Interpretation of p-values and Confidence Intervals

Jeffrey Michael Franc
MD, MSc, FCFP.EM, Dip Sport Med, EMDM

Medical Director, E/D Management
Alberta Health Services

Associate Clinical Professor of Emergency Medicine
University of Alberta

Visiting Professor in Disaster Medicine
Universita' Degli Studi del Piemonte Orientale

MedStatStudio

Objectives

1. Understand the correct definition of the p-value and recognize its limitations.

2. Understand the correct definition of confidence interval including the relationship between level of confidence and width of the interval

Quiz

MedStatStudio

Methodology of Research in Emergency and Disaster Medicine:
Quiz 5a
Interpretation

Classify the following as True or False

Part I

To study the difference in time to evacuate from a simulated earthquake site using helicopter or ground ambulance 10 simulations were performed. The mean difference in time to evacuation was 12.4 minutes (p-value=0.03)

1. True / False   There is a 3% chance that the observed difference in evacuation time was due to chance alone

2. True / False   The study should reject the null hypothesis of no difference in evacuation times at α=0.05 but not at **α=0.01**

3. True / False   If the study was repeated 100 times, we would expect to find the same results approximately 3% of the time

4. True / False   There is a 97% chance that the null hypothesis is false

5. True / False   There is a 3% chance of finding results at least this extreme if in fact the two evacuation times are equal

Part II

A study found the average concentration of carbon monoxide among survivors of a building fire to be 8.7% (95% confidence interval: 3.5% to 12.9%).

6. True / False   There is a 95% chance that the true mean concentration of carbon monoxide poisoning is between 3.5% and 12.9%

7. True / False   If we repeated the study 100 times, we would expect the interval to be different each time. About 95 of 100 intervals constructed would contain the true mean

8. True / False   The researcher should reject the null hypothesis that true mean concentration is 10% using **α=0.05**

9. True / False   If we take a survivor at random, we estimate there is a 95% chance that there concentration of carbon monoxide is between 3.5 and 12.9%

10. True / False   Increasing the sample size will decrease the width of the

Interpretation of Results

P-Values

Descriptive Statistics for Each Value of Crosstab Variable

	Obs	Total	Mean	Variance	Std Dev
ctas	90.0000	12405.0000	137.8333	1505.6461	38.8027
start	90.0000	2957.0000	32.8556	204.3946	14.2967

	Minimum	25%	Median	75%	Maximum	Mode
ctas	73.0000	106.0000	138.0000	159.5000	247.0000	92.0000
start	10.0000	24.0000	29.0000	38.5000	95.0000	29.0000

T-Test

	Method	Mean	95% CL Mean	Std Dev
Diff (Group 1 - Group 2)	Pooled	104.9778	96.3759 113.5796	29.2407
Diff (Group 1 - Group 2)	Satterthwaite	104.9778	96.3413 113.6143	

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	178	24.08	0.0000
Satterthwaite	Unequal	112.73	24.08	0.0000

ANOVA, a Parametric Test for Inequality of Population Means

(For normally distributed data only)

Variation	SS	df	MS	F statistic
Between	495915.02222	1	495915.02222	580.00377

P-Values

Who has used them before?

Who understands what they mean?

P-Values

What is a p-value?

How do we interpret it?

P-Values

What is a p-value?

p-value represents the smallest level of significance at which H_0 would be rejected

How do we interpret it?

P-Values

What is a p-value?

  p-value represents the smallest level of significance at which H_0 would be rejected

How do we interpret it?

  Reject the null hypothesis if the p-value is less than our α

P-values

What are the Advantages of reporting the p-value instead of just saying we did or did not reject H_0?

P-values

What are the Advantages of reporting the p-value instead of just saying we did or did not reject H_0?

