Hypothesis Testing

Type I error: Reject a true H_0

Type II error: Fail to reject a false H_0.

Ex) city wants to buy water pipes

wanted mean strength:

$\mu > 2400$ eb/ft

city buys 50 pipes to sample. They obtain $\bar{x} = 2460$. Suppose $\sigma = 200$ of the population.

This sample passes the safety threshold.

Q: should the city buy 10,000 pipes from this company?

Recall: a different sample of 50 pipes may have a different mean strength.

Need decision rule s.t.

$\alpha = P(\text{type I error})$ is small.

Let $\alpha = 0.05$.

Let $\tilde{H}_0$: $\mu \leq 2400$ lb/ft

H_1 : $\mu > 2400$ lb/ft

Note, it is simpler to replace $\tilde{H}_0$ with H_0, since we would not buy the pipes even if $\mu = 2400$.

H_0 : $\mu = 2400$ lb/ft (null)

H_1 : $\mu > 2400$ lb/ft (alternate)

Critical region: subset of sample space which corresponds to rejecting H_0. (also called rejection region)

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{by the CLT}$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

We want $P(\text{reject } H_0 \text{ when true}) = 0.05$

$$\Rightarrow P(Z > Z_\alpha \mid \mu = 2400) = 0.05$$

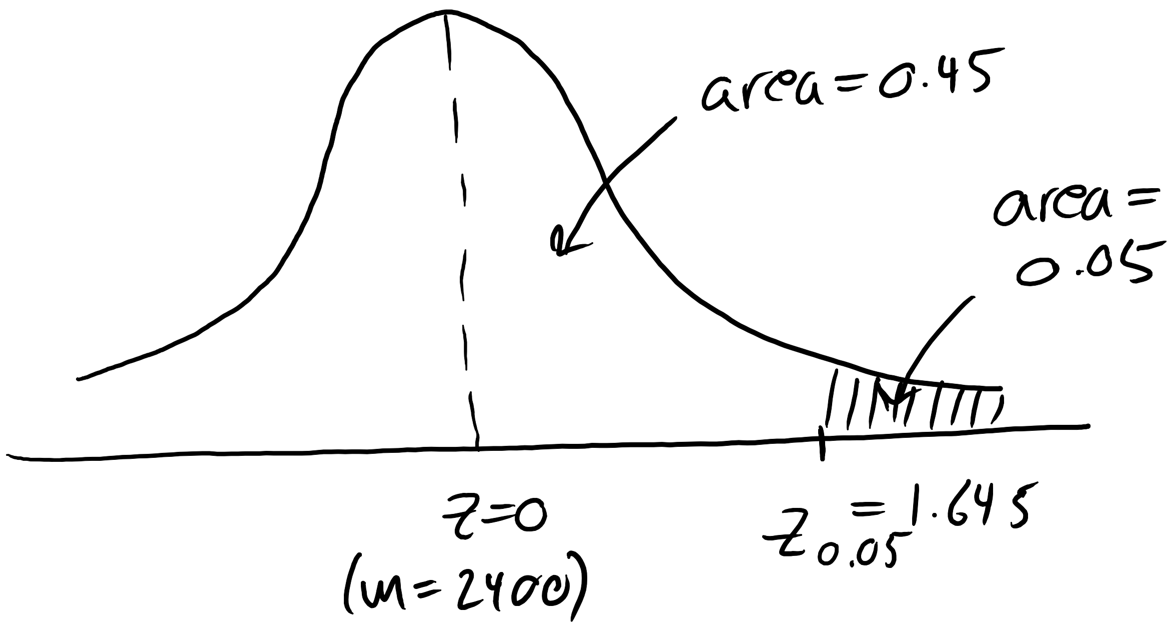
$$\Rightarrow P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > Z_\alpha \mid \mu = 2400\right) = 0.05$$

$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is called the test statistic.

$$P(Z > Z_\alpha) = 0.05$$

$$\Rightarrow P(0 < Z < Z_\alpha) = 0.5 - 0.05 = 0.45$$

$Z_\alpha = 1.645$ (from table)



shaded area is <u>rejection region</u>.
If our sample mean falls in there then the chance that the population mean is 2400 is only 0.05.

$$z_s = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{2460 - 2400}{200/\sqrt{50}}$$

$$= \frac{60}{28.12} \approx 2.12$$ (z_s "std normal statistic")

Since $z_s > z_\alpha = z_{0.05} = 1.645$

we <u>reject</u> H_0 in favor of <u>H_1</u> at the 0.05 confidence level.

what is $\beta = P(\text{type II error})$?

