FMS1203S: Randomness in scientific thinking

Week 4

Hypothesis Testing

Neyman-Pearson framework of hypothesis testing

The hypothesis testing procedure

- Two competing hypotheses are formulated, one is called the null hypothesis denoted by H_0, the other is called the alternative hypothesis denoted by H_1.
- A test statistic $T(X)$ is calculated from data X and a critical value c_α is determined.
- The decision rule: if $T(X) \geq c_\alpha$, the null hypothesis H_0 is rejected; otherwise H_0 is not rejected (which does not mean that H_0 is accepted or H_1 is rejected).

Neyman-Pearson framework of hypothesis testing

Type I and Type II errors

- Two errors might occur in the decision:
  - Type I: H_0 is indeed true but it is rejected;
  - Type II: H_0 is false (i.e. H_1 is true) but it is not rejected
- The two hypotheses are formulated in such a way that committing the Type I error is more serious than committing the Type II error.
- The critical value c_α is determined such that the probability of committing the Type I error is controlled at a level α.

Neyman-Pearson framework of hypothesis testing

Significance level and *p*-value

- The significance level is the probability of Type I error:

$$P(T(X) \geq c_\alpha | H_0) \leq \alpha.$$

- If x is the observed value of X, the value $P(T(X) \geq T(x)|H_0)$ is called the p-value, which is the smallest significance level at which H_0 can be rejected based on the observed data.

Power and sample size

- The power of a test at a particular parameter value under the alternative hypothesis is given by 1 minus the Type II error at that particular parameter value.
- While controlling the Type I error, to achieve a give power, a certain sample size is needed.

Readings for next week

Group one: Salsburg, David (2001). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, W.H. Freeman (Chapter 11, Hypothesis testing).

Group two: Copas, John (2005). The Downside of publication. *Significance*, 2(4), pp. 154–157.

Group three: Bland, M. and Altman, D. (2005). Do the left-handed die young? *Significance*, 2(4), pp. 166-170.

Group four: Bland, M. (2009). Keep young and beautiful: evidence for an "anti-aging" product? *Significance*, 6(4), pp. 182-183.

Group five: Cowan, G. (2009). Testing nature to the limit: the Large Hadron Collider *Significance*, 6(4), pp. 154-158.