

1. People of different ages were asked to stand on a "force platform" and maintain a stable upright position. The "wiggle" of the board in the forward-backward direction is recorded; more wiggle corresponds to less balance. The participants are divided into two age groups: young and elderly. The average wiggle among elderly people was 26.33 mm, and the average among young people was 18.125 mm. The bootstrap distribution for the difference in means is shown below, based on 100 bootstrap samples. Of the following choices, which is the **most accurate** 90% bootstrap confidence interval for the true difference in means?



- ☐ (2.5 mm, 18 mm)
☐ (5 mm, 15 mm)
☐ (3 mm, 17 mm)
☒ (3.75 mm, 15 mm)
2. Which of the following is **false** regarding paired data?
- ☐ In a paired analysis we first subtract the paired observations from each other, and then do inference on the differences.
☐ Two data sets of different sizes cannot be analyzed as paired data.
☐ Each observation in one data set has a natural correspondence with exactly one observation from the other data set.
☒ Each observation in one data set is subtracted from the average of the other data set's observations.

✓ **Correct**

This question refers to the following learning objective(s):

- Define observations as paired if each observation in one dataset has a special correspondence or connection with exactly one observation in the other data set.
- Carry out inference for paired data by first subtracting the paired observations from each other, and then treating the set of differences as a new numerical variable on which to do inference (such as a confidence interval or hypothesis test for the average difference).

It doesn't make any sense to subtract each observation in one data set from the average of the other data set's observations, we subtract the paired observations from each other.

3. Which of the following is **false** about bootstrap and sampling distributions?

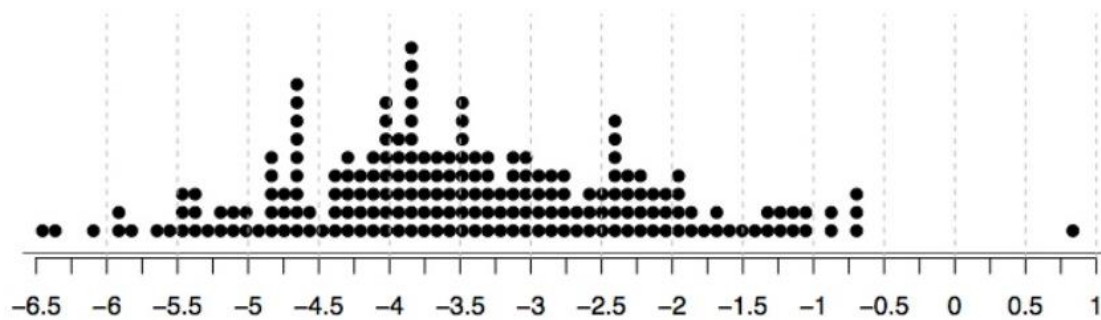
- ☐ Both distributions are comprised of sample statistics.
- ☒ Both distributions are created by sampling with replacement from the population.
- ☐ Bootstrap distributions are centered at the sample statistic, sampling distributions are centered at the population parameter.
- ☐ Both distributions get narrower as the standard deviation decreases.

✓ **Correct**

This question refers to the following learning objective(s): Describe how bootstrap distributions are constructed, and recognize how they are different from sampling distributions.

The bootstrap distribution is created by sampling with replacement from the original sample, not the population.

4. Researchers studying IQ scores of mothers and fathers of "gifted" children collected data from 36 gifted children and their parents. First, differences in IQ scores of the father and the mother were calculated for each child (calculated as father's IQ score - mother's IQ score). The dot plot below shows the bootstrap distribution of means of **200 bootstrap samples** taken from this original sample of differences in IQ scores. The mean of the bootstrap distribution is approximately -3.48 points and the bootstrap standard error is 1.3 points. Assume the usual conditions for constructing a bootstrap confidence interval are satisfied. Which of the following statements is **false**?

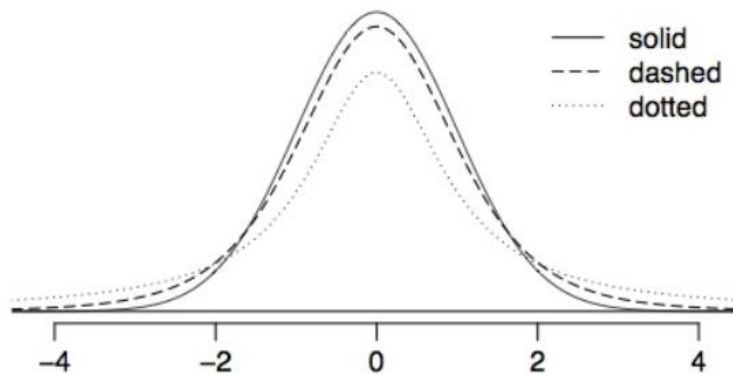


- ☐ A 90% bootstrap confidence interval for the difference in IQ scores would be approximately (-5.6, -1.3).
- ☒ Since 0 is apparently an unusual value for the statistic, then at the 5% significance level we would fail to reject a null hypothesis of that claims that the fathers' and mothers' average IQs are equal.
- ☐ It's likely that in the original sample, most mothers had higher IQ scores than the fathers.
- ☐ A 95% bootstrap confidence interval for the difference in IQ scores is approximately (-6, -0.9).
5. An insurance company wants to estimate (using a confidence interval) its average claim amount using data from 20 randomly selected claims. Which of the following is **false**?
- ☐ The critical t -score, t^* , has 19 degrees of freedom.
- ☐ The confidence interval can also be calculated using bootstrapping.
- ☐ If the distribution of the sampled claim amounts is not extremely skewed, a T interval is appropriate.
- ☒ A confidence interval based on this sample is not accurate since the sample size is small.

