Section 9.1 - Significance Tests: The Basics (Part 1) (pp 529-537)

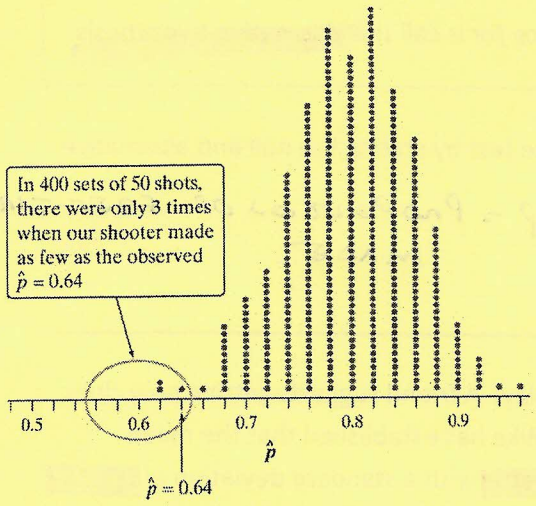
In the last chapter, we used data from random samples to *estimate population* parameters. In this chapter, we will use data from random samples and randomized experiments to *test a claim about a population* parameter. This type of inference is called a **significance test**. A significance test is a formal procedure for comparing observed data with a claim (called a **hypothesis**) whose truth we want to assess. The claim will be a statement about the population parameter.

The Reasoning of Significance Tests

Statistical tests deal with claims about a population. Tests ask if sample data give good evidence *against* a claim. Let's look at an example.

Suppose a basketball player is *claimed* to be an 80% free-throw shooter. Suppose he shoots 50 free-throw and makes 32 of them. What is the sample proportion $\hat{p}$? $\frac{32}{50} = \boxed{0.64}$

What can we conclude about the player's claim based on the sample data? PROBABLY SHOULD MAKE MORE SHOTS



A simulation was run to simulate 400 sets of 50 shots assuming the player really is an 80% free-throw shooter. This dotplot shows the results.

Based upon the distribution of $\hat{p}$, what would you now say about the player's claim and the result of the sample proportion of 0.64? Justify.

THERE IS STRONG EVIDENCE THAT HE IS NOT AN 80% FREE-THROW SHOOTER SINCE ONLY 3 TIMES OUT OF 400 DID WE HAVE $\hat{p}$ AS LOW AS 0.64.

There are two possible explanations for that the fact that our player only made $\hat{p}$ = 32/50 = 0.64 of his free-throws:

1. THE PLAYER'S CLAIM IS CORRECT & BY BAD LUCK, A VERY UNLIKELY OUTCOME OCCURRED.

2. THE POPULATION PROPORTION p IS ACTUALLY LESS THAN 0.80, SO THE SAMPLE RESULT IS NOT AN UNLIKELY OUTCOME.

We are going to learn very specific vocabulary in order to conduct significance testing but the basic idea is: *an outcome that would rarely happen if a claim is true is good evidence that the claim is not true.*

Stating Hypotheses

The first step in any significance test is *stating the hypothesis (claim) that we want to test.* This is called the **null hypothesis**. This is the claim we seek evidence *against*. The null hypothesis is abbreviated H_0.

The null hypothesis is typically a statement of *"no difference."*

- For the free-throw example: $H_0: p = 0.80$ USUALLY AN "=" SIGN

The claim we hope or suspect to be true instead of the null hypothesis is called the **alternative hypothesis**. It is abbreviated H_A. ALSO H_a OR H_1

- In our example: $H_A: p < 0.80$ DOES NOT HAVE AN "=" SIGN.

> The claim tested by a statistical test is called the **null hypothesis (H_0)**. The test is designed to assess the strength of the evidence *against* the null hypothesis. Often the null hypothesis is a statement of "no difference."
>
> The claim about the population that we are trying to find evidence *for* is call the **alternative hypothesis (H_A)**.

To sum up the free-throw example, we would begin a significance test by stating the null and alternative hypotheses:

$H_0: p = 0.80$

$H_A: p < 0.80$

p = PROPORTION OF FREE-THROWS MADE.

