Lesson 037 Testing Two Proportions Friday, December 1

Two Sample Tests of Proportions What if we have two samples of data from independent binomial distributions?

- - We take $X \sim Bin(n, p_1)$
 - We take $Y \sim Bin(n, p_2)$
 - We assume that $X \perp Y$
- We are interested in $p_1 p_2$.

Estimation of Differences in Proportions • The estimator $\hat{p}_1 - \hat{p}_2$ is unbiased for $p_1 - p_2$.

• The variance of $\hat{p}_1 - \hat{p}_2$ will be

$\operatorname{var}(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1 - p_1)}{p_2} + \frac{p_2(1 - p_2)}{p_2}$ N M

Sampling Distribution for Proportions

- As long as the normal approximation applies for both X and Y, we can use a normal approximation for the differences in proportions.
- Suppose that we wish to test $H_0: p_1 = p_2$ versus $H_1: p_1 \neq p_2$.

Estimation of Differences in Proportions Under the null hypothesis, we can estimate a shared

variance term.

$$\operatorname{var}(\widehat{p}_1 - \widehat{p}_2) =$$

Then, we can construct the usual test statistic.

$$\frac{\hat{p}_1 - \hat{p}_2}{-p)[n^{-1} + m^{-1}]} \sim N(0,1)$$

$$\frac{p_1 - p_2}{\sqrt{p(1-p)[n^{-1} + m^{-1}]}} \sim N(0,1)$$

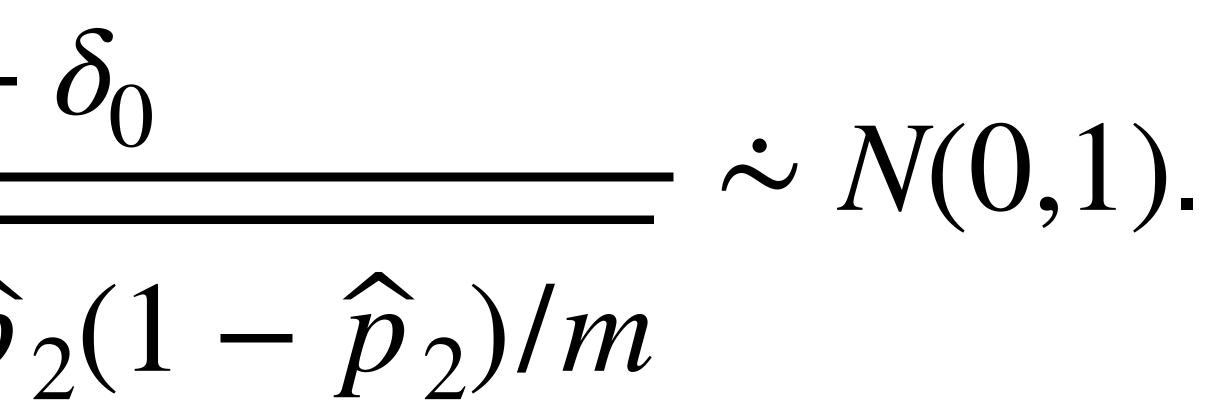
$$p(1-p)\left[\frac{1}{n}+\frac{1}{m}\right].$$

Tests of Specific Differences

- If we wish to test $H_0: p_1 p_2 = \delta_0$, for some $\delta_0 \neq 0$, this procedure does not work.
- There is not a shared variance.
- Some statisticians will use the estimated variance and appeal to the CLT, as we have for other situations.



Tests of Specific Differences In order to test the null hypothesis $H_0: p_1 - p_2 = \delta_0$ versus $H_1: p_1 - p_2 \neq \delta_0$ we can use: $\hat{p}_1 - \hat{p}_2 - \delta_0$ $\sqrt{\hat{p}_1(1-\hat{p}_1)/n+\hat{p}_2(1-\hat{p}_2)/m}$



 $H_0: \mu = \mu_0$?

N(0,1), assuming the population is normal.

N(0,1).