✓ **Correct**

Low sample size alone does not necessarily preclude an accurate confidence interval; the T interval may be used if conditions are met.

6. How does the shape of the t -distribution change as the sample size increases?



- ☐ It becomes flatter
- ☐ It becomes skewed
- ☐ It becomes wider
- ☒ It becomes more normal looking
7. Air quality measurements were collected in a random sample of 25 country capitals in 2013, and then again in the same cities in 2014. We would like to use these data to compare average air quality between the two years. Which of the following tests is the **most** appropriate?
- ☐ independent samples t-test with one-sided alternative hypothesis
- ☐ independent samples t-test with two-sided alternative hypothesis
- ☒ paired t-test with two-sided alternative hypothesis
- ☐ paired t-test with one-sided alternative hypothesis

✓ **Correct**

We're looking for a difference between the two years' averages, and the data are collected in the same cities in both years, creating dependency requiring paired analysis.

8. We are testing the following hypotheses:

$$H_0 : \mu_1 = \mu_2$$

$$H_A : \mu_1 \neq \mu_2$$

The sample sizes are $n_1 = 20$ and $n_2 = 40$. Your friend who is working on this hypothesis test calculates a Z statistic, $Z = 2.5$. Which of the following is true?

- ☐ She should be using an F statistic instead of a Z statistic.
- ☐ p-value is approximately 0.006.
- ☐ p-value is approximately 0.012.
- ☒ She should be using a T statistic instead of a Z statistic.

✓ **Correct**

This question refers to the following learning objective(s):

- Use a T -statistic, with degrees of freedom $df = n - 1$ for inference for a population mean.
- Use a T -statistic, with degrees of freedom $df = \min(n_1 - 1, n_2 - 1)$ for inference for difference between means of two population means using data from two small samples.
- Describe how to obtain a p-value and a critical t-score (t_{df}^*) for a confidence interval.

Since one of the samples is small, she should be using a T statistic instead.

9. When doing an ANOVA, you observe large differences in means between groups. Within the ANOVA framework this would most likely be interpreted as:

- ☒ Evidence strongly favoring the alternative hypothesis.
- ☐ Evidence strongly favoring the null hypothesis.
- ☐ Evidence revealing which group mean is different from the others.
- ☐ None of the above

✓ **Correct**

If there is more variability between groups compared to within groups, it's likely there is some difference between at least some of the means.

10. Which of the following is **not** a condition required for comparing means across multiple groups using ANOVA?

- ☐ The observations should be independent within and across groups.
- ☒ There should be at least 10 successes and 10 failures.
- ☐ The variability across the groups should be about equal.
- ☐ The data within each group should be nearly normal.

✓ **Correct**

This question refers to the following learning objective(s):

List the conditions necessary for performing ANOVA

- the observations should be independent within and across groups
- the data within each group are nearly normal
- the variability across the groups is about equal

and use graphical diagnostics to check if these conditions are met.

Success-failure condition is relevant for categorical variables.

11. Based on the ANOVA output below, what is the value of the F-statistic? Choose the closest answer.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
section	7	525.01	75.00		0.0767
Residuals	189	7584.11	40.13		

- ☐ 0.54
- ☐ 10.71
- ☐ 14.45
- ☒ 1.87
- ☐ 27.00

12. A study compared five different methods for teaching descriptive statistics. The five methods were traditional lecture and discussion, programmed textbook instruction, programmed text with lectures, computer instruction, and computer instruction with lectures. 45 students were randomly assigned, 9 to each method. After completing the course, students took a 1-hour exam. We are interested in finding out if the average test scores are different for the different teaching methods.

The p-value of the test is 0.0168. What is the conclusion of the test?

- ☐ Only two group means are significantly different from each other.
- ☐ At most two group means are significantly different from each other.
- ☐ All five group means are significantly different from each other.
- ☐ All five group means are equal to each other.
- ☒ At least two group means are significantly different from each other.

✓ **Correct**

This question refers to the following learning objective(s): Recognize that the null hypothesis in ANOVA sets all means equal to each other, and the alternative hypothesis suggest that at least one mean is different.

- $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$
- H_A : At least one mean is different

The p-value is low so we reject the null hypothesis.

13. Researchers studying people's sense of smell devised a measure of smelling ability. A higher score on this scale means the subject can smell better. A random sample of 36 people (18 male and 18 female) were involved in the study. The average score for the males was 10 with a standard deviation of 3.4 and the average score for the females was 11 with a standard deviation of 2.7. Which of the following is the correct standard error for the test evaluating whether the males and females have differing smelling abilities, on average? Choose the closest answer.

- ☐ 3.504
- ☒ 1.023
- ☐ 1.047
- ☐ 0.801
- ☐ 0.724

14. A study compared five different methods for teaching descriptive statistics. The five methods were traditional lecture and discussion, programmed textbook instruction, programmed text with lectures, computer instruction, and computer instruction with lectures. 45 students were randomly assigned, 9 to each method. After completing the course, students took a 1-hour exam. We are interested in finding out if the average test scores are different for the different teaching methods.

If the original significance level for the ANOVA was 0.05, what should be the adjusted significance level for the pairwise tests to compare all pairs of means to each other?

- ☐ 0.5
- ☐ 0.25
- ☐ 0.01
- ☒ 0.005
- ☐ 0.05

✓ **Correct**

This question refers to the following learning objective(s): Describe why conducting many t -tests for differences between each pair of means leads to an increased Type 1 Error rate, and we use a corrected significance level (Bonferroni correction, $\alpha^* = \alpha/K$, where K is the number of comparisons being considered) to combat inflating this error rate.

- Note that $K = \frac{k(k-1)}{2}$, where k is the number of groups.

$$K = \frac{5 \times 4}{2}, \alpha^* = 0.05/10 = 0.005$$