Example 2. We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it. The manufacturer claims that at most one in 10 items in the shipment is defective. At the 1% level of significance, do we have sufficient evidence to disprove this claim?

$n = 200$ (sample size)

24 items are defective

$\hat{p} = \frac{24}{200} = \frac{3}{25}$ - proportion of defective items (sample proportion)

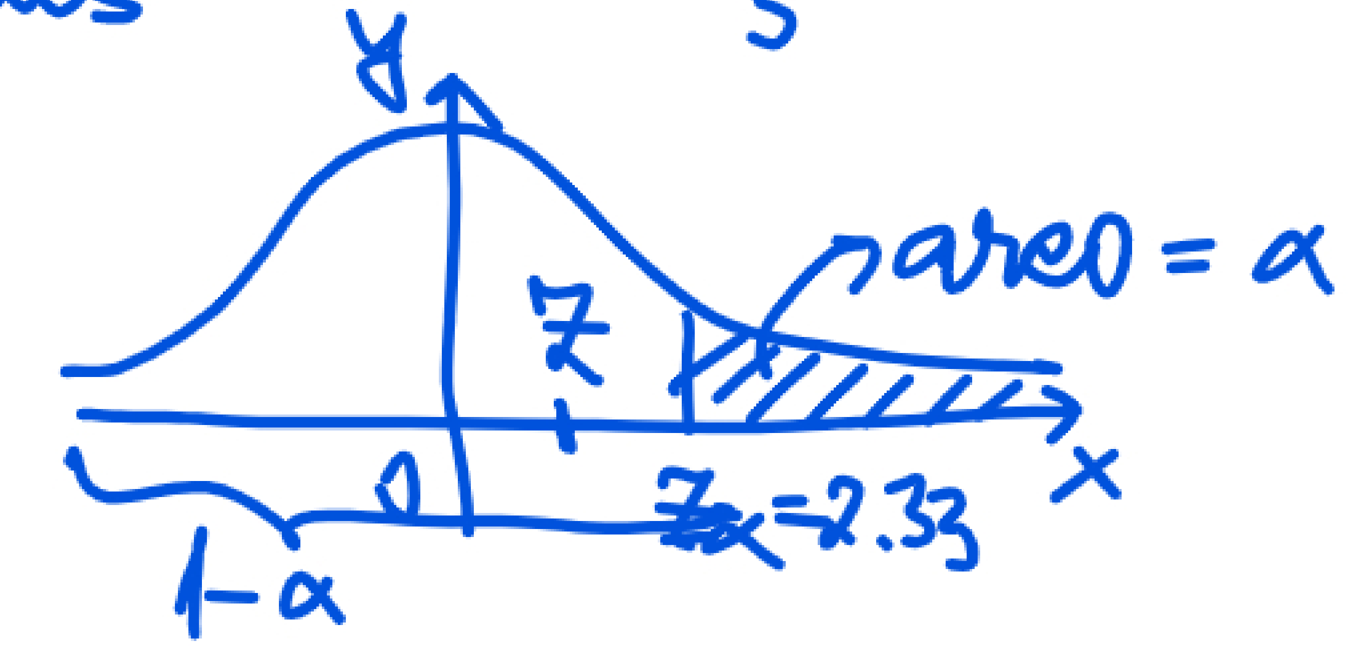
$H_0: p = \frac{1}{10}$ $(= 0.1)$ $\rightarrow$ $p_0 = \frac{1}{10}$

$H_a: p \boxed{>} \frac{1}{10}$ $\Rightarrow$ $R = (z_\alpha, \infty)$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

$$Z = \frac{\frac{3}{25} - \frac{1}{10}}{\sqrt{\frac{1}{10}\cdot\frac{9}{10}\cdot\frac{1}{200}}} = \frac{\frac{1}{50}}{\frac{3}{100\sqrt{2}}} = \frac{100\sqrt{2}}{150} = \frac{2\sqrt{2}}{3}$$

$$Z = \frac{2\sqrt{2}}{3} = 0.94$$



z_α = qnorm$(1-\alpha)$ = qnorm(0.99) = 2.33

$\alpha = 1\% = 0.01$

$Z = 0.94 \notin R = (2.33, \infty) \Rightarrow H_0$ is not rejected

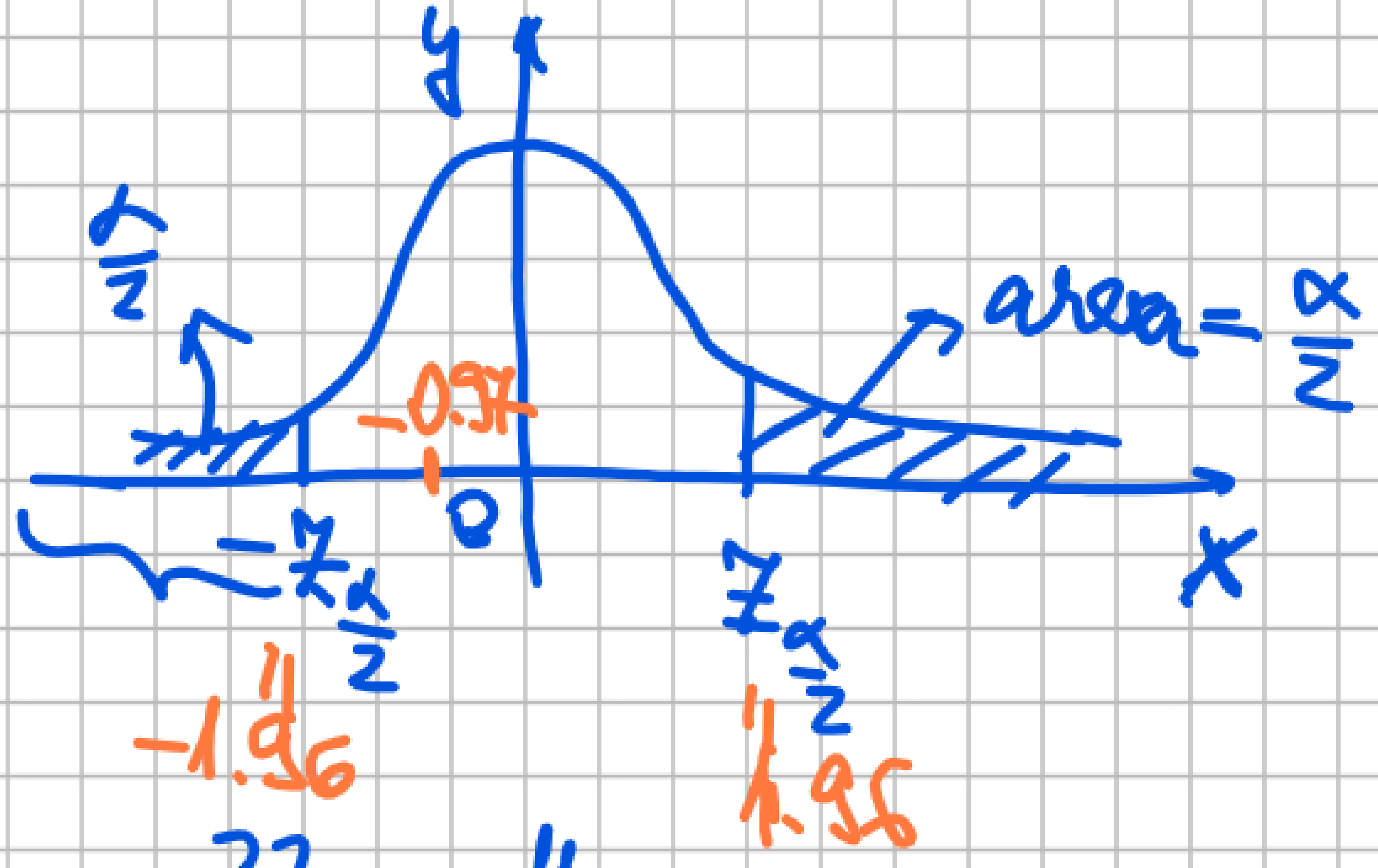
→ the manufacturer's claim is true (we don't have sufficient evidence to disprove the claim) → z.test() in the BSDA package

Example 3. A quality inspector finds 10 defective parts in a sample of 500 parts received from manufacturer A. Out of 400 parts from manufacturer B, she finds 12 defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the 5% level of significance, do we have enough evidence to disprove this claim?

- 2 populations: parts from A and parts from B
- 2 samples: 500 parts from A ($n = 500$)
  400 parts from B ($m = 400$)
- sample proportions of defective items: $\hat{p}_1 = \frac{10}{500}$, $\hat{p}_2 = \frac{12}{400} = \frac{3}{100}$
- $\alpha = 0.05$

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2 \Rightarrow R = (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$



• p_1, p_2 - population proportions

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n} + \frac{1}{m})}}$$

$$\hat{p} = \frac{n\hat{p}_1 + m\hat{p}_2}{n+m} = \frac{10+12}{500+400} = \frac{22}{900} = \frac{11}{450}$$

$$Z = \frac{0.02 - 0.03}{\sqrt{\frac{11}{450} \cdot \frac{439}{450} \cdot \left(\frac{1}{500} + \frac{1}{400}\right)}} = \frac{-0.01}{\frac{1}{450} \cdot \sqrt{439 \cdot 11 \cdot \frac{900}{400 \cdot 500}}} = \frac{-0.01 \cdot 450}{\frac{30}{200} \cdot \sqrt{\frac{439 \cdot 11}{5}}} = \frac{-0.01 \cdot \overset{15}{\cancel{450}} \cdot 200\sqrt{5}}{\cancel{30} \cdot \sqrt{11 \cdot 439}} \Leftrightarrow$$

$$Z = -\frac{30\sqrt{5}}{\sqrt{11 \cdot 439}} = -0.97$$

$-z_{\alpha/2} = \text{qnorm}(\alpha/2) = -1.96 \Rightarrow z_{\alpha/2} = 1.96 \Rightarrow R = (-\infty, -1.96) \cup (1.96, \infty)$

$\Rightarrow z \notin R \Rightarrow$ we cannot reject H_0

(we don't have enough evidence to disprove the claim)

(the quality is the same)

Example 4. (Unauthorized use of a computer account, continued). A long-time authorized user of the account makes 0.2 seconds between keystrokes. One day, the data in Example 9.19 on p. 260 are recorded as someone typed the correct username and password. At a 1% level of significance, is this an evidence of an unauthorized attempt?

