STAT 2593
Lecture 033 - Further Aspects of Hypothesis Testing

Dylan Spicker

Further Aspects of Hypothesis Testing

1. Describe the difference between statistical and practical significance.

2. Understand the equivalence between hypothesis testing and confidence intervals.

3. Describe shortcomings with NHST and how it is implemented in practice.

Practical Significance versus Statistical Significance

▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.

Practical Significance versus Statistical Significance

- ▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.
- ▶ The hypothesis test does not indicate whether observed differences *actually* matter.

Practical Significance versus Statistical Significance

- ▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.
- ▶ The hypothesis test does not indicate whether observed differences *actually* matter.
  - ▶ Often we will differentiate **practical significance** with **statistical significance**.

Practical Significance versus Statistical Significance

- ▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.
- ▶ The hypothesis test does not indicate whether observed differences *actually* matter.
  - ▶ Often we will differentiate **practical significance** with **statistical significance**.
  - ▶ An effect which is statistically significant may not be practically significant.

Practical Significance versus Statistical Significance

▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.

▶ The hypothesis test does not indicate whether observed differences *actually* matter.

  ▶ Often we will differentiate **practical significance** with **statistical significance**.

  ▶ An effect which is statistically significant may not be practically significant.

  ▶ Important to ask: "Even if this *were* true, would we care?"

Practical Significance versus Statistical Significance

▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.

▶ The hypothesis test does not indicate whether observed differences *actually* matter.

  ▶ Often we will differentiate **practical significance** with **statistical significance**.

  ▶ An effect which is statistically significant may not be practically significant.

  ▶ Important to ask: "Even if this *were* true, would we care?"

▶ Note: the p-value is **not** a measure of effect size, generally.

Practical Significance versus Statistical Significance

▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.

▶ The hypothesis test does not indicate whether observed differences *actually* matter.

  ▶ Often we will differentiate **practical significance** with **statistical significance**.

  ▶ An effect which is statistically significant may not be practically significant.

  ▶ Important to ask: "Even if this *were* true, would we care?"

▶ Note: the p-value is **not** a measure of effect size, generally.

  ▶ With a large sample, even very small differences will be detected.

Practical Significance versus Statistical Significance

▶ Our hypothesis tests indicate whether these is evidence *against* the null hypothesis.

▶ The hypothesis test does not indicate whether observed differences *actually* matter.

  ▶ Often we will differentiate **practical significance** with **statistical significance**.

  ▶ An effect which is statistically significant may not be practically significant.

  ▶ Important to ask: "Even if this *were* true, would we care?"

▶ Note: the p-value is **not** a measure of effect size, generally.

  ▶ With a large sample, even very small differences will be detected.

  ▶ In a small sample, even very large differences may not be detected.

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

▶ If you have a $100(1 - \alpha)\%$ confidence interval, this gives you a α level hypothesis test.

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

▶ If you have a $100(1-\alpha)\%$ confidence interval, this gives you a α level hypothesis test.

    ▶ Consider $H_0 : \theta = \theta_0$.

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

▶ If you have a $100(1 - \alpha)\%$ confidence interval, this gives you a α level hypothesis test.

   ▶ Consider $H_0 : \theta = \theta_0$.

   ▶ If θ_0 is in the CI, then you **fail to reject** the null hypothesis.

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

▶ If you have a $100(1 - \alpha)\%$ confidence interval, this gives you a α level hypothesis test.

  ▶ Consider $H_0 : \theta = \theta_0$.

  ▶ If θ_0 is in the CI, then you **fail to reject** the null hypothesis.

  ▶ Otherwise, you can reject the null (at α significance).

Confidence Intervals and Hypothesis Testing

▶ Every confidence interval can be viewed as a corresponding hypothesis test based on critical values.

▶ If you have a $100(1 - \alpha)\%$ confidence interval, this gives you a α level hypothesis test.

  ▶ Consider $H_0 : \theta = \theta_0$.

  ▶ If θ_0 is in the CI, then you **fail to reject** the null hypothesis.

  ▶ Otherwise, you can reject the null (at α significance).

▶ The mathematics is equivalent, it is simply two ways of expressing the same idea.

Critiques of NHST

- While NHST can be a powerful tool, there are many individuals who are critical of its use.

Critiques of NHST

- While NHST can be a powerful tool, there are many individuals who are critical of its use.

  - It will almost never be the case in the real world that effects are *exactly* equal to zero.

Critiques of NHST

▶ While NHST can be a powerful tool, there are many individuals who are critical of its use.

  ▶ It will almost never be the case in the real world that effects are *exactly* equal to zero.

  ▶ The NHST framework prioritizes p-values above all else.

Critiques of NHST

▶ While NHST can be a powerful tool, there are many individuals who are critical of its use.

  ▶ It will almost never be the case in the real world that effects are *exactly* equal to zero.

  ▶ The NHST framework prioritizes p-values above all else.

  ▶ The null hypothesis is often a *strawman*.

Critiques of NHST

► While NHST can be a powerful tool, there are many individuals who are critical of its use.

  ► It will almost never be the case in the real world that effects are *exactly* equal to zero.

  ► The NHST framework prioritizes p-values above all else.

  ► The null hypothesis is often a *strawman*.

  ► The framework itself is ripe for abuse.

Summary

▶ Practical significance should be considered in conjunction with statistical significance.

▶ Confidence intervals present a rephrased version of hypothesis tests.

▶ There are many prominent statisticians who think that the NHST framework should be abandoned; it is always worth considering alternatives.