Hypothesis Testing

A statistical hypothesis is an assertion or conjecture concerning one or more populations.

To prove that a hypothesis is true with absolute certainty (or to prove that it is false with absolute certainty), we would have to examine the entire population.

Instead, take a random sample from the population and use it as evidence that either supports or does not support the hypothesis.

H_0 = Null hypothesis. The hypothesis we wish to test.

The rejection of H_0 leads to the acceptance of an alternate hypothesis H_1. 2 possible outcomes:

1. reject H_0: in favor of H_1 because of sufficient evidence in the sample.
2. Fail to reject H_0: because of insufficient evidence

Note: There is no formal outcome that says accept H_0. H_0 often represents "status quo" in opposition to a new idea H_1.

Example: Jury Trial H_0 = defendant is innocent
H_1: " is guilty

H_0 is rejected if H_1 is supported by evidence beyond a reasonable doubt. Failure to reject H_0 doesn't imply innocence, only that the evidence was not sufficient to reject it. Therefore, failure to reject H_0 does not necessarily mean accept H_0.

Example: A company manufacturing RAM chips claims the defective rate of the population is 5%. Let p denote the true defective probability

$H_0: p = 0.05$

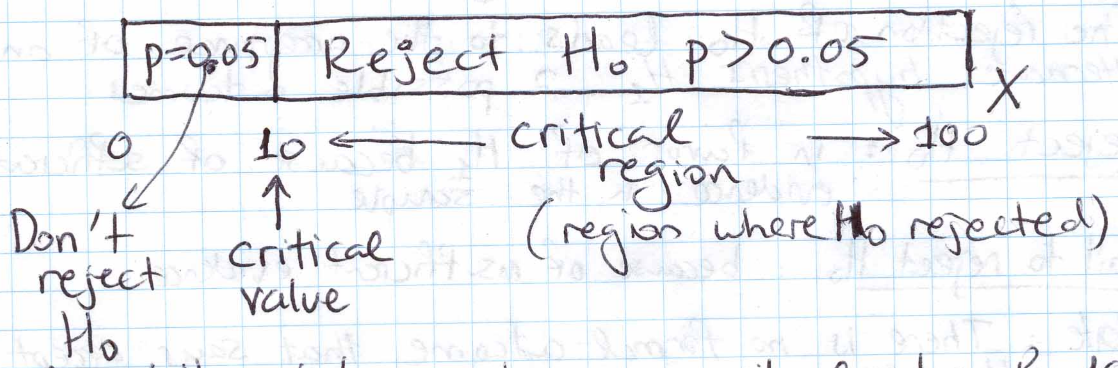
$H_1: p > 0.05$

Take a sample of 100 RAM chips off the production line and test.

Let X denote the number of defectives in the sample of 100. Reject H_0 if $X \geq 10$

(Note 10 is chosen arbitrarily in this case)

X is called the test statistic



Q: Why did we choose a critical value of 10 in this example?

A: Since we know this is a Bernoulli Process, the expected value of defectives is np. So if we believe that $p = 0.05$, we should expect $100 \times 0.05 = 5$ defectives in a sample of 100. Therefore, 10 defectives would be strong evidence that $p > 0.05$. Later we will learn how to choose the critical value based on the desired level of significance for the hypothesis test.

Possible situations

	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

<u>Defn</u>: Rejection of H_0 when it is true is called a Type I error.

<u>Example</u>: Convicting the defendant when he is innocent.

<u>Defn</u>: The probability of committing a type I error, denoted by α, is called the level of significance.

<u>Example</u> (RAM chips continued)

$$\alpha = P(\text{Type 1 error}) = P(\underbrace{X \geq 10}_{\text{critical region}} \text{ when } \underbrace{p = 0.05}_{H_0 \text{ is true}})$$

$$= \sum_{x=10}^{100} b(x; n=100, p=0.05) \quad \text{Binomial Distribution}$$

$$= \sum_{x=10}^{100} \binom{100}{x} 0.05^x \, 0.95^{100-x} = 0.0282$$

Level of significance is $\alpha = 0.0282$.

The lower the α, the less likely we are to commit a type I error. Would like small α values. (0.05 or smaller generally used)

Defn: Nonrejection of H_0 when it is false is called a type II error. The probability of committing a type II error is denoted β.

Note: β is impossible to compute unless we have a specific alternate hypothesis.