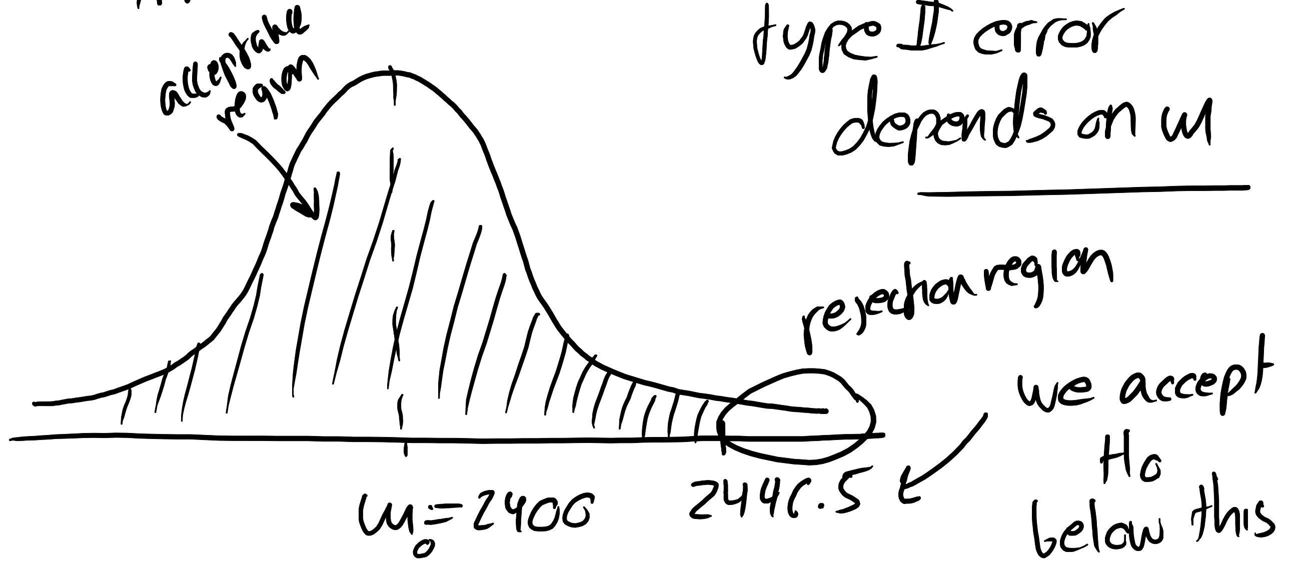
$P(\text{Accept } H_0 \mid H_0 \text{ is false})$

depends on actual mean strength of pipes

$= P(z < 1.645 \mid \mu > 2400)$

$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1.645 \mid \mu > 2400 \right)$

type II error depends on μ



$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = 1.645 = \frac{\bar{X} - 2400}{200/\sqrt{50}}$$

$$\bar{X} - 2400 = 46.53 \Rightarrow \bar{X} \approx 2446.5$$

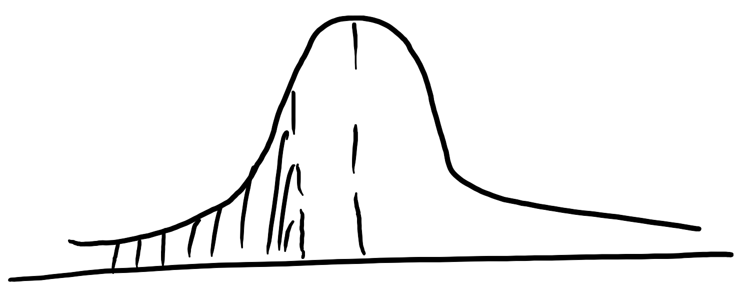
Accept H_0 when $\bar{X} < 2446.5$.

Suppose $m = 2450$. (actual mean strength)

$$\text{Then } P(\underbrace{\bar{X} < 2446.5}_{\text{accept } H_0} \mid \underbrace{m = 2450}_{H_0 \text{ is false}})$$

$$= P\left(\frac{\bar{X} - 2450}{200/\sqrt{50}} < \frac{2446.5 - 2450}{200/\sqrt{50}} \right)$$

$$= P(z < -0.12)$$



$$P(\text{type II}) = 0.5 - P(0 < z < 0.12)$$

$$= 0.5 - 0.0478 \approx 0.45$$

That is, if the actual mean strength of population is 2450 eb/ft, then it is likely that choosing a sample of size 50 would give us $\bar{X} < 2446.5$.

