

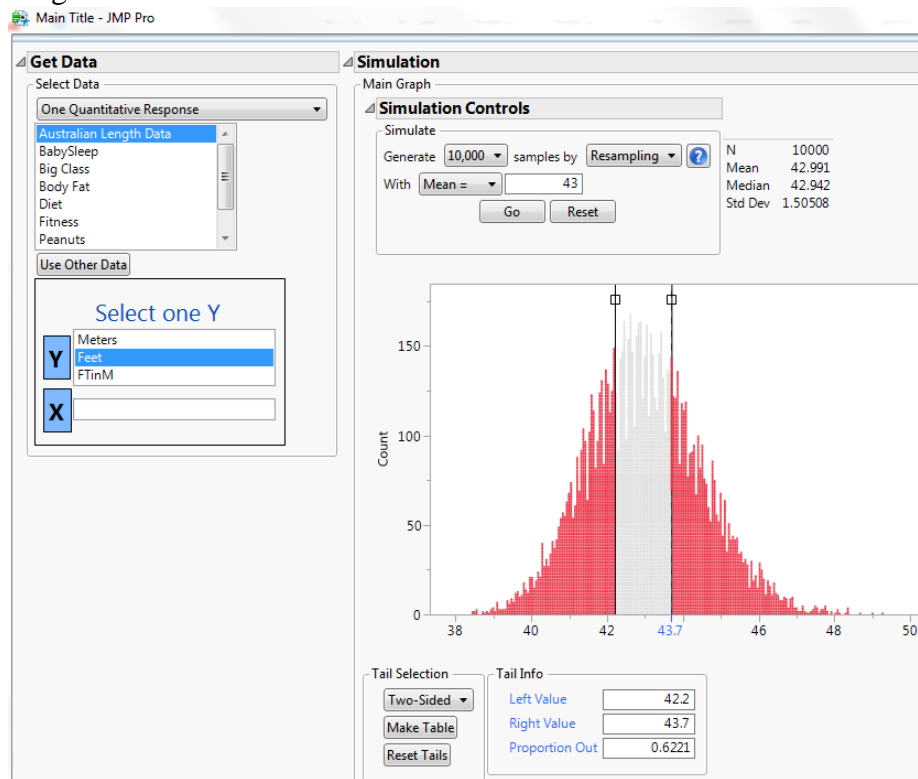
## Data Science – Fall 2016 Hypothesis Testing - Solutions

For the assignment, I will demonstrate both methods of inference, simulation-based methods and formal statistical inference.

- 1) Conduct the two relevant hypothesis tests and comment on whether students could accurately assess the length of the room.
  - a) Feet

Null model: The average guess was 43 feet

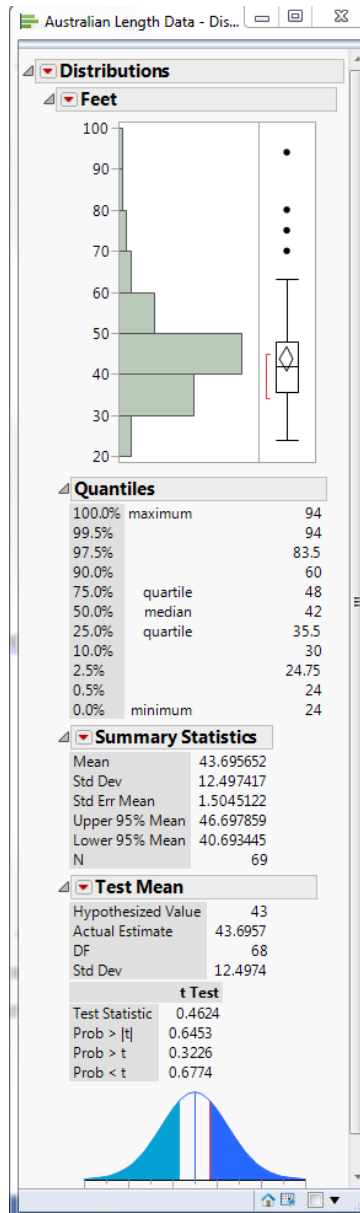
p-value using simulation method = 0.6221 from Randomization Add-in



p-value using t-test – 0.6453

Using Analyze -> Distribution -> Feet as Y

Red triangle -> Test Mean



Interpretation of p-value:

