This is a Quiz Preview. When this quiz is both published (by the quiz creator) and visible to students, they can find it here.

Lecture 04B: univariate data analysis

Help Center

Warning: You have already made the maximum number of submissions. Additional submissions will not count for credit. You are welcome to try it as a learning exercise.

To help you work on the quiz earlier, the quizzes will be based on the textbook material.

This quiz is based on material from pages 62 to 64, as well as earlier parts of the course will also be tested.

Videos on the new content will be posted on Thursday, during the day. All will be available on early Thursday evening, before 7:00pm.

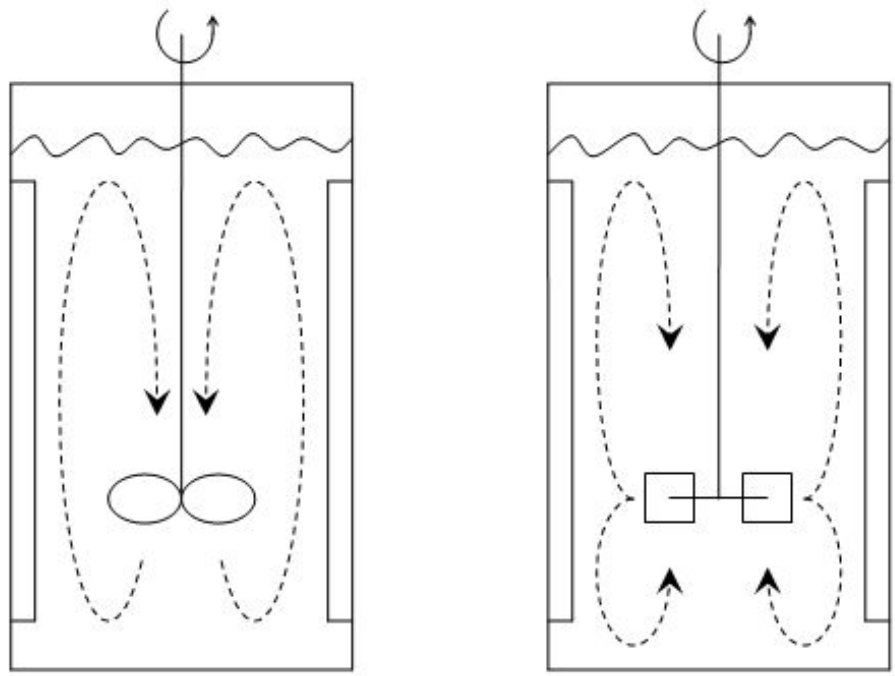
You have 1 attempt for the quiz. Please read the instructions carefully. Please double check your answers before submitting.

Solutions will be released when the quiz closes, at 09:25am, Friday, 30 January 2015.

- [ ] **In accordance with the Coursera Honor Code, I (Kevin Dunn) certify that the answers here are my own work.**

Question 1

You are convinced that a different impeller (mixing blade) shape for your tank will lead to faster, i.e. shorter, mixing times. The choices are either an axial blade (left) or a radial blade (right).



Credit: Wikipedia

You obtain the following confidence interval from various tests:

$-21 \text{ seconds} \leq \mu_{\text{Axial}} - \mu_{\text{Radial}} \leq 187 \text{ seconds}$

Check all the options that correctly apply on how you might be able to make the confidence interval narrower.

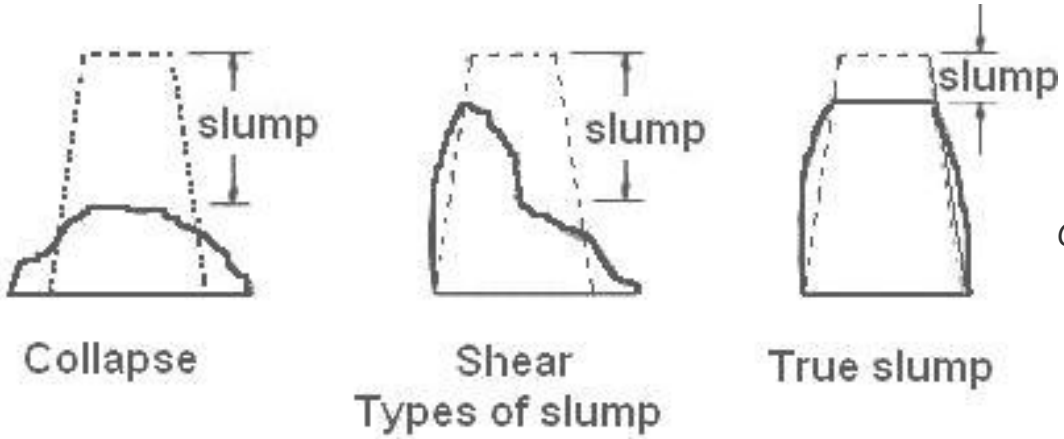
- [ ] Take more samples from either the axial impeller, or radial impeller, or both.
- [ ] If it is possible, we should make the standard deviation smaller (though this might not be feasible in practice).
- [ ] Use a smaller magnitude value for c_t (if z is t-distributed) or use a smaller value for c_n (if z is normally distributed).
- [ ] Using a higher level of confidence will get a narrow interval, because we will get closer bounds for the true population value.

