1.

a) There population being examined is represented by two samples. The two samples are houses with fireplaces and houses without fireplaces. The sample of houses with fireplaces consists of a 100 houses, which is large enough to meet the nearly normal assumption. The same goes for the sample of houses without fireplaces. The condition of independence is also assumed to have been met.

b)
F = Houses with fireplaces
N = Houses with NO fireplaces

H_0: F – N = 0
H_A: F – N ≠ 0

Assumption: The two population variances are unequal

Therefore, the test used is the **“two-sample t-test”**.

Variable	Fireplace	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3
Price ($)	0	100	0	125973	4361	43611	54210	98637	114423	147632
	1	100	0	190577	6701	67010	69814	138577	185005	239996

SE (F – N) = $\sqrt{[(67010^2/100) + (43611^2/100)]}$
= 7995.16

t = [(190,577 – 125973) – 0]/7995.16
= 8.08

Using technology:
Degrees of freedom = 170
p-value < 0.0001

The p-value is lower than the three alpha values most commonly used (0.05, 0.01, 0.001); therefore, the result is significant and we reject the null hypothesis.

c) The corresponding 95% confidence interval for the difference in mean house prices is:
The 95% critical value for a t with 170 degrees of freedom is around 1.973

190,577 – 125,973 ± 1.973 x 7995.16
64604 ± 15774.45

(49029.55, 80178.45)

d) The test in part b gave a p-value so low that we had to reject the null hypotheses. The confidence interval in part c also further confirms that we have to reject the null hypotheses since the value of zero is not covered by the interval in part c – meaning that these two samples are different.

2.

a)
μ_1: Average spending in December
μ_2: Average spending across January, February and March

H_0: $\mu_1 - \mu_2 = 0$
H_A: $\mu_1 - \mu_2 \neq 0$

Since we are comparing the same 200 customers across these months we have data that is called "paired data".

Therefore, the test used is the **"Paired t-Test"**.

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
December	200	0	731.8	47.5	671.2	-33.4	175.6	505.6	1165.1	2385.5
JantoMarAvg	200	0	579.2	32.9	464.8	6.8	196.5	456.7	917.3	2030.9

Using technology, the standard error of difference = 35.75
t = [(731.8 – 579.2) – 0]/35.75
 = 4.27

Using technology, the p-value of the one tailed hypothesis is 0.000015 which is less than the 0.05 level of significance. Therefore, we reject the null hypothesis and this means that the average spending in December is indeed different to that of January to March.

b) The 95% one-sided confidence interval is:
The corresponding critical value of t with 199 degrees of freedom in terms of one-tail probability is **1.653**

152.6 ± 1.653 x 35.75

152.6 ± 59.09

(93.51, 211.69)

This confirms the previous conclusion in part a. The interval does not include the value zero; therefore, the means are indeed different. Leading us to again reject the null hypothesis meaning that the average spending in December is different than that of January to March.

3.

a)

Observed Counts

	uOttawa	Carleton	UofT	Totals
0 to 9	84	91	70	245
10 to 19	43	24	35	102
20 or more	21	22	14	57
Total	148	137	119	404

Expected Counts

	uOttawa	Carleton	UofT	Totals
0 to 9	89.75	83.08	72.17	245
10 to 19	37.37	34.59	30.04	102
20 or more	20.88	19.33	16.79	57
Total	148	137	119	404

b)

H_0: The distributions for number of books read and the university attended are independent
H_A: The distributions for number of books read and the university attended are dependent

Chi^2

	uOttawa	Carleton	UofT	Totals
0 to 9	0.368	0.755	0.065	
10 to 19	0.848	3.242	0.819	
20 or more	0.000	0.369	0.464	
Total				6.930

$Chi^2 = 6.93$
$D_f = (R-1) \times (C-1) = (3-1) \times (3-1) = 4$

Using technology, at 0.01 significance level and 4 degrees of freedom, the p-value is 0.140 which is greater than 0.01. Therefore, the result is not significant and we fail to reject the null hypothesis. This means that we do not have a result that contradicts our null hypothesis.

c) The chi-squared approach is appropriate here since the data we have is counted, thus meeting the counted data condition.
There is no reason not assume the independence and randomization of the data since the use of appropriate sampling techniques was noted. Our sample size is 404 which is large enough. Finally, the expected cell frequency condition is met since in every cell on the expected count table the values are greater than 5.

4.

a)

The Mann-Whitney test is used since the 2 samples are assumed to be independent from one another.

The test for food safety at fairs is as follows:

H_0: $\mu_{w} - \mu_{m} = 0$

H_A: $\mu_{w} - \mu_{m} \neq 0$

Mann-Whitney Test and CI: WSF, MSF

```
       N  Median
WSF  196  2.0000
MSF  107  2.0000


Point estimate for η1 - η2 is 0.0000
95.0 Percent CI for η1 - η2 is (0.0000,0.0000)
W = 31996.5
Test of η1 = η2 vs η1 ≠ η2 is significant at 0.0025
The test is significant at 0.0009 (adjusted for ties)
```

The level of significance chosen is 0.05

P-value is 0.0025 which is < 0.05; therefore, we reject the null hypothesis meaning that the data for women and men is not the same.

The test for food safety at restaurants is as follows:
H_0: $\mu_{w} - \mu_{m} = 0$
H_A: $\mu_{w} - \mu_{m} \neq 0$

Mann-Whitney Test and CI: WSR, MSR

```
       N  Median
WSR  196  2.0000
MSR  107  2.0000


Point estimate for η1 - η2 is -0.0000
95.0 Percent CI for η1 - η2 is (0.0001,0.0001)
W = 32267.5
Test of η1 = η2 vs η1 ≠ η2 is significant at 0.0007
The test is significant at 0.0001 (adjusted for ties)
```

The level of significance chosen is 0.05

P-value is 0.0007 which is < 0.05; therefore, we reject the null hypothesis so the data for women and men is different.

b) It would not be appropriate to use the Wilcoxon Rank-Sum test/Mann-Whitney test because in this case the data is no longer 2 samples independent of each other, the data is actually paired data. Therefore, we have to use the Wilcoxon Signed-Rank test.

c)

μ_r = restaurant
μ_f = fair

H_0: $\mu_{r} - \mu_{f} = 0$
H_A: $\mu_{r} - \mu_{f} \neq 0$

Wilcoxon Signed Rank Test: Differences

```
Test of median = 0.000000 versus median ≠ 0.000000

                   N for  Wilcoxon          Estimated
               N    Test  Statistic      P     Median
Differences  303     157    10850.5  0.000     0.5000
```

The p-value is clearly below 0.05; therefore, we reject the null hypothesis meaning that restaurant food is indeed perceived as safer than the food at fairs.