However, if $\mu = 2500$ lb/ft

$$P\left(\frac{\bar{x} - 2500}{\sigma/\sqrt{n}} < \frac{2446.5 - 2500}{200/\sqrt{50}} \right)$$

$= P(z < -1.89)$

$= 0.5 - P(0 < z < 1.89) =$

$= 0.5 - .4706 = 0.0294$

Notice $\mu = 2500$ high above $\mu_0 = 2400$

If the actual μ is higher, than the P(type II error) is even lower.

Thus, the decision rule:

$$z_s > z_\alpha$$

used for $H_0: \mu = \mu_1$

$H_1: \mu > \mu_1$

is acceptable in practice.

P = power of test $= 1 - \beta$

standard decision rules with specified α:

$H_0: \mu = \mu_0$

$H_1: \mu > \mu_0 \Rightarrow$ $\underline{z > z_\alpha}$

$H_1: \mu < \mu_0 \Rightarrow$ $\underline{z < -z_\alpha}$

$H_1: \mu \neq \mu_0 \Rightarrow |z| > z_{\alpha/2}$

Ex) Let $\bar{x}$ measure mean test score of admitted students. Records for last few years indicate mean 115.

$H_0: \mu = 115$

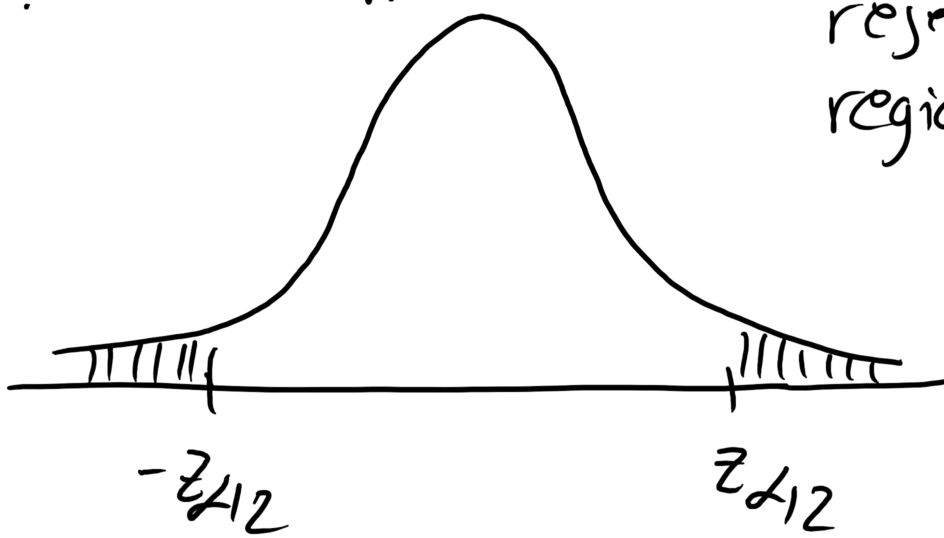
$H_1: \mu \neq 115$

want to test if incoming class is any different.

sample size $n = 50 \Rightarrow \bar{x} = 118$.

Assume $\underline{\sigma = 20}$.

Is this class better than previous years? "(different) or worse"



rejection regions

decision rule $|z_s| > z_{\alpha/2}$

$$z_{\alpha/2} = z_{0.05/2} = z_{0.025} = 1.96$$

$$p(0 < z < z_{0.025}) = 0.5 - 0.025$$

$$\bar{X} - 1.96\,\sigma_{\bar{X}} = \bar{X} - 1.96\frac{\sigma}{\sqrt{50}} = \bar{X} - 1.96\frac{20}{\sqrt{50}}$$

$\Rightarrow$ cut offs for critical region are:

$115 - 1.96(2.8) = 109.5$ and

$115 + 1.96(2.8) = 120.5$

$\bar{X} \in (109.5, 120.5)$ so we <u>do not</u> reject H_0.

$$z_s = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{118 - 115}{20/\sqrt{50}} = 1.06$$

and $|z_s| < z_{\alpha/2} = 1.96$.

We fail to reject H_0 since it is not true that $|z_s| > z_{\alpha/2}$!