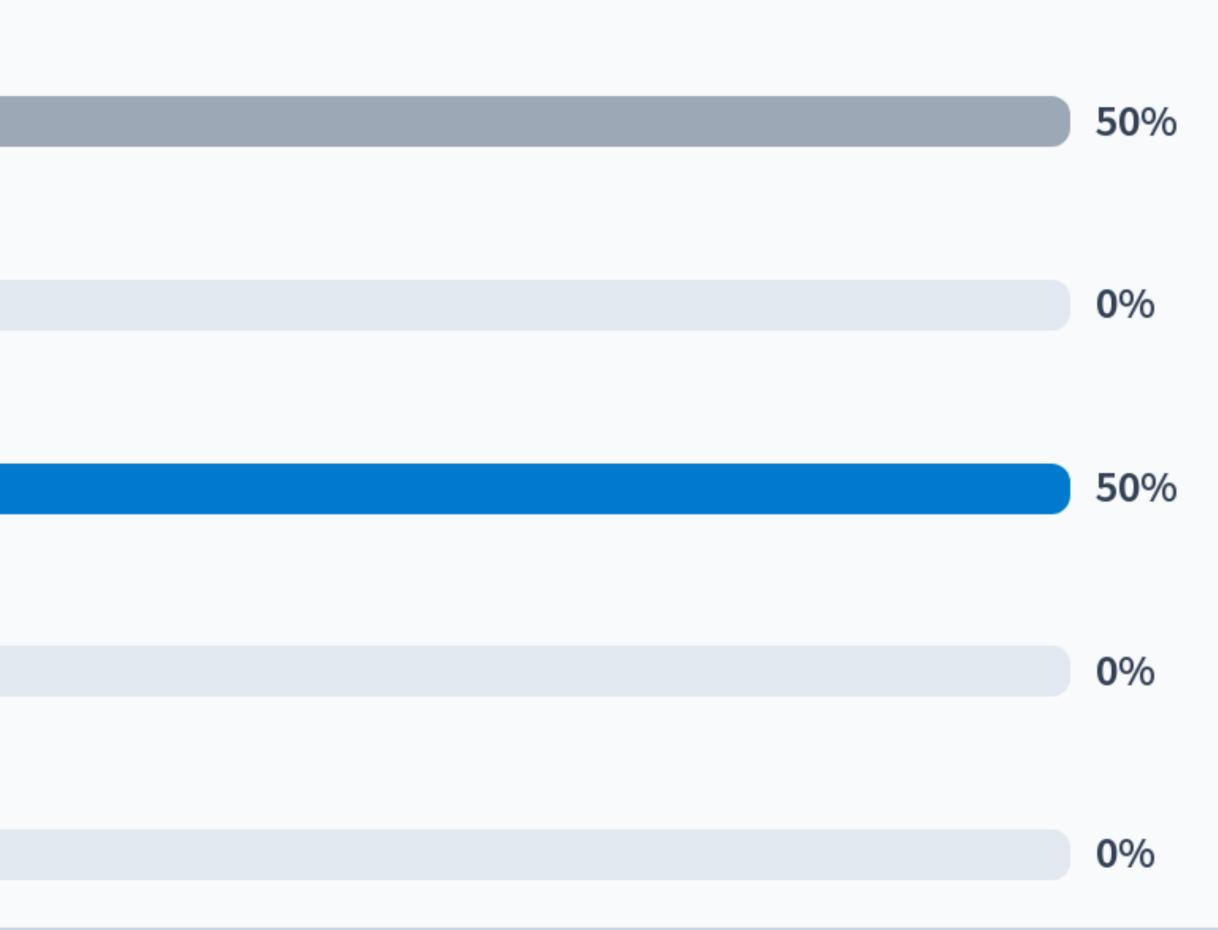
 $<\!\!< t_9$ assuming the population is normal.



 t_{10} assuming the population is normal.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app

Suppose that you have a sample of size 10 from a single population. You estimate the mean to be 15 and the standard deviation to be 3. What is the sampling distribution under the null,





