STAT 110: Practice Problem 15 Solutions
Fall 2017

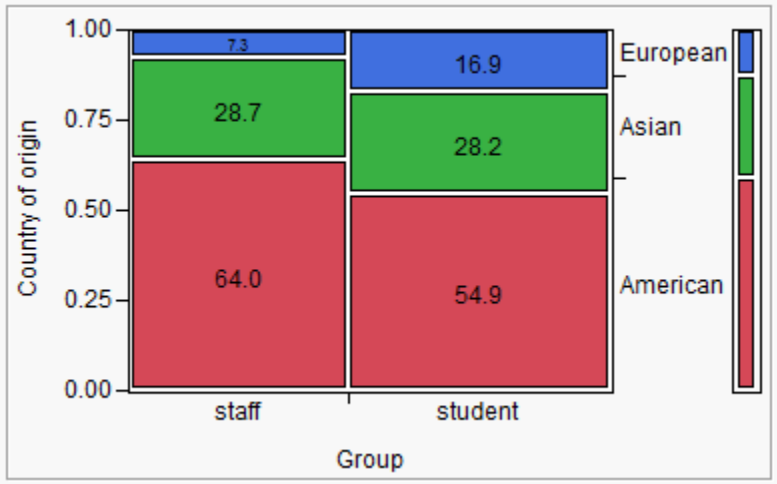
---

1. A European manufacturer of automobiles claims that their cars are preferred by the younger generation and would like to target university students in their next ad campaign. Suppose we test their claim with our own survey. Random samples of autos parked in the student lot and the staff lot at a large university classified the brands by country of origin, as seen in the following table.

	Student	Staff
American	107	105
European	33	12
Asian	55	47

Research Question: Are there differences in the national origins of cars driven by students and staff?

a. Use JMP to create a mosaic plot for these data, and describe the differences you see between student and staff preferences in this sample.



In our sample, the percentage of the staff that drives American cars (64%) is slightly higher than that of students (54.9%). Also, the percentage of the staff that drives European cars (7.3%) is slightly lower than that of students (16.9%).

b. Convert the research question into a null and alternative hypothesis.

H_o: There are no differences in the national origins of cars driven by students vs. staff
H_a: There are differences in the national origins of cars driven by students vs. staff

c. Calculate the expected counts "by hand" and then find the chi-square test statistic "by hand."

Observed counts:

Group	Count	American	Asian	European	
		Country of origin			
	staff	105	47	12	164
	student	107	55	33	195
		212	102	45	359

Overall, 195/359 = 54.3175% of the subjects are students. If the null hypothesis is true, we expect

- **54.3175% of the American cars to belong to students (.543175*212 = 115.15)**
- **54.3175% of the Asian cars to belong to students (.543175*102 = 55.40)**
- **The rest of the expected counts can be obtained by either this same logic or by subtraction**

Expected counts:

Group	Expected	American	Asian	European
		Country of origin		
	staff	96.8468	46.5961	20.5571
	student	115.153	55.4039	24.4429

Test Statistic =

$$\frac{(105-96.8468)^2}{96.8468}+\frac{(47-46.5961)^2}{46.5961}+\frac{(12-20.5571)^2}{20.5571}+\frac{(107-115.153)^2}{115.153}+\frac{(55-55.4039)^2}{55.4039}+\frac{(33-24.4429)^2}{24.4429}$$

= 7.83

d. Use JMP to verify the test statistic and to find p-value from these data.

Tests

N	DF	-LogLike	RSquare (U)
359	2	4.0788140	0.0122

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	8.158	0.0169*
Pearson	7.828	0.0200*

e. Write a conclusion in the context of the problem.
We have evidence that there are differences in the national origins of cars driven by students vs. staff (p-value = .02).

STAT 110: Practice Problem 15 Solutions
Fall 2017

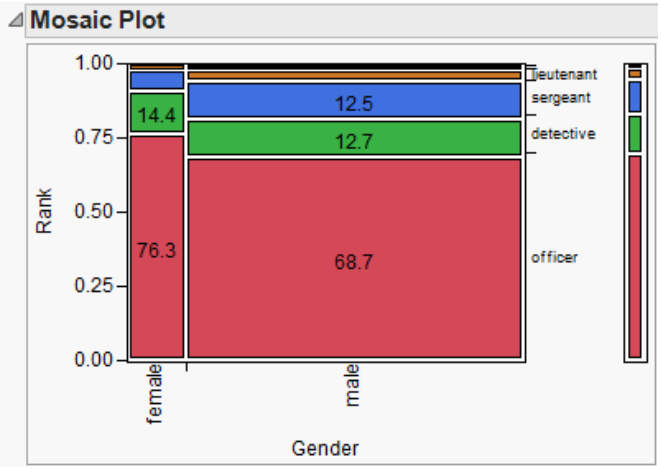
---

2. The following table shows the rank attained by male and female officers in the police department of a large U.S. city.

	Male	Female
Officer	21,900	4281
Detective	4058	806
Sergeant	3989	415
Lieutenant	1333	89
Captain	359	12
Higher ranks	218	10

Research Question: Do these data indicate that men and women are unequally represented at some levels of the department in this city?

a. Use JMP to create a mosaic plot for these data, and describe the differences you see between males and females in this sample.



In this sample, there are more female officers and detectives than male officers and detectives. For every other rank, there are slightly more males.

STAT 110: Practice Problem 15 Solutions
Fall 2017

Contingency Table

Gender	Count / Row %	officer	detective	sergeant	lieutenant	captain	higher	
		Rank						
female	Count	4281	806	415	89	12	10	5613
	Row %	76.27	14.36	7.39	1.59	0.21	0.18	
male	Count	21900	4058	3989	1333	359	218	31857
	Row %	68.74	12.74	12.52	4.18	1.13	0.68	
		26181	4864	4404	1422	371	228	37470

b. Convert the research question into a null and alternative hypothesis.

H_o: There is no association between gender and rank
H_a: There is an association between gender and rank

c. Calculate the expected counts "by hand" and then find the test statistic "by hand."

Observed counts:

Gende	Count	officer	detective	sergeant	lieutenant	captain	higher	
		Rank						
female		4281	806	415	89	12	10	5613
male		21900	4058	3989	1333	359	218	31857
		26181	4864	4404	1422	371	228	37470

Overall, 5613/37470 = 14.98% of the subjects are female. If the null hypothesis is true, we therefore expect

- **14.98% of all officers to be female (.1498*26181 = 3921.91)**
- **14.98% of all detectives to be female (.1498*4864 = 728.63)**
- **14.98% of all sergeants to be female (.1498*4404 = 659.72), etc.**

Expected counts:

Gende	Expected	officer	detective	sergeant	lieutenant	captain	higher
		Rank					
female		3921.91	728.626	659.718	213.015	55.5757	34.1544
male		22259.1	4135.37	3744.28	1208.98	315.424	193.846

Test Statistic =

$$\frac{(4281-3921.91)^2}{3921.91}+\frac{(806-728.626)^2}{728.626}+\frac{(415-659.718)^2}{659.718}+\cdots+\frac{(218-193.846)^2}{193.846}=300.3$$

d. Use JMP to verify the test statistic and to find p-value from these data.

Tests

N	DF	-LogLike	RSquare (U)
37470	5	177.62528	0.0049

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	355.251	<.0001*
Pearson	300.307	<.0001*

e. Write a conclusion in the context of the problem.
We have evidence that there is a relationship between gender and rank. More males appear to be in higher ranking positions.