Confidence Intervals for the difference between two means of two populations with known variances.

(case I – normal population or large sample size)

Then:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

standard normal

when $n_1, n_2 \geq 30$ or populations are normal

Recall that $P(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha$

$$\Rightarrow (\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

is a $(1-\alpha)\cdot 100\%$ CI for the difference between two population means.

Case II - σ not known

Then if $\bar{x}_1, \bar{x}_2, s_1, s_2$ are the values of the means and std deviations of independent random samples of size n_1 and n_2 from normal populations with equal variances, then:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2, n_1+n_2-2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2$$

$$< (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2, n_1+n_2-2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

is a $(1-\alpha)\cdot 100\%$ CI for difference in population means.

(as in the case of one population when σ not known, we replace normal distribution with t-dist.)

Tests of Statistical Hypothesis

Suppose $\sigma^2 = .25$ is known for a normal population.

$H_0: \mu = 1$ versus $H_1: \mu = 2$

Ex decision rule for sample of size $n=9$:

Reject H_0 when $\bar{x} < .6122$ (not a good decision rule)

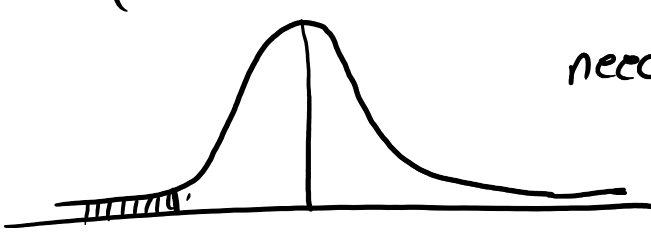
Calculate probabilities of type I and type II errors.

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$

$= P(\bar{x} < 0.6122 \text{ when } \mu = 1)$

$$= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{.6122 - 1}{.5/3} \right)$$

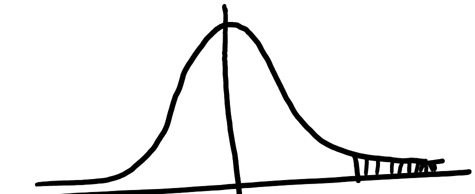
$= P(z < -2.33)$



-2.33

need to find shaded area

Equivalent area on the right:



2.33

$P(z < -2.33) = 0.5 - P(0 < z < 2.33)$

$P(Z < -2.33) = 0.5 - 0.4901 \approx 0.01$

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

$= P(\text{accept } H_0 \text{ when } H_0 \text{ is false})$

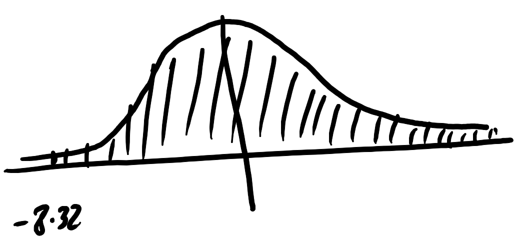
$= P(\text{accept } H_0 \text{ when } H_1 \text{ is true})$

$= P(\bar{X} > 0.6122 \text{ when } \mu = 2)$

$= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{.6122 - 2}{.5/3} \right)$

$= P(z \geq -8.32) \approx 1 !$

what is this probability:



Notice the point: bad decision rule detected only by one of two errors

Review of "standard" decision rules

Tests for means of normal populations with known variances. (or when $n \geq 30$)

	Decision Rule
$H_1: \mu \neq \mu_0$	$\lvert z \rvert > z_{\alpha/2}$
$H_1: \mu > \mu_0$	$z > z_\alpha$
$H_1: \mu < \mu_0$	$z < -z_\alpha$

Ex) $H_0: \mu = 3, \quad H_1: \mu < 3$ — use $\alpha = .05$ level of significance

$n = 10, \ \sigma = .25$, normal population

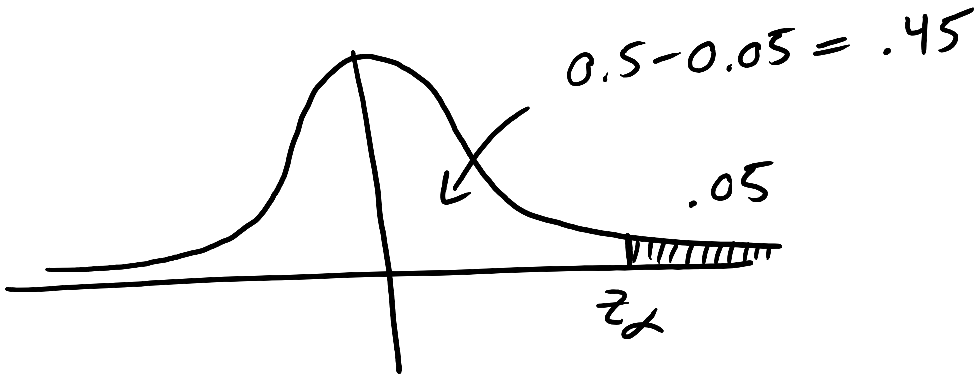
calculate test statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 3}{.25/\sqrt{10}}$$

Reject H_0 if $z < -z_\alpha$

Choose $\alpha = .05$ as specified.

what is $z_\alpha = z_{.05}$?