.24, .22, .26, .34, .35, .32, .33, .29, .19, .36, .30, .15, .17, .28, .38, .40, .37, .27

$\alpha = 0.01$

$H_0: \mu = 0.2$

$H_a: \mu \neq 0.2 \Rightarrow R(-\infty, -t_{\alpha/2}) \cup (t_{\frac{\alpha}{2}}, \infty)$

$(\mu > 0.2) \Rightarrow R = (t_\alpha, \infty)$

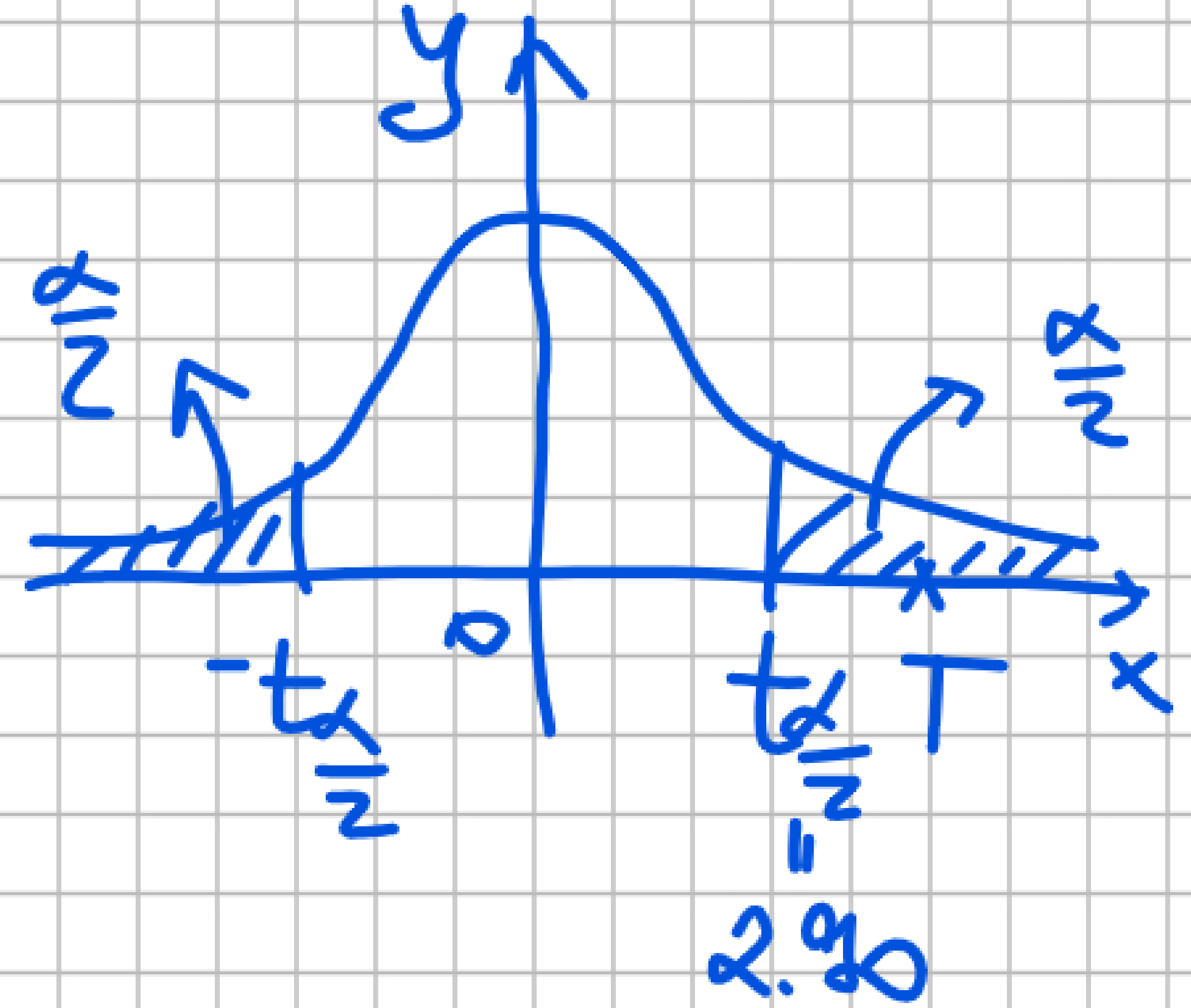
- population: time between keystrokes for a user
- sample size: $n = 18$
- μ - population mean (average time between keystrokes)

$T = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$ - has a Student t distribution with 17 degrees of freedom

$\mu_0 = 0.2$

$$\bar{X} = \frac{0.24+0.22+\dots+0.27}{18} = 0.29$$

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sqrt{\frac{1}{17}\sum_{i=1}^{18}(x_i-\bar{x})^2} = 0.07$$

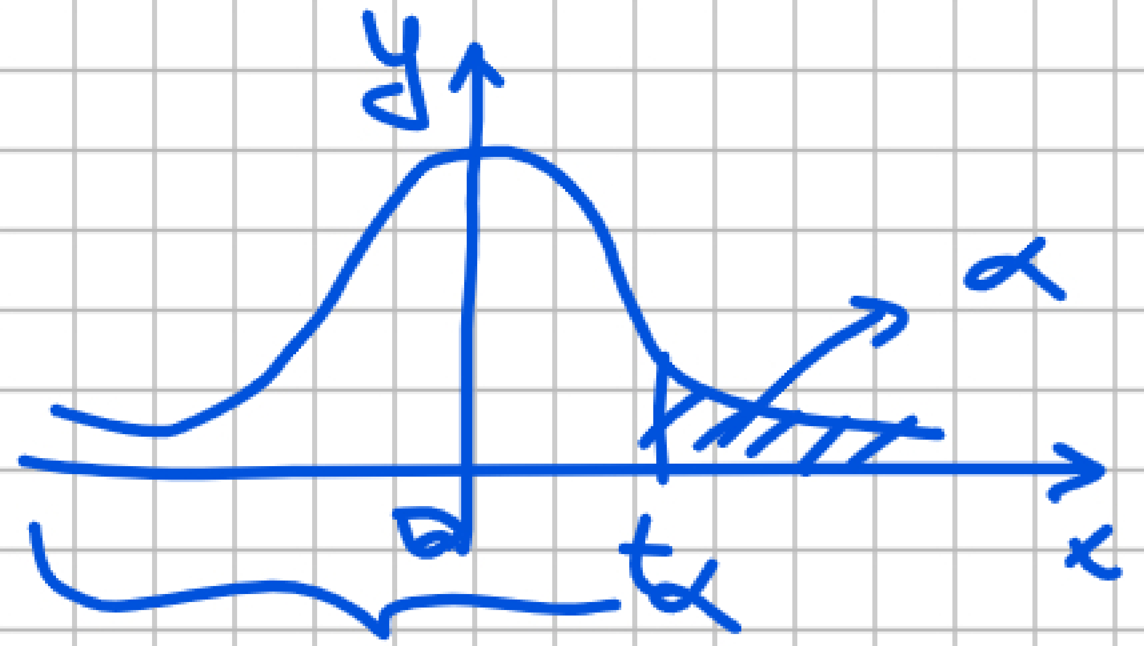


$$T = \frac{0.29-0.2}{\frac{0.07}{\sqrt{18}}} = \frac{0.09\cdot 3\sqrt{2}}{0.07\underline{396}} = \frac{27\sqrt{2}}{7.396} = \cancel{5.45}\ \left(\underline{\underline{5.16}}\right)$$

$$-t_{\frac{\alpha}{2}} = qt(\alpha/2, 17) = -2.90 \Rightarrow t_{\alpha/2} = 2.90$$

$$R = (-\infty, -2.90) \cup (2.90, \infty)$$

$\Rightarrow T \in R \Rightarrow H_0$ is rejected (the user is an unauthorized one)



- $H_0: \mu = 0.2$

  $H_a: \mu > 0.2 \Rightarrow R = (t_\alpha, \infty)$

  $t_\alpha = qt(1-\alpha, 17) = 2.57$

  $\Rightarrow R = (2.57, \infty)$

  $T = \underline{5.45} \in R \Rightarrow H_0$ is reject

In R: t.test()

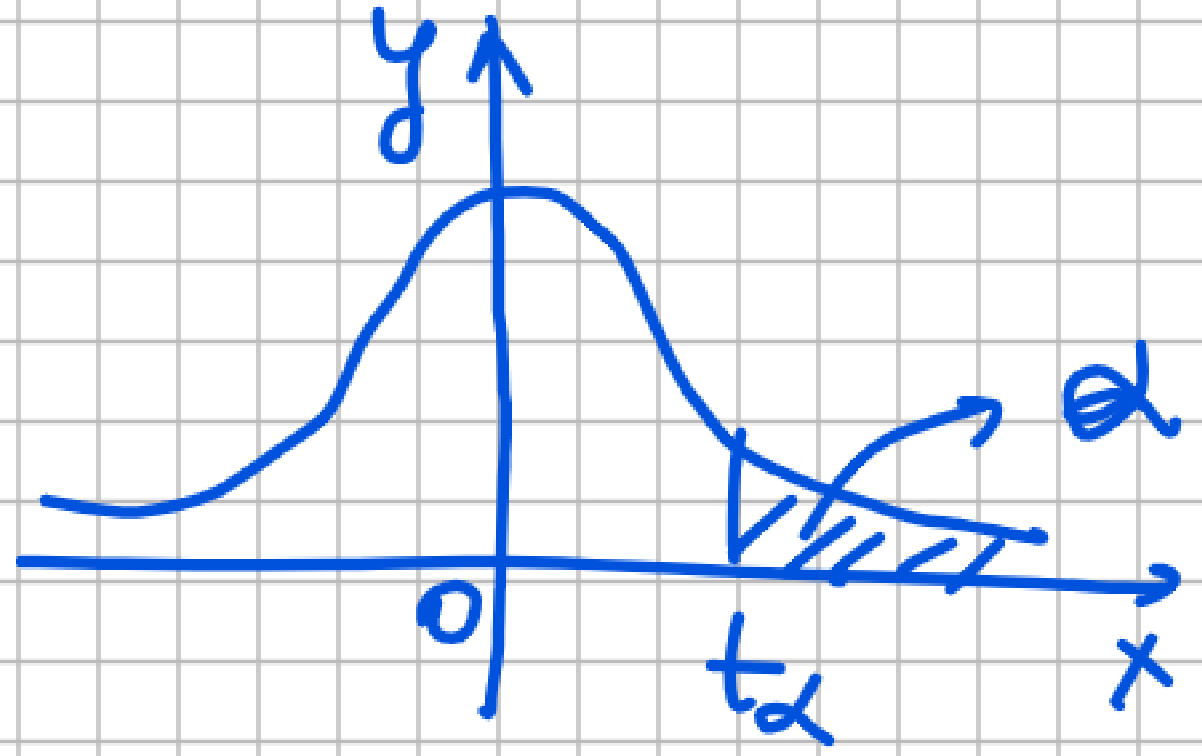
t.test(time, mu= 0.2, conf.level= 0.99)

p-value $= 7.81 \cdot 10^{-5} \ll \alpha = 0.01 \Rightarrow H_0$ is rejected

p-value $< \alpha \Rightarrow H_0$ is rejected

$> \alpha \Rightarrow H_0$ is not rejected

Example 5. (CD writer and battery life). Does a CD writer consume extra energy, and therefore, does it reduce the battery life on a laptop? The data collected is the following: eighteen users without a CD writer worked an average of 5.3 hours with a standard deviation of 1.4 hours; other twelve, who used their CD writer, worked an average of 4.8 hours with a standard deviation of 1.6 hours. Consider a level of significance $\alpha = 0.1$