- Clearly gives a rejection or non-rejection of null hypothesis

- Reader can rapidly see if results are significant

- Can tell if it was "close call"

- Allows reader to make choice about the level of signficance

- Conveys information about the strength of evidence

P-values

Disadvantages of using the p-value:

- Do not permit any direct statement about the direction of the effect

- Do not indicate strength of effect

- Very small differences in p-values may lead to binary accept/reject

  - (eg. 0.04 vs 0.06)

Myths and Truths

This is all you need to know!!

However, there are some myths and truths floating around!!

Scenario

Carlo and Robert are playing a dice game. Carlo rolls the number 6 three times in a row.

"This die is not fair! There is no way you can roll the number 6 three times in a row!," says Roberto

"Prove it!," says Carlo.

Carlo and Robert

1. Parameter of interest

    ????

Carlo and Robert

1. Parameter of interest

   p=probability of getting a 6

Carlo and Robert

1. Parameter of interest

   p=probability of getting a 6

2. Hypotheses

   ????

Carlo and Robert

1. Parameter of interest

   p_1=probability of getting a 6

2. Hypothesis

   p_1=1/6 (die is fair)

   $p_1 \neq$1/6 (die is not fair)

Carlo and Robert

3. Test Statistic:
 ???

Carlo and Robert

3. Test Statistic:

  X=number of 6s rolled in three rolls

Carlo and Robert

4. Rejection Region

Carlo and Robert

4. Rejection Region

- We want α<0.05
  - X=0   Probability=58%
  - X=1   Probability=35%
  - X=2   Probability=7%
  - X=3   Probability=0.46% (1/216)
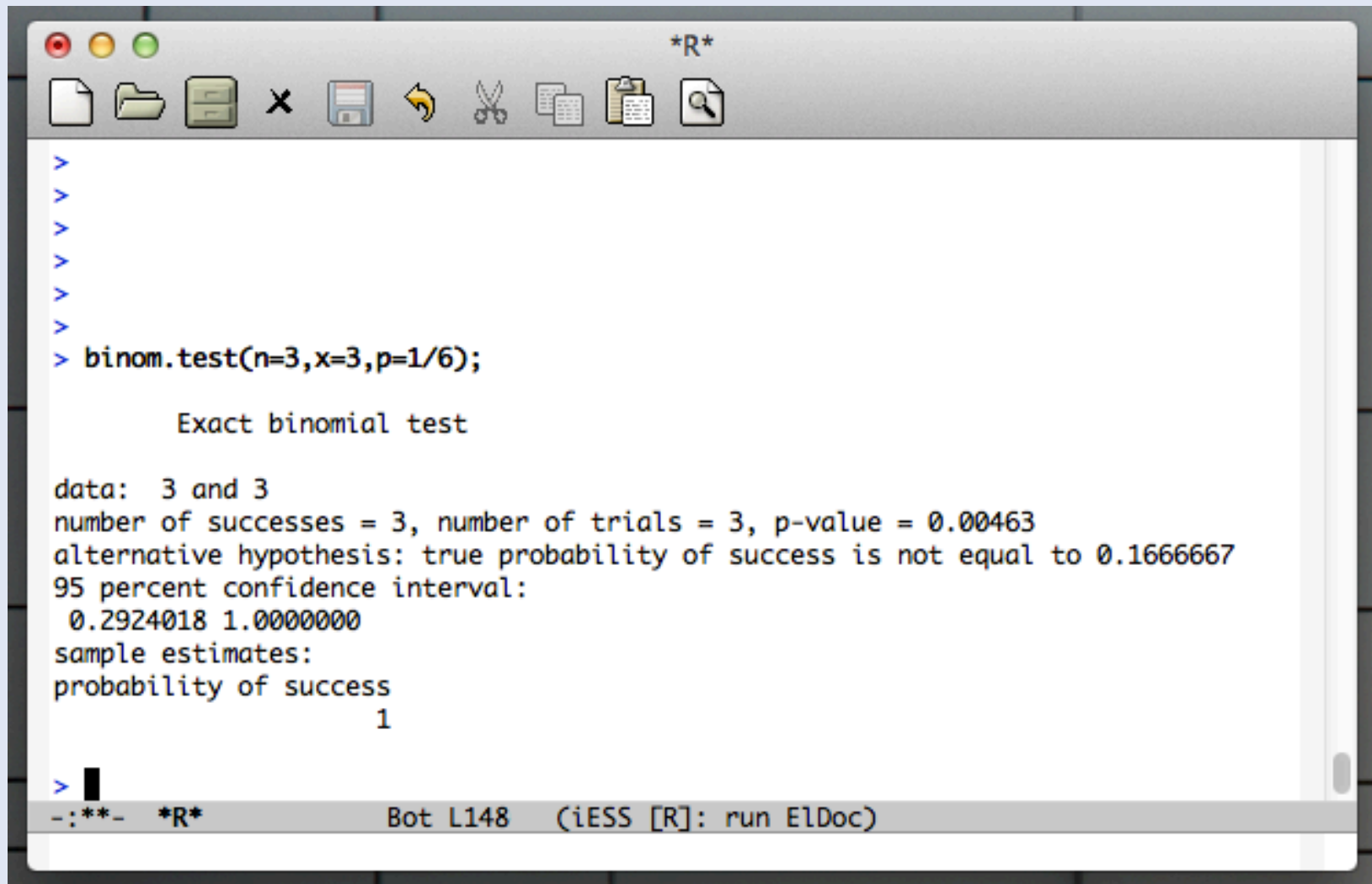
  Thus we reject if X=3

