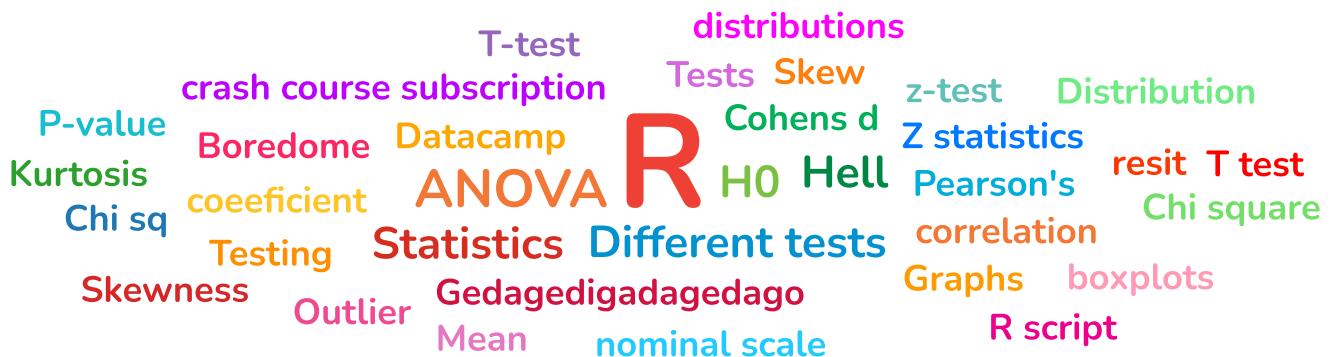


1_stats_II_Intro

Number of participants: 60

- cloud icon 1. What are some things you learned in Statistics for CSAI I? 50 respondents





2. If I measure the entire class's knowledge of statistics before and at the end of the course, and I wanted to see if you improved, what would be an appropriate test to make this comparison?

35 correct answers
out of 49 respondents

Factorial ANOVA

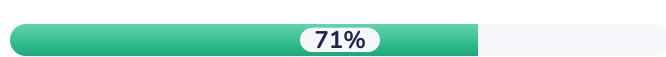


2%

1 vote



Paired-samples t-test



71%

35 votes

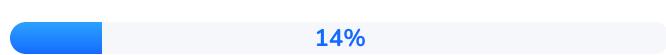
Chi-Square



8%

4 votes

Independent samples t-test



14%

7 votes

Correlation



4%

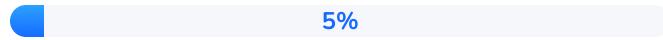
2 votes



3. When would you apply a Bonferroni correction?

10 correct answers
out of 43 respondents

The dependent variable is non-normal



2 votes

The independent variable is highly skewed



13 votes

To control for making multiple comparisons



10 votes

To account for heterogeneity of the variance



18 votes

What do you hope to gain from
4. Statistics for CSAI II (besides
credits/passing grade)?

