5.1 Is the Data Good Enough?

Random Sampling/Assignment

The statistical inference methods you have learned almost always start with

- Observational Studies: Participants were a simple random sample from the population.
  - Meaning: Every possible sample of size *n* from the population has an equal chance of being chosen as our sample.

- Experiments: Subjects were randomly assigned to the values of the explanatory variable.
  - Meaning: Each subgroup of the appropriate size had the same chance of being assigned to each value of the explanatory variable.
    (Unequal sample sizes among the groups can also be handled with careful calculations.)

Or other, more complex random methods

In some more advanced statistics courses . . .

We learn how to perform appropriate statistical inference on data from a more complex random sample.

- Must have appropriate randomness built in so that we can do (or model) forming an appropriate sampling dist'n of the summary statistic.
- Generally speaking, the analysis results in more complicated estimators of the variance of the summary statistic.
- Generally speaking, the reason for the more complex analysis is to reduce the variance of the summary statistic in order to obtain the information we need with smaller samples than what we would need from a simple random sample.

What if it is not a random sample?

- Maybe the data did not actually come from a simple random sample. When do people make such inferences?
- Can you convince yourself that the data are "equivalent to" what would have come from a simple random sample?
- Which of these scenarios would you bet on?
  1. For a professional basketball player, estimate the proportion of free-throws he will make this year from his free-throw proportion from the first ten games of this season.
  2. For your friend's 14-year old son on the school basketball team, estimate the proportion of free-throws he will make this year from his free-throw proportion from the first ten games of this season.

Conclusion

- If
  the data are randomly selected from the population

- It is appropriate
  to use that to infer results about the population.

What about experiments?

- Random assignment of subjects to the explanatory variable is not hard to do (unlike random sampling from pop'n.)
- Random assignment …. results in the various possible "confounding variables" being, on the average, randomly distributed among the treatments.
- Thus, we can use the results of such appropriately designed experiments to provide evidence for causation.

“Representative” Samples

In introductory statistics courses, we often emphasize that random sampling is needed to obtain “representative” samples.

That’s tempting to say, because they are useful for that, but doesn’t get to the heart of the matter.

Statistics is Based on Probability

Statistical Inference procedures

- Based on sampling dist'ns of summary statistics
- Sampling dist'ns come from replicating the sampling method MANY times
- Need a probability-based sampling method to create those sampling dist'ns

Chapter 5

Start of new section:
Difficulties about Making Decisions

Making Decisions (from previous lecture)

Consider

- The design of the study
- The implementation of the data collection (not just what was planned, but also what actually happened)
- The requirements of the method you used to make the probability statement(s) from statistical inference

Conclusion of a Hypothesis Test

- Do we stop with the *p*-value?

- Not if we want to make a decision
  (Make a decision = Decide whether to act as if the Ha claim is true.)

- For that, someone needs to decide
  how small the *p*-value must be for them to act as if the Ha claim is true. (i.e. as if the Ho is false.)

  The name we give to that chosen value is the "significance level."

Choosing a Significance Level 1

- There are two types of errors one can make in such a decision
  Rejecting Ho when it is really true.
  Failing to reject Ho when it is really false.
- If we choose a value that makes it fairly hard to reject Ho (for a given sample size)
  that makes it more likely to fail to reject Ho when we should.
- This “trade-off” between the two types of error is important.
- In many situations, those two types of error have differing consequences.

Choosing a Significance Level 2

- Choosing a significance level should be done from the perspective of the decision-maker
- *Either*
  --by balancing the "cost" of the two types of consequences
  or
  --by listening to accepted experts who are used to making real-life decisions in your field and have done this balancing for typical types of decisions.

Choosing a Sample Size

The sample size is crucial in how large a difference from the Ho value you can expect to detect.

- A small sample size will make it unlikely you can detect a relatively small difference from the Ho value.
- A large sample size will make it easy to detect a relatively small difference from the Ho value. (Some differences may be so small that they are not of practical significance.)