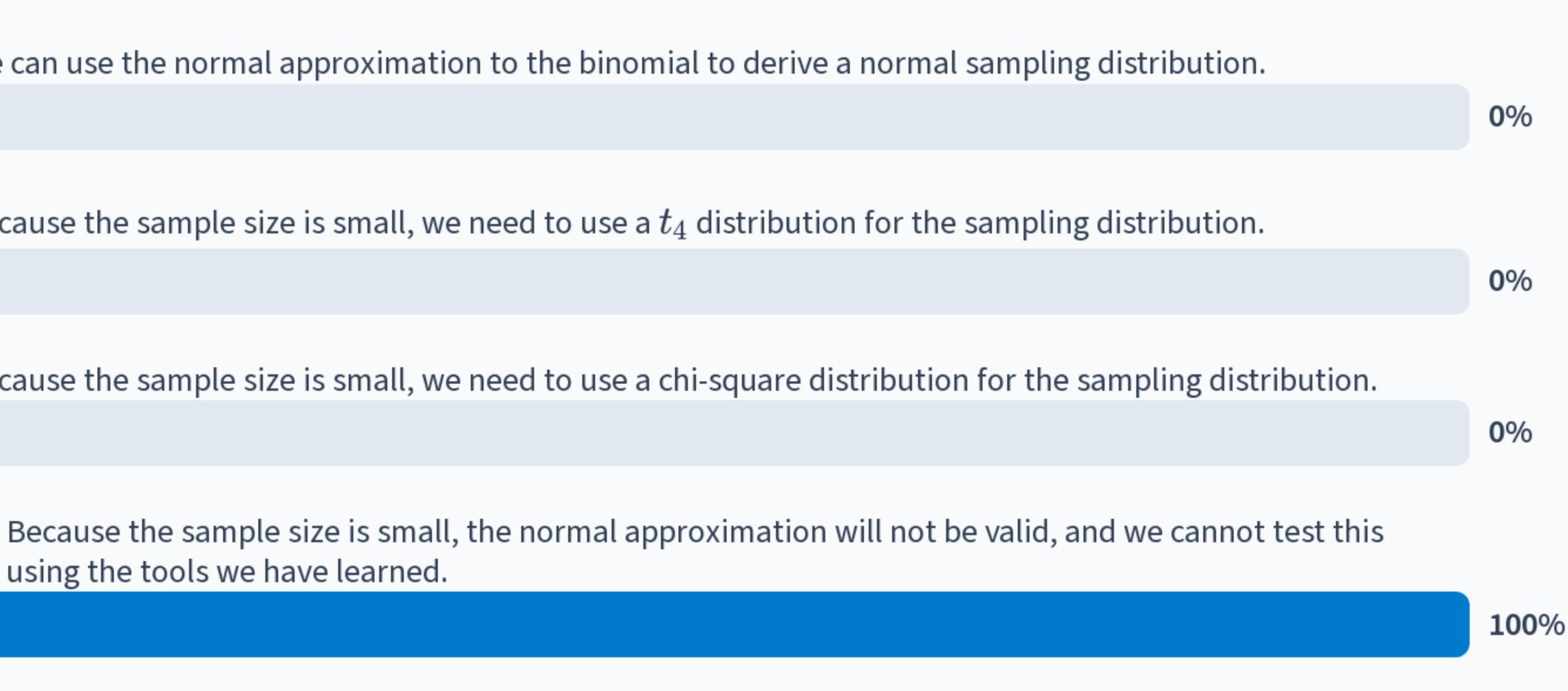
Suppose that you have $X\sim { m Bin}(5,p)$. You wish to test $H_0: p=0.5$. Which of the following is true?

We can use the normal approximation to the binomial to derive a normal sampling distribution.

Because the sample size is small, we need to use a t_4 distribution for the sampling distribution.

Because the sample size is small, we need to use a chi-square distribution for the sampling distribution.

 \checkmark using the tools we have learned.





Suppose we have data from two samples, and wish to test about their mean difference. Which of the following will typically correspond to the highest p-value, assuming that data remains the same?

 \checkmark A standard t-test using $\min\{m-1, n-1\}$ degrees of freedom, without pooling variance.

A standard t-test using the pooled variance estimator.

A
$$t$$
-test using $u=rac{\left(rac{s_1^2}{n}+rac{s_2^2}{m}
ight)^2}{rac{s_1^4/n^2}{n-1}+rac{s_2^4/m^2}{m-1}}$ degrees of freedom, r

A Z test, leveraging the normal approximation.

Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app



without pooling variance.

0%

100%

0%



Which of the following will produce the shortest confidence interval?

