One of the most powerful applications of statistics. Allows us to

- Formally devise a research question
- Quantify the extent to which our data supports one conclusion over another

Ex. Suppose a doughnut chain claims that the true average size of their doughnut holes is 2.2cm. Then suppose you buy a box whose doughnut holes are on average 1.7cm.

Is there any reason to doubt the company's figure?

- Is there any reason to doubt the *null hypothesis*? — H_0
- *Alternative hypothesis:* H_A. This simply states that "the null hypothesis is not true".

Ex. Suppose we claim that 50% of STAT 231 students are coffee drinkers.

Let $Y \sim \text{Bin}(25, \theta)$ — θ is the true proportion of STAT 231 students who are coffee drinkers

---

H_0: $\theta = 0.5$

H_A: $\theta \neq 0.5$

Values of y that would support H_0 are values close to 12.5

As y gets further away from 12.5, evidence against H_0 gets stronger

<u>Magnitude of difference:</u> $d = |y - 12.5|$

Larger d = larger "surprise"

<u>Test statistic / discrepancy measure:</u> a function of data $D = g(Y)$ that measures the amount of "agreement" between the data Y and H_0

- Function of Y -> random variable!

Ex. Suppose 10 of the 25 STAT 231 students drank coffee.

Then $d = |10 - 12.5| = 2.5$.

Using the binomial PDF, $P(Y=10) = 0.097$.

$P(D=2.5) = P(Y=10) + P(Y=15) = 2(0.097) =$ around 0.2

Note that we calculate $P(D=d)$ instead of $P(Y=y)$ because we are trying to quantify the difference from the null hypothesis.

Weirder example: Suppose we survey 600 STAT 231 students, with the same assumption that $\theta = 0.5$. Then we would be calculating $P(D=0.5) = P(Y=300) = 0.033$, which is much smaller than in the previous example.

As such, computing $P(D=d)$ itself does not directly quantify evidence against the null hypothesis

---

To test for a greater amount of "surprise", we might want to consider values of y such that $d \geq 2.5$ (or some other choice of bound)

=> $P(D \geq 2.5; H_0) = P(|Y - 12.5| > 2.5)$ where $Y \sim Bin(25, 0.5)$

$$
\begin{aligned}
P(D \geq 2.5) &= P(Y \leq 10) + P(Y \geq 15) \\
&= 1 - P(11 \leq Y \leq 14) \\
&= 0.4244
\end{aligned}
$$

p-values

<u>p-value:</u> $p = P(D \geq d)$

- NOT the probability of observing some data
- Is the probability of observing something *at least* as large / surprising

Steps for hypothesis testing:

1. Specify the null hypothesis
2. Define a test statistic $D(Y)$, for which $d = D(Y)$ is the corresponding observed value
3. Calculate $p = P(D \geq d)$
4. Draw a conclusion

"Moderately large" values like $p \geq 0.20$ are not enough evidence

Small enough values like $p = 0.05$ are ok — depends on researcher

Table 5.1: Guidelines for interpreting $p - values$

$p - value$	Interpretation
$p - value > 0.10$	No evidence against H_0 based on the observed data.
$0.05 < p - value \leq 0.10$	Weak evidence against H_0 based on the observed data.
$0.01 < p - value \leq 0.05$	Evidence against H_0 based on the observed data.
$0.001 < p - value \leq 0.01$	Strong evidence against H_0 based on the observed data.
$p - value \leq 0.001$	Very strong evidence against H_0 based on the observed data.

"Statistical significance" is usually meaningless

Proper way to use this: "statistically significant at the 0.05 significance level"

Hypothesis Testing for Gaussian Distributions

Ex.

- Suppose your favourite coffee shop, Smartbucks, produces lattes that should contain approximately 600 ml of coffee on average
- To test for quality maintenance, the manager at your local branch orders 25 lattes
- For simplicity, let's assume Y models the amount of coffee in a latte where $Y \sim G(\mu, \sigma)$
- How can we start constructing a hypothesis test for whether $\mu = 600$?

Sample mean : $\bar{y}$

$d = |\bar{y} - 600|$; $\bar{D} = |\bar{Y} - 600|$

↳ discrepancy measure : estimator

Large values of d would make us doubt the null hypothesis

Suppose the sample mean is 588. Then, d = 12.

Now suppose we want to compute the p-value for $P(\bar{D} \geq 12)$

$$
\begin{aligned}
&\Rightarrow P(|\bar{Y} - 600| \geq |588 - 600|) \\
&= P(|\bar{Y} - 600| \geq 12) \\
&= P(\bar{Y} \leq 588) + P(\bar{Y} \geq 612) \\
&= 1 - P(588 \leq \bar{Y} \leq 612) \\
&= 1 - P(-12 \leq \bar{Y} - \mu_0 \leq 12)
\end{aligned}
$$

$$= 1 - P\left(\frac{-12}{S/\sqrt{n}} \le \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \le \frac{12}{S/\sqrt{n}}\right)$$

$$\Rightarrow \bar{D} = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1); \quad d = \frac{|\bar{y} - \mu_0|}{S/\sqrt{n}}$$

$$P(D \ge d) = 2 - 2P(D \le d)$$ since T is symmetric

Ex. Suppose a multinational e-commerce company wants to determine whether customers who order an item with one day delivery receive their package, on average, in 24 hours.
They randomly select a sample of 12 purchases made with times of delivery recorded as follows:

20, 22, 24.5, 24.8, 25, 27, 23, 22.5, 24, 25, 23.5, 26

$\bar{y} = 23.9 \qquad s = 1.9$

$$p = P(|D| \ge d) \qquad \text{PCs: } D = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \qquad d = \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}$$

$$D \sim t(n-1)$$

$$\Rightarrow p = 2 - 2P(D \le d)$$

$$= 2 - 2P\left(\frac{\bar{Y} - 24}{S/\sqrt{n}} \le \underbrace{\frac{|23.9 - 24|}{1.9 \cdot \sqrt{12}}}_{0.18}\right)$$

$$= 2 - 2P(T \le 0.18) \qquad k = 11$$

$$= 0.36$$

If $p \ge 0.05$:

we want $$P\left(T \ge \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}\right) \ge 0.05$$

$$= 1 - P\left(T \le \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}\right) \le 0.95$$

$$= P(|T| \leq a) \leq 0.95$$

$$= P\left(\left|\frac{\bar{Y}-\mu_0}{S/\sqrt{n}}\right| \leq a\right) \leq 0.95$$

Then, by the properties of t-distributions, if we want $p \geq 0.05$, we must make sure that the null hypothesis mean falls within the 95% confidence interval

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

Hypothesis Testing for σ^2

Recall:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

let $u = \dfrac{(n-1)s^2}{\sigma_0^2}$; actual quantity

Unlike Gaussian, the chi-squared distribution is NOT symmetric, so we cannot use the same trick

> ▶ A manufacturer that produces optical glass for lenses can only tolerate a certain variability in its thickness, i.e. they want to achieve a requirement of $\sigma = 0.008$. Given a sample size of $n = 25$ and $s = 0.013$, test the hypothesis $H_0 : \sigma = \sigma_0 = 0.008$. Calculate the test statistic U, state its distribution, and compute the corresponding p-value. What can you conclude?

$H_0 : \sigma = \sigma_0 = 0.008$

$H_a : \sigma \neq 0.008$

$p = P(U \geq u)$; $U \sim \chi^2(n-1)$; $u = \dfrac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)$

$$= P\left[U \geq \frac{(n-1)s^2}{\sigma_0^2}\right]$$

$$= P\left[U \geq \frac{(24)(0.013^2)}{0.008^2}\right] = P(U \geq 63.375) \approx 0$$

Very low p-value -> sufficient evidence to reject H0

Ex.

▶ A hospital administrator wants to determine whether the average waiting time for patients in the emergency room is 15 minutes. From a sample of $n = 10$ patients, an assistant records the sample mean and standard deviation of waiting times to be $\bar{y} = 15.7$ and $s = 2.2$. Calculate the test statistic $|T|$, state its distribution, and compute the p-value of the test $H_0 : \mu = \mu_0 = 15$. What can you conclude?

$H_0 : \mu = \mu_0 = 15 \quad ; \quad H_a : \mu \neq 15$

$$\begin{aligned} p &= P(|D| \geq d) \\ &= 2 - 2P(D \leq d) \\ &= 2 - 2P\left(\frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \leq \frac{|\bar{y} - \mu_0|}{s/\sqrt{n}}\right) \quad ; \quad \frac{\bar{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \\ &= 2 - 2P\left(T \leq \frac{|15.7 - 15|}{0.2\sqrt{10}}\right) \\ &= 0.34 \end{aligned}$$

Hypothesis Testing and the Central Limit Theorem

Suppose we want to test $H_0 : \theta_0 = 0.5$ where θ represents the proportion of STAT 231 students who drink coffee.

Doing a survey, suppose we find that 3/25 students drink coffee.

Let Y = number of students who drink coffee.

$Y \sim Bin(n, \theta)$

$E(Y) = n\theta \quad ; \quad sd(Y) = \sqrt{n\theta(1-\theta)}$

Using the CLT : $\dfrac{Y - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}} \sim G(0, 1)$

Then $p = 2 - 2P(Z \le d)$, where $d = \dfrac{Y - n\theta_0}{\sqrt{n\theta_0(1-\theta_0)}}$: $Y = 3$

$\theta_0 = 0.5$

$n = 25$

This works for any distribution — just use calculated expressions for mean and sd

Likelihood Ratio Hypothesis Testing

Recall: Likelihood ratio test statistic

$$\Lambda(\theta_0) = -2\log\frac{L(\theta_0)}{L(\hat{\theta})} = -2\log(R(\theta_0)) \sim \chi^2(1) \text{ for large } n$$

Values of θ_0 closer to 1 -> more plausible

Closer to 0 -> more implausible

$R(\theta_0)$ is very small <=> $-2\log(R(\theta_0))$ is very large

Calculating the p-Value with the Likelihood Ratio Test Statistic

- The test statistic
$$\Lambda(\theta_0) = -2\log\frac{L(\theta_0)}{L(\hat{\theta})} = -2\log R(\theta_0)$$
is larger for data that are more surprising if $H_0 : \theta = \theta_0$ is true
- Since $\Lambda(\theta_0) \sim \chi^2(1)$ if H_0 is true, then we compute
$$\text{p-value} = P(W \ge \lambda(\theta_0)),\ W \sim \chi^2(1)$$

Hypothesis Testing with the Likelihood Ratio Test Statistic: Summary

- We now have the following process for using the likelihood ratio statistic when testing $H_0 : \theta = \theta_0$:
  1. Propose a model for your data and derive expressions for $L(\theta), \hat{\theta}$
  2. From a sample of data, compute the observed value of $\hat{\theta}$
  3. Compute the observed value of $\lambda(\theta_0)$, i.e. $-2\log\frac{L(\theta_0)}{L(\hat{\theta})} = -2\log R(\theta_0)$
  4. Calculate $2[1 - P(Z \le \sqrt{\lambda(\theta_0)})]$

$$p = P(W \ge -2\log(R(\theta_0)) \quad : W \sim \chi^2(1)$$

$$= 2 - 2P\left[Z \le \sqrt{-2\log(R(\theta_0))}\right] \quad : Z \sim G(0,1)$$