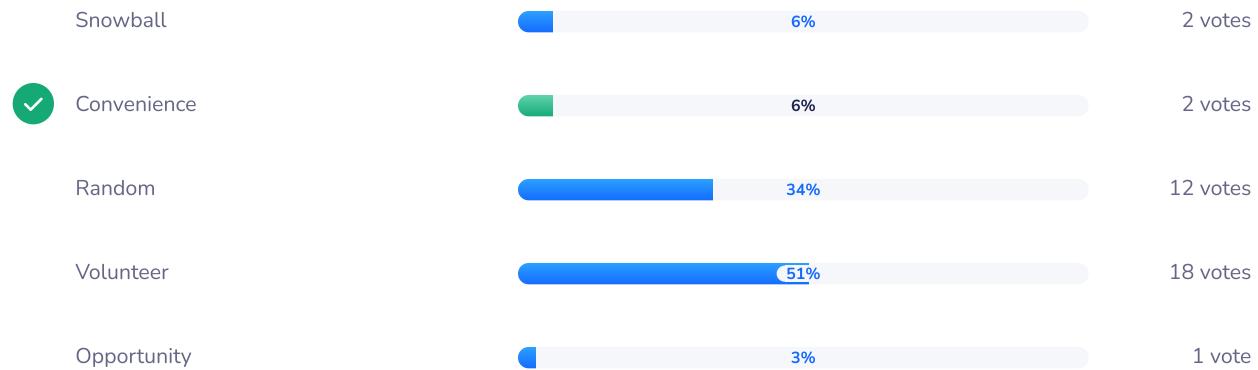
32 respondents

programming skills
Apply for future calculations Know something
Not crash R Fame money) R credits More R knowledge
boneferoni 😊 Statistic knowledge Knowledge pizza calzone Bonferoni
Enough knowledge Passing grade understand More stat knowledge
Gedagedigadagedago Morer understanding Happiness 😁
linear regression programming in R pass Kow

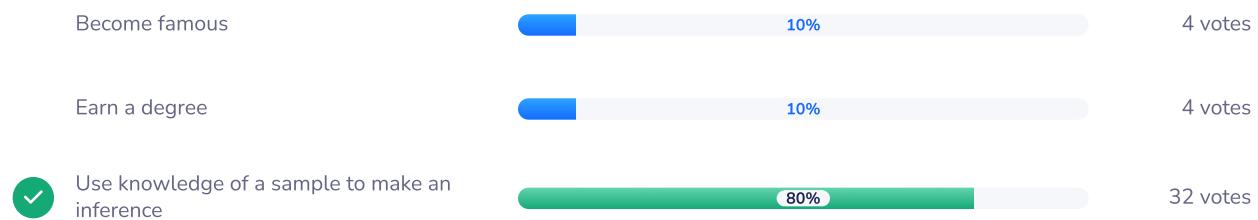
2_stats_ll_sampling

Number of participants: 52

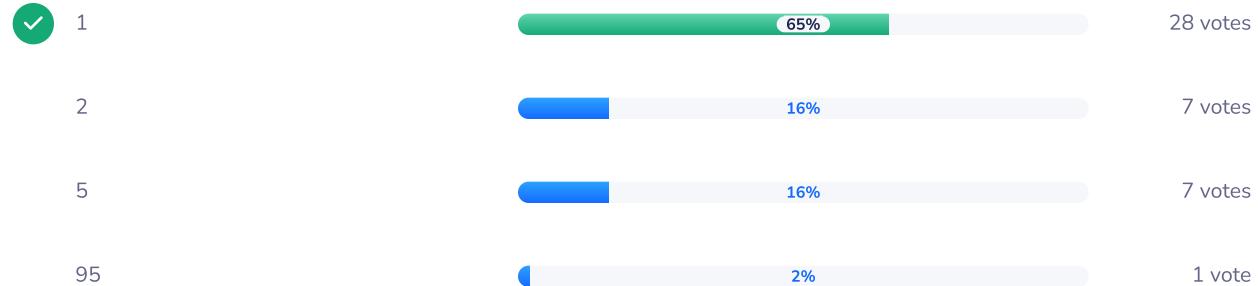
1. What type of sampling is most commonly used in cognitive and social science university settings?