- $n = 18$, $m = 12$ (the 2 sample sizes)

$\bar{X} = 5.3$

$\bar{Y} = 4.8$

$S_x = 1.4$, $S_y = 1.6$

$\alpha = 0.1$

$\sigma_1 = \sigma_2$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2 \Rightarrow R = (t_\alpha, \infty)$

$$T = \frac{\bar{X} - \bar{Y}}{S_p \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}}$$ has a Student t distribution with $n+m-2$ df

$$S_p = \sqrt{\frac{1}{n+m-2}\left((n-1)S_x^2 + (m-1)S_y^2\right)} = 1.48$$

$$T = \frac{5.3 - 4.8}{S_p \cdot \sqrt{\frac{1}{18} + \frac{1}{12}}} = 0.91$$

$$t_\alpha = qt(1-\alpha, \underbrace{18+12-2}_{28}) = 1.31$$

$T = 0.91 \notin R = (1.31, \infty) \Rightarrow$ we cannot reject H_0

sgr. 2

9.10. We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it.

(a) Construct a ~~96%~~ 95% confidence interval for the proportion of defective items in the whole shipment.

(b) The manufacturer claims that at most one in 10 items in the shipment is defective. At the ~~4%~~ 5% level of significance, do we have sufficient evidence to disprove this claim? Do we have it at the ~~15%~~ 1% level?

$\alpha = 5\% = 0.05$

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

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

$n = 200$ (sample size)

24 defective items

$\alpha = 0.05$

$H_0: p = \frac{1}{10}$ $\left(p_0 = \frac{1}{10}\right)$

$H_a: p \neq \frac{1}{10} \Rightarrow R = (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$

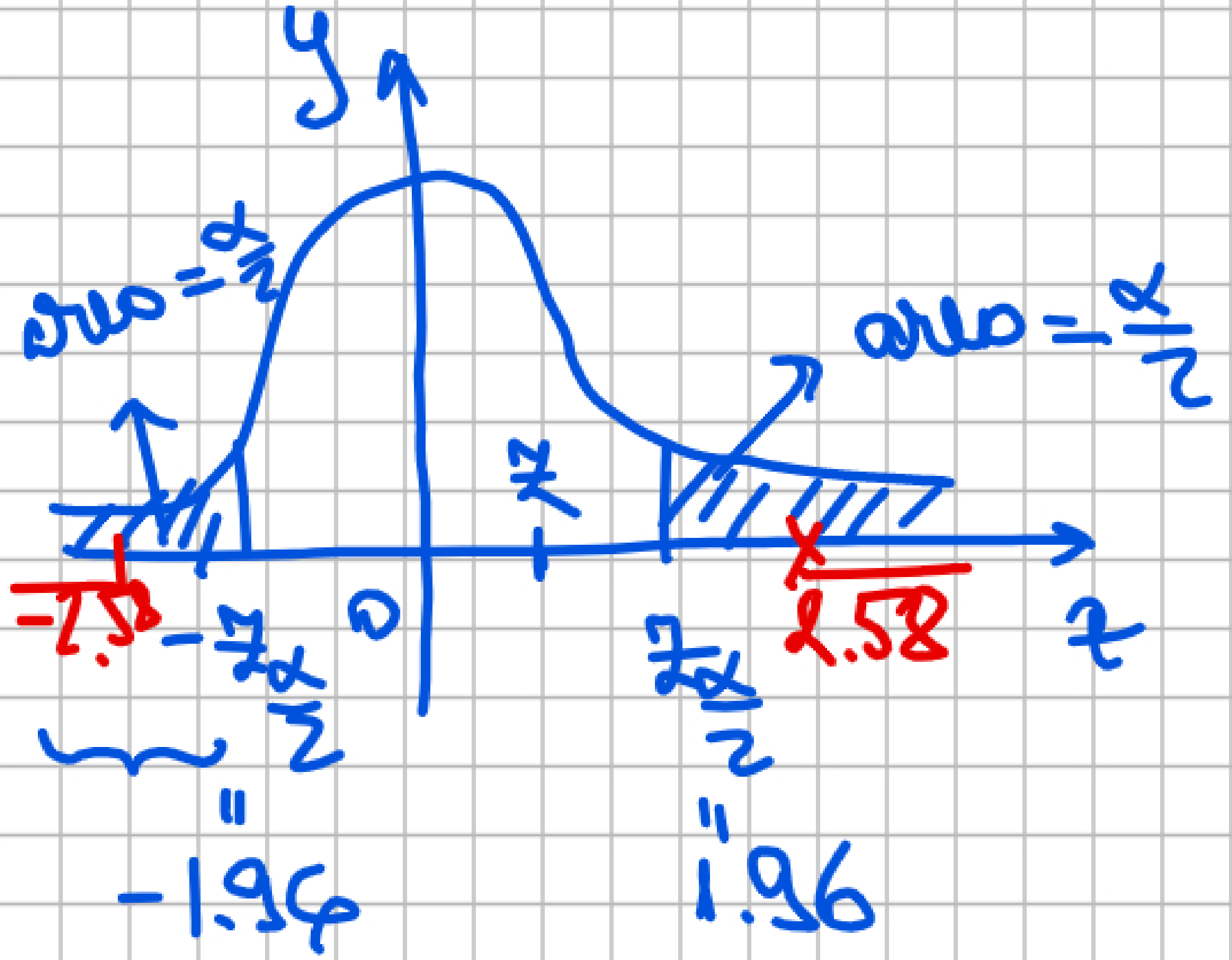
$p > \frac{1}{10} \Rightarrow R = (z_\alpha, \infty)$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

$\hat{p}$ - sample proportion (of defective items)

$\hat{p} = \frac{24}{200} = \frac{3}{25}$, $z = \frac{0.12 - 0.10}{\sqrt{\frac{0.10 \cdot 0.90}{200}}} = \frac{0.02 \cdot 10\sqrt{2}}{0.3} = \frac{0.2\sqrt{2}}{0.3} =$

$= \frac{2\sqrt{2}}{3} = 0.94$



$$-z_{\frac{\alpha}{2}} = qnorm(\alpha/2) = -1.96$$

$$R = (-\infty, -1.96) \cup (1.96, \infty)$$

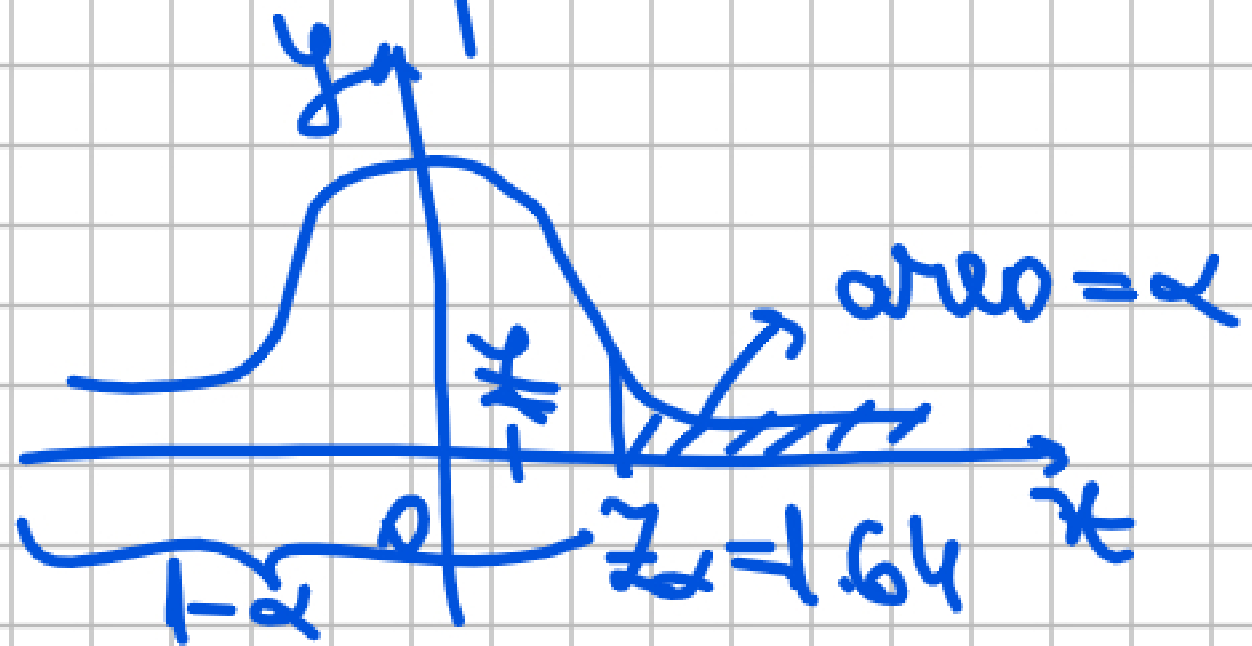
$$z = 0.94$$

$\Rightarrow z \notin R \Rightarrow$ we cannot reject H_0

$\Rightarrow$ the manufacturer's claim is true

(there isn't suff. evidence to disprove the claim)

- $H_a: p > 0.1 \Rightarrow R = (z_\alpha, \infty) = (1.64, \infty)$



$$z_\alpha = qnorm(1-\alpha) = 1.64$$

$z = 0.94 \Rightarrow z \notin R \Rightarrow H_0$ is not rejected

$\alpha = 1\% = 0.01$

$H_a: p \neq \frac{1}{10} \Rightarrow R = (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty) = (-\infty, -2.58) \cup (2.58, \infty)$

$-z_{\frac{\alpha}{2}} = \text{qnorm}(\alpha/2) = -2.58$

$z = 0.94 \Rightarrow z \notin R \Rightarrow$ we cannot reject H_0

$\alpha = 10\% \Rightarrow R = (-\infty, -1.64) \cup (1.64, \infty)$

(a) 95% confidence interval ≡ acceptance region at the 5% level of significance (for the two-tailed test)

$P\left(Z \in \left(-z_{\frac{\alpha}{2}}, z_{\frac{\alpha}{2}}\right)\right) = 1-\alpha \iff P\left(\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$

95% CI : $\left(\hat{p} - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}, \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}\right) = \left(0.3 - 1.96 \cdot \frac{0.03}{\sqrt{2}}, 0.3 + 1.96 \cdot \frac{0.03}{\sqrt{2}}\right) =$

$\alpha = 1 - 95\% = 5\% \ (=0.05)$

$= (0.26, 0.34)$

$$\frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \frac{0.3}{10\sqrt{2}} = \frac{0.03}{\sqrt{2}}$$

$(a, b) \qquad (a, \infty), (-\infty, b)$

9.11. Refer to Exercise 9.10. Having looked at the collected sample, we consider an alternative supplier. A sample of 150 items produced by the new supplier contains 13 defective items.

