

Oral Exam - Example Questions and Responses

Allison Theobald

These are a variety of example exam questions given to students for their oral and final midterm exams. Students answered one randomly selected question from each bank, and were informed of their assigned questions when they entered the “exam room.”

The numbered questions are what were given to students the Friday before exam week. The *follow-up questions* are a sample of the questions I asked students following their explanation. These questions are not exhaustive, but provide an outline of the types of questions I asked.

Statistical Inference - Question Bank

1. Explain to someone who has never taken a statistics course what a p-value is.

Possible follow-up questions:

- A p-value is often reported as a percentage. What does that percentage p-value mean?
 - How is a p-value calculated?
 - What does a small / large p-value suggest?
2. It is very likely researchers in your field are familiar with parametric tests (e.g. *t*-test, *F*-test) but may not be familiar with simulation based methods (e.g. randomization test, bootstrap confidence interval). Explain to someone in your field the similarities and differences between these two methods.

Possible follow-up questions:

- How does a t-test calculate a p-value / confidence interval?
 - Why would a researcher choose to use one method over the other?
 - How are the data “simulated” to obtain a permutation distribution?
 - What are similarities and differences between a permutation distribution and a t-distribution?
3. Large portions of the scientific community are pushing to eliminate p-value thresholds for publication of scientific research. What is the role of a p-value threshold (e.g. 0.05) in hypothesis testing? What are the benefits and costs of eliminating p-value thresholds?

Possible follow-up questions:

- What does a p-value of 0.05 mean?
- Does a p-value less than 0.05 demonstrate that the findings of a study are “significant”?
- What could replace a p-value threshold in determining the “significance” of a study?

Linear Models - Question Bank

1. Describe to someone who has never taken a Statistics course the difference between simple linear regression and multiple linear regression.

Possible follow-up questions:

- Are there differences in the conditions for these two models?
 - How does the interpretation of the slope coefficient change in a multiple linear regression?
 - What are the benefits of using a multiple linear regression instead of a simple linear regression?
2. Some textbooks list “lack of influential points” as a model condition for regression. Why is this a potentially important condition to consider? What are the implications of removing all “influential” observations from a dataset?

Possible follow-up questions:

- What is an “influential” point?
 - How would an influential point impact a linear regression?
 - How would you determine if a point was “influential”?
 - How does removing every “influential” point from a dataset change what can be inferred from the regression?
3. A one-way ANOVA compares the means of three or more groups, and summarizes these differences with an F-statistic. What are the components of an F-statistic and what does each term symbolize in terms of the variability of the responses?

Possible follow-up questions:

- What does the mean square of the groups represent? How is it calculated?
 - What does the mean square of the errors represent? How is it calculated?
 - If I were to obtain a large F-statistic, what does that indicate about the mean square of our groups relative to the mean square of the error?
 - What would you expect to see in a data visualization for a one-way ANOVA with a large F-statistic?
4. What are the similarities and differences between a two-way ANOVA model and a multiple linear regression?

Possible follow-up questions:

- How are the “slopes” in a two-way ANOVA model formed?
- What would you look for in a visualization of a two-way ANOVA model to decide if the “slopes” are different?
- What do parallel slopes / different slopes indicate about the relationship between the explanatory variables and the response?
- How does a two-way ANOVA model summarize the relationships between variables? How does this differ from a multiple linear regression?