Module 3F: Hypothesis Testing

Applied Epistemology: A Framework for how we know things scientifically

Agenda:

1. H_O/H_A: Our model of the test universe (the distribution of the variable)
2. Test & assumptions: are the assumptions met? Is the test valid?
3. Quantitative evidence: **p-value**, or critical value.
   - False positive =Type I (α), False Negative = Type II (β), Type III errors
   - Sensitivity, Specificity, Power → confusion matrix, ROC/AUC curve
   - Positive Predictive Power, Negative Predictive Power
   - Confusion Matrix
   - **ROC/AUC curve**
4. Conclusion & uncertainty/estimation

Two independent studies are performed to test the same null hypothesis.

What is the probability that one or both of the studies obtains a significant result and rejects the null hypothesis ***even if the null hypothesis is true***? Assume that in each study there is a **0.05** probability of rejecting the null hypothesis.

P[rejecting |ho is true] =P[study 1 reject|ho] + P[study 2 reject|ho] - P[study 1 reject|ho] * P[study 2 reject|ho]

= 0.05+0.05-0.025 = 0.0975

You can consider the previous question in one of two equally valid ways:

P(at least 1 study obtains significant results)

= 1-P(neither study obtains significant results)

$= 1 - (1-0.05)^2 = 0.0975$

P[1st study significant OR 2nd study is significant]

$= (0.05)+(0.05) -(0.05)^2 = 0.0975$

The experimenter thinks that they are using an alpha=0.05, but they are actually using an alpha =0.0975

223 admissions reasons, 223*12 hypothesis

Austin et al (2006): sifted through health care data for >10 million residents and **223** different reasons for admissions; **12 astrological signs.**

Conclusion: 72 conditions were significantly associated with a particular zodiac sign.

- **This is actually 223*12 hypothesis being tested (~2500 hypothesis)**
- You expect 134 statistically significant associations just due to change so 72 is < 134 (calculated with alpha =0.05)

<u>Bonferroni correction</u>

alpha*=alpha/num of hypothesis = 0.05/(223*12) = 0.0000187

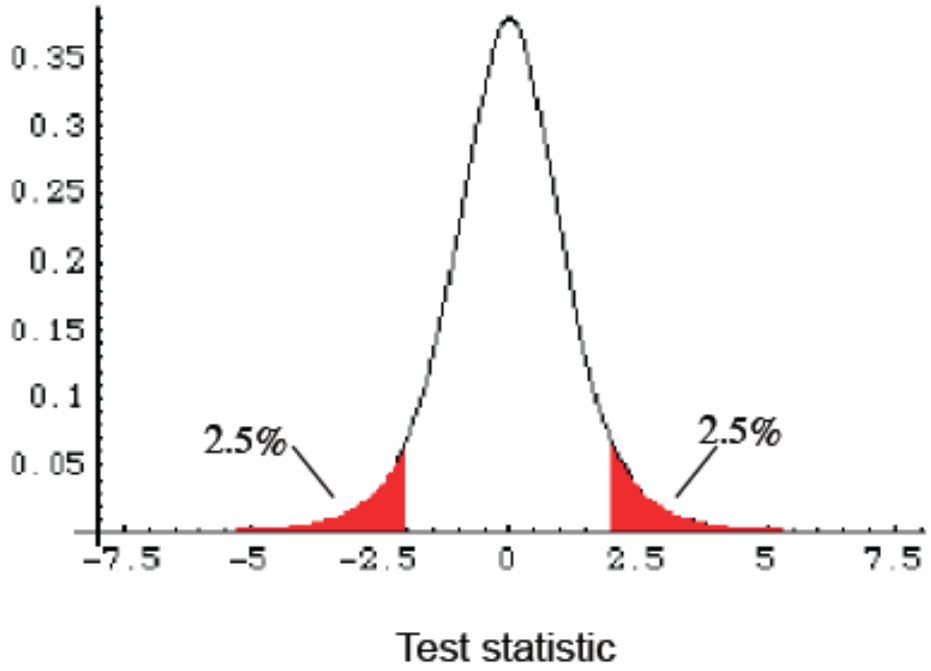
- (GWAS tests hundreds of thousands if not millions at a time)

https://www.sciencedirect.com/science/article/abs/pii/S0895435606001247

One tailed and two tailed tests:

- Most tests are two-tailed tests

- This means that a deviation in either direction would reject the null hypothesis
  - this means that α is divided into $\frac{\alpha}{2}$ on the one side and $\frac{\alpha}{2}$ on the other



One Tailed Tests:

Only used when the other tail is nonsensical

- **Example:**
  - Comparing grades on a multiple-choice test to random guessing

- **Example:**
  - Do daughters resemble their biological fathers?
    - Experiment involves a subject who examines photo of one girl and two adult men and guesses the father
    - If subjects pick father correctly > 0.5 then the hypothesis being tested would FTR
    - Wouldn't make sense that daughters would, on average, resemble their biological fathers less than other men.

- Some parting words & popular misconceptions:
  - FTR does not mean ACCEPT
    - We ***never*** accept the null hypothesis
  - If FTR the null hypothesis, we can conclude that the data is compatible with the hypothesis
- If the result is statistically significant there is a temptation to believe that the effect is large. DO NOT GIVE IN THIS TO ERRONEOUS BELIEF.
  - Nor does it mean that the effect is interesting
  - If the sample size is large (and measurements have little variation) then even inconsequential differences will be significant
- P-values are calculated from the data itself. In contrast, the alpha value is set by the experimenter prior to conducting the experiment. P-values and alpha are related BUT THEY ARE NOT THE SAME!

- Why use hypothesis testing at all?
  - Why don't we skip hypothesis testing since confidence intervals give us similar information **plus** gives us information about the actual magnitude of the parameter?
    - *Main purpose of hypothesis testing is to determine if sufficient evidence has been presented to support a scientific claim*
- Deploy these tools wisely
  - Just because something is statistically significant does not mean it is biologically important or interesting
  - **Almost any null hypothesis can be rejected with a large enough sample**