Is there significant evidence that the quality of items produced by the new supplier is higher than the quality of items in Exercise 9.10? What is the P-value?

p_1, p_2 – population proportions of defective items

p_1 – prop. of def. items for previous supplier

p_2 – –"– of the new supplier

2 samples: $n_1 = 200$, $\hat{p}_1 = \frac{24}{200} = \frac{3}{25} = 0.12$

$n_2 = 150$, $\hat{p}_2 = \frac{13}{150} = 0.09$

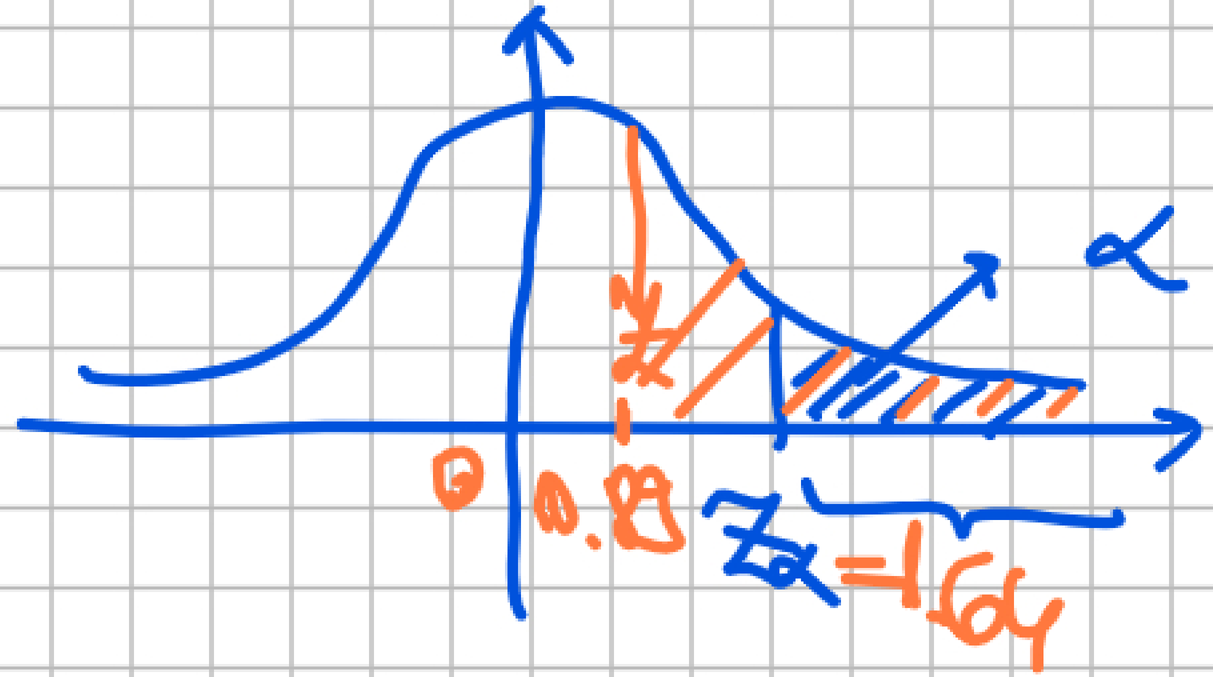
$H_0: p_1 = p_2$

$H_a: p_1 > p_2 \Rightarrow R = (z_\alpha, \infty)$

$\hat{p} = \frac{24+13}{200+150} = \frac{37}{350} = 0.106$

$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.12 - 0.09}{\sqrt{0.11 \cdot 0.89 \cdot \left(\frac{1}{200} + \frac{1}{150}\right)}} = 0.89$

$\hat{p}$ – pooled sample proportion



$z_\alpha = \text{qnorm}(1-\alpha) = 1.64$

$z = 0.89, \quad R = (1.64, \infty)$

$\Rightarrow z \notin R \Rightarrow H_0$ cannot be reject (there isn't enough evidence to support the claim that the new supplier has a better quality)

p-value = area under the PDF from z to ∞

$= 1 - \text{pnorm}(z) = \text{pnorm}(z, \text{lower.tail} = F) = 0.19$

If p-value $< \alpha \Rightarrow H_0$ is rejected

p-value $> \alpha \Rightarrow H_0$ is not rejected

p-val $= 0.19 > \alpha = 0.05 \Rightarrow$

H_0 is not rejected

9.9. Salaries of entry-level computer engineers have Normal distribution with unknown mean and variance. Three randomly selected computer engineers have salaries (in \$ 1000s):

$$30, 50, 70$$

(a) Construct a 90% confidence interval for the average salary of an entry-level computer engineer.

(b) Does this sample provide a significant evidence, at a 10% level of significance, that the average salary of all entry-level computer engineers is different from \$80,000? Explain.

(b) $n=3$ (one sample t-test)

$\alpha = 10\% = 0.1$

$H_0: \mu = 80\,000$

$H_a: \mu \neq 80\,000 \Rightarrow R = (-\infty, -t_{\frac{\alpha}{2}}) \cup (t_{\frac{\alpha}{2}}, \infty)$

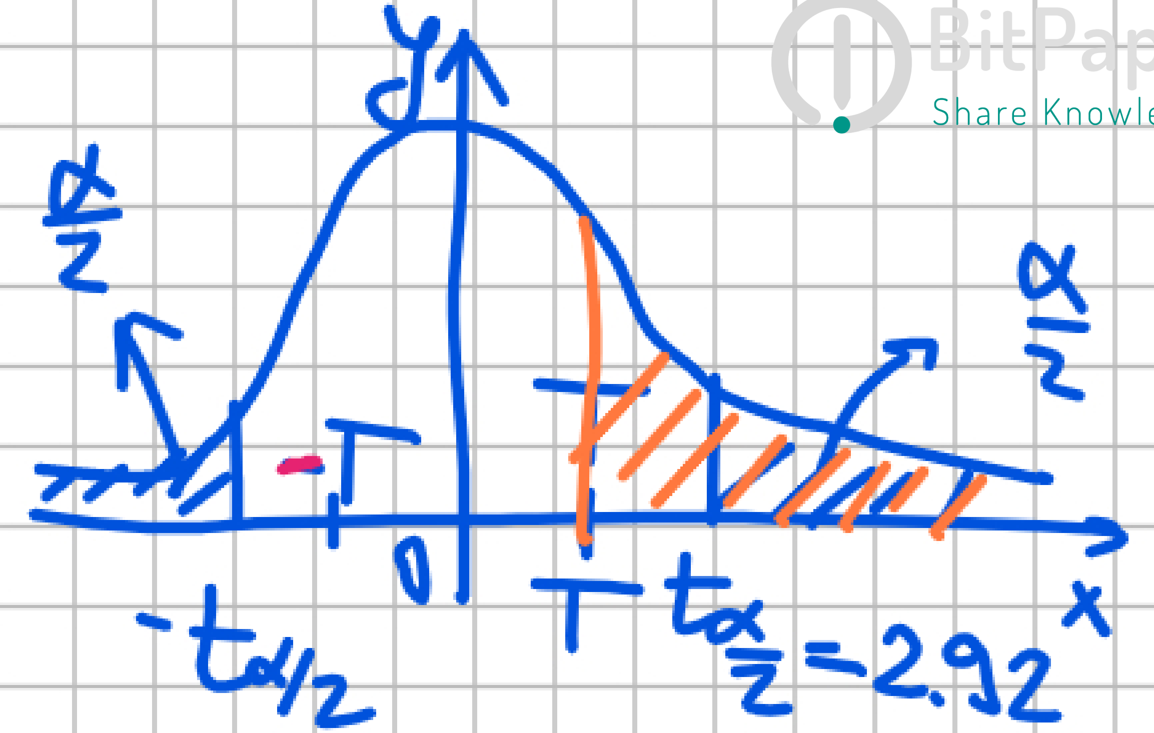
$T = \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}$ has a Student t distribution with $n-1$ degrees of freedom

$\mu_0 = 80000$, S - sample std. deviation

$\overline{X}$ - sample mean

$\overline{X} = 50\,000$

$s = 20\,000$



$$T = \frac{50\,000 - 80000}{\frac{20000}{\sqrt{3}}} = \frac{-30\,000\sqrt{3}}{20000} = -\frac{3\sqrt{3}}{2} = 2.60$$

$-t_{\frac{\alpha}{2}} = qt(\alpha/2, n-1) = -2.92$

$R = (-\infty, -2.92) \cup (2.92, \infty)$

$T = -2.60 \notin R \Rightarrow$ we cannot reject H_0 (the sample doesn't provide significant evidence that the salary is different from $80000 (at the 10% level of significance))

$$p\text{-value} = 2 * pt(\overset{2.60}{\overset{\shortparallel}{T}}, 2, lower.tail=F) = \underline{\underline{0.12}} > \alpha = 0.10 \Rightarrow H_0 \text{ cannot be rejected}$$

In R: t.test()

```
salary=c(30,50,70)
t.test(salary, mu = 80, conf.level= 0.90)
```

Exercise 91. The mean length of time required for returning students to register at our university has been 90 minutes, and the standard deviation is 15 minutes. A new "faster" registration procedure was introduced this semester. The students seem to think the registration people "blew it". The students collected a sample of size $n = 45$ and found the average registration time $\bar{x} = 95.6$. Does the sample show sufficient evidence, at the 0.05 level of significance, to support their contention?