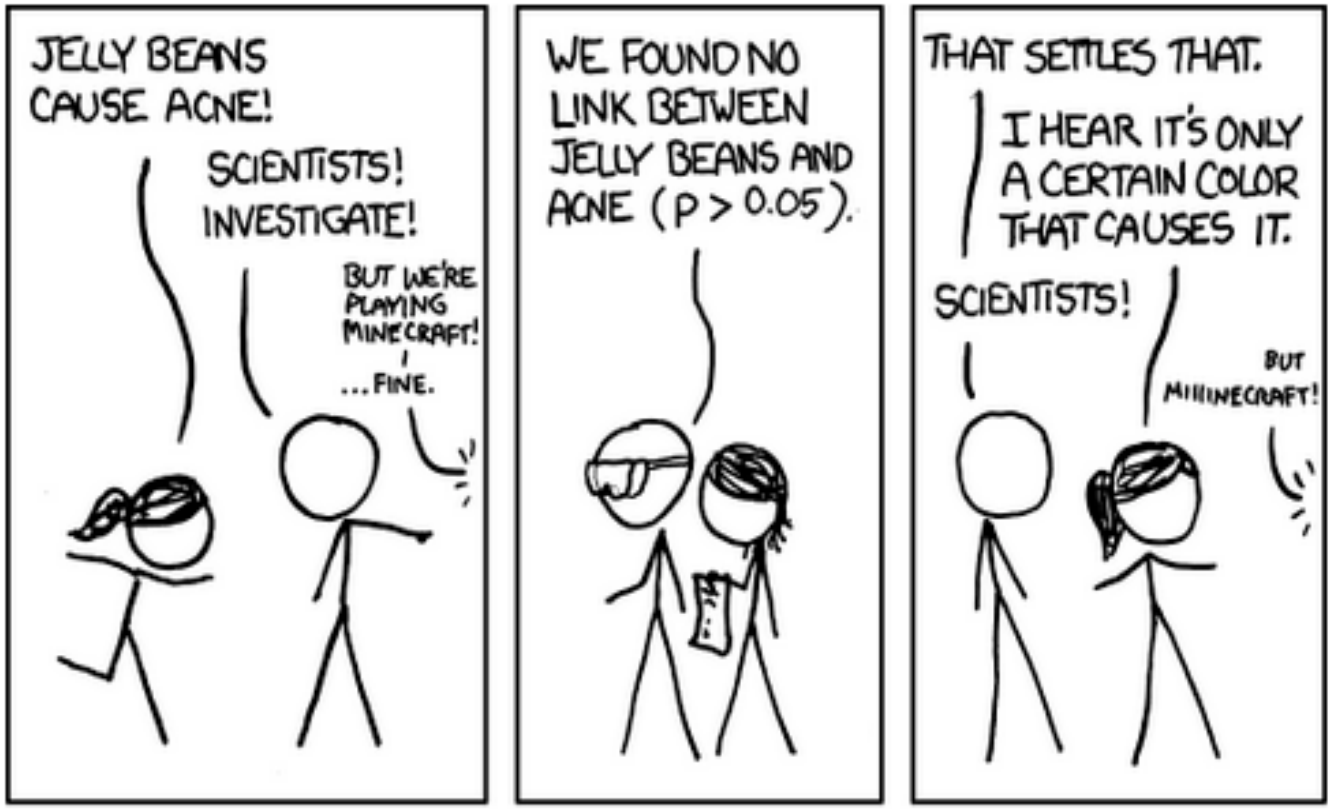
“Multiple” Testing

Hypothesis testing is designed to test a claim that you have some reason to think might be true.

