Hypothesis Testing

Links: Math 3339

(lecture 15, hypothesis testing for population mean, and population proportion)

What is a Hypothesis Test?

To assess the evidence provided by data (the data is from the sample) about some claim concerning a population, the reasoning is based on what would happen if we repeated the sample or experiment many times.
The **test of significance** answers the question: "Is the observed effect due to chance?"

Components of a Significance Test

- Null (typically denoted by H_0) and alternative hypothesis (typically denoted by H_a)
- Rejection region
- Test Statistic (either use *z-statistic* or *t-statistic*)
- P-value (which is a probability)
- Decision of test (to reject or not reject the null hypothesis H_0)
- Conclusion in the context of the test

Null Hypothesis of Significance Tests

A statement that is assumed to be true, we assume *"no effect"* or *"no difference"* for the parameter tested, it is denoted by H_0.
For a significant test of the mean, the null hypothesis is always equal to some value of what we assume the mean to be. The null hypothesis is always $H_0 : \mu = \mu_0$, where μ_0 is some value that is assumed to be the true mean.

Alternative Hypothesis of Significance Tests

The alternative hypothesis is the statement we hope or suspect is true *instead of* the null hypothesis, it is denoted by H_a. The test of significance is designed to assess the strength of the evidence *against* the null hypothesis.

There are three possible ways that we would want to test against the null hypothesis

1. Test to prove the mean is really lower than what is assumed, this is called the *left-tailed test*. $H_a : \mu < \mu_0$
2. Test to prove that the mean is greater than what is assumed, this is called the *right-tailed test*. $H_a : \mu > \mu_0$
3. Test to prove that the mean is not equal (either higher or lower) than what is assumed, this is called a *two-tailed test*. $H_a : \mu \neq \mu_0$

Decision

Since there are only two hypotheses, there are only two possible decisions. **Reject** the null hypothesis in favor of the alternative hypothesis (RH0), or **Fail to reject** the null hypothesis (FTRH0). However, we **never** say that we accept the null hypothesis.

Errors

A **type I error** is when you reject the null hypothesis when in fact it is true, this is the worst conclusion to make so we try to control for this error.
$P(Type\ I\ error) = \alpha$ (alpha is "the level of significance", the probability of making a type I error). By predetermining α (usually 0.05), we are saying that we make this type I error only 5% of the time.
If we do not reject H_0 then is is a **type II error**.

(from slide 34 of lecture 15 slides)

In the same way we can make an error in our decisions.

Our Decision	Correct Condition	
	H_0 is true	H_0 is false
Reject H_0	Type I Error	Correct
Fail to reject H_0	Correct	Type II Error

Thus by determining α, the level of significance, we try to control for the Type I error.

P(Type I error) = α
P(Type II error) = β
Power = $1-\beta$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Test Statistic

A value calculated based on sample data and the type of distribution, the test statistic is used to measure the difference between the data and what is expected on the null hypothesis.

For a hypothesis about the population mean if σ (the population standard deviation) is known, use the the *z-test statistic*

$z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$

For a hypothesis about the population if σ is not known, use the *t-test statistic*

$t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$

(using sample standard deviation instead)

In general, the test statistic is the difference between the sample mean and the mean that is expected (i.e. the population mean) divided by the standard error.

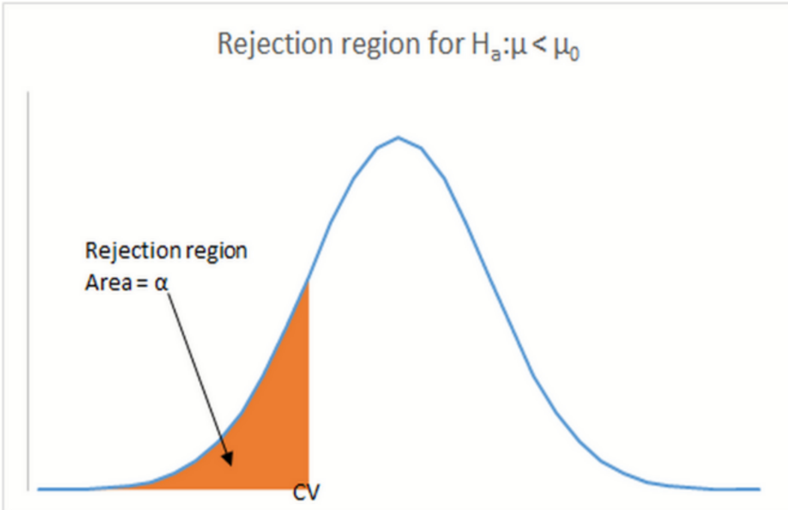
How to Make the Decision

Rejection Region

A rejection region is the set of values for which the test statistic leads to a rejection of the null hypothesis. The *critical value* is the boundary of the rejection region, based on the alternative hypothesis and the level of significance α. (We use the critical value to set up the rejection region).

(see slides 41 to 43 of lecture 15 slides for graph examples)

the level of significance (or α) is the shaded area under the curve (rejection region), to find the critical value you do `qnorm(alpha, 0, 1)`.



