Tony Lim
BIOSTAT 203A
Professor Hilary Aralis
13 Dec 2019

Final Project

Step 1

Completed in the attached SAS file.

Step 2

After conducting chi-square tests between the "stayed together in Wave 4" variable and the 18 HCMST variables with a Yes/No answer, there were five pairings that had statistically significant associations with p-values lower than 0.05 (Table 1). These pairings were between the "stayed together in wave 4" variable and

1. Respondents who identified as the same race as their partner, p-value = 0.0024;
2. Respondents whose age was within five years of their partner's age, p-value = 0.0007;
3. Respondents who met their partners through church/church activity, p-value = 0.0069;
4. Respondents who met their partners by personal ads/dating service via the internet, p-value < 0.0001; and
5. Respondents who met their partners at school, p-value = 0.0227.

After conducting a chi-square test for trend between the "stayed together in Wave 4" variable and relationship quality, it appears there is a statistically significant association with a Mantel-Haenszel chi-square test statistic of 230.52 and p-value < 0.0001 (Table 2).

Therefore, we reject the null hypothesis and suggest that there is a statistically significant association between the two variables for each of the six pairings (i.e., the six HCMST variables and the "stayed together in Wave 4" variable).

Table 1. Summary of chi-square tests between W4_TOGETHER and HCMST variables with a Yes/No answer

HCMST Variables	Stayed Together Wave 4	Count	Row %	Chi-Square Value	p-value
Are you and your partner the same race?	No	316	77.07	9.2384	0.0024
	Yes	1077	83.68	9.2384	0.0024
Is the difference between your age and your partner's age ≤ 5 years?	No	262	64.06	11.3666	0.0007
	Yes	938	72.77	11.3666	0.0007
Where did you meet: Bar/nightclub/dance club?	No	47	11.52	0.0136	0.9070
	Yes	151	11.73	0.0136	0.9070
Where did you meet: Church/church activity?	No	11	2.70	7.3008	0.0069
	Yes	79	6.14	7.3008	0.0069
Where did you meet: Other?	No	153	37.50	1.2230	0.2688
	Yes	444	34.50	1.2230	0.2688
Where did you meet: Personal ads/dating service via the Internet?	No	43	10.54	20.8341	<.0001
	Yes	57	4.43	20.8341	<.0001
Where did you meet: Private party?	No	50	12.25	1.7578	0.1849
	Yes	128	9.95	1.7578	0.1849
Where did you meet: School?	No	49	12.01	5.1948	0.0227
	Yes	215	16.71	5.1948	0.0227
Where did you meet: Social organization/health club/gym/volunteer-service activity?	No	21	5.15	0.7856	0.3754
	Yes	53	4.12	0.7856	0.3754
Where did you meet: Vacation/business trip?	No	6	1.47	0.2357	0.6274
	Yes	15	1.17	0.2357	0.6274
Where did you meet: Work?	No	54	13.24	3.2777	0.0702
	Yes	219	17.02	3.2777	0.0702
Who introduced you: Classmates?	No	16	3.94	0.0834	0.7728
	Yes	55	4.27	0.0834	0.7728
Who introduced you: Co-workers?	No	23	5.67	2.4381	0.1184
	Yes	103	8.00	2.4381	0.1184
Who introduced you: Family?	No	29	7.14	1.8185	0.1775
	Yes	120	9.32	1.8185	0.1775
Who introduced you: Introduced self or partner introduced self?	No	175	43.10	0.0788	0.7789
	Yes	545	42.31	0.0788	0.7789
Who introduced you: Mutual friends or acquaintances?	No	134	33.00	0.7852	0.3756
	Yes	395	30.67	0.7852	0.3756
Who introduced you: Neighbors?	No	6	1.48	0.0000	0.9969
	Yes	19	1.48	0.0000	0.9969
Who introduced you: Other?	No	40	9.85	0.0000	0.9962
	Yes	127	9.86	0.0000	0.9962