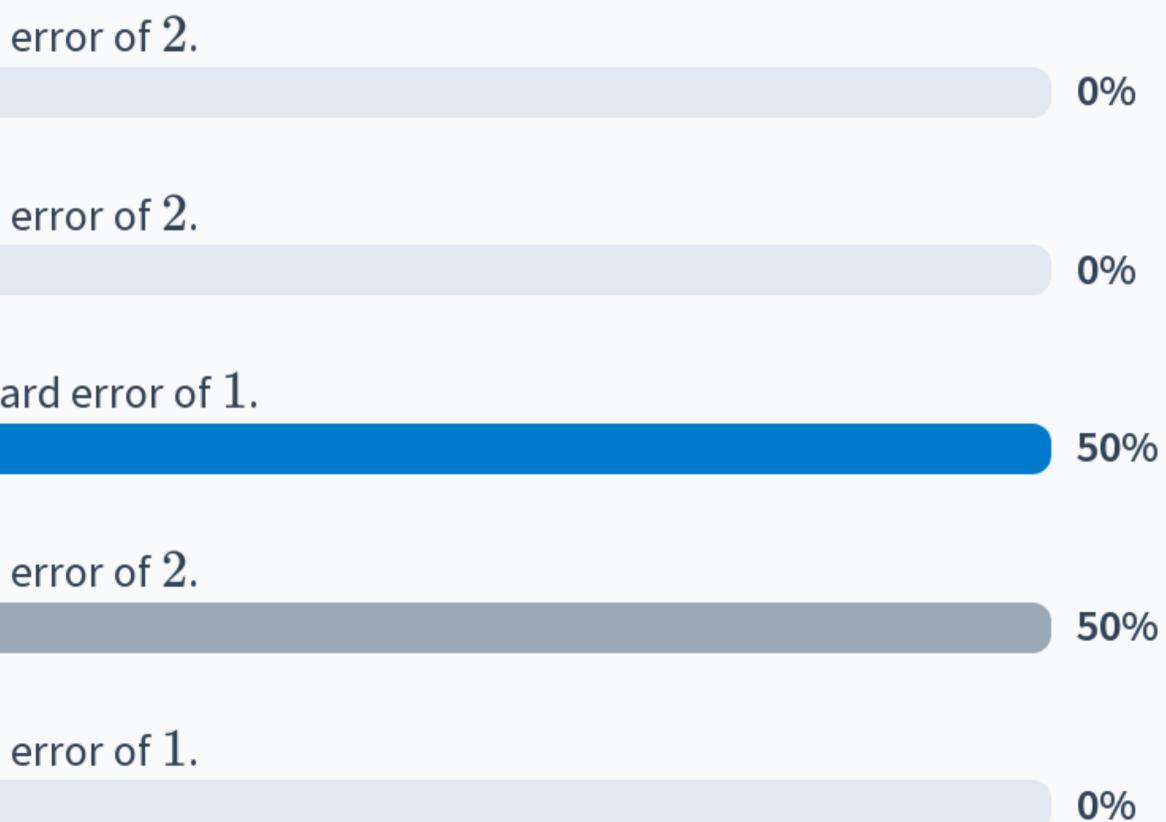
A 90% confidence interval computed with a standard error of 2.

A 95% confidence interval computed with a standard error of 2.

 ${\color{black} {f O}}$ A 90% confidence interval computed with a standard error of 1.

A 95% confidence interval computed with a standard error of 2.

A 99% confidence interval computed with a standard error of 1.





A new process is being used by a supplier. Despite assurances to the contrary, you think that the new yield (μ) is worse than the old process (with a yield of μ_0). Larger yields are preferred. What should be the null hypothesis to test this?

(A) $H_0: \mu = \mu_0.$

(B) $H_0: \mu \leq \mu_0$.

(C) $H_0: \mu \geq \mu_0.$

Start the presentation to see live content. For screen share software, share the entire screen. Get help at pollev.com/app







Suppose that you are considering two processes for producing a particular component. Process 1 is expensive, but is known to work well. Process 2 is cheaper, but may not produce components within spec. What is the relevant null hypothesis?

 $H_0: \mu_1 - \mu_2 = 0.$

 $H_0: \mu_1 - \mu_2 \ge 0.$

 $H_0: \mu_1 - \mu_2 \leq 0.$

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

0%

0%



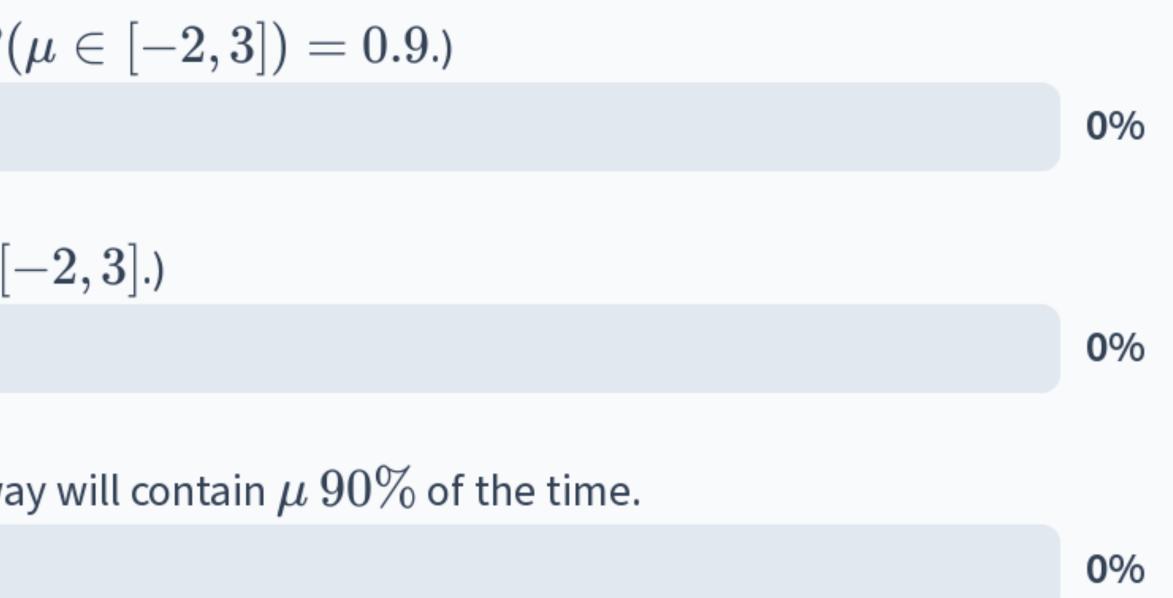
A 90% confidence interval for μ is found to be [-2,3]. What is the correct interpretation of this statement?