In this example the alternative hypothesis is left-tailed, so if the test statistic were to be less than the critical value it would be in the rejection region (and thus the null hypothesis would be rejected).
Same would apply for right-tailed (test statistic greater than critical value is rejection of H_0) where you would do `qnorm(1-alpha)` or (`qt(1-alpha, n-1)` if [t-curve]). For the two-tailed test where you'd have to do `qnorm(alpha/2, 0, 1)` and `qnorm(1-alpha/2, 0, 1)` to find both critical values (same for `qt()` but with degrees of freedom `n-1`).
(or for two tailed $\pm$`qnorm()` or $\pm$`qt()` with $1-\alpha$)

P-values
One approach is to announce how much evidence against H_0 we will require to reject it. We compare the *P-value* with a level that says "this evidence is strong enough." This decisive level is the **significance level** denoted by α, which is usually given, but if it is not we assume that $\alpha = 0.05$. When we compare the *P-value* to α, we have to choose from these two decisions

- **Reject** H_0 if the *P-value* is as small or smaller than α, thus we say that the data are statistically significant at level α (i.e. if $P - value \leq \alpha$ then reject H_0).
- **Do not reject** H_0 if the *P-value* is larger than α.

The P-value is the probability (assuming H_0 is true) that the test statistic would take a value as extreme or more extreme (in the way of H_a) than that actually

observed. The smaller the P-$value$, the stronger the evidence **against** H_0 provided by the data.
To calculate the P-$value$ for the mean we will use the sampling distribution of the means which by the central limit theorem is the Normal distribution. Hence this probability is the same as the area under a normal curve depending on the alternative hypothesis.

if $H_a : \mu < \mu_0$, then P-$value$ = $P(Z < \ test \ statistic)$
if $H_a : \mu > \mu_0$, then P-$value$ = $P(Z > \ test \ statistic)$
if $H_a : \mu \neq \mu_0$, then

$$\begin{aligned} &P(Z < -test\, statistic \text{ or } Z > +test\, statistic) \\ &\quad = P(Z < -test\, statistic) + P(Z > +test\, statistic) \\ &\quad = 2P(Z > |test\, statistic|) \text{ (or } 2P(Z < -test\, statistic)) \end{aligned}$$

If we use t as the test statistic replace Z with T.

Assumptions of the Tests

For a z-$test$:

1. An SRS of size n from the population.
2. Known population standard deviation, σ.
3. Either a Normal population or a large sample ($n \geq 30$)

For a t-$test$:

1. An SRS of size n from the population.
2. Unknown population standard deviation.
3. Either a Normal population or a large sample ($n \geq 30$)

Steps of a Significance Test

1. Check assumptions (What kind of test?)
2. State the null and alternative hypothesis.
3. Graph the rejection region, labeling the critical values.
4. Calculate the test statistic.
5. Find the p-$value$. If this answer is less than the significance level, α, we can reject the null hypothesis in favor of alternative hypothesis.

6. Give your conclusion using the context of the problem. When stating the conclusion give results with a confidence of $(1-\alpha)(100)\%$.

If alpha (level of significance) is Not Provided
If the *P-value* for testing H_0 is less than:

- 0.1 we have *some evidence* that H_0 is false.
- 0.05 we have *strong evidence* that H_0 is false (we can reject null hypothesis).
- 0.01 we have *very strong evidence* that H_0 is false.
- 0.001 we have *extremely strong evidence* that H_0 is false.

If the *P-value* is greater than 0.1, we *do not have any evidence* that H_0 is false (fail to reject null hypothesis).
(see slides 50 to 59 of Lecture 15 slides for Hypothesis Testing Examples)

Hypothesis Tests for Population Proportions

For these inferences, p_0 represents the given population proportion and the hypothesis will be:
$H_0 : p = p_0$
$H_a : p \neq p_0$ (two tailed) or $p < p_0$ (left tailed) or $p > p_0$ (right tailed)

Conditions:

1. The sample must be a simple random sample from the population of interest.
2. The population must be at least 10 times the size of the sample.
3. The number of successes and the number of failures must each be at least 10 (both $n\hat{p} \geq 10$ and $n(1-\hat{p}) \geq 10$).

Recall that the statistic used for proportions is: $\hat{p} = \frac{\#\ of\ successes}{\#\ of\ observations} = \frac{X}{n}$
For tests involving proportions that meet the above conditions we use the *z-test statistic*:
$z = \frac{\hat{p}-p_0}{\sqrt{(p_0(1-p_0))/n}}$
"difference between the sample successes and what is believed to the population successes divided by the standard error"

use the steps similar to the significance test ones (since it is a z-statistic use `qnorm()` for critical value and `pnorm()` for *p-value*).

(see slides 62 to 69 of Lecture 15 slides for examples)

Summary Chart

The following table gives you a step by step approach for the significance tests:

Parameter	μ given σ	μ **not** given σ	p proportions
1. Null hypothesis	$H_0 : \mu = \mu_0$		$H_0 : p = p_0$
2. Alternative	Choose either $<$, $>$, or $\neq$ in place of $=$ in H_0.		
3. Rejection Region Depending on H_a.	$z_\alpha, -z_\alpha, \pm z_{\frac{\alpha}{2}}$	with df = n - 1 $t_\alpha, -t_\alpha, \pm t_{\alpha/2}$	$z_\alpha, -z_\alpha, \pm z_{\frac{\alpha}{2}}$
4. Test statistic	$z = \frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	$z = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
5. P-value	pnorm(z)	pt(t,n-1)	pnorm(z)
This is the area under the density curve shaded according to H_a.			
6. Decision	**Reject** H_0 if P-value $\leq \alpha$ **Fail** to reject H_0 if P-value $> \alpha$		