(at the 0.05 confidence level = P(type I error))

I Standard tests of hypothesis involving population means

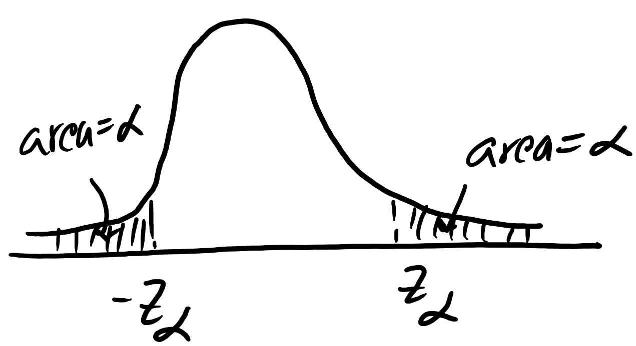
(a) large samples / known variance σ from normal population

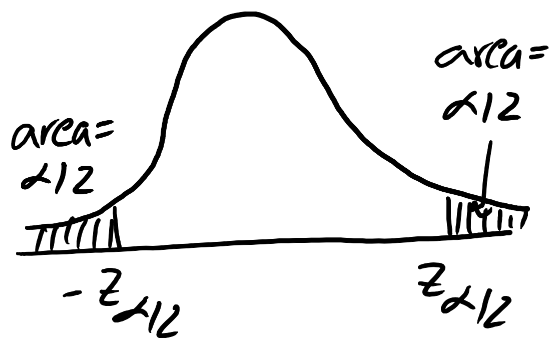
$H_0: \mu = \mu_0$

$$Z_s = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

alt. hypothesis	decision rule (to reject H_0)
$H_1: \mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$
$H_1: \mu > \mu_0$	$z > z_\alpha$
$H_1: \mu < \mu_0$	$z < -z_\alpha$

$\pm z_\alpha$ and $z_{\alpha/2}$ correspond to the following pictures:





so $P(0 < z < z_\alpha) = 0.5 - \alpha$

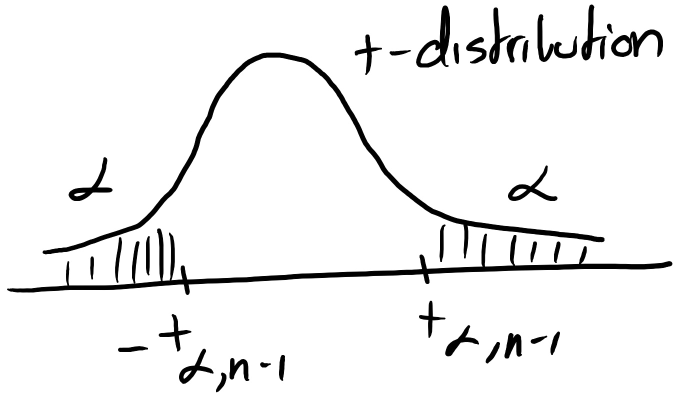
(b) small samples / unknown variances
population need not be normal but should not be explicitly not bell shaped.

$$t_s = \frac{\bar{x} - m}{s/\sqrt{n}} \sim \underbrace{T_{n-1}(0,1)}_{\text{student-t distribution}}$$

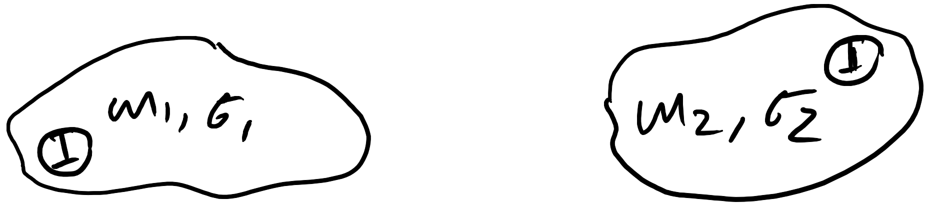
$H_1: m \neq m_0$ $\quad\quad |t_s| < t_{\alpha/2, n-1}$

$H_1: m > m_0$ $\quad\quad t_s > t_{\alpha, n-1}$

$H_1: m < m_0$ $\quad\quad t_s < -t_{\alpha, n-1}$



Two population tests



(c) large samples / normal populations / known variances σ_1, σ_2

$H_0 : \mu_1 - \mu_2 = d$

$$z_s = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

decision rules

$H_1 : \mu_1 - \mu_2 \neq d$ $\qquad |z| > z_{\alpha/2}$

$H_1 : \mu_1 - \mu_2 > d$ $\qquad z > z_\alpha$

$H_1 : \mu_1 - \mu_2 < d$ $\qquad z < -z_\alpha$

Ex) Compare avg protein content in two brands of dog food. Suppose 50 packages of brand A and 60 packages of brand B are sampled. $n_1 = 50;\ n_2 = 60$