- population: registration time

  $\sigma = 15 \Rightarrow z\text{-test}$
- sample: $n = 45$

  $\bar{X} = 95.6$

$\alpha = 0.05$

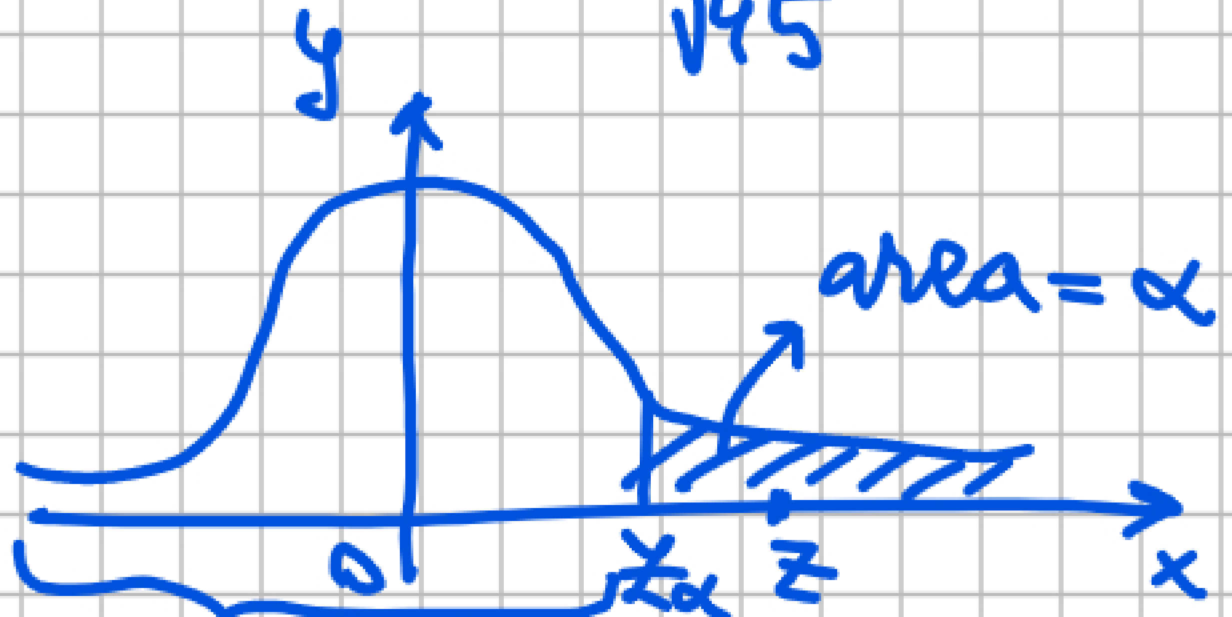
$H_0: \mu = 90 \quad (\mu_0 = 90)$

$H_a: \mu > 90 \quad \Rightarrow R = (z_\alpha, \infty)$

μ - population mean (mean registration time)

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \text{ - has a } N(0,1)$$

$$z = \frac{95.6 - 90}{\frac{15}{\sqrt{45}}} = \frac{5.6 \cdot \cancel{3}\sqrt{5}}{\cancel{15}_{5}} = \frac{5.6 \cdot \sqrt{5}}{5} = 2.50$$



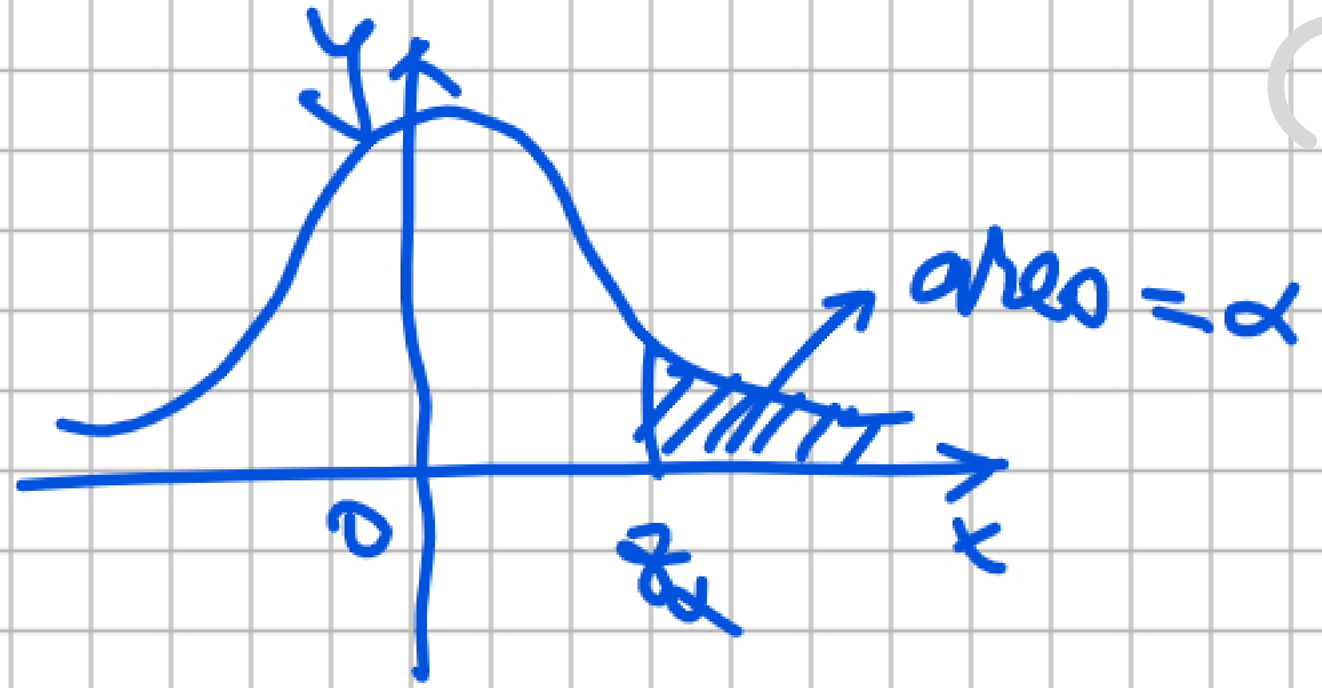
$z_\alpha = \text{qnorm}(1-\alpha) = 1.64$

$R = (1.64, \infty)$

$z = 2.50 \in R \Rightarrow H_0$ is reject

(the sample shows suff. evidence to support the claim)

Example 2. We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it. The manufacturer claims that at most one in 10 items in the shipment is defective. At the 1% level of significance, do we have sufficient evidence to disprove this claim?



$n = 200$

24 defective items

$\alpha = 1\% = 0.01$

$H_0: p = \frac{1}{10} = \underline{10\%}$

$H_a: p > \frac{1}{10} \Rightarrow R = (z_\alpha, \infty)$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1)$$

• population: items produce

p – proportion of defective items

$\hat{p}$ – sample proportion

$\hat{p} = \frac{24}{200} = \frac{12}{100} = \underline{12\%}\ (0.12)$

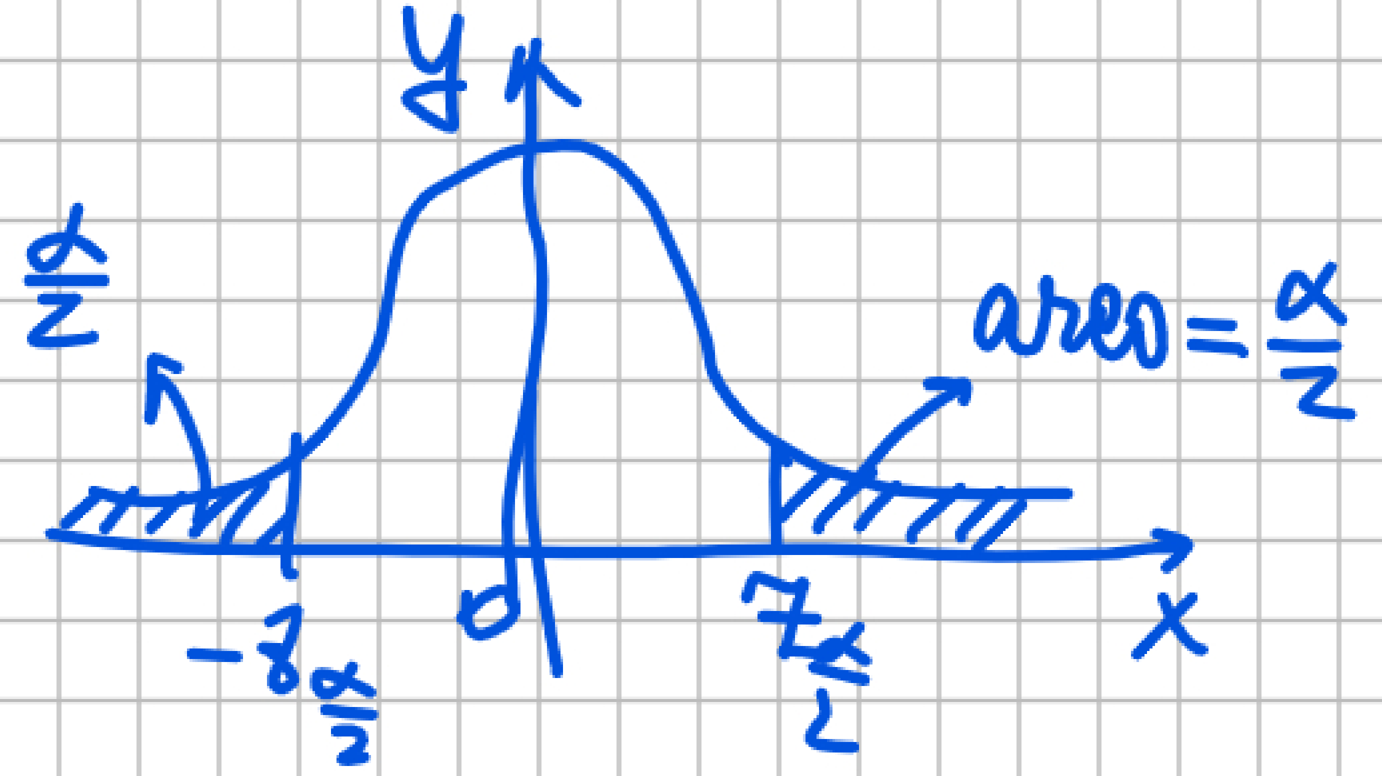
$$Z = \frac{0.12 - 0.1}{\sqrt{\frac{0.1 \cdot 0.9}{200}}} = \frac{0.02 \cdot \sqrt{200}}{\sqrt{0.1 \cdot 0.9}} = \underline{0.94}$$

$z_\alpha = \text{qnorm}(1-\alpha) = 2.33 \rightarrow R = (2.33, \infty)$

$Z \notin R \Rightarrow H_0$ is not rejected (we don't have suff. evidence to disprove the claim)

$\underline{\alpha} = P(\text{Type I error})$
$= P(H_0 \text{ is rejected} \mid H_0 \text{ is true})$

Example 3. A quality inspector finds 10 defective parts in a sample of 500 parts received from manufacturer A. Out of 400 parts from manufacturer B, she finds 12 defective ones. A computer-making company uses these parts in their computers and claims that the quality of parts produced by A and B is the same. At the 5% level of significance, do we have enough evidence to disprove this claim?