The probability that μ is in the interval is 0.9 (i.e., $P(\mu \in [-2,3]) = 0.9$.)

 μ must take on a value within the interval (i.e., $\mu \in [-2,3]$.)

Repeated often enough, intervals computed in this way will contain $\mu\,90\%$ of the time.



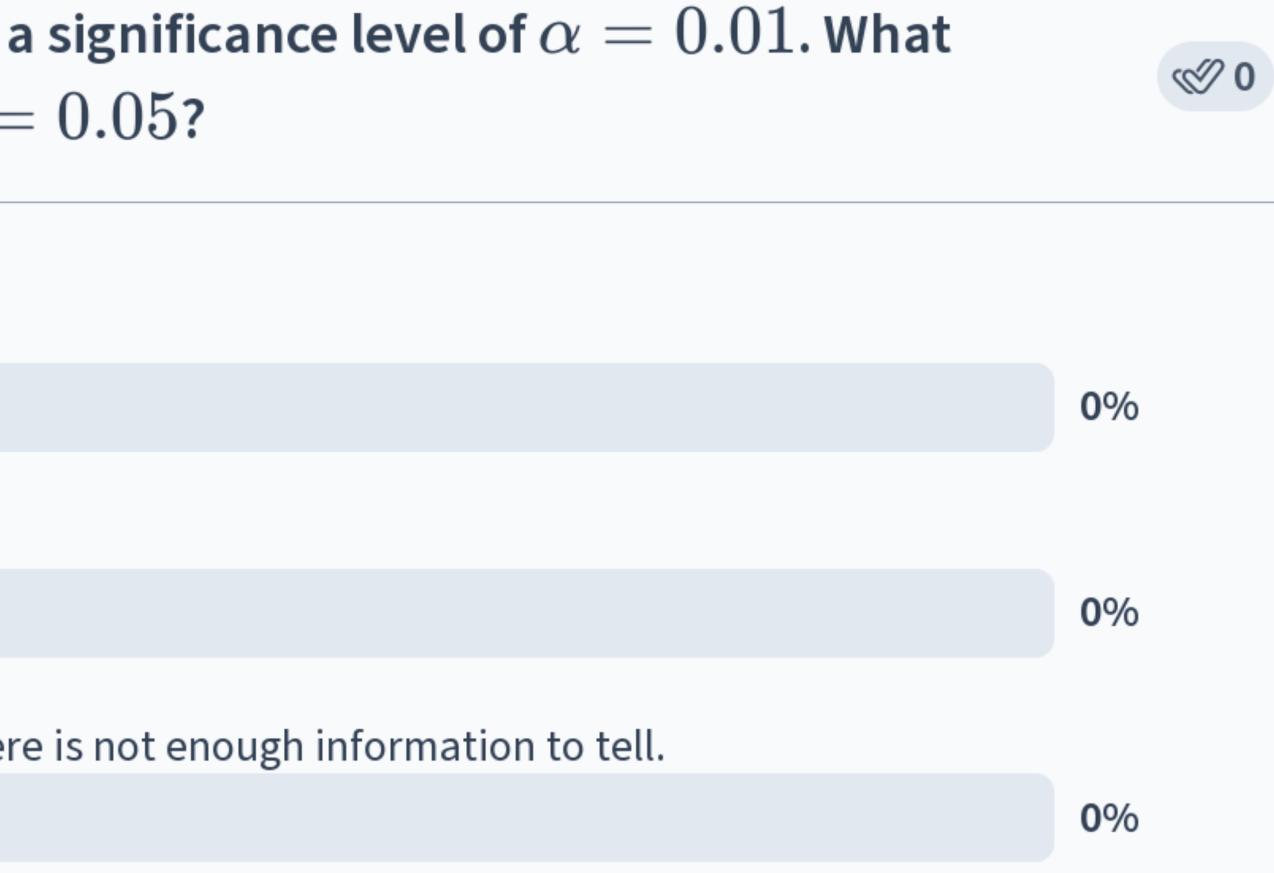


Suppose that we reject the null hypothesis at a significance level of lpha=0.01. What would have been the conclusion at a level lpha=0.05?

We would fail to reject the null hypothesis.