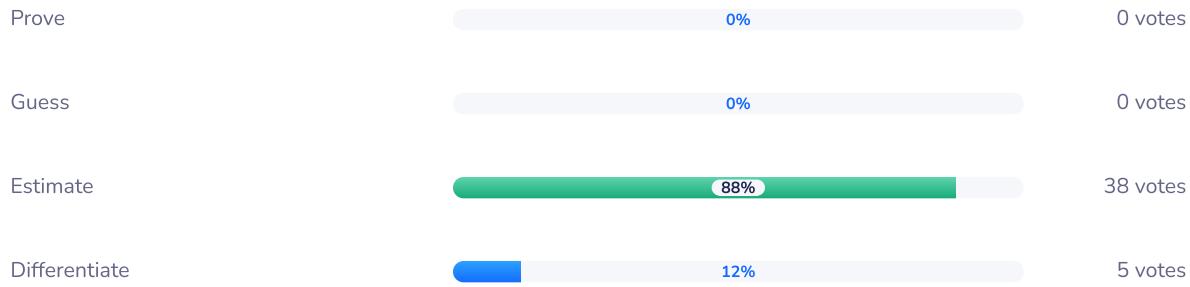
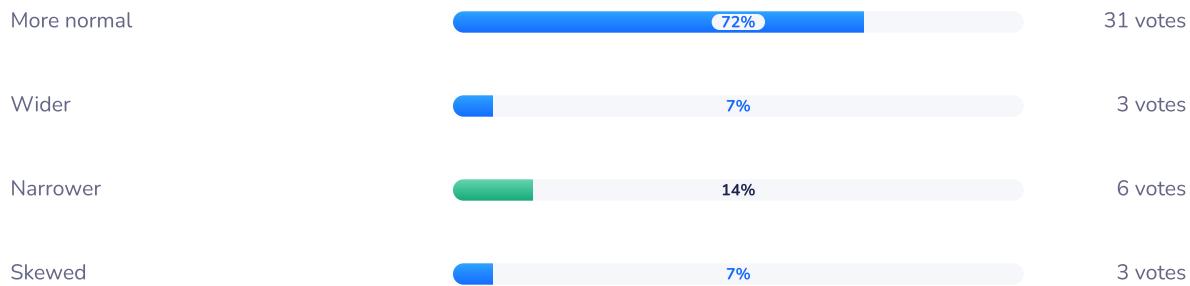
2 correct answers
out of 35 respondents

2. What is the common underlying goal of most empirical work utilizing statistics in CSAI?

32 correct answers
out of 40 respondents

3. If I replicate a study 20 times, how many studies will have a mean not included in the 95% CI?

28 correct answers
out of 43 respondents

 4. We use sample statistics to do which of the following regarding population parameters?**38 correct answers**
out of 43 respondents 5. As we increase the sample size, the sampling distribution of the mean becomes which of the following?**6 correct answers**
out of 43 respondents

3_stats_II_hypothesis_testing_intro_

Number of participants: 46

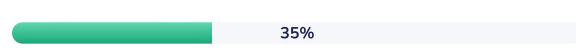
1. Which of the following is true about Pearson's correlation coefficient?

13 correct answers
out of 37 respondents

It ranges from -2 to 2



8 votes

 A value of 0 means there is no relationship between the two variables

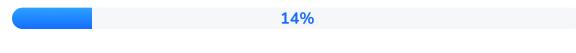
13 votes

A strong negative correlation implies that when the value of one variable decreases, the value of the other variable also decreases



11 votes

Pearson's correlation can only apply to a sample, and not a whole population



5 votes

2. What test statistic should be used when the population standard deviation is unknown?

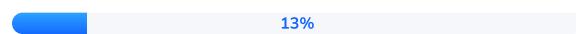
19 correct answers
out of 39 respondents

Z-test



14 votes

Cohen's D



5 votes

 T-test

19 votes

Beta test



1 vote

 3. Which of the following does NOT need to be known in order to compute the P-value?**9 correct answers**
out of 40 respondents

Knowledge of whether the test is one-tailed or two-tailed



20%

8 votes

The value of the test statistic



40%

16 votes



The alpha level



23%

9 votes

All of the above



18%

7 votes

4_more_correlation

Number of participants: 43

Which of the following test statistics

1. is used to test a hypothesis of a Pearson's correlation?