$P(0 < z < z_\alpha) = .45 \Rightarrow z_\alpha = 1.65$

$\Rightarrow$ Reject H_0 if $z < -1.65$

$$z = \frac{\bar{x} - 3}{.25/\sqrt{10}} = \frac{2.8 - 3}{.25/\sqrt{10}} = -2.53$$

We would reject H_0 since

$$z = -2.53 < -1.645 = z_\alpha$$

Test for means of normal populations with unknown variances

Use student-t distribution

$H_1: \mu \neq \mu_0 \Rightarrow |t| > t_{\alpha/2, n-1}$

$\mu > \mu_0 \Rightarrow t > t_{\alpha, n-1}$

$\mu < \mu_0 \Rightarrow t < -t_{\alpha, n-1}$

Ex) Restaurant example from last time ; $\alpha = .05$

$H_0: \mu = 100$ versus $H_1: \mu < 100$

sample $\left\{ \begin{array}{l} n = 9, \quad s = 10 \\ \bar{x} = 95 \end{array} \right.$

Buyer wants to test owners claim on # customers per day.

Reject H_0 if $t < -t_{\alpha, n-1}$

$\Rightarrow t_{\alpha, n-1} = t_{0.05, 8} = 1.860$ from table

statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{95 - 100}{10/\sqrt{9}} = -1.5$

Since $t = -1.5 > -1.860 = -t_{\alpha, n-1}$

at the .05 significance level, the buyer cannot reject the owners claim that the avg # of customers is 100.

Tests for the difference between means

Suppose normal populations P_1, P_2 (or large samples) with σ_1^2, σ_2^2 known.

$H_0: \mu_1 - \mu_2 = d$

$H_1: \mu_1 - \mu_2 > d \Rightarrow$ [reject H_0 if $z > z_\alpha$] standard decision rule

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

Ex) Compare avg protein content of two brands

$\bar{x}_1 = 11$ $\quad \sigma_1 \approx s_1 = 1$ $\quad n_1 = 50$ $\quad d = .5$

$\bar{x}_2 = 9$ $\quad \sigma_2 \approx s_2 = .5$ $\quad n_2 = 60$ $\quad \alpha = .01$

Since our samples are large, we take population std deviation to be the sample std dev (must be specified in problem to do this)

$H_0: \mu_1 - \mu_2 = .5$

versus

$H_1: \mu_1 - \mu_2 > .5$

$z_\alpha = z_{.01} = 2.33$

$0.5 - 0.01 = .49$

$P(0 < z < z_\alpha) = .49$

$\Rightarrow z_\alpha = 2.33$

Calculate statistic:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{11 - 9 - .5}{\sqrt{.02 + .0042}} = 9.649 > 2.33$$

Thus, at the .01 level of significance, we should reject H_0 in favor of H_1, concluding that protein in brand A exceeds that in brand B by more than .5g.

Note: For small samples and unknown variances the t-distribution can be used.

Tests for proportions

statistic $\quad z = \frac{x - np}{\sqrt{np(1-p)}} \sim N(0,1)$ for large n

Ex) A QC engineer suspects that the proportion of defective units has increased from the set limit of .01.
To test his claim he randomly selected 100 of these items and found that the prop. of defective units in sample is .02.
Test claim at $\alpha = .05$ level of significance.

$\Rightarrow H_0: p = .01$ versus $H_1: p > .01$

p_0 corresponds to H_0

Reject H_0 if $z = \dfrac{X - np_0}{\sqrt{np_0(1-p_0)}} > z_\alpha$

$z_\alpha = z_{.05} = 1.65$ (from before)

$X = \underbrace{n\hat{p}}_{\text{sample proportion}} = 100(.02)$

$\Rightarrow \dfrac{n\hat{p} - np_0}{\sqrt{np_0(1-p_0)}}$

$\Rightarrow z = \dfrac{100(.02) - 100(.01)}{\sqrt{100(.01)(1-.01)}} = 1.005$

Thus, we <u>cannot</u> reject H_0 at $\alpha = .05$ confidence level.

Note: from end of last class

$$\frac{n\hat{p} - np_0}{\sqrt{n p_0(1-p_0)}} = \frac{n(\hat{p}-p_0)}{\sqrt{n}\sqrt{p_0(1-p_0)}} = \frac{\sqrt{n}(\hat{p}-p_0)}{\sqrt{p_0(1-p_0)}}$$

$$\frac{n}{\sqrt{n}} \frac{\sqrt{n}}{\sqrt{n}} = \frac{n\sqrt{n}}{n} = \sqrt{n}$$

$$\Rightarrow z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

proportion claim testing statistic

(This is the form given in the book)

Ex) Let P be a normal population with variance $\sigma^2 = .25$ and unknown mean μ.
Sample of size $n=9$ is drawn.
Let $H_0: \mu = 1$ vs $H_1: \mu = 2$

Suppose decision rule is:

Reject H_0 when $\bar{x} > 1.39$

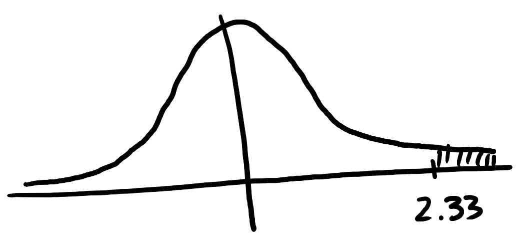
Find probabilities of type I, type II errors.