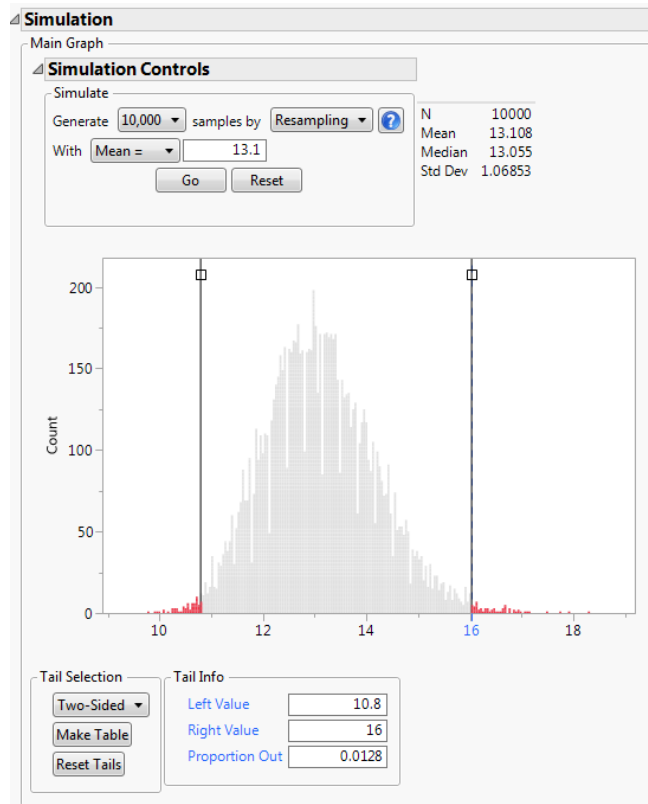
If the true guess were 43 feet, there is a 62% or 65% (based on method) chance of getting a mean that is off by 0.7 (43.7-43) feet or greater

Decision – There is a large chance (since 62% and 65% are both greater than 5%) that we could get this result if the true mean guess were 43 feet, so it's reasonable to conclude that the average guess was close to the true size

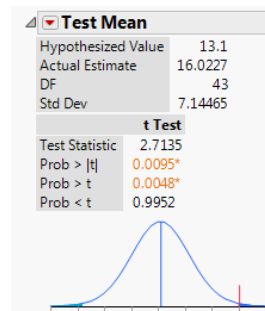
b) Meters

Null model: The average guess was 31.1 meters

p-value using simulation method = 0.0128 from Randomization Add-in



p-value using t-test – 0.0095



Interpretation of p-value:

If the true guess were 13.1 meters, there is about a 1% chance of getting a mean that is off by 2.9 meters or greater

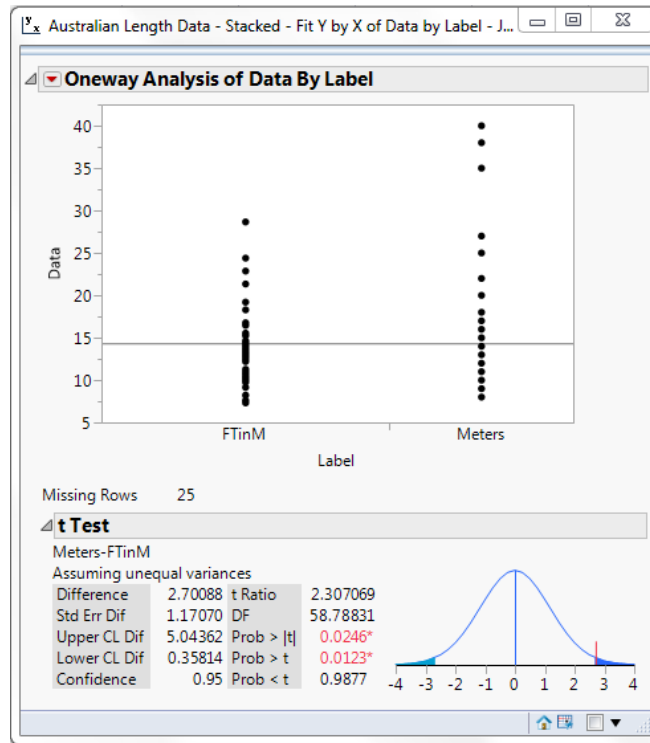
Decision – There is a small chance (1% is less than 5%) that we could get this result if they guessed an average of 13.1 meters (the true size), so it's reasonable to conclude that the average guess different from the true size

- 2) Based on the results in part 1, do you think that the students guessed the same lengths for the room, on average? Conduct the hypothesis test to compare the two groups to see whether the students' guesses were equivalent.

