THE POWER TO KNOW.®

Misunderstanding Statistical Inference

Mark Bailey
JMP Training & Service
SAS Education

Introduction

- Inference in the form of some hypothesis test is very common.
  - How many such tests are performed each day?
- Many people use statistical inference without understanding how it works or realizing what it means.
  - How does this confusion affect their conclusions?

Topics

- Review generic structure of a hypothesis test.
  - Demonstrate sampling distribution for sample average and p-value under null hypothesis and under alternative hypothesis.
- Review a short selection of common misconceptions.
  - Demonstrate when difference test is wrong approach and equivalence test should be used.

Simple Example

- New and improved microwave oven popcorn is supposed to yield more than 6.5 cups of popped corn on the average.
- How would you confirm this claim?
- Obtain samples of the new product, run the process, measure the response, and decide based on this experience.
  - Perhaps you also obtain a sample of the old product and include it in the comparison.

But What Does It Mean?

- Here is an answer, p > t: 0.003 (3/1000)
- Here is an answer, p > F: 0.06 (6/100)
- These answers confuse, even terrify some users instead of enlightening them.
- Users waste a lot of time guessing what it means.
- They look to someone else for an answer, and accept it on faith, because they do not know for themselves.

Meaning is Lost

- The hypothesis test involves
  - an abstract statement of the original question and
  - an elaborate, formal inference process, often detached from the process under study.
- This terminology and thought process is different from that of the user's main discipline.
- Show me the popcorn!

Element 1: The Hypotheses

- The mean yield of popped corn is greater than 6.5 or it is not.
  - What is my position on this claim?
- Assume one case to be true (*null hypothesis*) but consider the other possibility (*alternative hypothesis).*
- H0: $\mu \leq 6.5$, H1: $\mu > 6.5$
  - What is my position on this claim?

Element 2: The Significance

- What if H0 is true?
- Sample variation causes uncertainty in estimation and decisions.
- It is possible to obtain samples that cause us to decide against H0.

Element 2: The Significance

- Confidence is the probability of correctly not rejecting H0.
- Significance is the probability of an incorrect decision, often $\alpha = 0.05$
- Are these concepts intuitive?
- Is the meaning here (confidence, significance) the same as it is in common usage?

Element 3: The Power

- What if H1 is true?
  - What is the minimum deviation from H0?
- Sample variation causes uncertainty in estimation and decisions.
- It is possible to obtain samples that cause us to decide to not reject H0.

Element 3: The Power

- Power is the probability of correctly rejecting H0.
  - H0: $\mu \leq 6.5$, H1: $\mu > 6.5$
  - $\delta = 0.25$, $\sigma = 0.2$
  - $n = 10$, $1-\beta = 0.95$
- Are these concepts intuitive?
- Is the meaning here (power) the same as it is in common usage?

Element 4: The Sample Statistic

- Most users learn to use a sample average for comparison.
  - Compare sample average to hypothesized value, to another sample average, or compare several sample averages.
- The original sample statistic is first transformed for the sake of the statistical process.
  - To use a standard approximate sampling distribution of the statistic.
- Do users understand this step and the reason for it?

Element 5: The P-Value

- What if H0 is true?
  - How likely is your sample statistic or one even more extreme?
- What if H1 is true?
  - How likely is your sample statistic or one even more extreme?
- Demonstrate both cases with simulation.
- Is this phenomenon intuitive or expected?

Misconceptions: The Set Up

- Professionals may use hypothesis tests with no training, with only informal training, or with formal training in school or professional short course.
- Regardless, many users approach inference as if it is magic, and apply it mechanically.
  - Frustration leads to acquiescence.

Misconceptions: The Set Up

- Failure to understand and appreciate the context and the elements and flow in the process of a statistical hypothesis test.
  - Intelligent professionals in fields outside of statistics who want quick answers.
- Real-world questions are transformed into an abstract construct for the sake of the calculation.
  - Original question: Is A a better choice than B?
  - Statistical answer: The probability of obtaining this sample statistic theta, or one even more extreme, assuming the null hypothesis is true…

Misconception: Significance Always 0.05

- Reject the null hypothesis if $p < 0.05$ in all cases.
  - They believe that 0.05 level comes from theory.
- They believe that $\alpha = 0.05$ makes their decision objective.
- Many students react negatively to idea of using a smaller or larger alpha level.
- They do not like to choose significance for themselves.

Misconception: Power

- In almost a decade of teaching a variety of professionals, I have rarely encountered one who knows what power is, thinks about it, and uses it to estimate an adequate sample size.
- The majority of these students have used a hypothesis test prior to my training.
- Many of these students use the same amount of testing in every situation.
  - *We always run the test in triplicate.*

Misconception: Significance = Power

- Many users require the power to be the same as the confidence.
- For example, confidence is 95%, and power is, too.
  - When they are not the same, users realize they do not know what power is about.
- Fail to appreciate that significance assumes null hypothesis and power assumes the alternate hypothesis.

Misconception: Tests are Infallible

- Hypothesis tests *prove* an idea.
  - Some users really believe this statement.
- In spite of involving α in the decision, $p < \alpha$, the user accepts the conclusion as certain truth.

Misconception: Tests Work Both Ways

- Failure to reject H0 is the same as *proving* H0.
- This mistaken idea usually indicates that the hypotheses should be reversed.
  - Instead of using a hypothesis test of a difference, they should use a test of equivalence.
- Demonstrate this case with simulation.

Remedy: Context

- You can't just throw data at a calculation and expect meaning to arise.
  - Statistics are stupid.
- The calculation is the last step.
- Focus on the process, not the calculation.
  - Let's not 'put the cart before the horse.'
- Focus on the model, the whole picture.
  - Key elements presented as assumptions.

Remedy: Context

- Avoid ‘rules of thumb’ that short-cut statistical thinking about the problem.
  - Why hurry or rush?
- Help the user relate all elements of a hypothesis test to their process of making a decision.
- Isolate the elements in question from other influences so that any remaining uncertainty is essentially due to sampling variation.

Remedy: Alternatives

- Consider permutation tests or bootstrapping methods that might be less abstract, more direct than approximate methods (e.g., t distribution).
- Consider posing solution in Bayesian terms.

Contact Information

- Mark Bailey
  SAS Institute Inc.
  (856) 616-0644
  mark.bailey@sas.com