STAT 231 — Lecture 25

Bartosz Antczak *Instructor: Michael Wallace* *November 8, 2017*

Last Time

We learned how to calculate a CI for a distribution for unknown σ and μ. For μ, our CI is:

$$\overline{y} \pm a\frac{s}{\sqrt{n}}$$

where a comes from the t_{n-1} distribution and s^2 is the sample variance.
For σ^2, our CI is:

$$\left(\frac{(n-1)s^2}{b}, \frac{(n-1)s^2}{a}\right)$$

where a, b are from the $\chi^2(n-1)$ distribution.

25.1 Chapter 5: Hypothesis Tests

From the bad/good movie example from last time, we see that the mean length for *bad films* is 86.2 minutes and for *good films* 97.7 minutes. From our sample, can we assume that a characteristic about good movies is that they're longer than bad movies? A question like this leads to a new idea: **hypothesis tests**.

25.1.1 Structure of Hypothesis Test

Statistical tests of hypotheses are conducted very similar to a North American criminal trial (i.e., we assume defendant is innocent until enough evidence is shown to plausibly assume they're not).
Our "default" state (i.e., the initial state we assume our hypothesis to be in) is our **null hypothesis**, denoted at H_0. If it is shown that H_0 is not a plausible hypothesis, then we assume the **alternate hypothesis** H_1 to be true.
The **key idea** of hypothesis testing is to collect data, and based on the data we decide how plausible H_0 is.

How to Interpret Plausibility

To answer the question "how close is close", we use a **test statistic** or **discrepancy measure**, defined more formally as: *a function of the data $D = g(Y)$ that is constructed to measure the degree of 'agreement' between the data Y and the null hypothesis H_0.* We usually define D so that $D = 0$ represents the best possible agreement between the data and H_0.

Example 25.1.1. *Tossing a Coin*

Let Y represent the number of heads in 25 trials (tosses). We assume Y to have a binomial distribution with $n = 25$ and θ be the probability of a head. We define:

- H_0: $\theta = 0.5$
- H_1: $\theta \neq 0.5$

For this example, we can consider the discrepancy measure as $D(Y) = |Y - 12.5|$, since $E[Y] = 12.5$ if our coin isn't biased. Small observations of D would suggest no evidence against H_0; whereas large observations of D would suggest the opposite.
So let's suppose that 15 heads come up. We have

$$d(15) = |15 - 12.5| = 2.5$$

What can we deduce from this? To answer that, we must ask what is the probability of observing D greater than or equal to $d = 2.5$ assuming that H_0 is true? We answer this by solving:

$$P(D \geq 2.5; H_0) = P(|Y - 12.5| \geq 2.5)$$

where $Y \sim binomial(25, 0.5)$. We know from STAT 230 how to calculate this:

$$\begin{aligned} P(|Y - 12.5| \geq 2.5) &= P(Y \leq 10) + P(Y \geq 15) && (25.1) \\ &= 1 - P(11 \leq Y \leq 14) && (25.2) \\ &= 1 - \sum_{i=11}^{14} \binom{25}{y} (0.5)^{25} && (25.3) \\ &= 0.4244 && (25.4) \end{aligned}$$

So how do we interpret this? The calculated value is defined as the **p-value**. The $p-$value determines how likely we are to have observed data as extreme or more than the particular observed outcome, assuming H_0 is true.

Interpreting the p-value

Suppose our $p-$value is very small (say for instance, $p = 0.001$), then we can conclude that our observed outcome would occur 1 time out of 1000 assuming H_0 were true, which is highly unlikely, which means that there is evidence suggesting that H_0 is false.
The opposite (a large $p-$value) can be shown to deduce that H_0 is true. From this, we see that:

- p is small $\implies$ evidence against H_0 being true
- p is large $\implies$ no evidence against H_0 being true exists