Carlo and Robert

5. Calculate Test statistic

- X=3

Carlo and Robert

Carlo and Robert

6. Decide if H0 is rejected:

Reject

Carlo and Robert

7. State Conclusion in Context

We reject the null hypothesis the chance of rolling a six was 1/6. (p=0.0046) Thus we conclude that the dice was not fair

Myth or Truth

The p-value is the probability that the null hypothesis is true

Myth or Truth

The p-value is the probability that the null hypothesis is true

FALSE

Myth or Truth

A small p value indicates that the results are replicable

Myth or Truth

A small p value indicates that the results are replicable

FALSE

Myth or Truth

The smaller the p-value the more
likely that the treatment effect is
large

Myth or Truth

The smaller the p-value the more
likely that the treatment effect is
large

FALSE

Myth or Truth

If one rejects the null hypothesis, the p-value is the probability that one has committee a type I error.

Myth or Truth

If one rejects the null hypothesis,
the p-value is the probability that
one has committee a type I error.

FALSE

Truth or Myth

The p-value is the probability that the expected results occurred by chance.

Truth or Myth

The p-value is the probability that the expected results occurred by chance.

FALSE

Myth or Truth

The p-value is the probability of
  falsely rejecting the null
  hypothesis

Myth or Truth

The p-value is the probability of falsely rejecting the null hypothesis

FALSE

Myth or Truth

Increasing the sample size with decrease the p-value

Myth or Truth

Increasing the sample size with
  decrease the p-value

FALSE

Questions?

p-values

Confidence Intervals

Confidence Intervals

Who has used them before?

Who understands what they mean?

Confidence Intervals

Mean difference in time to triage was 105 seconds. (95% confidence interval 96 – 113 seconds)

Confidence Intervals

Mean difference in time to triage was
105 seconds. (95% confidence
interval 96 – 113 seconds)

What does this mean??

Confidence Intervals

Mean difference in time to triage was 105 seconds. (95% confidence interval 96 – 113 seconds)

What does this mean??

A confidence interval gives the probability that if the experiment is repeated and a sample is drawn repetitively the interval will contain the true value of the mean

Confidence Intervals

- Remember values above or below interval are not excluded (but are unlikely)

- Our best guess for the value is the point estimate

Confidence Intervals

Advantages of confidence intervals:

Confidence Intervals

Advantages of confidence intervals:

- Gives a range of values that represent a good estimate of what the parameter might be.