We would reject the null hypothesis.

We may fail to reject or reject the null hypothesis; there is not enough information to tell.

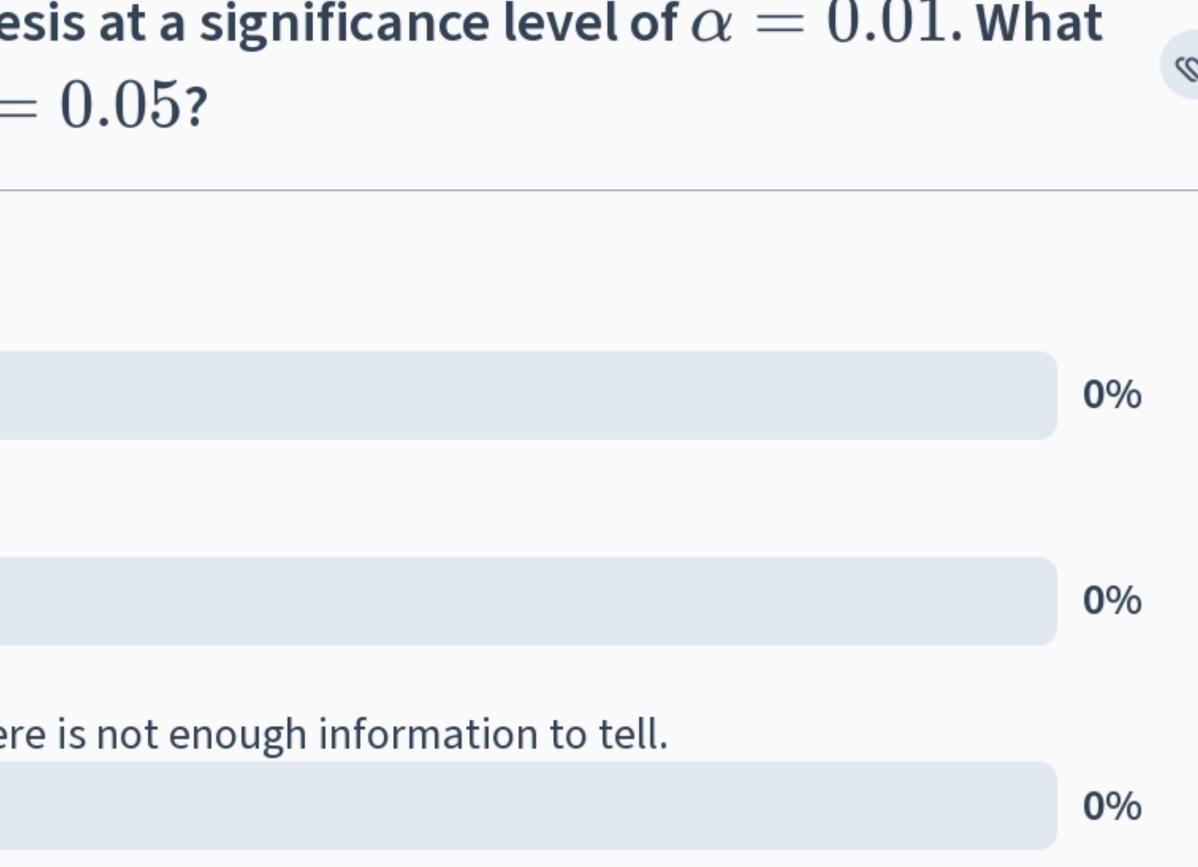


Suppose that we fail to reject the null hypothesis at a significance level of lpha=0.01. What would have been the conclusion at a level lpha=0.05?

We would fail to reject the null hypothesis.

We would reject the null hypothesis.

We may fail to reject or reject the null hypothesis; there is not enough information to tell.



