Homework 7: Chi-Square 2 sample test of independence (Due Friday Apr. 9)

Example Problem
An occupancy rate is merely the proportion of a building's rentable space that is occupied. It can be expressed as a ratio of occupied units divided by the total number of units. Suppose you would like to test the hypothesis that small apartment project occupancy differs depending on proximity to public transportation. You randomly sample several small apartment buildings within walking distance to a city's mass transit system and find a total of 250 units, with 240 being occupied (sample 1). You also randomly sample several small apartment buildings that are too far from mass transit to be considered within walking distance (sample 2). The second sample consists of 200 units, with 180 being occupied.

Sample proportions are p_1 = .96 (240/250) and p_2 = .90 (180/200). We will test the hypothesis that occupancy rates differ based on proximity to mass transit by investigating whether the sample proportion differences are sufficient to support this proposition with 95% confidence. That is, does the proportion of occupied small apartments differ between near and distant apartments?

The statistical null hypothesis is

H_0: *occupancy rate is not associated (is independent of) with proximity to public transportation*

and the alternative research hypothesis is

H_A: *occupancy rate is associated (is dependent upon) proximity to public transportation*.

The 2×2 contingency table for this example is presented in Table 1.

Table 1
Apartment Occupancy by Proximity to Mass Transit

Proximity to Mass Transit			
	Within Walking Distance	Not Within Walking Distance	Total
Occupied	240	180	420
Unoccupied	10	20	30
Total	250	200	450

Cell frequencies and expected frequencies are shown below. No expected frequency is less than 5, so we can apply the χ^2 test.

Calculating the expected frequencies:

420 of the total 450 apartments (both types) are occupied, which equals 0.933 (93.3%) (Table 1). If the proportions between the two types (near and distant) are equal we would expect 0.933 of each. There are 250 near and 200 distant; 0.933 of 250 (near) is 233.25 (Table 2, column 3). 0.933 of 200 (distant) is 186.6. 233.25 and 186.6 represent the expected numbers of occupied near and distant apartments if the proportion occupied is equal between the two samples.

But there are actually 240 (higher than expected 233.25) near and only 180 (lower than expected 186.6) distant. The question we are trying to answer is: does the observed occupancy rate significantly differ from the expected occupancy rate. If we conclude yes, we are saying that $p1 \neq p2$; we are rejecting the H_0. We would then be saying that the difference is real, and that occupancy rate is higher near mass transit.

What do we need to do to determine whether or not the difference between expected and observed is significant ($\alpha = 0.05$)? We need to calculate X^2 and see if the value from our test is sufficiently large that we are beyond the X^2 critical (or in the rejection region of the X^2 curve) given df.

Degrees of freedom (df) = <u>n rows (2 in Table 1) -1</u> x <u>n columns (2 in Table 1) -1</u> = 1

Table 2
Chi Square Calculations for Proximity to Mass Transit Analysis (f = frequency*)*

Cell	f_o	f_e	$(f_o - f_e)$	$(f_o - f_e)^2$	$\frac{(f_o - f_e)^2}{f_e}$
1	240	233.25	6.75	45.56	.195
2	10	16.75	-6.75	45.56	2.72
3	180	186.60	-6.60	43.56	.233
4	20	13.40	6.60	43.56	3.25
					$\chi^2 = \Sigma = 6.40$

The *p*-value for a χ^2 of 6.40 with 1 degree of freedom is < 0.05

So, we reject the H_0

Exercise 1:
1) Phrase the results of this test using the guidelines provided by Cronk, pg. 90.
2) Assuming you are employed by a developer who is looking to build a new small-unit apartment complex, what recommendation would you make regarding where to buy property?
3) Use statistics to bolster your recommendation.

Exercise 2:
A random sample of local home owners who recently purchased a home yielded 60 respondents in the first time home buyer and age 50 or less category (sample 1), 42 of whom had purchased a multistory home. Additionally, the sample included 40 respondents in the 50+ age category (sample 2), 15 of whom had purchased a multistory home.

Defining a “success” as having purchased a multistory home, we can divide each sample into “success” and “failure” sub-categories: you will do a Chi-Square test following the steps below. Is purchase of multistory homes dependent upon age?

1) What are the null and alternative hypotheses for this test?
2) Create a 2 x 2 contingency table (like Table 1) displaying the data.
3) Create a second table showing the observed and expected frequencies as well as the derivation of X^2 (like Table 2).
4) Determine using the Chi-Square table and your X^2 test statistic whether or not you should reject the H_0.
5) Phrase your results appropriately
6) What might you conclude about age and number of stories in a home based on your results?

TABLE C: χ^2 CRITICAL VALUES

	Tail probability p α										
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4