Mathematics 231

Lecture 24

Liam O'Brien

Announcements

- Reading
  - Today — M&M 6.3 — 494-399
    M&M 6.4 — 401-410
  - Next class — Power calculations

Topics

- Hypothesis Testing
  - Summary of hypothesis testing
  - Errors: Type I and Type II errors
  - Hypothesis testing for a population mean
  - Hypothesis testing for a population proportion

Hypothesis Testing: Summary

1. State H_0 and H_A.
2. Calculate the value of the test statistic measuring the evidence against H_0.
3. Compute the p-value.
4. Compare p-value to significance level α and draw a conclusion:
   a) If p-value $< \alpha$, then reject H_0 in favor of H_A.
   b) If p-value $> \alpha$, then do not reject H_0.
5. Report the conclusion carefully.

P-Values and One- vs Two-Sided Alternatives

- Remember that the p-value is the probability of observing a test statistic as extreme, or more extreme, than the one we did observe ***given*** the null hypothesis is true.
- The calculation of the p-value depends on the form of our alternative hypothesis.
- If $H_A: \mu \neq \mu_0$, then p-value$=2P(Z>|z|)$
- If $H_A: \mu < \mu_0$, then p-value$=P(Z<z)$
- If $H_A: \mu > \mu_0$, then p-value$=P(Z>z)$

Hypothesis Testing: 2 Types of Errors

- When conducting a test of hypothesis there are two types of errors we can make.
- That is, the data from our sample do not always lead to the correct conclusion.
- There is a useful analogy with the judicial system.

Decision based on sample	Truth about the population: H_0 true	Truth about the population: H_a true
Reject H_0	Type I error	Correct decision
Accept H_0	Correct decision	Type II error

Type I and Type II Errors

- When conducting a test of hypothesis, there are two types of errors we can make.
- Type I error: Rejecting H_0 when it is true.
- Probability of a Type I error is denoted by α.
- Type II error: Not rejecting H_0 when H_A is true.
- Probability of a Type II error is denoted by β.

Why Use α=0.05?

- Fisher said, "…it is convenient to draw the line at about the level at which we can say 'Either there is something in the treatment, or a coincidence has occurred such as does not occur more than once in twenty trials…"
- "Personally [I prefer] to set a low standard of significance at the 5% point, and ignore entirely all results which fail to reach that level."

Dear P-Value Worshipper

- "While the significance level ($p = 0.09$) obtained…did not reach significance at $\alpha = 0.05$, it does constitute a trend. The fact that the significance level approached 0.05 is encouraging and of considerable interest, in light of the small sample size… In fact, the term "statistically significant," which is dependent on the choice of the type I error level, is not necessarily synonymous with a finding being significant from a subject-matter perspective."

- me

Type II Errors and Power

- If the probability of committing a type II error is β, then 1 - β is the **power** of the test.
- Power = P(reject H_0 | H_A is true)
- Would like to minimize type I and type II errors, but there is a trade off: as α decreases, β increases and vice versa.
- Ordinarily, we fix the probability of committing a type I error (say, $\alpha = 0.05$) and we take what we can get when it comes to β.

Power

- If the power is low then there is little chance that we will detect a significant different even if one exists. The power of a test can be increased by:
  - Increasing the sample size
  - Reducing σ
  - Considering alternatives, μ_A, that are farther away from the null, μ_0
  - Raising the significance level, α
- The most practical way to increase the power is to increase the sample size, n.

Power

- The power is different for each **specific** alternative hypothesis.
- Curves can be drawn for a specific case.
- Tables can be generated for a variety of alternatives.
- Consider the power of a hypothesis test for a single population mean (μ known).

A Note on "Critical Values"

- Back in the olde days, p-values weren't easy to calculate.
- After the test statistic was calculated, it was compared to "critical" value(s).
- The critical values define the rejection region under the null hypothesis, and thus depend on the significance level, α.
- If the test statistic falls within the rejection region, then the null hypothesis can be rejected at the α level.