Example continued: Can't compute β for $H_1: p > 0.05$ but can compute it for testing $H_0: p = 0.05$ against the alternate hypothesis that $p = 0.1$ for instance.

$$\beta = P(\text{Type II error}) = P(X < 10 \text{ when } p = 0.1)$$

$$= \sum_{x=0}^{9} b(x; n=100, p=0.1) = 0.4513$$

↑ H_1 (H_0 false)

This high probability suggests we are likely to fail to reject H_0 if the true p is 0.1. This might be OK if we only want to make sure we don't fail to reject if the true p is 0.15

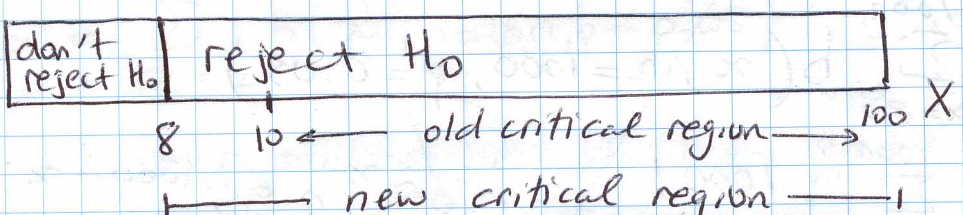
$$\beta = P(\text{Type II error}) = P(X < 10 \text{ when } p = 0.15)$$

$$= \sum_{x=0}^{9} b(x; n=100, p=0.15) = 0.0551$$

Effect of critical value

Moving the critical value provides a trade-off between α and β. A reduction in β is always possible by increasing the size of the critical region, but this increases α.

Example continued : Lets change the critical value from 10 to 8. So reject H_0 if $X \geq 8$.



$$\alpha = \sum_{x=8}^{100} b(x; n=100, p=0.05)$$ Probability of rejecting H_0 when it is true

$= 0.128$ (note larger than before when the critical value was set at 10)

testing against alternate hypothesis $p = 0.1$

$$\beta = \sum_{x=0}^{7} b(x; n=100, p=0.1)$$ Probability of not rejecting H_0 when $p = 0.1$

$= 0.206$ (lower than before)

testing against alternate hypothesis $p = 0.15$

$$\beta = \sum_{x=0}^{7} b(x; n=100, p=0.15)$$ Probability of not rejecting H_0 when $p = 0.15$

$= 0.012$ (again lower than before)

Effect of Sample size

Both α and β can be reduced simultaneously by increasing the sample size.

Example continued : Sample size $n = 150$, critical value 12. Reject H_0 if $X \geq 12$ (X is the # of defectives in sample of 150)

$$\alpha = \sum_{x=12}^{150} b(x; n=150, p=0.05) = 0.074$$

(was 0.128 for $n = 100$ and critical value 8)

testing against alternate hypothesis $p = 0.1$

$$\beta = \sum_{x=0}^{11} b(x; n=150, p=0.1) = 0.171$$

(was 0.206 for $n = 100$ and critical value 8)

Factorials of very large numbers problematic to compute accurately even with MATLAB. Thankfully, the Binomial Distribution can be approximated by the normal distribution - Details Section 6.5

Theorem : If X is a binomial random variable with n trials and probability of success of each trial p, then the limiting form of the distribution of $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ as $n \to \infty$ is the standard normal distribution.

Approximation is good when n is large and p is not extremely close to 0 or 1.

Lets recompute our α with the normal approximation.

$$\alpha = P(\text{Type I error}) = \sum_{x=12}^{150} b(x; n=150, p=0.05)$$

$$\approx P\left(Z \geq \frac{12 - 150 \times 0.05}{\sqrt{150 \times 0.05 \times 0.95}}\right) = P(Z \geq 1.69)$$

$$= 1 - P(Z \leq 1.69) = 1 - 0.9545 = 0.0455$$

not too bad.

What if we increase the sample size to $n = 500$ and the critical value to 40?

The normal approximation should be even better since n larger

$$\alpha \approx P\left(Z \geq \frac{40 - 500 \times 0.05}{\sqrt{500 \times 0.05 \times 0.95}}\right) = P(Z \geq 3.08)$$

$$= 1 - P(Z \leq 3.08) = 1 - 0.999 = 0.001$$

Very unlikely to commit type I error.