- Gives information about strength of the effect

- Gives information about direction of the effect

- Result is given in units of data measurement

- Easier to assess for clinical effect

Myths and Truths

This is all you need to know!!

However, there are some myths and truths floating around!!

Demo

Probability lesson

Myth or Truth

The confidence interval gives the probability that the true mean is in the the interval.

Myth or Truth

The confidence interval gives the probability that the true mean is in the the interval.

FALSE

Myth or Truth

A 95% confidence interval means that if the study were repeated 100 times we would expect the same result about 95 times.

Myth or Truth

A 95% confidence interval means that
  if the study were repeated 100
  times we would expect the same
  result about 95 times.

FALSE

Myth or Truth

A 95% confidence interval means that if we were to repeat the study 100 times, we would expect the interval we calculate to surround the mean about 95 times.

Myth or Truth

A 95% confidence interval means that if we were to repeat the study 100 times, we would expect the interval we calculate to surround the mean about 95 times.

TRUE

Myth or Truth

A confidence interval of 95% means that if one were take another observation at random, there is a 95% chance it would fall in the interval.

Myth or Truth

A confidence interval of 95% means that if one were take another observation at random, there is a 95% chance it would fall in the interval.

FALSE

Myth and Truth

Increasing the sample size will decrease the width of the confidence interval

Myth and Truth

Increasing the sample size will
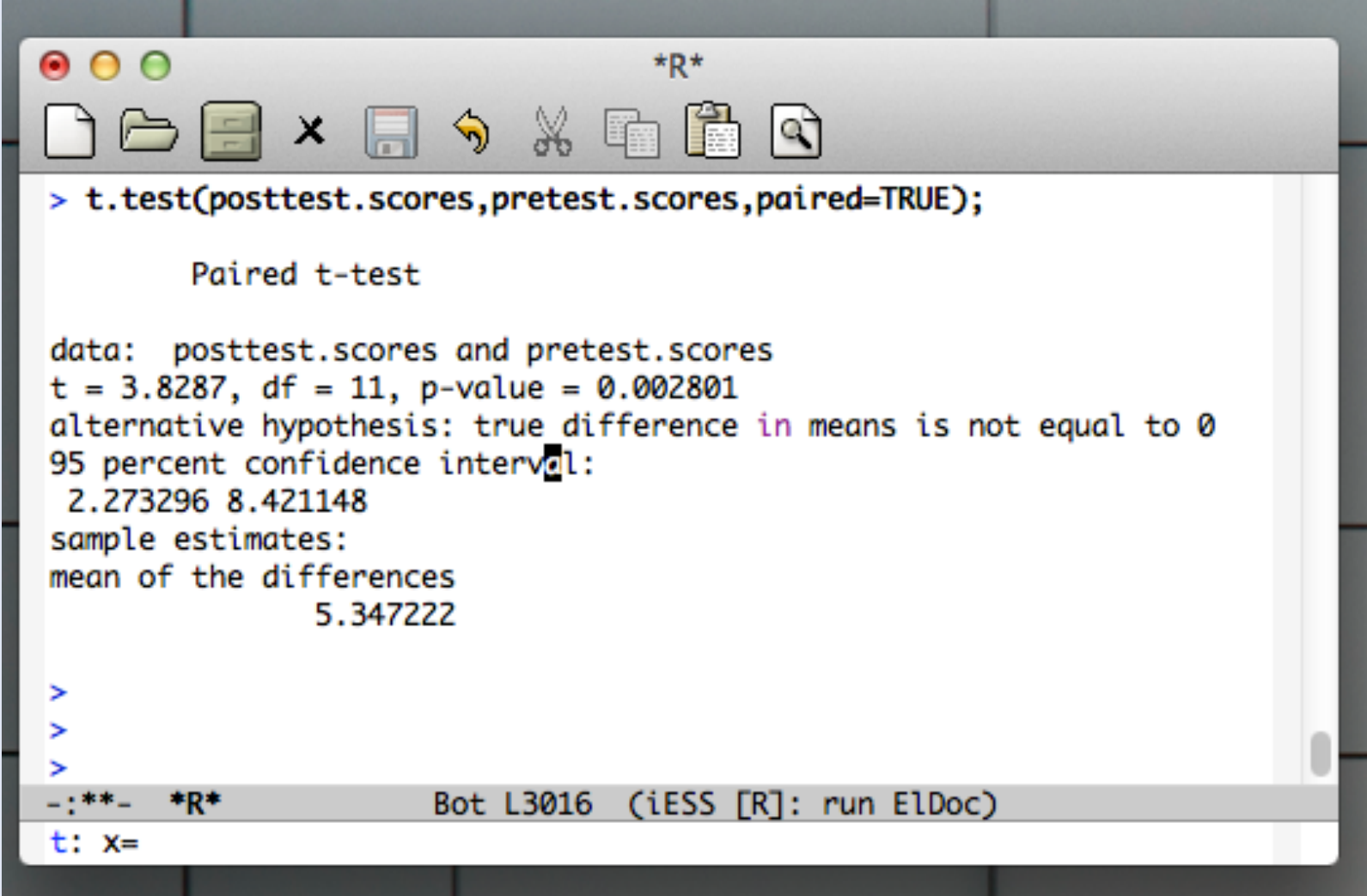 decrease the width of the
  confidence interval