Example - Mike is an avid golfer who would like to improve his play. A friend suggests getting new clubs and lets Mike try out his 7-iron. Based on years of experience, Mike has established that the mean distance that balls travel when hit with his old 7-iron is $\mu = 175$ yards with a standard deviation of $\sigma = 15$ yards. (All golfers probably know this, right?) He is hoping that this new club will make his shots more consistent (less variable), so he goes to the driving range and hits 50 shots with the new 7-iron.

(a) Describe the parameter of interest in this setting. PARAMETER = POP STD DEV σ OF DISTANCE W/ NEW 7-IRON.

(b) State appropriate hypotheses for performing a significance test.

$H_0: \sigma = 15$
$H_A: \sigma < 15$

DISCUSS (H_A 1st)

It should be noted that *hypotheses should express the hopes or suspicions we have **before** we see the data.* In test situations, build your hypotheses based on wording of the question not by looking at the data set and going with what you think is true.

Also, *hypotheses are in terms of population parameters not sample statistics*. We are going to using sample statistics to test the hypotheses concerning the parameters.

The alternative hypothesis is **one-sided** if it states that a parameter is *larger than* the null hypothesis value or if it states that a parameter is *smaller than* the null value.

It is **two-sided** if it states that the parameter is *different from* the null hypothesis value (it could be larger or smaller)

Application - In each of the following (a) describe the parameter of interest; (b) state the appropriate hypotheses for a significance test.

(a) According to the web site sleepdeprivation.com, 85% of teens are getting less than 8 hours sleep a night. Janine wonders whether this result holds true in her large high school. She asks an SRS of 100 students at her school how much sleep they got on a typical night. In all, 75 of the responders said less than 8 hours.

(A) p = PROP. OF STUDENTS WHO GET LESS THAN 8 HRS/NIGHT.

(B) $H_0: p = 0.85$
$H_A: p \neq 0.85$

(b) As part of the 2010 census marketing campaign, the U.S. Census Bureau advertised "10 questions, 10 minutes -- that's all it takes." On the census form itself, we read, "The U.S. Census Bureau estimates that, for the average household, this form will take about 10 minutes to complete, including the time for reviewing instructions and answers. " We suspect that the actual time it takes to complete the form may be longer than advertised.

(A) μ = MEAN AM'T OF TIME THAT IT TAKES TO COMPLETE CENSUS FORM

(B) $H_0: \mu \leq 10$
$H_A: \mu > 10$

Interpreting *P*-Values

To determine whether random chance is a plausible explanation for why the sample statistic is different from the hypothesized parameter value, *we calculate a* ***P-value***.

The ***P*-value** is a measure of inconsistency between the hypothesized value of a population parameter and the observed sample statistic.

It is the probability , assuming the H_0 is true, of the statistic (such as $\hat{p}$ or $\bar{x}$) would taking on a value as extreme or more extreme than the one actually observed.

In other words, the *P*-value measures how likely it is to get a value of the sample statistic at least as extreme as the observed sample statistic in the direction specified by the alternative hypothesis by *random chance* alone, given that the hypothesized parameter is correct.

If the *P*-value is **large**, then it is plausible that the difference between the sample statistic and the hypothesized parameter value could have been due to chance alone.

However, if the *P*-value is **small**, then we can essentially rule out random chance as a plausible explanation and conclude that the hypothesized parameter value is incorrect. (The smaller the *P*-value, the stronger the evidence is *against* the null hypothesis.)

"P VALUE LOW H_0 MUST GO."

Example (cont) - When Mike was testing the new 7-iron, the hypotheses were

H_0: $\sigma = 15$

H_A: $\sigma < 15$

where σ is the true standard deviation of the distances over which Mike ~~hits golf balls~~ using the new 7-iron. Based on 50 shots with the new 7-iron, the standard deviation was s_x = 10.9 yards.

A significance test using the sample data produced a *P*-value of 0.002.

(a) Interpret the *P*-value in this context.

IF THE TRUE STD DEV IS 15 yds, THEN THERE IS AN APPROXIMATE PROB. OF 0.002 THAT THE SAMPLE STD DEV WOULD BE 10.9 YDS OR LOWER BECAUSE OF CHANCE ALONE.

