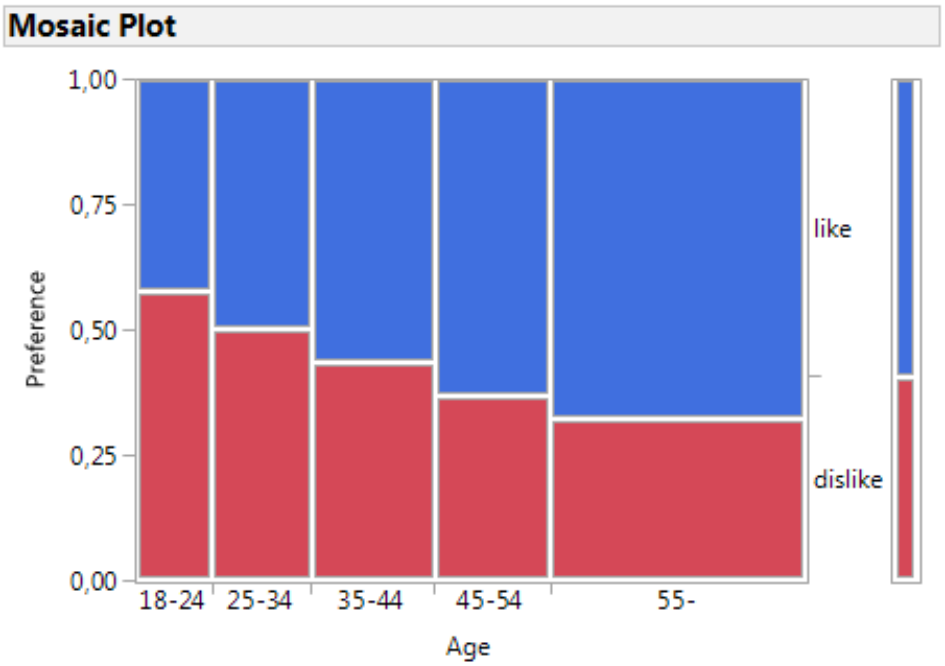
2. The Pearson χ^2-test statistic equals 45.156. Under the null it is approximately χ^2-distributed with 4 degrees of freedom and this gives us a p-value of less than 0.0001. Thus we reject the null hypothesis of no dependence and conclude that the probability of disliking the "Coconut Eclair" depends on age.

Tests

N	DF	-LogLike	RSquare (U)
1444	4	22,510405	0,0231

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	45,021	<,0001*
Pearson	45,156	<,0001*

From the mosaic plot or the estimated probabilities (42%, 50%, 57%, 63%, and 68%), it is clear the older you are the more likely it is that you like the "Coconut Eclair".



Contingency Table

	Preference	
Row %	dislike	like
18-24	57,58	42,42
25-34	50,00	50,00
35-44	43,40	56,60
45-54	37,05	62,95
55-	32,36	67,64

Age

Problem 5

1. As the two standard deviations are very different, a "not assuming the same standard deviations" t-test is the right test to use. Here the standard error of the difference of means equals 44.0876. The test statistic equals 1.1273, which is approximately t-distributed under the null with 26.492 degrees of freedom. The coresponding p-value equals 0.2697 (or a p-value larger than 20% -using the tables, comparing with 1.315, the 90%-quantile of the t-distribution with 26 degrees of freedom). Thus we cannot reject the null hypothesis, so we conclude that it is possible that the food supplement does not affect muscular power.

Problem 6

1. The test statistic for testing the hypothesis of no difference between the four groups equals 2.4649. Under the null hypothesis, it is F-distributed with 3 and 170 degrees of freedom, and this gives us a p-value of 0.0641. Hence we cannot reject the null hypothesis and conclude that verbal intelligence may not depend on age.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
age	3	29,72836	9,90945	2,4649	0,0641
Error	170	683,44869	4,02029		
C. Total	173	713,17704			

2. Estimates from the model:

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4,3260617	0,857667	5,04	<,0001*
performance.score	0,5223366	0,080468	6,49	<,0001*
age[a]	0,4602377	0,67375	0,68	0,4955
age[b]	-0,417033	0,348592	-1,20	0,2332
age[c]	0,5947651	0,32893	1,81	0,0724

After correcting for the effect of performance intelligence, verbal intelligence in age group 'a' is estimated to be 0.460 higher than in age group 'd' (the reference group). The verbal intelligence in age group 'c' lies 0.595 higher than in 'd' whereas it is 0.417 lower in age group 'c'.

The test statistic for testing the hypothesis of no difference between the four groups when the effect of performance intelligence is taken into account equals 2.7995. Under the null hypothesis, it is F-distributed with 3 and 169 degrees of freedom, and this gives us a p-value of 0.04166. Hence we reject our hypothesis of no difference and may conclude that once the performance intelligence is taken into account the age groups differ in verbal intelligence.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
performance.score	1	1	136,39349	42,1356	<,0001*
age	3	3	27,18647	2,7995	0,0417*

Age group 'c' has the largest average verbal intelligence when corrected for performance intelligence; group 'a' only differs very little (0.134) from group 'c'. Group 'b' has the smallest verbal intelligence and group 'd' is roughly midway between 'b' and 'a'/'c'. Though we are able to conclude that there are significant differences between the age groups, it does not seem possible[1] to conclude which of the age groups differ.

[1] even if we change reference group, none of the differences become significant, not even if we do not correct for multiple testing