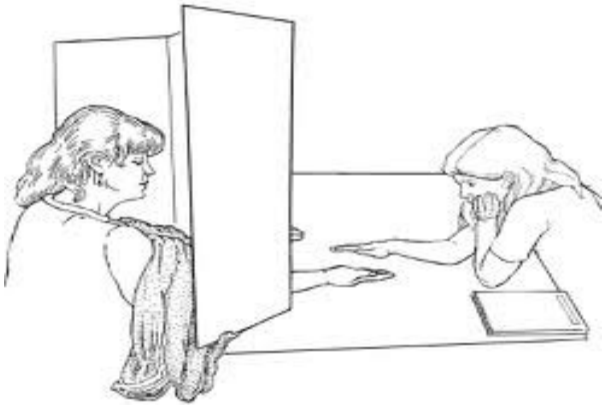
Chapter 26: Tests of Significance

- So far: we draw a sample, and then look at the sample values and make an inference about the population (box).
- Rely on the error estimate, and that the observed values can't be too far off from the box values.
- If the observed value is “off”, we can say that it is just “chance error”....or can we?

Example: Therapeutic Touch

- In 1996 Emily Rosa conducted an experiment where she investigated the claims of Therapeutic Touch practitioners.



Therapeutic touch

- The test consisted of Emily placing one of her hands a few inches above a therapist's right or left hand, as determined by the flip of a coin.
- If the therapist could sense which hand, better than 50 percent of the time, that would support the theory.
- Fifteen TT practitioners got 10 trials each.
- Successful 70 times.

Emily Rosa's experiment

- Total number of tries = 150
- Overall success rate = 70/150=46.7%
- Expected success rate if guessing = 50%!!
- Set up a box: 150 draws from a 0-1 box.
- EV of sample percentage = 50%, Box SD = 0.5
- SE of sample percentage is about 4%
- Observed value=46.7%
- Z-score of 44% = (46.7%-50%)/4 = -0.825
- What is area to right of -0.825? About 80%!!
- Cannot conclude that they were not guessing.

What happened?

- Emily got data, and checked to see if it was “significant”.
- Since she played the role of a skeptic, she assumed they were guessing, and then conducted an experiment to see if the data would prove her wrong.
- It did not.
- She could **not** conclude that they were not guessing.

Tests of Significance

- The observed z-score is called the ***z-statistic.***
- The ***null hypothesis*** says that nothing is going on and any difference that we see between the expected value and the observed value is due to chance.
- The z-statistic measures the difference and it tells us that how many SEs away from the EV is the observed value, **if the null hypothesis were true**.
- If z is very large, 2 possibilities – either null hyp is true and the sample is rare OR the null is false.
- The larger the z, the less plausible that null is true.

P-Value

- The P-value is also called the ***observed significance level.***
- P-value is the probability of the observed statistic or more extreme data IF the null were to be true.
- The larger the P-value, the less extreme the data and there is no evidence to reject the null hypothesis.
- The smaller the P-value, the more extreme the observed data, and evidence **against** the null.
- P-value is NOT the probability that the null is true.

Logic of Significance Tests

- The **null hypothesis** says that there is no effect other than chance.
- The **alternative hypothesis** says that there is an effect other than chance.
- To prove there is some kind of effect, we have to **disprove** the possibility that the result is due the chance. We must show that the evidence leads us to **reject the null hypothesis.**

P-value

- P-value can be thought of as a conditional probability: P(data | null)
- We would like to know P(null | data) *which is not the same!*
- Usually, if the P-value < 5%, the data is called “statistically significant”, as in – it tells us something. But there is some subjectivity here, and no magic cut-off that we can use to reject the null.
- At what temperature would you say “It's cold!”?

Test of significance

1. Set up the null hypothesis in terms of a box model.

2. Pick a test statistic to measure the difference between the data and what is expected by the null hypothesis.

3. Compute the P-value using the observed z-statistic.

4. Decide if the null is to be rejected or not.

Examples

- HS student tosses coin 17,950 times. Gets 9207 heads. Is the coin fair?
  - Set up box. How many draws?
  - EV of # of H?
  - z = $(9207-8950)/SE$

Example

- A study in Punjab, India, reports that in 1993, in a particular hospital, 57% of the 550 live births that year were boys. The baseline for this region is 52% boys. Is this sample proportion of 57% evidence of a higher proportion of male births?