$\bar{x}_1 = 11g \ ; \ \bar{x}_2 = 9g$

$s_1 = 1g \ ; \ s_2 = 0.5g$

A difference of .5g is considered to be not sufficiently important in comparing the two brands. Thus, a decision was made to test:

$H_0: \mu_1 - \mu_2 = 0.5$ vs $\alpha = 0.01$ confidence level

$H_1: \mu_1 - \mu_2 > 0.5$

Let $\sigma_1 \approx S_1 = 1$; $\sigma_2 \approx S_2 = .5$ (assumption based on sample size)

$$z_s = \frac{\bar{X}_1 - \bar{X}_2 - d}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} = \frac{11 - 9 - .5}{\sqrt{.02 + .0042}} \approx 9.65$$

$$z_\alpha = z_{0.01} = 2.33$$

$$(P(0 < z < z_\alpha) = 0.5 - 0.01 = 0.49)$$

Since $z_s > z_\alpha$ we reject the null hypothesis in favor of H_1 at the .01 confidence level.

Thus brand A dog food exceeds brand B by more than .5g.

(d) small samples / unknown variances / approx normal populations

$$t_s = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-1}$$

	decision rule
$H_0: \mu_1 - \mu_2 = d$	
$H_1: \mu_1 - \mu_2 \neq d$	$\lvert t_s \rvert > t_{\alpha/2,\, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 > d$	$t_s > t_{\alpha,\, n_1+n_2-2}$
$H_1: \mu_1 - \mu_2 < d$	$t_s < -t_{\alpha,\, n_1+n_2-2}$

Tests for proportions

$$z_s = \frac{S_n - np_0}{\sqrt{np_0(1-p_0)}}$$

when n is large and p_0 not too small

$z_s \sim N(0,1)$

$H_0: p = p_0$

$H_1: p \neq p_0 \qquad |z| > z_{\alpha/2}$

$H_1: p > p_0 \qquad z > z_\alpha$

$H_1: p < p_0 \qquad z < -z_\alpha$

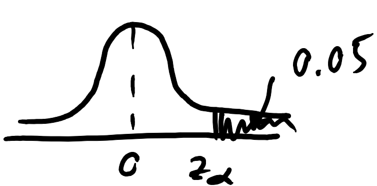
Ex) A Samsung QC engineer suspects that the proportion of defective units has increased from set limit of .01. To test his claim he randomly selects 100 units and found that the proportion of defective units in the sample was $\underbrace{.02}_{p_1}$. (test at $\alpha = 0.05$ level)

$H_0: p = \underbrace{.01}_{p_0}$ vs $H_1: p > 0.01$

Reject H_0 if
$$\frac{S_n - np_0}{\sqrt{np_0(1-p_0)}} > z_\alpha$$

$z_\alpha = z_{0.05} = 1.645$

$S_n = np_1$
$= 100 \times 0.02$



$$z_s = \frac{(.02)(100) - (100)(.01)}{\sqrt{100(.01)(1-0.1)}} = 1.005$$

Since $z_s \not> z_\alpha$ we cannot reject H_0.
at $\alpha = 0.05$ conf. level.

Tests for variances

$$\chi_s^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 \neq \sigma_0^2$ $\quad \chi_s^2 > \chi^2_{\alpha/2, n-1}$

or $\chi_s^2 < \chi^2_{1-\alpha/2, n-1}$

$H_1: \sigma^2 > \sigma_0^2$ $\quad \chi_s^2 > \chi^2_{\alpha, n-1}$

$H_1: \sigma^2 < \sigma_0^2$. $\quad \chi_s^2 < \chi^2_{1-\alpha, n-1}$

Ex) According to manufacturer specifications, the cooling cycle for certain type of equipment should have std deviation of at most 2 seconds. A QC engineer want to test if this is satisfied using the equipment for 10 cooling cycles, he obtains

$s = 3$ (sample std deviation)

At the .01 level of significance, test

$H_0: \sigma = 2$ vs $H_1: \sigma > 2$

$$X_s^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(10-1)\cdot 9}{2^2} = \frac{9\cdot 9}{4} = 20.25$$

$$X^2_{\alpha, n-1} = X^2_{.01, 9} = 21.67$$

since $X_s^2 \not> X^2_{\alpha, n-1}$,

we cannot reject H_0 at the .01 significance level.