

## Stress Survey

The New York Times recently reported (1/30/00) results from a national survey of college freshmen. In the survey, 30.2% said they felt “frequently overwhelmed by all I have to do.” Women in the survey were twice (39%) as likely as men (20%) to indicate that they were overwhelmed. How do these results compare to those of undergraduates here at Wharton?

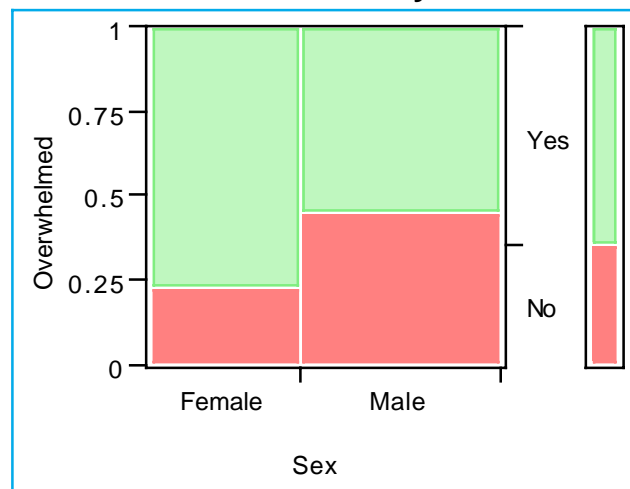
This question was part of the small “survey” conducted in class. For this example, I am going to treat the 70 responses as a sample of Wharton undergraduates, but you ought to think about whether you believe that this assumption is reasonable. The data for this example is available from my class web page

[www-stat.wharton.upenn.edu/~bob/stat102](http://www-stat.wharton.upenn.edu/~bob/stat102)

in the class survey data file.

A single table summarizes the survey responses for both male/female and perception of stress. You can reproduce this figure and table by using the “Fit Y by X” tool and picking the column “Sex” as X and the column “Overwhelmed” as Y.

**Overwhelmed By Sex**



Overwhelmed	Sex		
	Female	Male	
No	7	18	25
Yes	23	22	45
	30	40	70

Overall, 45 out of 70 (64%) felt stressed, with 23 out of 30 women (77%) and 22 out of 40 men (55%) replying “yes” for this question.

Do these results differ significantly from the national results? For the overall level of stress, does the survey percentage 64% differ significantly from that reported in the *Times*?

It is quite a bit higher, but is it significantly different? If we build a 95% confidence interval for the population proportion who say “yes” based on our survey, we get the interval

$$[0.64 \pm 2 \sqrt{(0.64)(0.36)/70}] = [0.64 \pm 2(0.06)] = [0.64 \pm 0.12]$$

Since this interval does not include the reported national value, we can conclude (with 95% confidence and a host of assumptions) that the perceived level of stress is higher here at Wharton. Formulated as a test (one or two-sided?), the sample proportion 0.64 lies

$$z = (0.64 - 0.30)/0.06 = 5.7$$

standard errors away from the hypothesized (national) proportion – clearly a significant difference. For the sake of comparison, here’s the JMP output for this same problem, using the “Distribution of Y” command with Y set to the column “Overwhelmed.”

Level	Count	Probability	StdErr	Prob	Cum Prob
No	25	0.357		0.057	0.35714
Yes	45	0.643		0.057	1.00000
Total	70				

In order to get JMP to do the interval, we have to use the associated “dummy variable” which is coded as 0s and 1s. Use the same “Distribution of Y” command with Y set to the column “Dummy”. Notice that the mean of the dummy column is just the proportion saying “yes” to the stress question. The interval is slightly different from that found above due to rounding.

### Dummy

Mean	0.643
Std Error Mean	0.058
Upper 95% Mean	<b>0.758</b>
Lower 95% Mean	<b>0.528</b>
N	70

What about the difference between men and women here at Wharton – is this difference (22% = 77% – 55%) comparable to the national value (19% = 39% – 20%)? We can again use a confidence interval, this time for the difference in two proportions. The standard error is (see Definition 10.1, page 394)

$$SE = \sqrt{((0.77)(0.23)/30 + (0.55)(0.45)/40)} = \sqrt{(0.0059 + 0.0062)} = 0.11 ,$$

and the 95% confidence interval (using the z value 1.96) is

$$0.22 \pm 1.96(0.11).$$

Clearly, the national value 0.19 is well inside the interval; it’s not significantly different. With zero right on the edge of the interval, the observed difference in stress between men and women at Wharton is barely significant (using 1.96, not 2). JMP will do the arithmetic if we use the “Fit Y by X” command with “Sex” as X and “Dummy” as Y. Carried out to more digits by JMP, the t-interval indicates that the difference is not quite significant: zero is just inside the interval. (But, is this interval appropriate?)

### t-Test

	Difference	t-Test	DF	Prob> t
Estimate	0.216667	1.893	68	0.0626
Std Error	0.114440			
Lower 95%	-0.01169			
Upper 95%	0.445028			
Assuming equal variances				