9 correct answers
out of 27 respondents

d  26% 7 votes

r  26% 7 votes

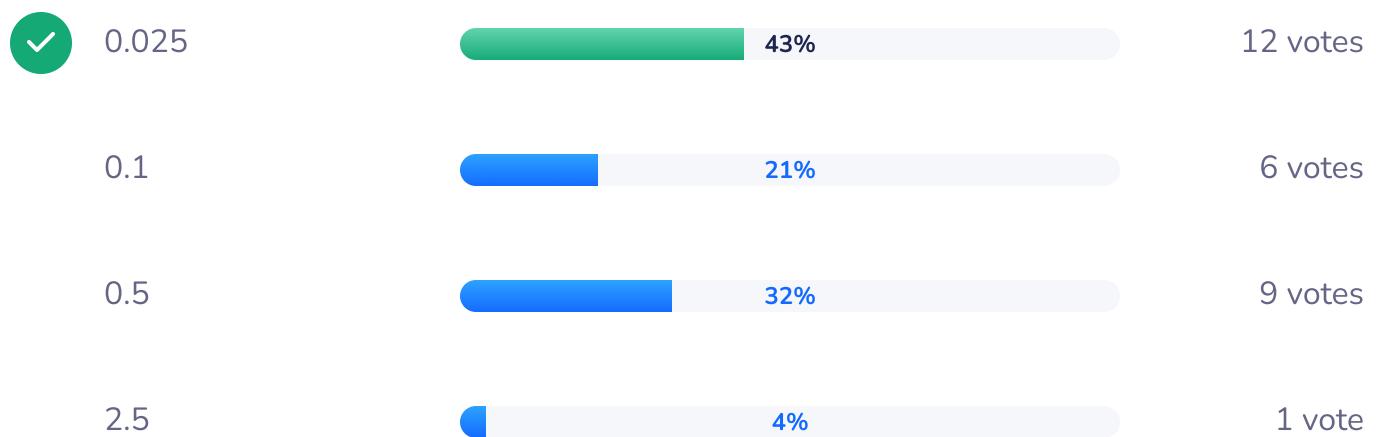
t  33% 9 votes

F  15% 4 votes

A researcher analyzes the relationship between annual income (in euros) and years of education completed, finding a covariance of 500 euros per year.

2. **Given that the standard deviation of annual income is 10,000 euros and the standard deviation of years of education is 2 years, what is the correlation between income and education?**

12 correct answers
out of 28 respondents





3. Which of the following would be a good reason to calculate a partial correlation coefficient?

12 correct answers
out of 29 respondents

The effect size is large



28%

8 votes

The R² is marginal



24%

7 votes

The data are ordinal



7%

2 votes



To account for influence of a third variable



41%

12 votes

stats_II_intro_to_regression

Number of participants: 32

Which of the following types of correlation would you use if you want to examine the

-  1. **relationship between x and y, while controlling for the relationship of z on both x and y?**

9 correct answers
out of 28 respondents

Pearson correlation



21%

6 votes

 Partial correlation



32%

9 votes

Spearman correlation



11%

3 votes

Semi-partial correlation



36%

10 votes



2. What correlation would be most appropriate for examining the relationship between ranking in a gaming competition and age, when the sample size is small?

12 correct answers
out of 25 respondents

Pearson correlation



12 votes

Partial correlation



0 votes



Spearman correlation



12 votes

Semi-partial correlation



1 vote



3. Which letter is always used to specify the outcome variable in linear regression models?

9 correct answers
out of 27 respondents

b



2 votes

β



9 votes

x



7 votes



Y



9 votes

5_stats_II_More_Regressio

Number of participants: 19

If you observed a -0.03 correlation

1. between hours of daylight and mood,
which are you most likely to conclude?



4 correct answers

out of 15 respondents

There is a negative
relationship
between hours of
daylight and mood.



40%

6 votes

There is a positive
relationship
between hours of
daylight and mood.



7%

1 vote

There is no
relationship
between hours of
daylight and mood.



27%

4 votes

Not enough
information is
provided.



27%

4 votes

2. In a regression equation, b0 describes which of the following?

8 correct answers
out of 17 respondents

The x intercept



12%

2 votes



The y intercept



47%

8 votes

The correlation coefficient



24%

4 votes

The regression coefficient of the predictor



18%

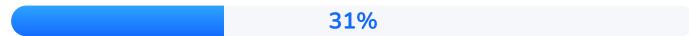
3 votes



3. All regression models must use a straight line to describe the data.

11 correct answers
out of 16 respondents

True



31%

5 votes



False



69%

11 votes



4. The slope of a regression line can be:

15 correct answers
out of 16 respondents

Positive

0%

0 votes

Negative

6%

1 vote

Zero

0%

0 votes

Undefined

0%

0 votes



All options are true

94%

15 votes



5. The R-squared value tells you which of the following (select the most correct)?

16 correct answers
out of 16 respondents

If there is a relationship between your predictor and your outcome

0%

0 votes

How much error is in your model

0%

0 votes



How much of the variance in the outcome variable your overall model accounts for

100%

16 votes

If your regression assumptions are met.