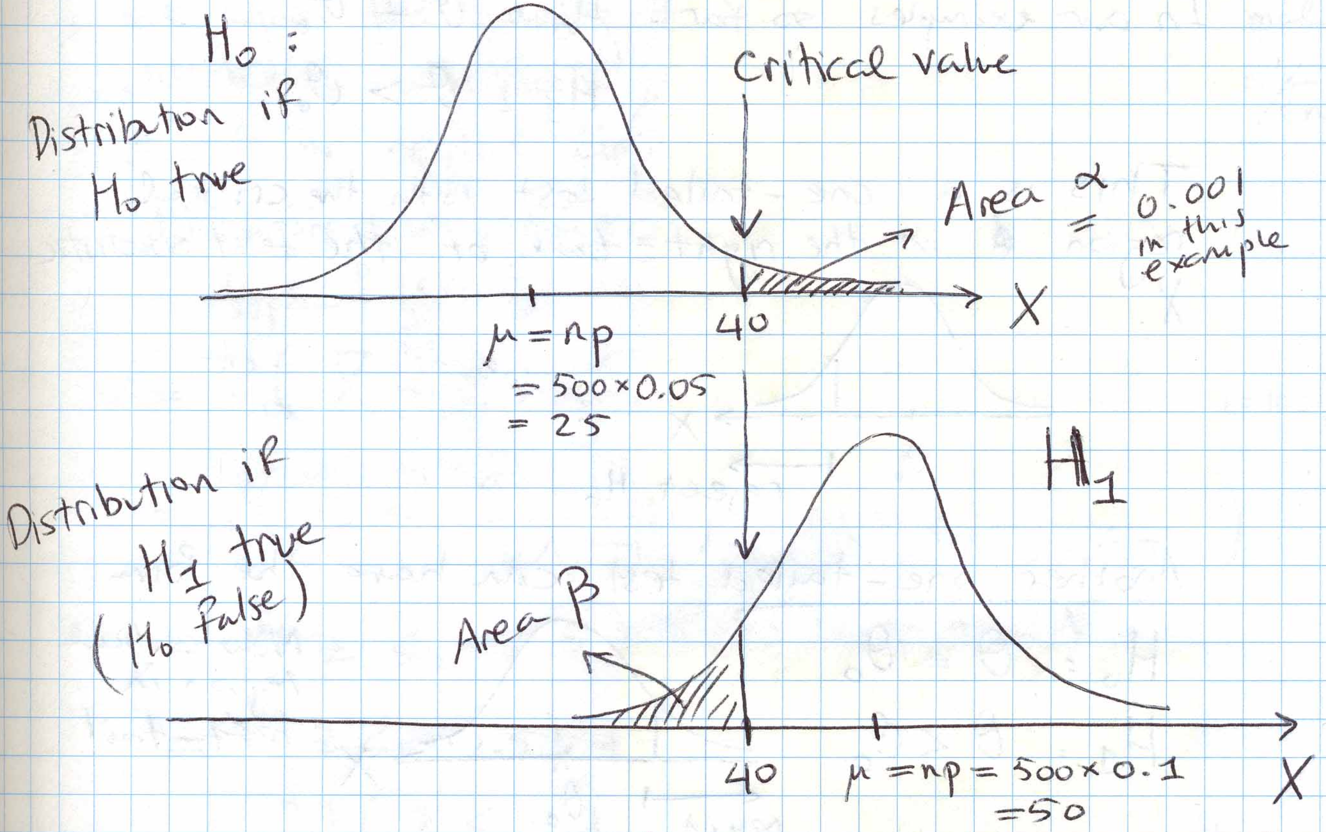
testing against alternate hypothesis $p = 0.1$

$$\beta = \sum_{x=0}^{39} b(x; n=500, p=0.1)$$

$$\approx P\left(Z \leq \frac{39 - 500 \times 0.1}{\sqrt{500 \times 0.1 \times 0.9}}\right) = P(Z \leq -1.49)$$

$$= 0.0681$$

Visual Interpretation with normal approximation



Defn: The power of a test is the probability of rejecting H_0 given that a specific alternate is true. Power $= 1 - \beta$.

Properties of hypothesis testing

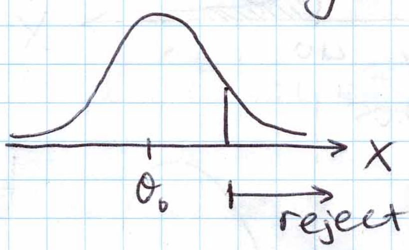
1. α and β are related. Decreasing one generally increases the other
2. α can be set to a desired value by adjusting the critical value. α typically set at 0.05 or 0.01
3. Increasing n decreases α and β both.
4. β decreases as the distance between the true value and hypothesized (H_1) value increases.