- 2 populations : items produced by A
  items produced by B
- 2 samples : $n=500$, $m=400$

$$\left.\begin{array}{l}\hat{p}_1 = \frac{10}{500} = \frac{2}{100} = 0.02 \\ \hat{p}_2 = \frac{12}{400} = \frac{3}{100} = 0.03\end{array}\right\} \text{sample proportions of defective parts}$$

$\alpha = 5\% = 0.05$

$H_0: p_1 = p_2$

$H_a: p_1 \neq p_2 \Rightarrow R = (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$

$-z_{\frac{\alpha}{2}} = \text{qnorm}\left(\frac{\alpha}{2}\right) = -1.96 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$

$R = (-\infty, -1.96) \cup (1.96, \infty)$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim N(0,1)$$

$$\hat{p} = \frac{10+12}{500+400} = \frac{22}{900} = 0.024$$

$$Z = (0.02 - 0.03)/\sqrt{0.024 \cdot 0.876 \cdot \left(\frac{1}{500} + \frac{1}{400}\right)}$$

$Z = -0.97$

$R = (-\infty, -1.96) \cup (1.96, \infty)$

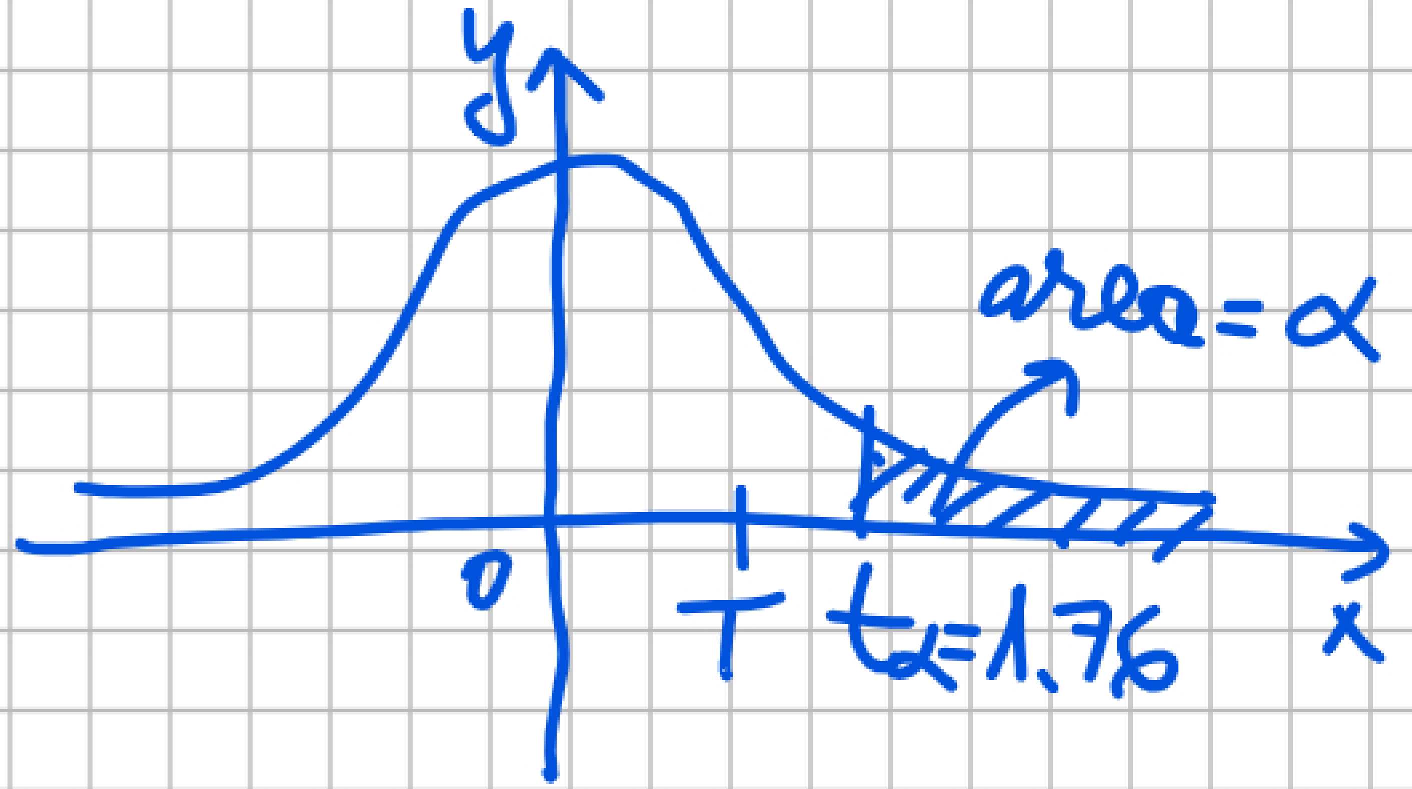
$Z \notin R \Rightarrow H_0$ is not reject

(At the 5% level of significance we don't have suff. evidence to disprove the claim)

Exercise 104. In a large cherry orchard, the average yield has been 4.35 tons per acre for the last several years. A new fertilizer was tested on 15 randomly selected 1-acre plots. The yields from these plots follow:

3.56 5.00 4.88 4.93 3.92 4.25 5.12 5.13 4.79 4.45 5.35 4.81 3.48 4.45 4.72

At the 0.05 level of significance, does this sample show sufficient evidence to claim that there was a significant increase in production?



$n = 15$

$\alpha = 0.05$

$H_0: \mu = 4.35 \quad (\mu_0 = 4.35)$

$H_a: \mu > 4.35 \Rightarrow R = (t_\alpha, \infty)$

$T = \dfrac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$ has a Student t distribution with $n-1$ degrees of freedom

s - sample standard deviation

$\bar{x} = 4.59$

$s = 0.57$

$T = \dfrac{4.59 - 4.35}{\frac{0.57}{\sqrt{15}}} = 1.63$

$t_\alpha = qt^{14}(1-\alpha) = 1.76$

$R = (1.76, \infty)$

$T \notin R \Rightarrow H_0$ is not rejected

(the sample doesn't show suff. evidence to support the claim at 5% level of sign.)

Example 6. (Comparison of two servers, continued). A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results. Is server A faster? Formulate and test the hypothesis at a level $\alpha = 0.05$.

	Server A	Server B
Sample mean	6.7 min	7.5 min
Sample standard deviation	0.6 min	1.2 min

$n = 30$, $m = 20$

$\alpha = 0.05$

$\bar{X} = 6.7$, $\bar{Y} = 7.5$ (sample means)

$s_X = 0.6$, $s_Y = 1.2$ (sample std. deviations)

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2 \Rightarrow R = (-\infty, -t_\alpha)$

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$ has a Student t distribution with ν df

$$\nu = \left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2 \Big/ \left(\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}\right)$$

$$T = \frac{6.7 - 7.5}{\sqrt{\frac{0.6^2}{30} + \frac{1.2^2}{20}}} = -2.76$$

$\nu = 25.40 \to 25$

$-t_\alpha = qt(\alpha, \nu) = -1.71$

$R = (-\infty, -1.71)$

$T \in R \Rightarrow H_0$ is rejected

(we have enough evidence at the 5% level of signif. to support the claim that A is faster) sgr. 1

Exercise 93. Waiting times, in hours, at a popular restaurant are believed to be approximately normally distributed with a variance of 2.25 hours during busy periods(*). (a) A sample of 20 customers revealed a mean waiting time of 1.52 hours. Construct the 95% confidence interval for the estimated of the population mean. (b) Suppose that the mean of 1.52 hours had resulted from a sample of 32 customers. Find the 95% confidence interval. What effect does a larger sample size have on the confidence interval?

(*) and a mean of 1.50 h.

c) Does the sample show sufficient evidence to support the claim of the restaurant owner at the 5% level of significance?

c) $H_0: \mu = 1.50$ $(\mu_0 = 1.50)$

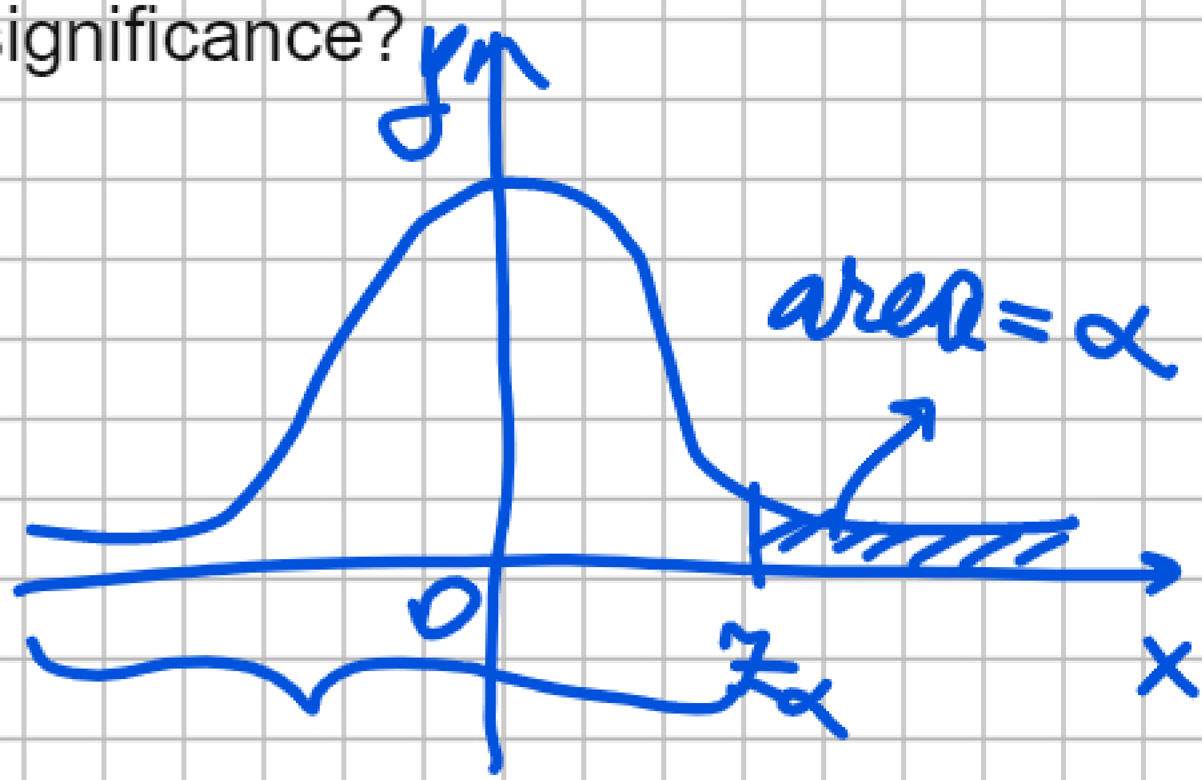
$H_a: \mu > 1.50 \Rightarrow R = (z_\alpha, \infty)$

