Chapter 22: What is a Test of Significance?

Thought Question

Assume that the statement “If it’s Saturday, then it’s the weekend” is true. Which of the following statements will also be true?

- If it’s the weekend, then it’s Saturday.
- If it’s not the weekend, then it’s not Saturday.

Assume that the statement “If A is true, then B is true” is true. Which of the following statements will also be true?

- If B is true, then A is true.
- If B is false, then A is false.
Reasoning of Statistical Tests of Significance

**Goal:** to determine how consistent observed sample data is with a claimed value of the population proportion $p$.

**Terminology**

- **Significance test:** asks if sample data gives good evidence *against* a claim.

**Example: Coffee**

People of taste are supposed to prefer fresh-brewed coffee to the instant variety. But perhaps many coffee drinkers just need their caffeine fix. A skeptic claims that coffee drinkers can’t tell the difference. Let’s do an experiment to test this claim.

Each of 50 subjects tastes two unmarked cups of coffee and says which he or she prefers. One cup in each pair contains instant coffee; the other, fresh-brewed coffee. The statistic that records the result of our experiment is the proportion $\hat{p}$ of the sample who say they like the fresh-brewed coffee better. We find that 36 of our 50 subjects choose the fresh coffee.

*Calculate $\hat{p}$.*

\[ \hat{p} = \frac{36}{50} = 0.72 \]

How strong is the evidence from the sample that the majority of the population prefer fresh coffee?  

Claim: The skeptic claims that coffee drinkers can’t tell fresh from instant, so that only half will choose fresh-brewed coffee, i.e., $p = 0.5$.

*Assume that the claim $p = 0.5$ is true. What is the sampling distribution of the sample proportion $\hat{p}$ if we repeat the experiment many times?*

The sampling distribution of $\hat{p}$ is Normal with mean $= p = 0.5$ and standard deviation $= \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.5)(0.5)}{50}} \approx 0.0707$. 
How common is a $\hat{p}$ value of 0.72 or larger if $p = 0.5$?

Very unusual.

How common is a $\hat{p}$ value of 0.56 or larger if $p = 0.5$?

Quite common.

What is the probability that a sample gives a $\hat{p}$ value of 0.72 or larger?

0.001

What is the probability that a sample gives a $\hat{p}$ value of 0.56 or larger?

0.20

Does a $\hat{p}$ value of 0.56 provide strong evidence against the claim that $p = 0.5$?

No - this outcome or more extreme will happen 20% of the time if $p = 0.5$, which is not that unlikely.

Does a $\hat{p}$ value of 0.72 provide strong evidence against the claim that $p = 0.5$?

Yes - this outcome or more extreme will happen 0.1% of the time if $p = 0.5$, which is very unlikely.
Hypotheses and $P$-values

**Structure of the Hypotheses**

The claim can take two forms.

- The proportion has *not* changed from a previously determined value $p_0$.

- The proportion *has changed* from a previously determined value $p_0$.

**Null Hypothesis $H_0$**

The claim being tested in a test of significance. The test is designed to assess the strength of the evidence against $H_0$. Usually $H_0$ is a statement of “no effect” or “no difference”.

**Alternative Hypothesis $H_a$**

The name of the statement suspected to be true instead of $H_0$.

**The Statistical Evidence About the Hypothesis: $P$-value**

The probability, assuming that $H_0$ is true, that the sample outcome would be as extreme or more extreme than the actually observed outcome.
Steps to Testing Significance

1. State the null and alternative hypotheses, $H_0$ and $H_a$.

2. Identify the sampling distribution of $\hat{p}$ when $H_0$ is true.

3. Calculate the value of $\hat{p}$ on which the test will be based.

4. Find the $P$-value for the observed value of $\hat{p}$.

5. Interpret what the $P$-value indicates about additional samples.

6. Draw a conclusion based on the $P$-value.
Example: Coins

In 400 flips, a coin lands heads 228 times. Is the coin a fair coin?

