Probability and stats mop-up unit

- Statistical inference and related concepts (Monday)
- Hypothesis testing (**today**)
- Benford's law (next week Monday)
- Bayesian inference (next week Wednesday)
- Begin graph theory (next week Friday)

Today:

- Finishing MLE and confidence intervals from last time
- Definition of statistical hypothesis testing
- Hypothesis testing scenario

Inference often reduce to numerical estimates of parameters, in the form of either single points or confidence intervals. But not always. In many experimental situations, the conclusion to be drawn is not numerical and is more aptly phrased as a choice between two conflicting theories, or **hypotheses***.*

The process of dichotomizing the possible conclusions of an experiment and then using the theory of probability to choose one option over the other is known as **hypothesis testing***. The two competing propositions are called the* **null hypothesis** *(H_0) and the* **alternative hypothesis** *(H_1).*

How we go about choosing between H_0 and H_1 is conceptually similar to the way a jury deliberate in a court trial. The null hypothesis is analogous to the defendant: Just as the latter is presumed innocent until "proven" guilty, so is the null hypothesis "accepted" unless the data argue overwhelmingly to the contrary. Mathematically, choosing between H_0 and H_1 is an exercise in applying courtroom protocol to situations where the "evidence" consists of measurements made on random variables.

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 343

The **null hypothesis** H_0 is the default or "status quo" theory, such as "there is no relationship between A and B" or "A does not cause B."

The **alternative hypothesis** H_1 is the complement of the null hypothesis.

Based on evidence, we either **reject** the null hypothesis or **retain** it (or **fail to reject** it).

A **type I error** is rejecting H_0 when it is in fact true. A **type II error** is retaining H_0 when H_1 is in fact true.

A car company is researching fuel additives to increase gas mileage. As a pilot study, they send thirty cars fueled with a new additive on a road trip from Boston to Los Angeles. Without the additive, those cars are known to average 25.0 mpg.

Suppose it turns out that the thirty cars average $\overline{x} = 26.3$ *mpg With the additive. What should the company conclude?*

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 344

Rephrase in random variable terminology: Let $x_0, x_1, \ldots x_{29}$ *denote the mileages recorded by each of the cars during the cross-country test run. We assume that the* X_i*s are normally distributed with unknown mean* μ *Furthermore, suppose that prior experience with road tests of this type suggests that* $\sigma = 2.4$ *mpg. Thus*

$$f_X(x;\mu) = \frac{1}{\sqrt{2\pi}(2.4)} e^{-\frac{1}{2}\left(\frac{x-\mu}{2.4}\right)^2}$$

The two competing hypotheses, then, can be expressed as statements about μ*. In effect, we are testing*

$$\begin{array}{llll} H_0: & \mu & = & 25.0 \quad \textit{(Additive is not effective)} \\ & & & \quad \textit{versus} \\ H_1: & \mu & > & 25.0 \quad \textit{(Additive is effective)} \end{array}$$

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 344

Values of the sample mean $\overline{x}$ less than or equal to 25.0 are certainly not grounds for rejecting the null hypothesis; averages a bit larger than 25.0 would also lead to that conclusion because of the commitment to give H_0 the benefit of the doubt.

On the other hand, we would probably view a cross-country average of, say 35.0 mpg as exceptionally strong evidence against the null hypothesis.

Somewhere between 25.0 and 35.0 there is a point—call it $\overline{x}*$—were the credibility of H_0 ends.

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 344

Suppose we set $\overline{x}*$ to 25.25. Is that a good decision rule? Consider the probability of a type I error:

$$
\begin{aligned}
P(\textit{Reject } H_0 \mid H_0 \textit{ is true}) &= P(\overline{X} \geq 25.25 \mid \mu = 25.0) \\
&= P\left(\frac{\overline{X}-25.0}{2.5/\sqrt{30}} \geq \frac{25.25-25.0}{2.4/\sqrt{30}}\right) \\
&= \int_{25.25}^{\infty} \mathcal{N}\left(25.0, \frac{2.4}{\sqrt{30}}\right) dx \\
&= 0.2843
\end{aligned}
$$

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 344–345

Suppose, on the other hand, we set $\overline{x}*$ to 26.5. Is that a good decision rule? Consider the probability of a type I error:

$$
\begin{aligned}
P(\textit{Reject } H_0 \mid H_0 \textit{ is true}) &= P(\overline{X} \geq 26.5 \mid \mu = 25.0) \\
&= P\left(\frac{\overline{X}-25.0}{2.5/\sqrt{30}} \geq \frac{26.5-25.0}{2.4/\sqrt{30}}\right) \\
&= \int_{26.5}^{\infty} \mathcal{N}\left(25.0, \frac{2.4}{\sqrt{30}}\right) dx \\
&= 0.0003
\end{aligned}
$$

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 345

What value should we aim at for $P(\overline{X} \geq \overline{x}* \mid H_0 \text{ is true})$? While there is no way to answer that question definitively or mathematically, researchers who use hypothesis testing have come to the consensus that the probability of a type I error should be somewhere in the neighborhood of .05.

If we set

$$P\left(\overline{X} \geq \overline{x}* \mid H_0 \text{ is true}\right) = .05$$

we then compute

$$\begin{aligned}
P\left(\frac{\overline{X}-25.0}{2.5/\sqrt{30}} \geq \frac{\overline{x}*-25.0}{2.4/\sqrt{30}}\right) &= .05 \\
\frac{\overline{x}*-25.0}{2.4/\sqrt{30}} &= 1.64 \\
\overline{x}* &= 25.718
\end{aligned}$$

Adapted from Larsen and Marx, An Introduction to Mathematical Statistics, 2018 pg 346–347

For next time:

Take quiz on Canvas