$(\mu \neq 1.50)$

$\alpha = 5\% = 0.05$

σ - is known $\Rightarrow$ z-test

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$



$\sigma^2 = 2.25$ (population variance)

- population: waiting times at a restaurant
- sample: $n = 20$

$\bar{X} = 1.52$ (sample mean)

$$z = \frac{1.52 - 1.50}{\frac{\sqrt{2.25}}{\sqrt{20}}} = \frac{0.02 \cdot 2\sqrt{5}}{1.5} = \frac{4\sqrt{5}}{150} = 0.06$$

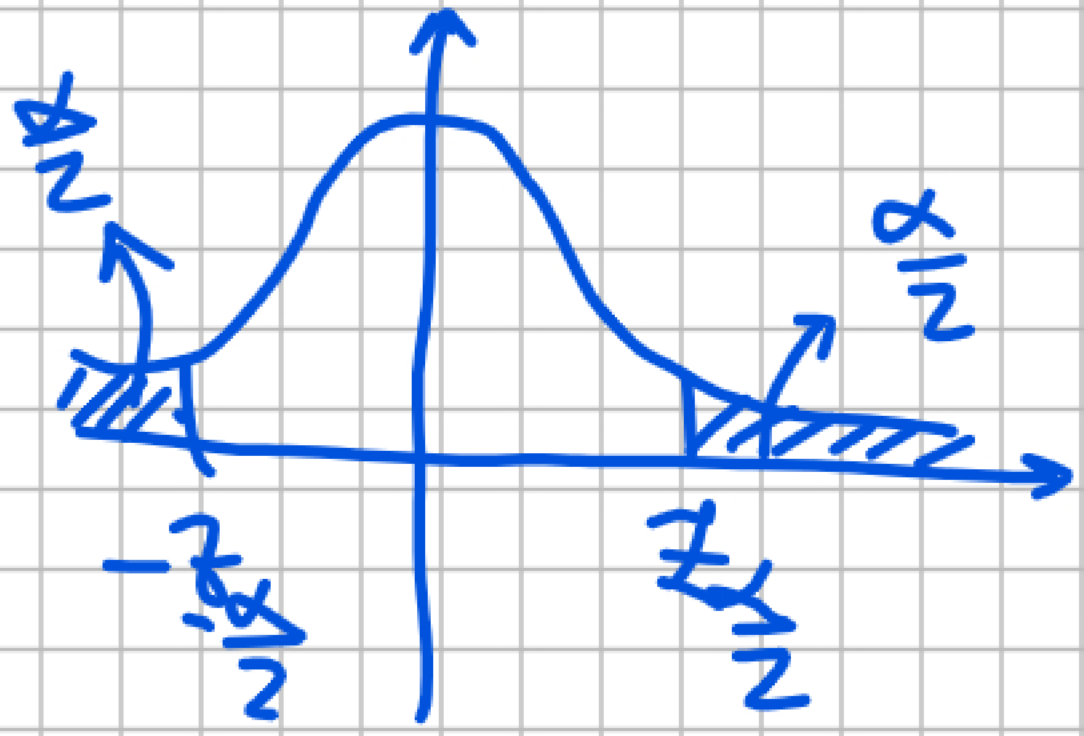
$z_\alpha = \text{qnorm}(1-\alpha) = 1.64$

$\rightarrow R = (1.64, \infty)$

$\Rightarrow Z = 0.36 \notin R = (1.64, \infty) \Rightarrow H_0$ is not rejected

(the sample does not show suff. evidence to disprove the claim of the restaurant owner at the 5% level of signif.)

a) 95% CI : $\left[\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = [0.86, 2.18]$



$-z_{\frac{\alpha}{2}} = \text{qnorm}(\alpha/2) = -1.96 \Rightarrow z_{\frac{\alpha}{2}} = 1.96$

$\alpha = 1 - 0.95 = 0.05$

$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 1.52 - 1.96 \cdot \frac{1.5}{\sqrt{20}} = 0.86$

9.16. A sample of 250 items from lot A contains 10 defective items, and a sample of 300 items from lot B is found to contain 18 defective items.

(a) Construct a 98% confidence interval for the difference of proportions of defective items.

(b) At a significance level $\alpha = 0.02$, is there a significant difference between the quality of the two lots?

c) the manufacturer of lot A claims that the proportion of defective items produced is at most 2%. Do we have enough evidence to disprove this claim at the 5% level of significance?

c) • population : items produced by company A (lot A)

p - proportion of defective items of the population

• sample : $n=250$

10 defective items in the sample

$\hat{p}$ - sample proportion of defective items

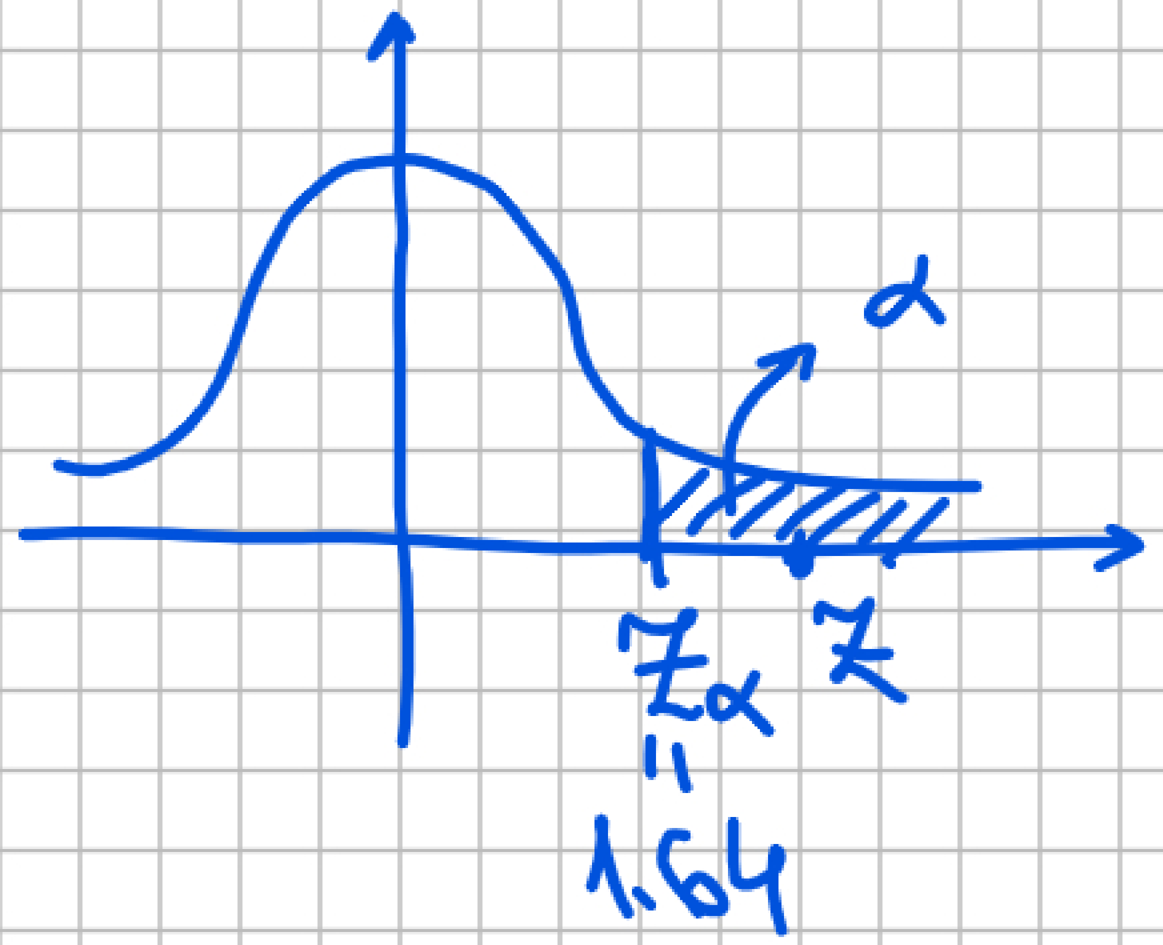
$\alpha = 5\% = 0.05$

$H_0 : p = 0.02 \ (p_0 = 0.02)$

$H_a : p > 0.02 \Rightarrow R = (z_\alpha, \infty)$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0,1), \quad \hat{p} = \frac{10}{250} = \frac{1}{25} = 0.04$$

$$z = \frac{0.04 - 0.02}{\sqrt{0.02 \cdot 0.98 / 250}} = 2.26$$



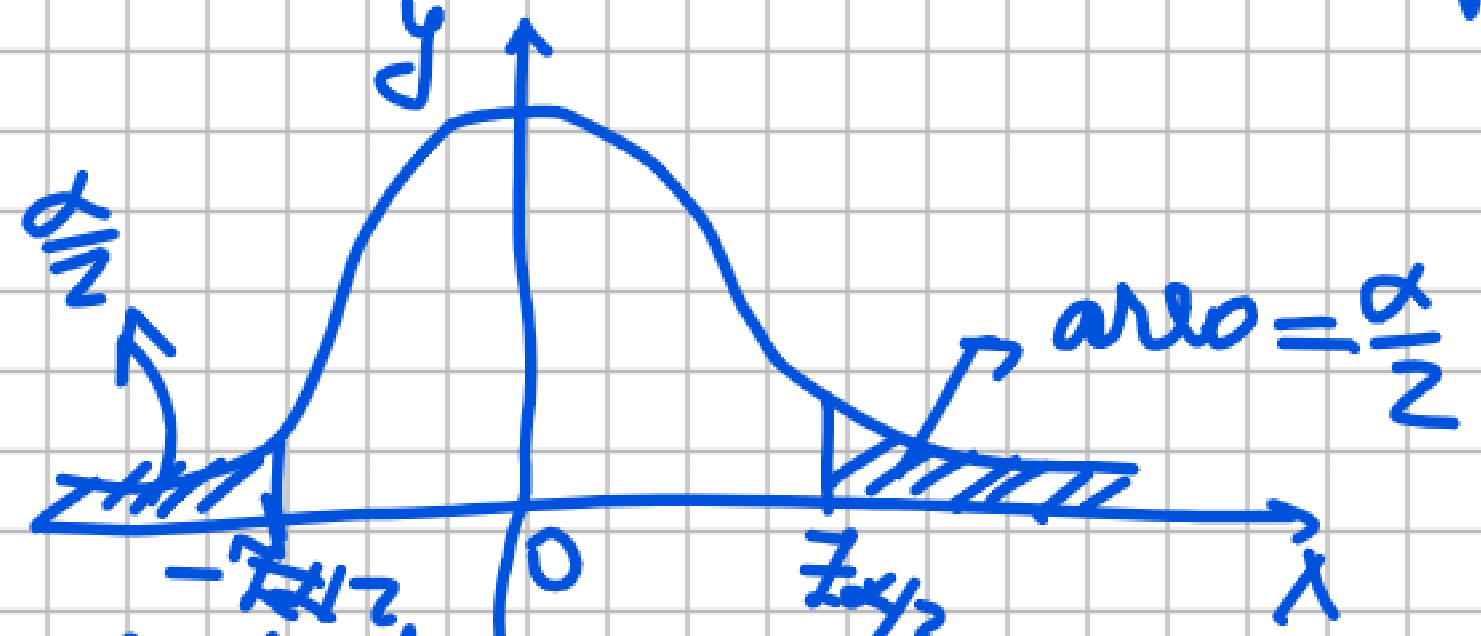
$z_\alpha = \text{qnorm}(1-\alpha) = 1.64$