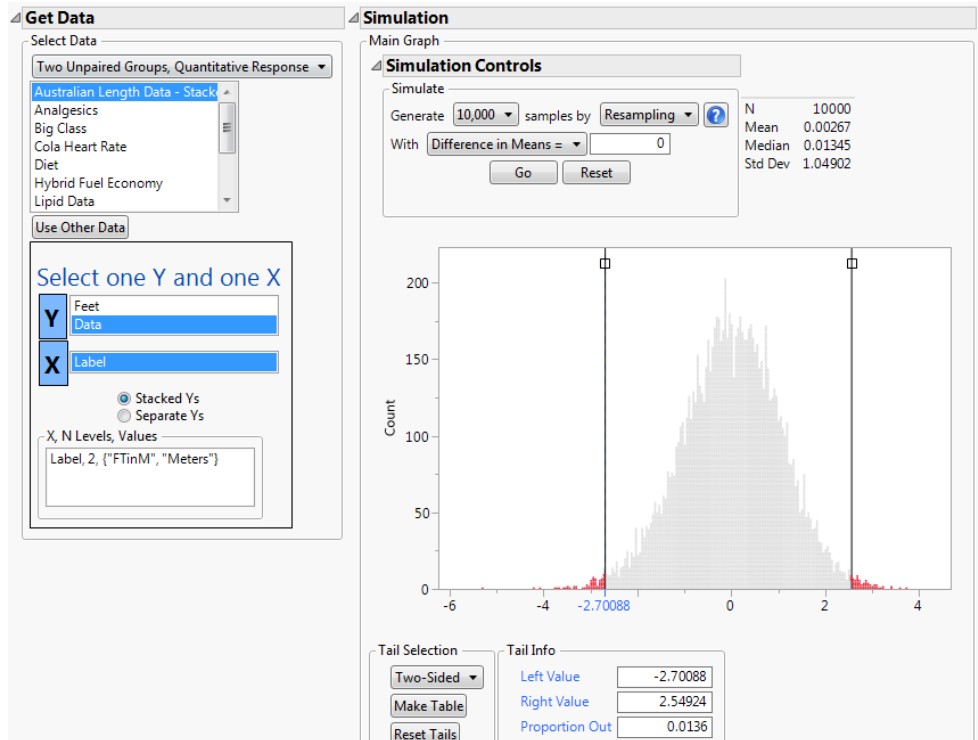
We have to modify the data to get it into a usable format for JMP. In class, we chose Table -> Stack, and selected FtInM (Feet converted to Meters) and M which was correct, but I failed to uncheck "Stack by Row". A subset of the resulting dataset should look like this (at the bottom is all the Meter measurements).

	Feet	Label	Data
1	24	FTinM	7.317073171
2	25	FTinM	7.62195122
3	27	FTinM	8.231707317
4	30	FTinM	9.146341463
5	30	FTinM	9.146341463
6	30	FTinM	9.146341463
7	30	FTinM	9.146341463
8	30	FTinM	9.146341463
9	30	FTinM	9.146341463
10	32	FTinM	9.756097561
11	32	FTinM	9.756097561
12	33	FTinM	10.06097561
13	34	FTinM	10.36585366
14	34	FTinM	10.36585366
15	34	FTinM	10.36585366
16	35	FTinM	10.67073171
17	35	FTinM	10.67073171
18	36	FTinM	10.97560976
19	36	FTinM	10.97560976
20	36	FTinM	10.97560976
21	37	FTinM	11.2804878
22	37	FTinM	11.2804878
23	40	FTinM	12.19512195

We then do a Fit Y by X and assign “Data” to Y and “Label” to X (you might want to change these names when you stack them) then choose the t-test option (in the red triangle). The p-value to test whether these are the same is 0.0246.



Using the randomization method, we get the following output, with p-value = 0.0136.



To formalize this statistical inference:

Null model – For each room size guess, it doesn't matter whether you call it a guess in meters or a guess in feet (converted to meters). Note that if it does matter how you assign them, then they are better at guessing using one method compared to the other. Alternatively, we could frame this as the average guess is the same no matter what method they used.

P-value – 0.0246 using the t-test method and 0.0136 using the simulation method

Decision – since the p-value is small (both less than 0.05), we say that there is small chance we could guess this result if the unit they are guessing in doesn't matter. Therefore, we conclude that the units in which they guess do matter.

In summary - they appear to guess the room length better using feet than using meters (third test). When they guess in meters they are far off (second test) whereas when they guess in feet they get it right, on average (first test).