*Chapter Ten notes* ***for Agri-food research***

Summary from page 304

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a P-value passes a specific threshold.
4. Proper inference requires full reporting and transparency.

*Summary Tests of null hypotheses – default assumptions about statistical models – form a major part of statistical practice.*

Like “this new variety of barley beats the industry benchmark yield”, for example.

*A P-value is a measure of the incompatibility between the observed data and a null hypothesis: formally it is the probability of observing such an extreme result, were the null hypothesis true.*

“Worth another look” is another way to express this.

**Practical significance also matters! Large studies may give rise to results that are statistically but not practically significant.**

**And small studies may produce interesting results, given background knowledge and science, that are not statistically significant.**

*Traditionally, P-value thresholds of 0.05 and 0.01 have been set to declare ‘statistical significance’.*

1. Set up a question in terms of a null hypothesis that we want to check. This is generally given the notation H0.
2. Choose a test statistic that estimates something that, if it turned out to be extreme enough, would lead us to doubt the null hypothesis (often larger values of the statistic indicate incompatibility with the null hypothesis).
3. Generate the sampling distribution of this test statistic, were the null hypothesis true.
4. Check whether our observed statistic lies in the tails of this distribution and summarize this by the P-value: the probability, were the null hypothesis true, of observing such an extreme statistic. The P-value is therefore a particular tail-area.
5. ‘Extreme’ has to be defined carefully – if say both large positive and large negative values of the test statistic would have been considered incompatible with the null hypothesis, then the P-value has to take this into account.
6. Declare the result statistically significant if the P-value is below some critical threshold.

Perhaps the most challenging component in null-hypothesis significance testing is Step 3 – establishing the distribution of the chosen test statistic under the null hypothesis.

In class we are focusing on the t-test, the test statistic “t” and the t-distribution.

*These thresholds need to be adjusted if multiple tests are conducted, for example on different subsets of the data or multiple outcome measures.*

You need to make the thresholds more extreme, to account for false positives you know will occur if you test multiple null hypotheses that are true.

Fisher envisaged the sort of situation seen in the early examples in this chapter, with a single set of data, a single summary outcome measure and a single test of compatibility. But in the last few decades P-values have become the currency of research, with vast numbers appearing in the scientific literature – a study scraped around 30,000 t-statistics and their accompanying P-values from just three years of papers in eighteen psychology and neuroscience journals.

*There is a precise correspondence between confidence intervals and P-values: if, say, the 95% interval excludes 0, we can reject the null hypothesis of 0 at P < 0.05.*

A two-sided P-value is less than 0.05 if the 95% confidence interval does not include the null hypothesis (generally 0). A 95% confidence interval is the set of null hypotheses that are not rejected at P < 0.05.

*Neyman–Pearson theory specifies an alternative hypothesis, and fixes Type I and Type II error rates for the two possible kinds of errors in a hypothesis test.*

Two types of mistake are therefore possible: a Type I error is made when we reject a null hypothesis when it is true, and a Type II error is made when we do not reject a null hypothesis when in fact the alternative hypothesis holds.

This is important – see the Module 2 notes for our class.

*Separate forms of hypothesis tests have been developed for sequential testing.*

Do not worry about this.