One-~~sided~~ tailed vs. two-tailed tests

In our examples so far $H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

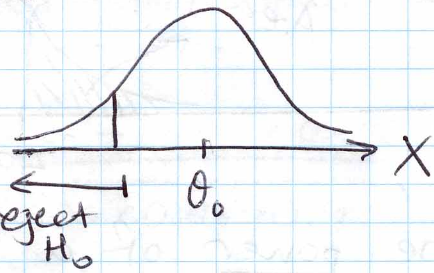
This is a one-tailed test with the critical region in the right-tail of the test statistic X.



Another one-tailed test can have the form

$H_0: \theta = \theta_0$

$H_1: \theta < \theta_0$

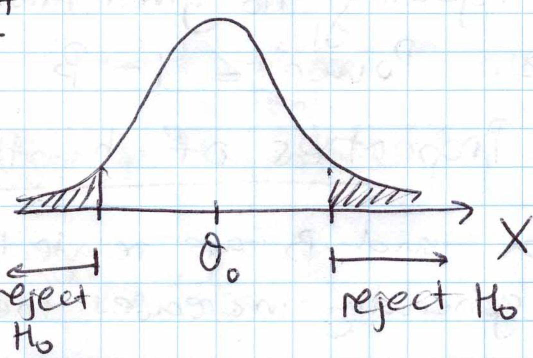


Now critical region in left-tail.

Two-sided test

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$



Example: Production line of resistors that are supposed to be 100 ohms. Assume $\sigma = 8$

$H_0: \mu = 100$

$H_1: \mu \neq 100$

Let $\bar{X}$ be the sample mean for a sample of size $n = 100$

Reject H_0	Do not reject H_0	Reject H_0

98 102

The test statistic is the sample mean in this case.
We know the sampling distribution of $\bar{X}$ is a normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$ due to the central limit theorem. $(n = 100 > 30)$
Therefore we can compute the probability of a type I error as

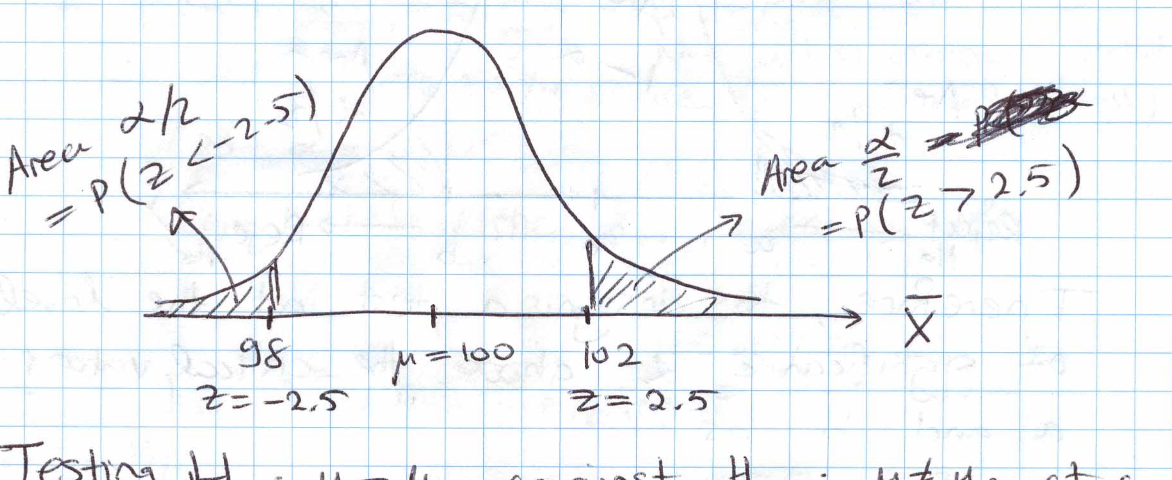
$$\alpha = P(\bar{X} < 98 \text{ when } \mu = 100) + P(\bar{X} > 102 \text{ when } \mu = 100)$$

$$= P\left(Z < \frac{98-100}{8/\sqrt{100}}\right) + P\left(Z > \frac{102-100}{8/\sqrt{100}}\right)$$

$$= P(Z < -2.5) + P(Z > 2.5)$$

$$= P(Z < -2.5) + (1 - P(Z < 2.5))$$

$$= 2P(Z < -2.5) = 2 \times 0.0062 = 0.0124$$



* Testing $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$ at a significance level α is equivalent to computing a $100(1-\alpha)\%$ confidence interval for μ and rejecting H_0 if μ_0 is outside the confidence interval.