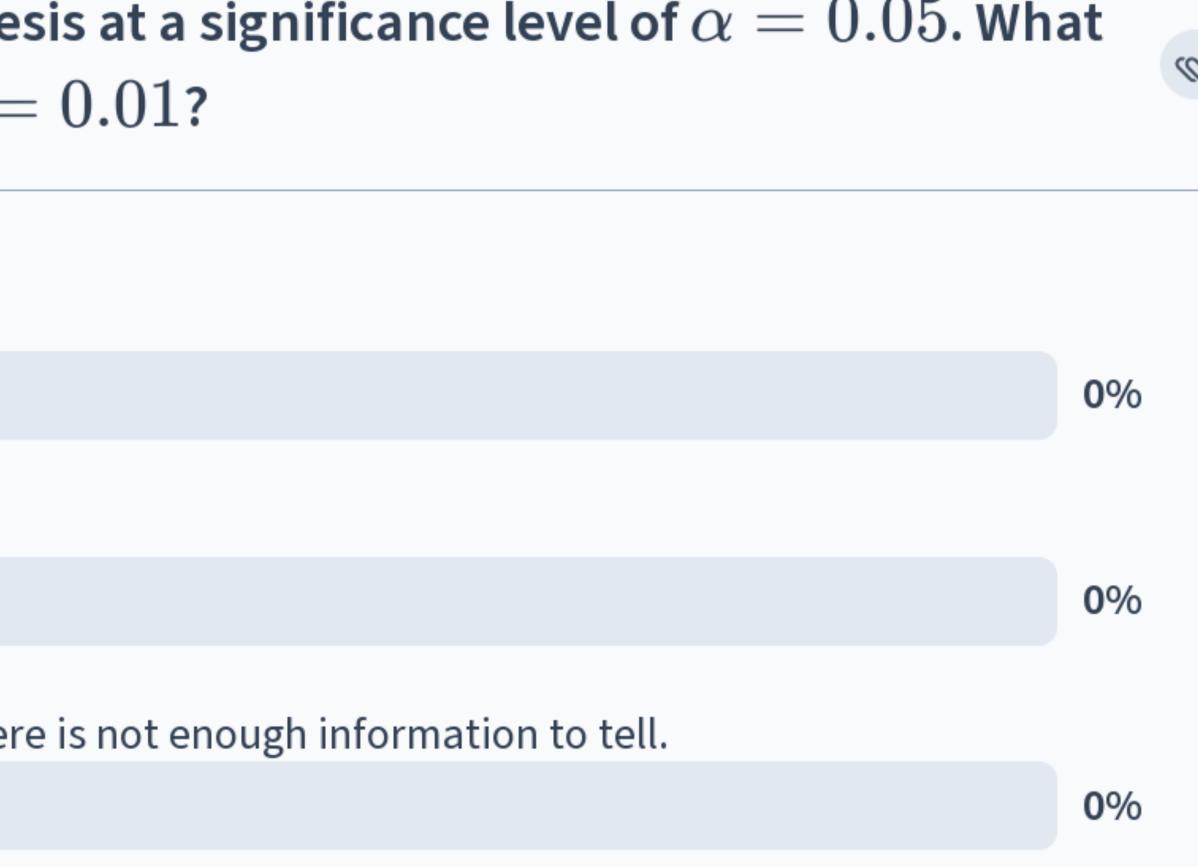


Suppose that we fail to reject the null hypothesis at a significance level of lpha=0.05. What would have been the conclusion at a level lpha=0.01?

We would fail to reject the null hypothesis.

We would reject the null hypothesis.

We may fail to reject or reject the null hypothesis; there is not enough information to tell.