Hypotheses:

$H_0: p = 0.5 \quad H_a: p \neq 0.5$

Distribution of $\hat{p}$ when $H_0$ is true:

The sampling distribution is normal with mean $\mu = p = 0.50$ and standard deviation

$$
\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.5)(0.5)}{400}} = 0.025.
$$

Observed value of $\hat{p}$:

$\hat{p} = 228/400 = 0.57$

$P$-value:

standard value $= (0.57 - 0.5)/0.025 = 2.8$

The $P$-value is $1 - 0.9974 = 0.0026$.

Interpretation:

The probability, assuming that $p = 0.5$, of observing a sample proportion at least as large as 0.57 is 0.0026.

Conclusion:

This value of $\hat{p}$ provides strong evidence against the claim that the coin is fair.
Reasoning of Statistical Tests of Significance

**Basic Statistical Belief**

The sample is not unusual but is common.

**Two Types of Statistical Evidence**

- Large $P$-value
  - indicates the statistic ($\hat{p}$) or worse occurs frequently; i.e., if $H_0$ is true, the observed sample would be a common sample;
  - is consistent with the basic statistical belief;
  - does not provide evidence against $H_0$;
  - conclude that there is not sufficient statistical evidence to reject $H_0$; i.e., do not reject $H_0$.

- Small $P$-value
  - indicates the statistic ($\hat{p}$) or worse rarely occurs in a random sample; i.e., if $H_0$ is true, the sample would be a rare sample (If A is true, then B is true);
  - does not agree with the basic statistical belief:
    * Two possible scenarios
      1. $H_0$ is true and the sample is an unusual one;
      2. The sample is not unusual (B is false). Then $H_0$ is probably false (A is false).
  - provides evidence against $H_0$;
  - conclude there is sufficient statistical evidence to reject $H_0$; i.e., the evidence indicates $H_a$ is true.
Example: Poll

A politician believes he can vote any way he wants on a controversial bill without suffering any fallout provided he has the support of 40% (or more) of his electorate. Suppose a random sample of 225 constituents reveals support from only 81 voters. Is this significant evidence to indicate that the politician should be concerned about political fallout from this vote?

\[ H_0: p = 0.40 \quad H_a: p < 0.40 \]

**Distribution of \( \hat{p} \) when \( H_0 \) is true:**

The sampling distribution is normal with mean \( \hat{p} = 0.40 \) and standard deviation \( \sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.4)(0.4)}{225}} = 0.03266 \).

**Observed value of \( \hat{p} \):**

\( \hat{p} = \frac{81}{225} = 0.36 \)

**P-value:**

standard value = \( \frac{0.36 - 0.4}{0.03266} = -1.22 \)

The \( P \)-value = 0.11.

**Interpretation:**

The probability, assuming that \( p = 0.4 \), of observing a sample proportion no larger than 0.36 is 0.11.

**Conclusion:**

This value of \( \hat{p} \) does not provide strong evidence against the claim that 40% of the electorate support the politician. The politician should be not be too concerned about political fallout from this vote.
Example: DVDs

A manufacturer of DVDs believes that, under adequate quality control, no more than 6% of the DVDs should be returned as faulty. For a random sample of 250 sales of these DVDs, it was found that 22 were returned as faulty. Determine the $P$-value for a test of the hypothesis that the percentage of all DVDs returned as faulty is at most 6% and state what this indicates about the percentage of faulty discs under this production process.

$H_0: \ p = 0.06 \quad \quad H_a: \ p > 0.06$

Distribution of $\hat{p}$ when $H_0$ is true:

The sampling distribution is normal with mean $= p = 0.06$ and standard deviation

$$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.06)(0.94)}{250}} = 0.015.$$

Observed value of $\hat{p}$:

$$\hat{p} = \frac{22}{250} = 0.088$$

$P$-value:

standard value $= (0.088 - 0.06)/0.015 \approx 1.87$

The $P$-value $= 1 - 0.97 = 0.03$.

Interpretation:

The probability, assuming that $p = 0.06$, of observing a sample proportion at least as large as 0.088 is 0.03.

Conclusion:

This value of $\hat{p}$ provides strong evidence against the claim that no more than 6% of the DVDs are faulty.