Question 2

A concrete slump test is used to test for the fluidity, or workability, of concrete. It's a crude,

but quick test often used to measure the effect of polymer additives that are mixed with the concrete to improve workability.

The concrete mixture is prepared with a polymer additive. The mixture is placed in a mold and filled to the top. The mold is inverted and removed. The height of the mold minus the height of the remaining concrete pile is called the “slump”.



Credit:

Wikipedia

You run several tests to acquire data. The goal is see which one of your additives, either A or B, provides a **smaller slump**.

Which of the following are good experimental practices, to ensure that you get independent data?

(*please check all that correctly apply*)

- [ ] Ideally, all the additive A experiments should be done by 1 person, then all the additive B experiments done by another person so they have no relationship with each other.
- [ ] Clean the slump test equipment thoroughly between runs.
- [ ] In an ideal situation where you can do many trials, use a variety of suppliers for the cement, binder, and other ingredients, to ensure you test the additive performance across a variety of conditions.
- [ ] Perform all the additive A samples first, then all the additive B samples next.

Question 3

Continue on with the concrete slump example in the prior question. Which are valid assumptions in order to calculate a confidence interval of the form shown below?

$$\_\_\_\_\_ \leq \mu_B - \mu_A \leq \_\_\_\_\_$$

(*check all that correctly apply*)

- [ ] Assume the data points are independent.
- [ ] Assume the variances for sample A and B are the same.
- [ ] To calculate the variance of the difference in sample averages, $\mathcal{V}\{\bar{x}_B - \bar{x}_A\}$, we must assume the averages from each sample are independent.
- [ ] Assume the data are uniformly distributed.
- [ ] Assume the data for sample A and sample B are from the central limit theorem.

Question 4

Here are the data that were collected for the slump tests, with the two additives: A and B.

Samples for A	Samples for B
12	13
15	7
12	9
6	12
17	12
	5

Based on these data, we would like to calculate a confidence interval:

$$\_\_\_\_\_ \leq \mu_B - \mu_A \leq \_\_\_\_\_$$

To do so, we calculate some intermediate values.

Please *check all that correctly apply*:

- [ ] There are $5 + 6 = 11$ degrees of freedom
- [ ]

The standard deviations for each sample can be pooled, so we have $s_A = 4.16$, $s_B = 3.2$, and the pooled variance is 3.66.

- [ ] The critical values for a 90% confidence interval would have 5% in each tail.
- [ ] The z-value should be assumed to be normally distributed.
- [ ] The z-value should be assumed to be t-distributed.
- [ ] If we were to calculate the z-value, the denominator would be 2.22, which is the overall standard deviation.
- [ ] The sample means are $\bar{x}_A = 12.4$ and $\bar{x}_B = 10.5$

Question 5

Once again, here are the data that were collected for the slump tests, with the two additives: A and B.

Samples for A	Samples for B
12	13
15	7
12	9
6	12
17	12
	5

Based on these data, this time we would like to calculate a **95% confidence interval**:

_____ $\leq \mu_B - \mu_A \leq$ _____

Please *check all that correctly apply*:

- [ ] The lower bound in the confidence interval is -7.75
- [ ] The critical t value is 1.83 for this level of confidence.
- [ ] The confidence interval shows that there is no statistical difference between additive A and additive B.
- [ ] The lower bound in the confidence interval is 7.75
- [ ] The upper bound in the confidence interval is -2.28

- [ ] The critical t value is 2.26 for this level of confidence.
- [ ] The upper bound in the confidence interval is 2.28

Question 6

Based on the statistical analysis in the foregoing questions, which additive do you recommend to get the smaller slump value? You have to pick A or B.

Please ensure your calculations in the prior questions are accurate, since there is no partial credit here: as in real-life, you have to pick A or B -- one is appropriate, the other isn't.

- ( ) A
- ( ) B

Admin Help

Question 7

Check all options that *correctly apply* with regards to paired testing.

Please note:

- that if I take 2 samples, one from system A and another from system B, these two values are called my *within-pair* values,
- and if I collect another 2 samples, then the first pair of values can be compared to the second pair of values: this is called a *between-pair* comparison
- So I can collect many such paired values, and in the end I will have n pairs, but $2n$ raw samples.

- [ ] Any systematic difference *within* pairs is eliminated (for example, if both sample A and sample B from a pair had a laboratory mis-calibration, causing the recorded value to be too low by 10 units).
- [ ] The raw data (the individual sample values) **do not** need to be normally distributed.
- [ ] My friend and I are both randomly selected for a clinical trial to test a drug's efficiency.

  My friend receives the placebo and I get the actual drug. Later it switches around, my friend gets the drug and I get the placebo. For the purpose of analyzing data from the clinical trial, my friend and I are considered pairs.

- [ ] **In accordance with the Coursera Honor Code, I (Kevin Dunn) certify that the answers here are my own work.**

Submit Answers Save Answers

You cannot submit your work until you agree to the Honor Code. Thanks!