

Statistically Grounded: Understanding power and p -values

Which location leads to the most profit in a small business?

Your friend, Joe, just inherited a coffee truck from his uncle. He loves coffee and has decided to start his own coffee company. He was able to get a loan from his bank, and after painting and repairing the truck, he has \$20,000 left for supplies and daily life expenses.

Joe's truck carries various items, but for the moment he is focusing his attention on his coffee. He has a fantastic coffee blend that people really seem to like. He can make it for about \$0.50 a cup. In addition, after a careful business plan, Joe estimates his daily business and life expenses at \$300 a day.

Estimating Profits and Loss:

- 1) Assume Joe sold 150 cups of coffee at \$2.00 a cup. Calculate his profit or loss for that day.
- 2) Assume Joe sold 100 cups of coffee at \$4.00 a cup. Calculate his profit or loss for that day.

Play the Statistically Grounded game: Joe knows you are taking a great statistics course and wants to use your knowledge to help him make key business decisions. There are multiple choices that Joe could make that may influence how much coffee he sells:

Name: <input type="text" value="Player Name"/>		Group Name: <input type="text" value="Group Name"/>		save data
Location	Music	Price	Time	Weather
<input checked="" type="checkbox"/> Business District <input type="checkbox"/> Near the Zoo <input checked="" type="checkbox"/> City Park <input type="checkbox"/> City Hall	<input checked="" type="checkbox"/> No Music <input type="checkbox"/> Hip Hop <input type="checkbox"/> Alternative	<input type="checkbox"/> \$2.00 per cup <input checked="" type="checkbox"/> \$3.00 per cup <input type="checkbox"/> \$4.00 per cup <input type="checkbox"/> \$5.00 per cup	<input type="checkbox"/> Morning (8am - 10am) <input checked="" type="checkbox"/> Lunch (10am - 2pm) <input type="checkbox"/> Afternoon (2pm - 6pm) <input type="checkbox"/> Evening (6pm - 10pm)	<input checked="" type="radio"/> Relatively Constant (recommended) <input type="radio"/> Variable Weather Patterns (temperature and amount of sunshine will change each day)
Where should Joe park his truck? Some locations may have more sales than others.	Does playing a certain type of music from your truck influence sales?	Charging less may increase the number of cups Joe sells, but how will it affect profits?	Each day, Joe will work every time that you select. During this study, Joe will only work during one four hour time frame each day. <i>Weather will be relatively constant for this game.</i>	

The on-line game allows you to design a study to evaluate whether these conditions influence sales. Go to the web site <http://statgames.tietronix.com/statisticallygrounded/>.

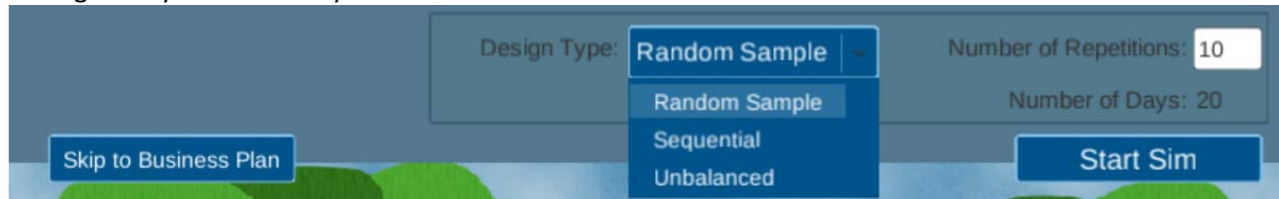
Check **Yes** in the **Save Data** box and type in the appropriate information to record your data:

Name: _____ (This will be on the web, do not use a player name that will identify you. However, make sure you print your player name on this worksheet.)

Group Name: _____ (Provided by the instructor)

In this study, we are assuming that Joe is primarily interested in determining whether he should sell coffee in either the **Business District** or the **City Park** during the lunch hour. Check both locations as shown in the image above. In addition, make sure to select **No Music**, **\$3.00 per cup** and the **Lunch Time**.

Select the design type, **Random Sample** and number of simulations to be **10**, as shown below. For this activity **Random Sample** means that both conditions (**Business District** and **City Park**) are tested the same number of times and that they will be tested in random order. In this study, both locations will be tested 10 times, totaling 20 days for the study.



- 1) Select **Start**, then click on the **Game Data** button. *Note, some computers may require you to remove the pop-up blocker to see the data.* Find your data by clicking on the "**Download CSV**" link corresponding to your **Player Name**. Copy and paste your data into an appropriate statistical software package and create one or two graphs that display your results.

- 2) Using **Sales** as the response variable, analyze the data. Can you conclude that there is a difference between locations? Be sure to include a hypothesis test and a 95% confidence interval for the difference between locations (**Business District - City Park**). Write a short report (300 words or less) that provide a suggestion to Joe. Did your study provide evidence that there was a difference in locations? Tell Joe what he should do. Clearly describe the study, how the data was analyzed, and what conclusions were drawn. Graphs should be included in this report.

- 3) Compare your study results to others in the class. Did every group in the class come to the same conclusion about which **Location** is best? Did every group in the class get a p-value less than 0.05?

- 4) Create two separate histograms of the class data found in question 3, one histogram of the business district means and a second histogram of the park means.

- 5) This activity was designed so that the distribution of cups sold is normally distributed with a standard deviation of 10. However, the mean for the business district is 83 cups sold while the mean for the park is 74 cups sold (assuming it is during the lunch time frame, with no music, and priced at \$3 a cup). Do the samples collected in your class appear to fit these assumptions?

- 6) When conducting a two-sample t -test, the **p-value** is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Write a short paragraph to Joe giving a practical explanation of p -values and alpha levels in the context of this study. More specifically, explain why the p-values tended to be inconsistent even though the business district is truly better than the park.

- 7) Compare the student data in question 3. How often was the Business District mean greater than the City Park mean? Explain why confidence intervals can provide more useful information than p-values.