Only displaying respondents who answered 'Yes' to HCMST variables

Table 2. Summary of chi-square test between W4_TOGETHER and relationship quality

Stayed Together Wave 4	Relationship Quality	Count	Row %	Mantel-Haenszel Chi-Square Value	p-value
No	Very poor	8	1.96	230.5162	<.0001
No	Poor	19	4.65	230.5162	<.0001
No	Fair	79	19.32	230.5162	<.0001
No	Good	181	44.25	230.5162	<.0001
No	Excellent	122	29.83	230.5162	<.0001
Yes	Very poor	3	0.23	230.5162	<.0001
Yes	Poor	7	0.54	230.5162	<.0001
Yes	Fair	60	4.65	230.5162	<.0001
Yes	Good	349	27.08	230.5162	<.0001
Yes	Excellent	870	67.49	230.5162	<.0001

As M (the number of chi-square tests) increases, so does the proportion of replications that resulted in at least one statistically significant result (Figure 1). The calculated proportions represent the expected probability of making a type I error (rejecting a true null hypothesis/erroneously concluding at least one significant association even though there should not be an association since the data came from random samples from binomial distributions). Therefore, for situations where multiple hypothetical tests are being conducted, the probability of making at least one type I error increases, and some of the significant results may actually be false positives.
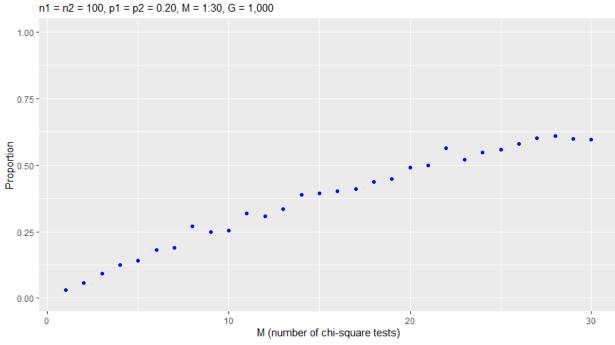
Figure 1. Proportion of replications with at least 1 statistically significant result depending on M
n1 = n2 = 100, p1 = p2 = 0.20, M = 1:30, G = 1,000

Step 4

When using the function created in Step 3 to mimic the results obtained in Step 2, we see that in the 1,000 element vector, the number of statistically significant results obtained in the 18 chi-square tests ranged from 0 statistically significant results to 4 statistically significant results (44.2% of the elements to 0.5% of the elements, respectively) (Figure 2, Table 3).