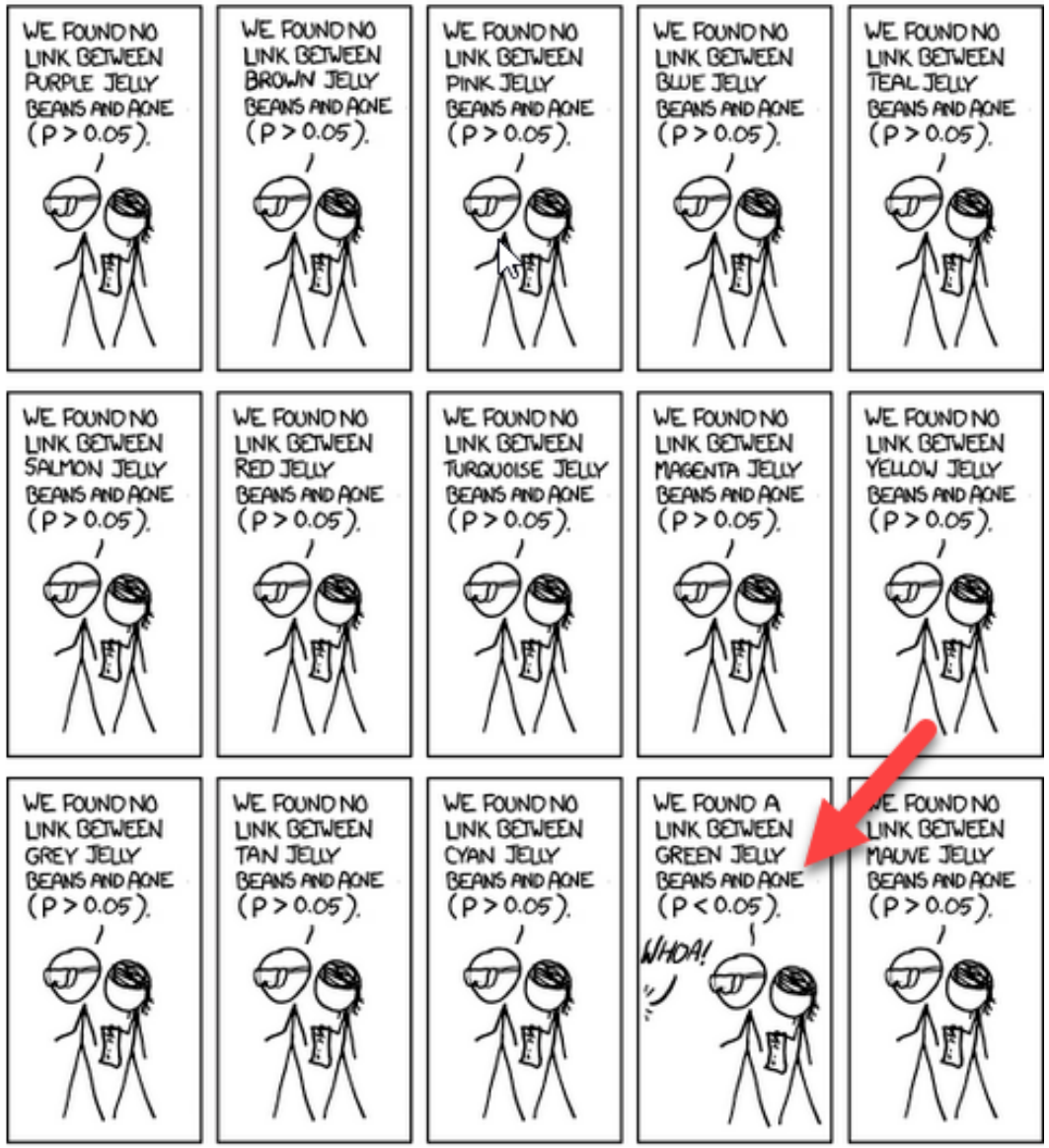
- If you use it to simply “shop around”
  - to try to find something that gives a small *p*-value,
  - then you’ll find some,
  - even if there are no real effects,
  - just because these *p*-values are probabilities and, occasionally, something with low probability happens just by chance.
- Mathematical Statistics comment:
  You can “correct” your *p*-value by using a Bonferroni correction

. .

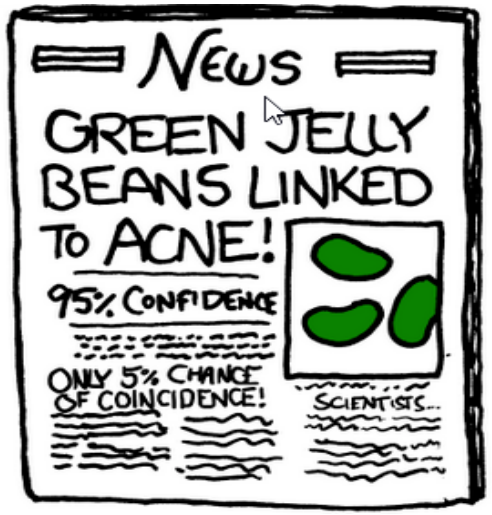
Example: https://xkcd.com/882/



Jellybeans, continued



What NOT to Do!



What to Do

Simplistic answer:

- Replicate the study with that focus.

- Why?
  If there really is not an effect, then it is not likely a second study will show a small *p*-value.

What to Do? Better answer

- Think about the reasons you had for testing jellybeans initially and decide whether the resources available are best used for
- replicating this study
  or
- exploring other hypotheses about what causes acne

Summary

- To choose a significance level, someone should balance the "costs" of the consequences of the two types of error.
- Sample size matters.
- Not all statistically significant results are significant in the real-life context.
- Beware when interpreting results from multiple tests.
- Replication of a study can be useful to help avoid rejecting a true Ho.