TRUE

Confidence Interval Width

What influences the width of the confidence interval?

- Sample size
- Standard deviation of the study group
- Level of confidence selected

Relation Between p-value and confidence interval


```
> t.test(posttest.scores,pretest.scores,paired=TRUE);

        Paired t-test

data:  posttest.scores and pretest.scores
t = 3.8287, df = 11, p-value = 0.002801
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.273296 8.421148
sample estimates:
mean of the differences
          5.347222

>
>
>
```

-:**-   *R*   Bot L3016 (iESS [R]: run ElDoc)
t: X=

Relation Between p-value and confidence interval

Difference in Pre and Post-Test Scores

H_0: $\mu_D=0$ $\qquad\qquad$ $\alpha=0.05$

H_A: $\mu D \neq 0$

p-value less than level of significance		Confidence interval does not include the value in null hypothesis
p=0.0028	**=**	95% CI: (2.73 to 8.42)

Relation Between p-value and confidence interval

Difference in Pre and Post-Test Scores

H_0: $\mu_D=0$　　　　　　$\alpha=0.05$

H_A: $\mu D \neq 0$

There was a significant improvement in test scores after the curriculum (p<0.003). The mean improvement in test score was 5.3 points (95% confidence interval: 2.3 to 8.4).

Statistical vs Clinical Relevance

Remember…neither p-values nor confidence intervals give clinical relevance.