0%

0 votes

8_stats_ll_interactions

Number of participants: 19

Given the output below from the performance R package comparing Model 1 and Model 2, which of hte model has the better fit? Model | AIC | BIC | R2 | RMSE | Sigma | LogLik | F | p

 1.

----- 1 | 280.6 | 290.7 |
0.638 | 1.05 | 1.06 | -136.3 | 58.2 |
<.001 2 | 276.3 | 288.1 | 0.659 | 1.02 |
1.03 | -134.1 | 49.9 | <.001 -----

7 correct answers

out of 15 respondents

Model 1



27%

4 votes



Model 2



47%

7 votes

Not enough
information
provided



27%

4 votes



- 2. Create a list of THREE variables. Two of these variables (X, Z), should be ones you think might interact and have an effect on Y.**

21 answers

1 X

Amount of time working

study

Attendance

sleep quality

Sun

age

air temperature

2 z

Salary

sleep

stress levels

Study time

Rain

height

rainfall

3

Y

Job type

work productivity

grade

Grade

Growth

forest growth

weight

**In contrast to simple linear regression
where we are trying to estimate the
best fitting _____ that has the**

-  3. **least squares, multiple regression
determines the best fitting
_____ that has the least squared
error.**

9 respondents

In contrast to simple linear regression where we are trying to estimate the best fitting **line (5)** that has the least squares, multiple regression determines the best fitting **plane (3)** that has the least squared error.

9_stats_II_MR_categories

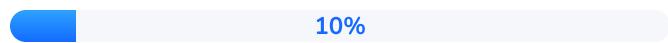
Number of participants: 18

Take a look at the R output below. It shows a simulated effect of the number of student in attendance for a given lecture, the level of engagement for students, and the effect of those on students final grades. Which of the following best describes the interpretation of this output, specifically regarding the interaction term? Call: lm(formula = grades ~ num_students * engagement, data = data) Residuals: Min 1Q Median 3Q Max -12.3516 -4.0132 0.1589 3.8344 11.5488 Coefficients: Estimate Std.

-  1. Error t value Pr(>|t|) (Intercept)
70.2346 2.3014 30.530
0.00000000000002 ***
num_students 1.5781 0.1432 11.014
0.00000000000002 *** engagement
2.9178 0.3832 7.615 0.000000000173
*** num_students:engagement -0.5153
0.0664 -7.759 0.000000000121 *** ---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error:
5.067 on 96 degrees of freedom
Multiple R-squared: 0.731, Adjusted R-squared: 0.725 F-statistic: 87.2 on 3 and 96 DF, p-value:
0.000000000000022

3 correct answers
out of 10 respondents

The effect of num_students on grades decreases as engagement increases, meaning larger classes have less impact on grades when students are more engaged.



10%

1 vote

The effect of engagement on grades decreases as num_students increases, meaning that in larger classes, the positive impact of engagement on grades diminishes.



3 votes

Both num_students and engagement independently increase grades, but the interaction suggests that higher engagement only has an effect when the class size is small.



5 votes

The interaction term indicates that grades are lower in larger classes regardless of engagement levels, meaning class size alone determines grades.



1 vote

Take a look at the R output below (same as before). It shows a simulated effect of the number of student in attendance for a given lecture, the level of engagement for students, and the effect of those on students final grades. Which of the following would be a good way to better understand the interaction effect? Call: lm(formula = grades ~ num_students * engagement, data = data) Residuals:

Min 1Q Median 3Q Max -12.3516

-4.0132 0.1589 3.8344 11.5488

Coefficients: Estimate Std. Error t

2. value Pr(>|t|) (Intercept) 70.2346
2.3014 30.530 0.0000000000000002 ***
num_students 1.5781 0.1432
11.014 0.0000000000000002 ***
engagement 2.9178 0.3832 7.615
0.000000000173 ***
num_students:engagement -0.5153
0.0664 -7.759 0.000000000121 *** ---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error:
5.067 on 96 degrees of freedom
Multiple R-squared: 0.731, Adjusted R-squared:
0.725 F-statistic: 87.2 on 3 and 96 DF, p-value:
0.000000000000022

2 correct answers

out of 8 respondents

Examine the AIC of the model

13%

1 vote

Probe the simple slopes to test whether they are different from each other

50%

4 votes



Calculate and plot
the simple slopes


50%

4 votes

Use gvlma() to
check the model
assumptions


25%

2 votes

In a study on student learning techniques, participants are divided into four groups based on their preferred study strategy: Note-taking, Highlighting, Flashcards, and Summarizing. To include these groups as a categorical predictor in a multiple regression model, the researchers need to use a coding system with how many dummy variables, and why?