Therefore, I am not confident about the presence of statistically significant associations between respondents remaining with their partner in wave 4 and the variables in Step 2 since our function produced statistically significant associations even though there should not be any since our function used random samples from binomial distributions. Some of the statistically significant associations found in Step 2 may be spurious.

```
> step4 <- step3(412, 1292, 0.3, 0.3, 18, 1000)
   [1] 1 1 0 3 0 1 0 1 0 3 1 0 1 0 1 0 1 1 0 1 1 0 1 0 1 0 1 0 0 1 0 1 0 1 3 0 0 1 0 2 0 2 1 2 1 0 0 2 0 2 1 1 1 1 1 3 1 2 1
  [57] 0 0 1 2 1 0 1 0 0 0 1 0 0 1 1 3 0 0 1 1 0 2 1 0 0 2 2 1 0 0 1 2 0 0 0 0 1 0 3 1 3 0 2 0 0 1 1 2 0 0 0 2 0 2 0 0
 [113] 0 0 0 1 2 0 1 1 1 1 3 1 0 0 1 0 0 3 0 1 0 0 0 0 0 1 0 0 1 1 2 2 1 2 1 1 0 3 1 2 3 2 0 1 1 1 1 1 1 0 0 2 1 0 0
 [169] 1 0 3 1 1 1 0 0 0 0 1 0 2 0 4 0 0 1 1 2 1 0 2 0 0 2 2 0 1 0 1 1 0 0 0 1 2 2 2 0 0 1 2 1 1 0 1 0 2 1 1 1 0 0 2 1
 [225] 1 1 2 0 0 0 0 1 1 1 0 0 1 0 2 0 1 0 1 2 3 0 3 0 1 0 2 0 0 1 2 1 1 1 2 1 0 1 1 0 0 1 1 0 1 1 1 2 0 2 1 1 0 0 1 0
 [281] 0 2 0 0 0 2 2 1 2 1 1 0 0 2 0 0 1 1 1 0 0 1 0 2 1 0 1 1 0 2 1 0 0 1 0 0 1 0 0 2 1 1 0 0 0 2 2 0 0 0 1 1 1 0 0 2
 [337] 0 0 1 1 0 0 2 2 2 2 2 1 0 0 0 1 1 0 1 0 1 1 2 1 0 3 1 1 2 0 0 1 1 2 1 0 1 0 2 1 0 1 0 0 1 2 0 0 1 2 0 2 1 1 1 1
 [393] 0 1 0 0 1 0 0 0 1 1 1 0 0 1 2 1 0 1 0 2 0 0 0 0 1 0 0 0 2 1 0 2 1 1 0 2 0 2 0 0 2 1 0 1 0 2 2 1 1 1 1 0 1 2 0 0 0
 [449] 0 3 0 0 1 1 1 2 0 1 1 0 1 0 1 0 1 0 0 1 0 2 2 0 0 0 0 2 2 3 0 0 3 0 2 1 2 2 0 1 2 1 4 0 1 1 0 2 0 1 0 0 2 2 1 0
 [505] 0 0 1 1 1 2 2 0 1 0 1 1 0 0 1 0 1 0 0 0 0 0 2 0 1 1 0 0 1 2 0 1 2 0 1 2 0 1 0 1 1 1 1 0 1 1 1 1 0 0 0 0 1 3
 [561] 1 1 0 0 0 0 1 0 1 1 1 0 0 1 0 0 1 0 0 0 0 0 2 1 1 0 1 3 1 2 1 0 1 1 0 0 1 0 0 1 2 1 0 1 0 0 2 1 0 0 0 0 1 1 0 2 0
 [617] 0 2 1 3 2 0 1 2 1 0 0 0 0 2 1 1 4 1 0 1 0 1 2 0 1 1 1 2 1 2 2 0 0 2 1 0 0 3 1 4 0 0 0 1 1 1 0 1 0 0 0 2 0 1 0 1
 [673] 2 1 0 1 2 0 0 1 1 0 0 0 0 1 1 1 0 0 0 0 2 1 1 1 1 0 2 2 0 0 1 0 0 0 0 0 0 2 0 1 1 1 0 1 1 1 1 0 1 1 0 2 1 0 0 0
 [729] 0 0 1 1 0 0 2 2 0 0 0 1 4 1 1 1 1 0 1 2 0 0 1 1 0 0 1 0 0 0 0 0 1 1 1 2 2 2 2 0 1 0 0 1 1 0 0 0 0 1 2 0 0 1 0 1
 [785] 0 2 3 1 0 2 0 1 1 0 1 1 1 1 3 2 1 0 1 0 0 0 1 0 1 2 0 2 2 0 2 0 0 0 3 2 2 0 0 2 0 0 0 0 0 0 1 1 1 2 2 1 2 1 1 1
 [841] 1 2 1 2 0 0 0 2 2 0 0 1 1 2 1 0 0 1 1 0 0 0 2 1 1 1 0 0 0 0 2 1 1 1 0 1 2 3 1 1 0 1 0 2 0 0 0 0 1 1 0 0 1 0 0 0
 [897] 2 1 1 0 0 2 2 0 1 1 0 0 2 1 0 1 0 0 0 1 0 1 1 1 0 1 0 1 1 1 2 0 1 1 0 0 2 0 2 0 0 0 0 0 2 2 0 1 0 1 0 2 1 0 0 0
 [953] 0 1 0 0 0 1 2 2 2 0 0 1 0 1 1 0 1 0 2 1 0 1 2 2 0 0 0 0 3 1 1 0 0 0 2 1 2 0 1 1 1 0 0 3 0 1 1 0
```

Figure 2. Vector containing the number of statistically significant results for 1,000 replications of 18 chi-square tests produced by using the function created in Step 3 to mimic results obtained in Step 2

Value	Count	%
0	442	44.2%
1	363	36.3%
2	162	16.2%
3	28	2.8%
4	5	0.5%
Total	1,000	100.0%

Table 3. Counts and percentages of statistically significant results in the vector