Ch. 17 – Hypotheses and Test Procedures

Def'n: A null hypothesis is a claim about a population parameter that is assumed to be true until it is declared false.

An alternative hypothesis is a claim about a population parameter that will be true if the null hypothesis is false.

In carrying out a test of H_0 vs. H_A, the hypothesis H_0 is "rejected" in favour of H_A only if sample evidence strongly suggests that H_0 is false. If the sample does not contain such evidence, H_0 is "not rejected" or you "fail to reject" it.
NEVER "accept" H_0 or H_A…for different reasons.

Ex17.1) H_0: $\mu = 2.8$ H_A: $\mu \neq 2.8$

↑ ↑

pop'n characteristic hypothesized value or "claim"

Def'n: A two-tailed test has "rejection regions" in both tails.

A one-tailed test has a "rejection region" in one tail.

A lower-tailed test has the "rejection region" in the left tail.

An upper-tailed test has the "rejection region" in the right tail.

Ex17.2)

a) H_0: $\mu = 15$ H_A: $\mu = 15$
b) H_0: $\mu = 123$ H_A: $\mu = 125$
c) H_0: $\mu = 123$ H_A: $\mu < 123$
d) H_0: $\mu \geq 123$ H_A: $\mu < 123$
e) H_0: $p = 0.4$ H_A: $p > 0.6$
f) H_0: $p = 1.5$ H_A: $p > 1.5$
g) H_0: $\hat{p} = 0.1$ H_A: $\hat{p} \neq 0.1$

	Two-Tailed Test	**Lower-Tailed Test**	**Upper-Tailed Test**
Sign for H_0	=	= or ≥	= or ≤
Sign for H_A	≠	<	>
"Rejection region"	In both tails	In the left tail	In the right tail

Ex17.3)

Def'n: A test statistic is the function of the sample data on which a conclusion to reject or fail to reject H_0 is based. For example, Z and t are test statistics.

The *P-value* is a measure of inconsistency between the hypothesized value for a pop'n characteristic and the observed sample. Assuming H_0 is true, the *P-value* can be defined as the probability of obtaining a test statistic value at least as inconsistent with H_0 as what actually resulted. Keep in mind that we *want* to be inconsistent with H_0 to reject it. **Thus, the smaller the P-value, the more likely we reject H_0.**
The significance level (denoted by α) is a number such that we reject H_0 if the P-value is less than or equal to that number.

The "significance level approach":

$$\text{reject } H_0 \text{ if } p\text{-value} \le \alpha$$
$$\text{do not reject } H_0 \text{ if } p\text{-value} > \alpha$$

Common choices for α are 0.01, 0.05, and 0.1, depending on the nature of the test.

PROBLEMS:
a) If you're comparing to $\alpha = 0.05$, are the P-values 0.045 and 0.000 001 "different"?
b) If we use a "cut-off" like $\alpha = 0.05$, does it make sense to conclude differently between P-values of 0.049 and 0.051?
Solution: ALWAYS report your P-value! That way a reader may draw their own conclusions. Moreover, use the "judgment approach" for rejection. Here, there's a tendency of avoiding "cut-off" points and going toward some "acceptable" guidelines:

$0.01 > P\text{-value} > 0$ → strong to convincing evidence against H_0
$0.05 > P\text{-value} > 0.01$ → moderate to strong evidence against H_0
$0.10 > P\text{-value} > 0.05$ → suggestive to moderate evidence against H_0, yet inconclusive
$1 > P\text{-value} > 0.1$ → weak evidence against H_0

Steps of a Significance Test:
1. *Assumptions*: Specify variable/parameter. What assumptions apply? Do they hold?
2. *Hypotheses*: State the null/alternative hypotheses. (Select α for the test.)
3. *Test statistic*: Use the appropriate formula for the given situation.
4. *P-value*: Determine an exact value or range.
5. *Conclusion*: Make a decision and conclude within the context of the problem.

Significance Tests About Proportions
Recall the 3 rules from Chapter 15. They collectively imply that when n is large,

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} \text{ is approximately } N(0, 1).$$

Assumptions: categorical variable, random sample, $np_0 \ge 15$ and $n(1 - p_0) \ge 15$.

Ex17.4) Recall the survey of random people from Ex16.3). Suppose p is 0.240. Does the sample disprove this claim? Test the claim using both approaches and $\alpha = 0.01$.

(Note that one-tailed tests with this example were also discussed in class.)

Summary

Each hypothesis test should include:
- clear null & alternative hypotheses
- assumptions (stated and checked)
- appropriately-used test statistic (show the formula, identify its distribution)
- calculation of both the test statistic and P-value (exact or range)
- conclusion in the context of the problem

Ch. 18 – Errors in Hypothesis Testing

In any hypothesis test, there is 1 of 2 choices: reject or not reject. There is also 1 of 2 choices as the test applies to reality: H_0 is true or H_0 is false.

		Actual situation	
		H_0 is true	H_0 is false
Decision	Do not reject H_0	Correct Decision	Type II or β error
	Reject H_0	Type I or α error	Correct Decision

Def'n: A Type I error occurs when a true null hypothesis is rejected. The value of α represents the prob. of committing this type of error; that is,

$$\alpha = P(H_0 \text{ is rejected} \mid H_0 \text{ is true})$$

The value of α represents the *significance level* of the test.

A Type II error occurs when a false null hypothesis is not rejected. The value of β represents the prob. of committing a Type II error; that is,

$$\beta = P(H_0 \text{ is not rejected} \mid H_0 \text{ is false})$$

The value of $1 - \beta$ is called the *power of the test*. It represents the probability of NOT making a Type II error. Or, power = $P(\text{rejecting } H_0 \mid H_0 \text{ is false})$.

Ex18.1) H_0: "innocent until proven guilty"

		Actual situation	
		Innocent	Guilty
Jury's decision	Find not guilty		
	Find guilty		

These 2 errors are dependent. For a fixed sample size, lowering α will raise β and vice versa. Decreasing α and β simultaneously requires increasing the sample size. Further information on the errors are not covered in this course.