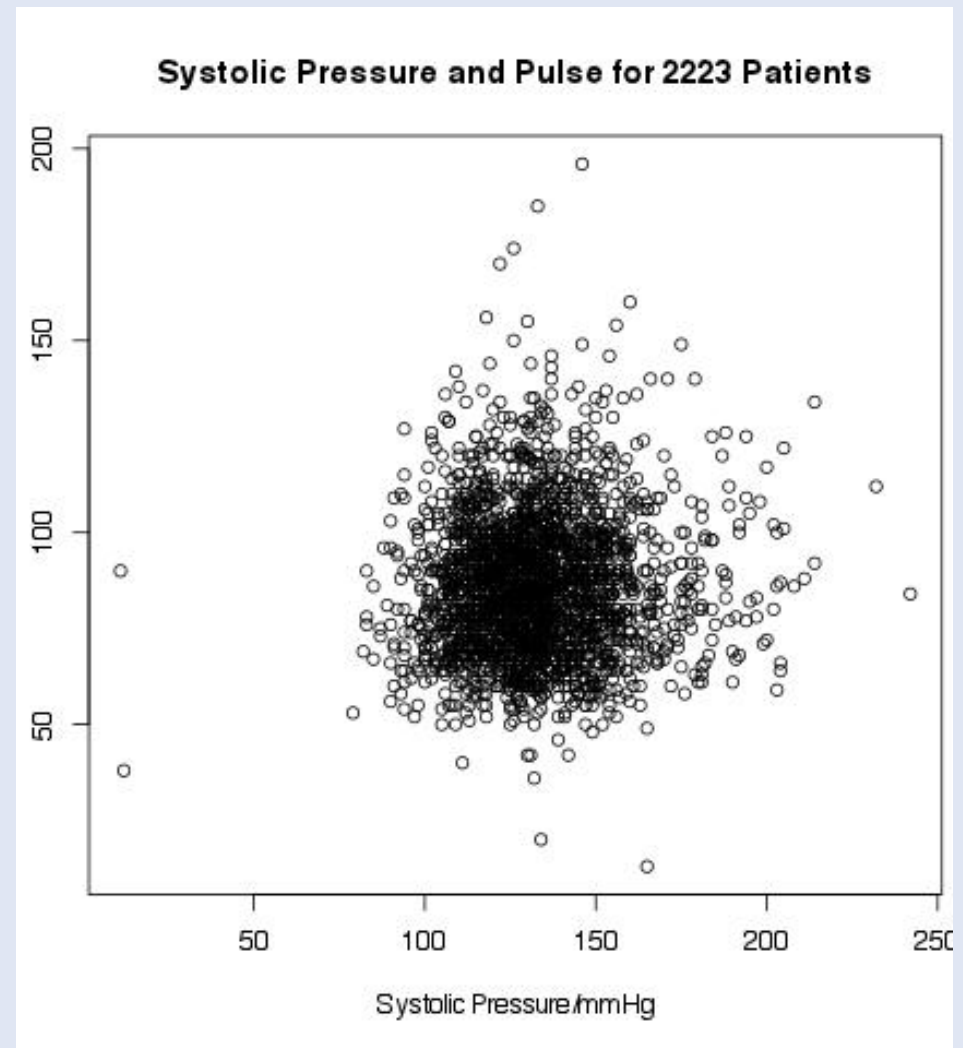# Statistical vs Clinical Relevance

H_0: ρ=0 (True Correlation is zero)
H_A: ρ≠0

Questions?

Confidence Intervals

Quiz Answers

1.

True / False

There is a 3% chance that the observed difference in evacuation time was due to chance alone

Quiz Answers

2.

True / False

The study should reject the null hypothesis of no difference in evacuation times at α=0.05 but not at **α=0.01**

Quiz Answers

3.

True / False

If the study was repeated 100 times, we would expect to find the same results approximately 3% of the time

Quiz Answers

4.

True / False

There is a 97% chance that the null
 hypothesis is false

Quiz Answers

5.

True / False

There is a 3% chance of finding results at least this extreme if in fact the two evacuation times are equal

Quiz Answers

6.

True / False

There is a 95% chance that the true mean concentration of carbon monoxide poisoning is between 3.5% and 12.9%

Quiz Answers

7.

True / False

If we repeated the study 100 times, we would expect the interval to be different each time. About 95 of 100 intervals constructed would contain the true mean

Quiz Answers

8.

True / False

The researcher should reject the null hypothesis that true mean concentration is 10% using **α=0.05**

Quiz Answers

9.

True / False

If we take a survivor at random, we estimate there is a 95% chance that there concentration of carbon monoxide is between 3.5 and 12.9%

Quiz Answers

10.

True / False

Increasing the sample size will decrease the width of the confidence interval

Bonus Question

What was wrong with the experiment of Carlo and Robert?

Why did they commit the type-1 error?

Was the stated p-value accurate?

What are possible consequences

Math Lesson

How to calculate a confidence interval