$z = 2.26 \in R = (1.64, \infty) \Rightarrow H_0$ is rejected

(the sample shows suff. evidence to disprove the claim of the manufacturer at the 5% level of signif.)

b) $H_0: p_1 = p_2$

$H_a: p_1 \neq p_2 \Rightarrow R = (-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$



- 2 populations: lot A & lot B $\rightarrow$ proportions of def. items (p_1 & p_2)
- 2 sample: $n = 250 \rightarrow 10$ defective items

  $m = 300 \rightarrow 18$ defective items

$\alpha = 0.02$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim N(0,1)$$

$\hat{p}_1, \hat{p}_2$ - sample prop. of defective items

$\hat{p}_1 = \frac{10}{250}$, $\hat{p}_2 = \frac{18}{300}$

$\hat{p}$ - pooled proportion of defective items

$$\hat{p} = \frac{10+18}{250+300} = \frac{28}{550} = 0.051$$

$\hat{p}_1 = 0.04$, $\hat{p}_2 = 0.06$

$$Z = \frac{0.04 - 0.06}{\sqrt{0.051 \cdot 0.849 \cdot \left(\frac{1}{250} + \frac{1}{300}\right)}}$$

$Z = -1.06$

$-z_{\frac{\alpha}{2}} = \text{qnorm}\left(\frac{\alpha}{2}\right) = -2.33$

$R = (-\infty, -2.33) \cup (2.33, \infty)$

$Z \notin R \Rightarrow H_0$ is not rejected

(the samples do not show suff. evidence to say the quality is different at the 2% level of significance)

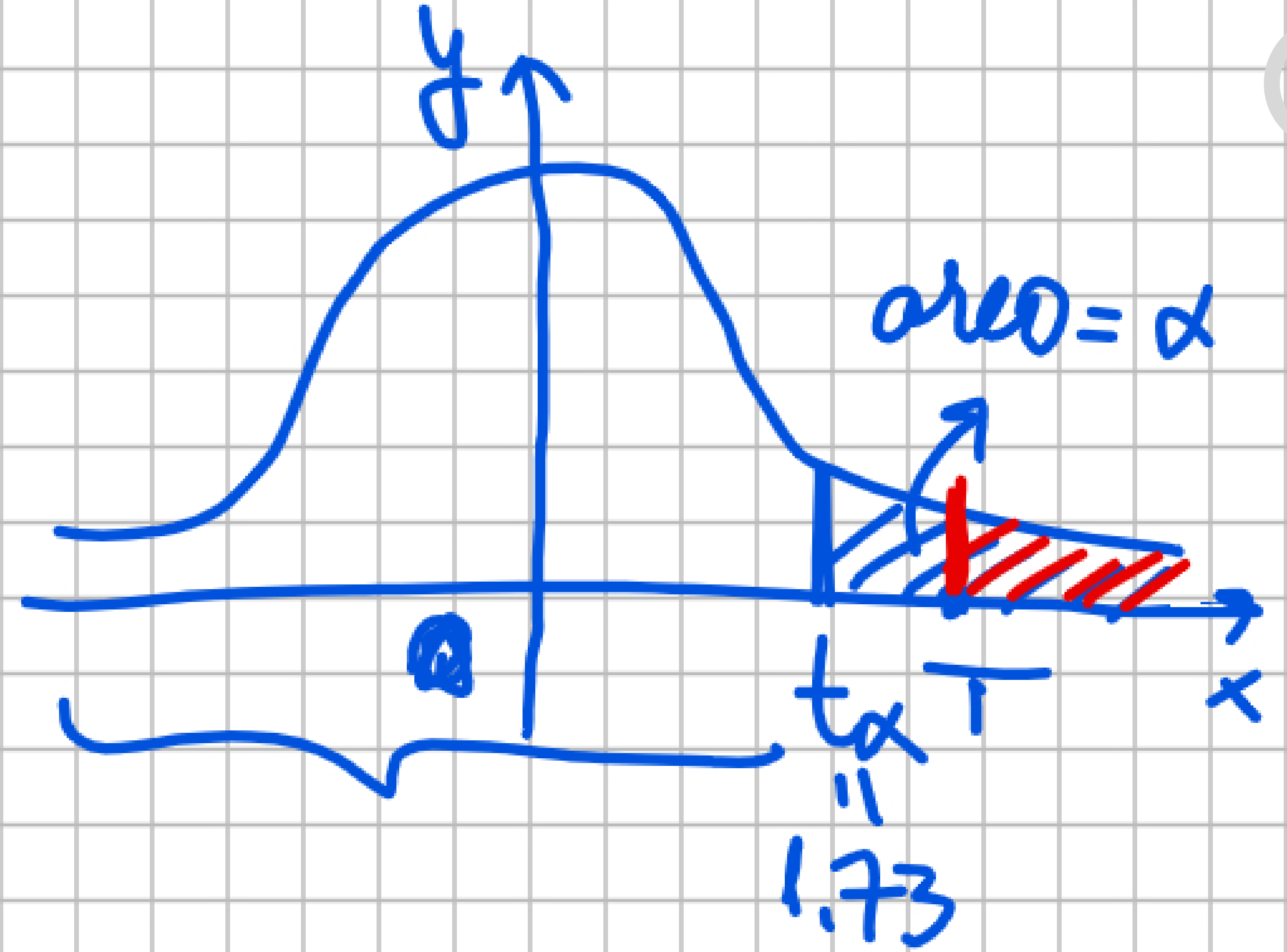
Exercise 102. It has been suggested that abnormal human males tend to occur more in children born to older-than-average parents. Case histories of 20 abnormal males were obtained and the ages of 20 mothers were:

31 21 29 28 34 45 21 41 27 31 43 21 39 38 32 28 37 28 16 39

The mean age at which mothers in the general population give birth is 28 years.

a. Compute the mean and the standard deviation of the sample.

b. Does the sample show sufficient evidence to support the claim that abnormal male children have older-than-average mothers? Use $\alpha = 0.05$.



$m = 20$

$\bar{X} = \frac{31 + \dots + 39}{20}$

S - std. deviation of the sample

σ - unknown $\Rightarrow$ t test

$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{m}}}$ has a Student t distribution with $m-1$ degrees of freedom

$H_0: \mu = 28$ $(\mu_0 = 28)$

$H_a: \mu > 28$ $\Rightarrow R = (t_\alpha, \infty)$

$\bar{X} = 31.45$, $s = 8.05$

$T = 1.92$

$t_\alpha = qt(1-\alpha, m-1) = 1.73 \rightarrow R = (1.73, \infty)$

$T \in R \Rightarrow H_0$ is rejected

(the sample shows suff. evidence to support the claim at 5% level of significance)

In R: t.test()

z.test() in the BSDA package

t.test(age, mu=28, alternative="greater", conf.level=0.95)

If p-value $< \alpha \Rightarrow H_0$ is rejected

p-value $> \alpha \Rightarrow H_0$ is not rejected

p-value = pt(T, n-1, lower.tail = F)

= 1 - pt(T, n-1)

p-val = 0.035 < 0.05 = $\alpha \Rightarrow$

H_0 is rejected

Example 5. (CD writer and battery life). Does a CD writer consume extra energy, and therefore, does it reduce the battery life on a laptop? The data collected is the following: eighteen users without a CD writer worked an average of 5.3 hours with a standard deviation of 1.4 hours; other twelve, who used their CD writer, worked an average of 4.8 hours with a standard deviation of 1.6 hours. Consider a level of significance $\alpha = 0.1$

- populations: users with a CD writer
  users without a CD writer
- 2 samples: $n=18$, $m=12$

$\bar{X} = 5.3$, $\bar{Y} = 4.8$

$S_x = 1.4$, $S_y = 1.6$ - sample std. deviations

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2 \Rightarrow R = (t_\alpha, \infty)$

2-sample t test ($\sigma_1 = \sigma_2$)

$$T = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}} \sim \text{Student t-distr. with } n+m-2 \text{ df}$$

S_p - pooled std. deviation

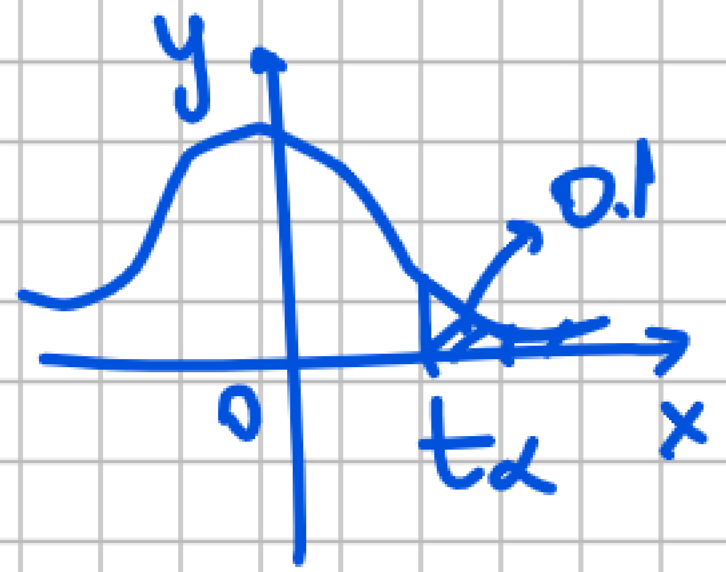
$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

$$S_p^2 = \frac{17 \cdot 1.4^2 + 11 \cdot 1.6^2}{28} = 2.20$$

$S_p = 1.48$

$$T = \frac{5.3 - 4.8}{1.48 \cdot \sqrt{\frac{1}{18} + \frac{1}{12}}} = 0.91$$

$\alpha = 0.1$



$t_\alpha = qt(1-\alpha, 28) = 1.31$

$T = 0.91$

$R = (1.31, \infty) \Rightarrow T \notin R \Rightarrow H_0$ is not rejected

(We don't have enough evidence to say the ad writer reduced the battery life at the 10% level of signif.)