(b) Do the data provide convincing evidence against the null hypothesis? Explain.

YES. SINCE THE P-VALUE IS LOW, RANDOM CHANCE IS NOT A PLAUSIBLE REASON THAT THE STD DEV WAS LOWER THAN 15 YDS. THUS THERE IS CONVINCING EVIDENCE THAT THE TRUE STD DEVIATION FOR THE NEW 7-IRON IS SMALLER.

Application - Complete questions 1 and 11 on p. 546.

(1) $H_0: p = 0.12$ p = PROPORTION OF LEFTIES @ SCHOOL.

$H_A: p \neq 0.12$

(11) 16/100 = 0.16 P-VALUE = 0.2184

(A) IF PROP OF LEFTIES IS REALLY 0.12, THERE IS A 21.84% CHANCE OF FINDING A SAMPLE OF 100 PEOPLE WITH A VALUE OF $\hat{p}$ THAT IS AS FAR FROM 0.12 AS THE SAMPLE VALUE IN EITHER DIRECTION.

(B) NO, SOMETHING THAT HAPPENS OVER 20% OF THE TIME JUST BY CHANCE ~~WHEN H_0 IS TRUE~~ IS NOT STRONG EVIDENCE.

Statistical Significance

The final step in performing a significance test is to draw conclusions about the competing claims you were testing. We will make one of two decisions based on the strength of the evidence against the null hypothesis (and in favor of the alternative hypothesis)

Reject H_0 or *Fail to reject H_0*

Example - In the 7-iron example, the estimated *P*-value of 0.002 is strong evidence against the null hypothesis H_0: $\sigma = 15$. For that reason, we would *reject the null hypothesis H_0 in favor of the alternative* H_A: $\sigma < 15$. It appears that variation with the new driver is smaller.

In summary, our conclusion in a significance test comes down to

"P-VALUE LOW — H_0 MUST GO."

***P*-value small → reject H_0 → conclude H_A (in context)**

***P*-value large → fail to reject H_0 → cannot conclude H_A (in context)**

There is no rule for how small a *P*-value must be to reject the null hypothesis -- it is a matter of judgment and depends on the circumstances. But we can compare the *P*-value with a fixed value that we regard as decisive.

This value is called the **significance level** and we use the Greek letter α denote it. An $\alpha = 0.05$ is considered **statistically significant** and an $\alpha = 0.01$ is considered **highly statistically significant.**

If the *P*-value is smaller than alpha, we say the data are **statistically significant at level α.** In that case, we reject the null hypothesis H_0 and conclude that there is convincing evidence in favor of the alternative hypothesis H_A.

P-value $< \alpha \rightarrow$ reject $H_0 \rightarrow$ conclude H_A (in context)

P-value $> \alpha \rightarrow$ fail to reject $H_0 \rightarrow$ cannot conclude H_A (in context)

The conclusion to a significance test should always include **three** components:

(1) an explicit comparison of the *P*-value to a stated significance level *OR* an interpretation of the *P*-value as a conditional probability;
(2) a decision about the null hypothesis: reject or fail to reject H_0;
(3) an explanation of what the decision means in context.

Application - Complete questions 3 and 13 on pp. 546-547.

(3) $H_0: \mu = 115$
$H_A: \mu > 115$

μ = MEAN SCORE ON THE SSHA FOR STUDENTS ~~AT~~ LEAST 30 YRS OLD AT THE TEACHERS COLLEGE.

(13) $\bar{x} = 125.7$ P-VALUE = 0.0101
$S_x = 29.8$

(A) IF THE MEAN SCORE FOR THE SSHA FOR OLDER STUDENTS AT THIS SCHOOL IS REALLY 115, THERE IS A 1.01% CHANCE OF FINDING A SAMPLE OF 45 STUDENTS W/A MEAN SCORE OF AT LEAST 125.7.

(B) REJECT THE NULL IF $\alpha = 0.05$; SINCE $0.0101 < 0.05$
FAIL TO REJECT NULL IF $\alpha = 0.01$
SINCE $0.0101 > 0.01$

HW: Read pp. 529-535; complete problems 2, 12, 4, 14.