3.

1 dummy variable, because only the most effective method needs to be compared with the others.

0%

0 votes

2 dummy variables, because g represents the number of groups, and two dummy variables will account for all four methods.

0%

0 votes

3 dummy variables, because there are four groups (study strategies), so $g-1 = 4-1 = 3$ dummy variables are needed to account for all methods in the model.

38%

3 votes

4 dummy variables, one for each group, because all groups need their own

63%

5 votes

3 correct answers
out of 8 respondents

variable in the regression model.

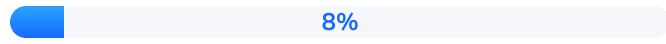
10_stats_II_polynomials

Number of participants: 15

-  **What does the intercept represent in
1. multiple regression using dummy
coding for three groups?**

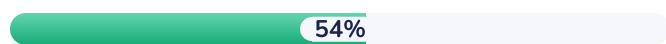
7 correct answers
out of 13 respondents

The grand mean



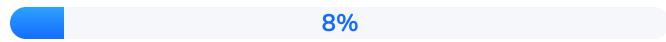
1 vote

 The mean of the
reference group



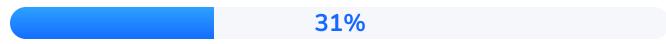
7 votes

The change in the
mean for a given
category relative to
the reference group



1 vote

The mean of the
base group adjusted
to differences in
group sizes



4 votes

2. You predict that there is a quadratic relationship between your x variable and y variable. What order polynomial model should you test?

11 correct answers
out of 12 respondents

1  0 votes

2  11 votes

3  0 votes

4  1 vote

What do you think about the current
3. projected winnow of the US 2024
election?

13 respondents

I didn't expect this
outcome



15%

2 votes

I expected this
outcome



23%

3 votes

Using my
knowledge of stats,
I knew the polls
were very close and
the margin of error
indicated either
candidate had a
possible path to
victory