$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$

$= P(\bar{x} > 1.39 \text{ when } \mu = 1)$

$$= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} > \frac{1.39 - 1}{.5/3} \right)$$

$= P(z > 2.33)$ since population is normal



$= 0.5 - P(0 < z < 2.33)$

$= .01$

$\beta = P(\text{accept } H_0 \text{ when } H_0 \text{ is false})$

$= P(\text{accept } H_0 \text{ when } H_1 \text{ is true})$

$= P(\bar{x} < 1.39 \text{ when } \mu = 2)$

$$= P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{1.39 - 2}{.5/3} \right)$$

$= P(z < -3.67) = .001$

power of test $= 1 - \beta$

The test is relatively accurate.

Comparing two population means

Recall statistic $z = \dfrac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

For large sample sizes (when stated so), can take $\sigma_1 \approx s_1$ and $\sigma_2 \simeq s_2$.

Ex) At a University it is speculated that male students spend more per week on unhealthy foods than females.

Sociologist randomly selects 200 students.

$\Rightarrow$ 125 men and 75 women

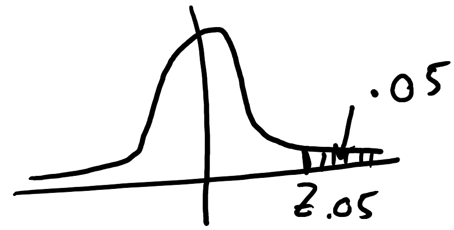
unhealthy spending by men is $\bar{x}_1 = 20$, $s_1 = 12$

unhealthy spending by women is $\bar{x}_2 = 9$, $s_2 = 3$

Test the claim below at .05 level of significance using $\sigma_1 \approx s_1$ and $\sigma_2 \approx s_2$.

$H_0: \mu_1 - \mu_2 = 10$ versus $H_1: \mu_1 - \mu_2 > 10$

Reject H_0 if $z > z_\alpha$



$z_\alpha = z_{.05} = 1.65$

$P(0 < z < z_{.05}) = 0.5 - .05$
$= .45$
usd table

$$z = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \simeq \frac{20-9-10}{\sqrt{\frac{12^2}{125} + \frac{3^2}{75}}}$$

$$= .88 \not> 1.65 = z_\alpha$$

At the .05 significance level, H_0 cannot be rejected.

Find the value of d for which the result will be significant at the .05 level:

$$z = \frac{\bar{x}_1 - \bar{x}_2 - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > 1.65 \Rightarrow \underline{d = 9}$$

Ex] (small sample)

Given the 8 observations of mpg ratings of new cars:

$\{31, 29, 26, 33, 40, 28, 30, 25\}$

Test the following Hypothesis:

$H_0: \mu = 35$ versus $H_1: \mu \neq 35$

at the $\alpha = .01$ significance level.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 30.25 \qquad (n=8)$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]$$

$$\approx 22.21 \implies s = \sqrt{22.2} = 4.71$$

Use studen-t distribution

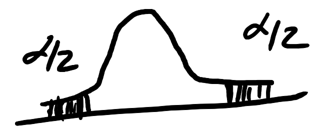
t-statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{30.25 - 35}{4.71/\sqrt{8}}$

$\Rightarrow t = -2.85$

Decision rule: $|t| > t_{\alpha/2, n-1}$ (double sided)

$\frac{\alpha}{2} = .01 = 0.005$

note the "$\frac{\alpha}{2}$" area in each end



$n-1 = 8-1 = 7$

$t_{\frac{\alpha}{2}, n-1} = t_{0.005, 7} = 3.499$

Since $|t| \not> t_{\alpha/2, n-1}$, we cannot reject the null Hypothesis H_0.

Ex) (two sided test for proportions)

A sports magazine reports watchers of Monday night football games evenly divided between men and women.

Out of a random sample of 400 people who regularly watch the Monday night game, 220 are men. Using .10 confidence level, can we conclude that the magazine report is false?

$H_0: p = .50$ where p is the true population proportion

$H_1: p \neq .50$

statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$ for large n

Use two sided decision rule:

Reject H_0 if $|z| > \underline{\underline{z_{\alpha/2}}}$

$\frac{\alpha}{2}$ due to double sided decision rule

$z_{\alpha/2} = z_{0.05} = 1.645$

$$z = \frac{.55 - .50}{\sqrt{[(.50)(.50)]/400}} \approx \frac{.05}{.025} = 2.0$$

since $|z| = 2.0 > z_{\alpha/2} = 1.65$ we reject

H_0 in favor of H_1.