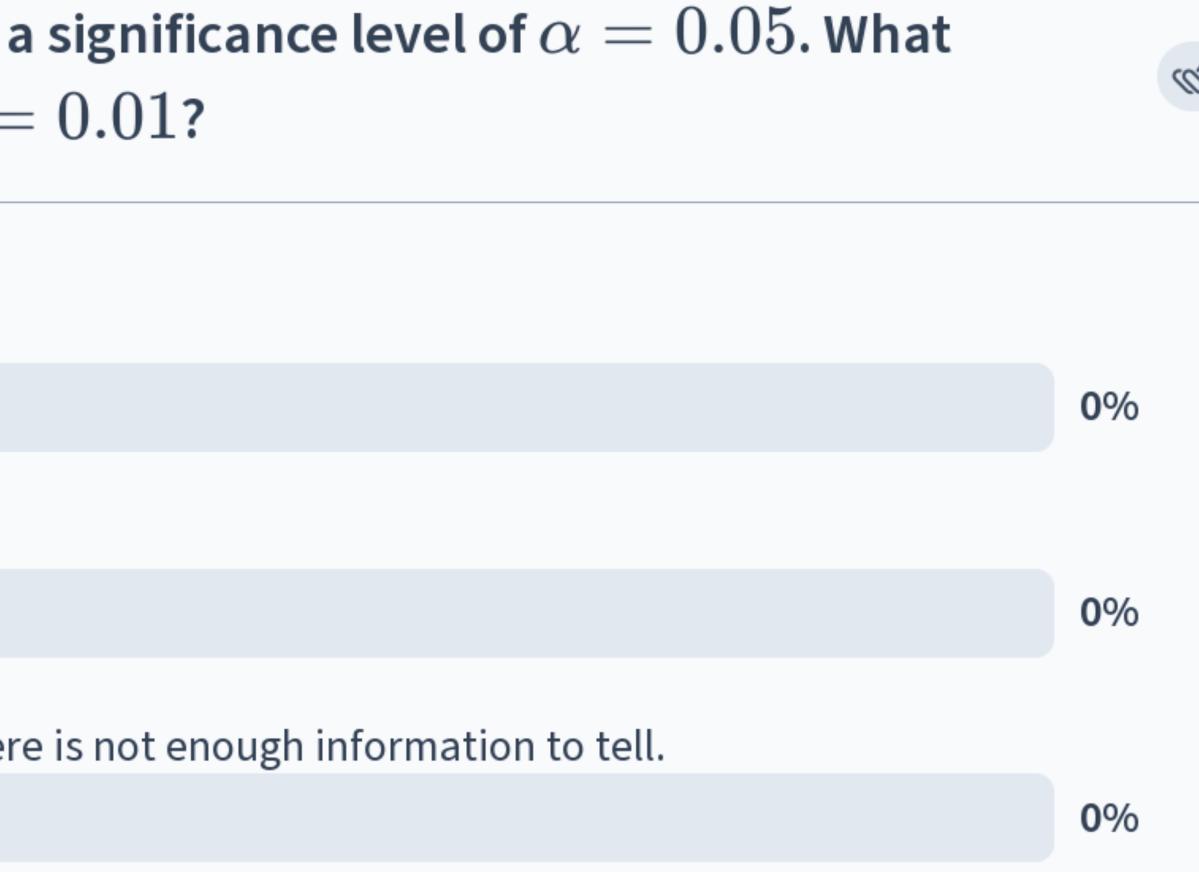


Suppose that we reject the null hypothesis at a significance level of lpha=0.05. What would have been the conclusion at a level lpha=0.01?

We would fail to reject the null hypothesis.

We would reject the null hypothesis.

We may fail to reject or reject the null hypothesis; there is not enough information to tell.





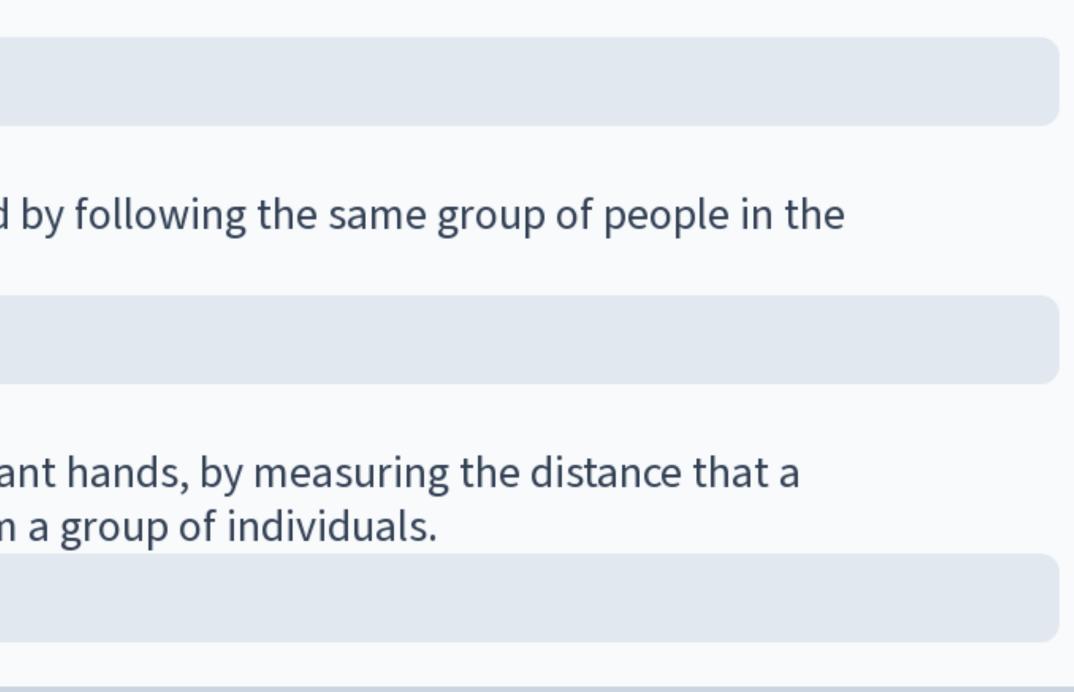
Which of the following is not a suitable candidate for using a hypothesis test with paired data?

Comparing the wear patterns of two brands of tires by installing one brand on a set of cars, driving until they wear down, then doing the same with the other brand.

Comparing the efficacy of a medication by recruiting two groups which are matched on demographic and health factors.

Determining whether seasonal changes impact mood by following the same group of people in the summer and winter.

Comparing the strength of dominant and non-dominant hands, by measuring the distance that a weighted ball travels when pushed by each hand from a group of individuals.





0%

0%

0%