31%

4 votes

I don't have an
opinion on this

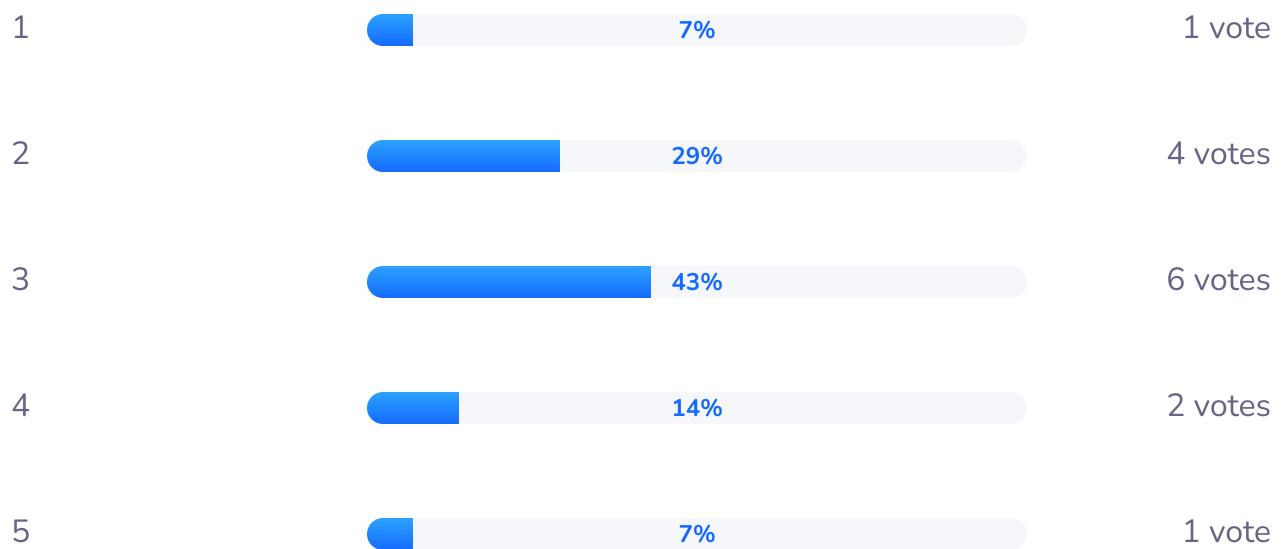


31%

4 votes

4. **On a scale of 1 (very low) -5 (very high), how would you rate your current understanding of the material covered so far in this course?**

14 respondents



11_stats_II_mixed_effects

Number of participants: 13

The following R code provides an example of polynomial regression of the second order using orthogonal transformations. `lm(formula = memory_score ~ poly(study_time, 2, raw = TRUE))`



1. The following R code provides an example of polynomial regression of the second order using orthogonal transformations. `lm(formula = memory_score ~ poly(study_time, 2, raw = TRUE))`

3 correct answers
out of 10 respondents

True



7 votes



False



3 votes



**In the equation below, which of the
2. following is true? $\text{memory_score} \sim$
 $\text{practice_time} + (1 | \text{participant})$**

4 correct answers
out of 10 respondents

The model is a
multiple linear
regression
predicting memory
score with practice
time and participant



20%

2 votes

The model is a
mixed effect model
with a random
effect for
`practice_time`



10%

1 vote

 The model is a
mixed effect model
with a random
effect for participant



40%

4 votes

The model includes
a first order
orthogonal
polynomial



30%

3 votes

12_stats_II_growth_curve

Number of participants: 10

Given the equation below, what is the correct random effect structure? Linear

1. mixed model fit by REML ['lmerMod']
Formula: Reaction ~ Days + (Days | Subject)

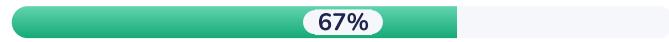
6 correct answers
out of 9 respondents

Random slopes for Subject, random intercepts for Days



2 votes

Random intercepts and random slopes for Days for each level of Subjects



6 votes

Random intercepts for Days, Random intercepts for subjects



0 votes

Random slopes and intercepts for both Days and Subject



1 vote

Given the information below, select the option below that is TRUE. Linear mixed model fit by REML ['lmerMod']

Formula: Reaction ~ Days + (Days |

Subject) Data: sleepstudy REML

criterion at convergence: 1743.63

Scaled residuals: Min 1Q Median 3Q

Max -3.9546 -0.4634 0.0231 0.4634

5.1793 Random effects: Groups Name

Variance Std.Dev. Corr Subject

(Intercept) 612.10 24.74 Days 35.07

5.92 0.07 Residual 654.94 25.59

Number of obs: 180, groups: Subject,

18 Fixed effects: Estimate Std. Error t

value (Intercept) 251.41 6.82 36.86

Days 10.47 1.55 6.75



2.

0 correct answer

out of 0 respondent

The random intercept variance of 612.10 indicates that there is no variation in baseline reaction times across subjects.

0%

0 votes

The correlation between the random effects for (Intercept) and Days suggests a strong negative relationship between baseline reaction times and the effect of Days across subjects.

0%

0 votes

The residual variance of 654.94 indicates that this model explains all

0%

0 votes

of the variability in Reaction times.

The fixed effect of Days suggests that each additional day is associated with an average increase of 10.47 in Reaction time across all subjects.

0%

0 votes



 **3. A growth curve model is a type of mixed effect model.**

8 correct answers
out of 9 respondents

True

89%

8 votes

False

11%

1 vote



What is the main fixed factor in an 4. unconditional growth curve model besides the intercept?

2 correct answers
out of 9 respondents

A random slope



11%

1 vote



Time



22%

2 votes

The main
independent
variable of interest



22%

2 votes

Curve variable



44%

4 votes