# Reading Quiz 4
## ASA Statement on p-values

NAME:

April 24, 2026

Please answer the following to the best of your ability. Partial credit will be given. For the first few questions please answer TRUE if the authors of the ASA statement on p-values would largely agree with the statement or FALSE if they would largely disagree.

1. **True** or **FALSE**: p-values give a probability that the data was randomly generated under the null hypothesis.
2. **True** or **FALSE**: A way to think about p-values is as a indicator of "incompatibility" of the data and the statistical model.
3. **True** or **FALSE**: p-values do not measure the size of an effect. As such, a small p-value for the the difference in means between two groups may still be associated with no meaningful difference between the two groups' means.
4. **True** or **FALSE**: While p-values do not measure the size of an effect, they are influenced by the size of the effect. As such, a large estimated effect for the the difference in means between two groups will be associated a small p-value.
5. **True** or **FALSE**: p-values give a statement on the explanation of the origination of the data.
6. **True** or **FALSE**: p-values by themselves do offer a good measure of evidence regarding a specific model or hypothesis.
7. **True** or **FALSE**: For hypothesis testing only, p-values by themselves do offer a good measure of evidence regarding a specific model or hypothesis.
8. **True** or **FALSE**: Journals are increasing leery of p-values, with one going as far as to ban them
9. **True** or **FALSE**: p-values are just one part of a statistical analysis and not the final stopping point.
10. **True** or **FALSE**: p-values can be used to judge/ascertain whether a model passes it's assumptions (eg independence)

Please answer 2 of the 5 following questions. Do not answer more than 2. If you answer more than 2, the lowest two scoring questions will be taken as well as another 2 point deduction so it is in your best interest to answer the two you are most confident about.

1. Point 3 of the paper was: "Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold." Why do you believe they listed (Scientific conclusions) and (business or policy decisions) specifically? Why not just say "Decisions"?

2. "Valid scientific conclusions based on p-values and related statistics cannot be drawn without at least knowing how many and which analyses were conducted, and how those analyses (including p-values) were selected for reporting." Why does knowing the original number of hypothesis tested matter? If helpful, give an example.

3. Earlier in the course we discussed the process of picking a linear regression model through several iterations of candidate models (eg the lego lab on multiple regression or as motivation for log-log models). How does that square away with the ASA's statement that one shouldn't pick a model based on statistical results?

4. I once said there was still a place where the use of p-values and the reject/fail to reject at .05 paradigm made sense. What was it? How does that fit into the ASA statement?

5. Describe why p-values are a theoretical probability and not an empirical probability (at least as far as we have seen in this class).

1. FALSE
2. TRUE
3. TRUE
4. FALSE
5. FALSE
6. FALSE
7. FALSE
8. TRUE
9. TRUE
10. FALSE

1. Because it is those two broad categories that are the audience such that situations exist in which p-values are used but for which a threshhold makes sense. For example, quality control can use p-values extensively but they are after a 5% rejection rate. Their (QC) decisions therefore aren't intended to be covered. (Side note: A finite list of items generally imply only those items are associated with the rest of the sentence and are not major examples, which was a common mistake. If they were examples or the important ones they (the authors) would need to indicate that)
2. This ties into "cherry picking" where the researcher may run 100 different tests for checking to see if two subgroups are the same and then only report the three tests that confirm their hypothesis. Knowing the number of original tests let's us understand how believable these results are.
3. Several points but one is that we actually didn't have statistical results at that point. We proposed a model and saw it was deficient for the assumptions of the model. Before ever looking at p-values we had finialized our model.

   More so if a model that fails the assumptions and doesn't fit the data well then clearly results based on it are suspect. A geneal rule in statistics is to be reasonable and using a model that is clearly deficient to say meaningless results is not reasonable.
4. See question 1 but it's quality control.
5. An emprircal statistic or probability is one that is built off of the data directly with minimal assumptions or statsitical modeling (eg the empirical maximum of 20 students' gpa's is just the highest ovserved gpa for those 20 students). The p-value, alteranteively, is not calculated from a distribution that we have (eg the sample distribution) and instead is a theoretical probabililty built using assumptions and statistical models to approximate reality.