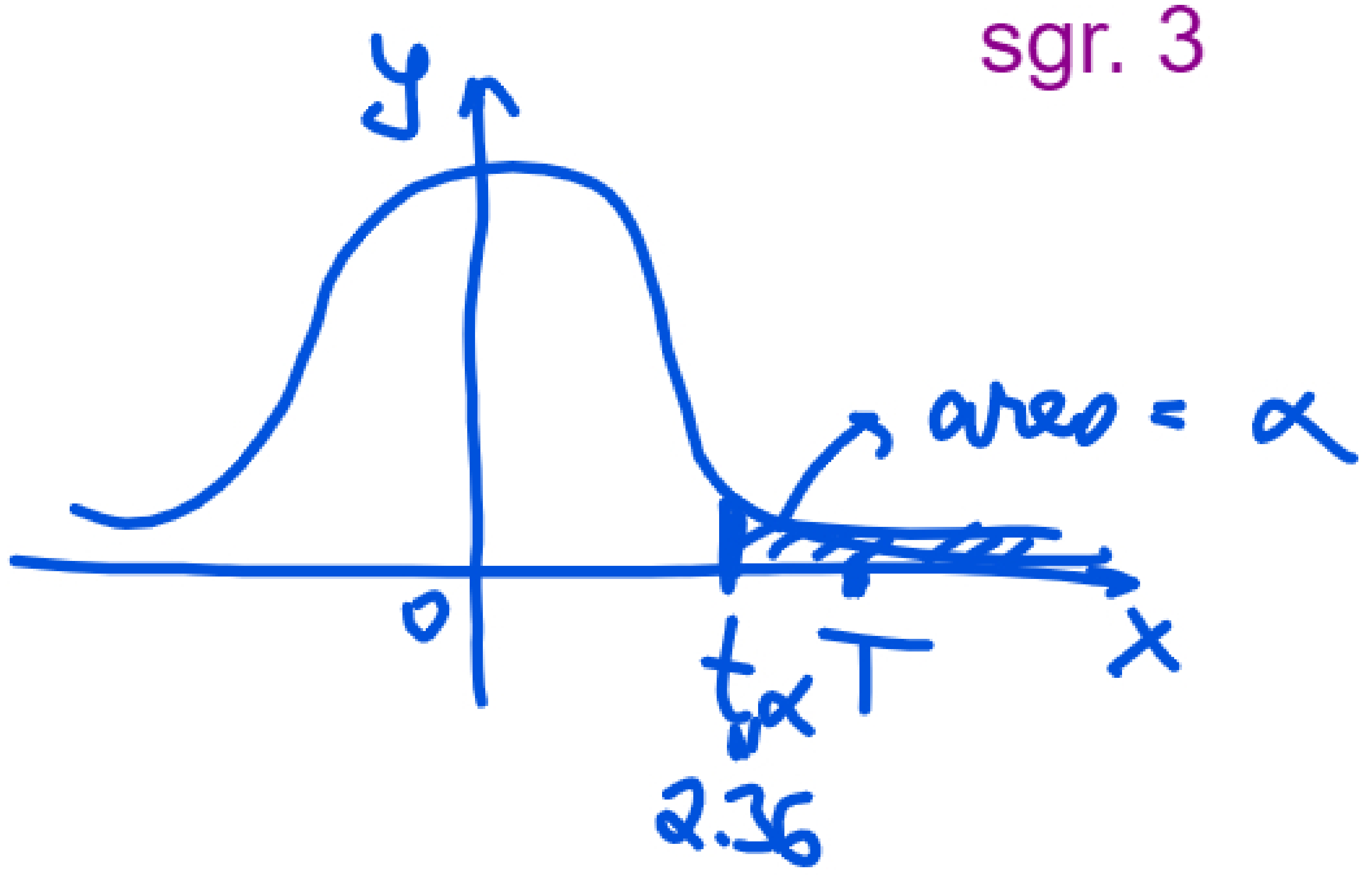
Lab 12.

sgr. 3



9.7. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?

$\bar{X} = 37.7$

$s = 9.2$

$n = 100$

$H_0: \mu = 35$ $(\mu_0 = 35)$

$H_a: \mu > 35 \Rightarrow R = (t_\alpha, \infty)$

σ (population std. deviation) is unknown $\Rightarrow$ t-test

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

$T = \dfrac{37.7 - 35}{\frac{9.2}{\sqrt{100}}} = \dfrac{2.7 \cdot 10}{9.2} = 2.93$

$t_\alpha = qt(1-\alpha, n-1) = 2.36$

$R = (2.36, \infty)$

$\Rightarrow T \in R \Rightarrow H_0$ is rejected

The data provides signif. evidence that the mean nr. of conc. users is greater than 35.

9.14. Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

Example 9.21 (Comparison of two servers). An account on server A is more expensive than an account on server B. However, server A is faster. To see if it's optimal to go with the faster but more expensive server, a manager needs to know how much faster it is. A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results,

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

95% ci is $[-1.4, -0.2]$.

a) $H_0: \mu_1 = \mu_2$

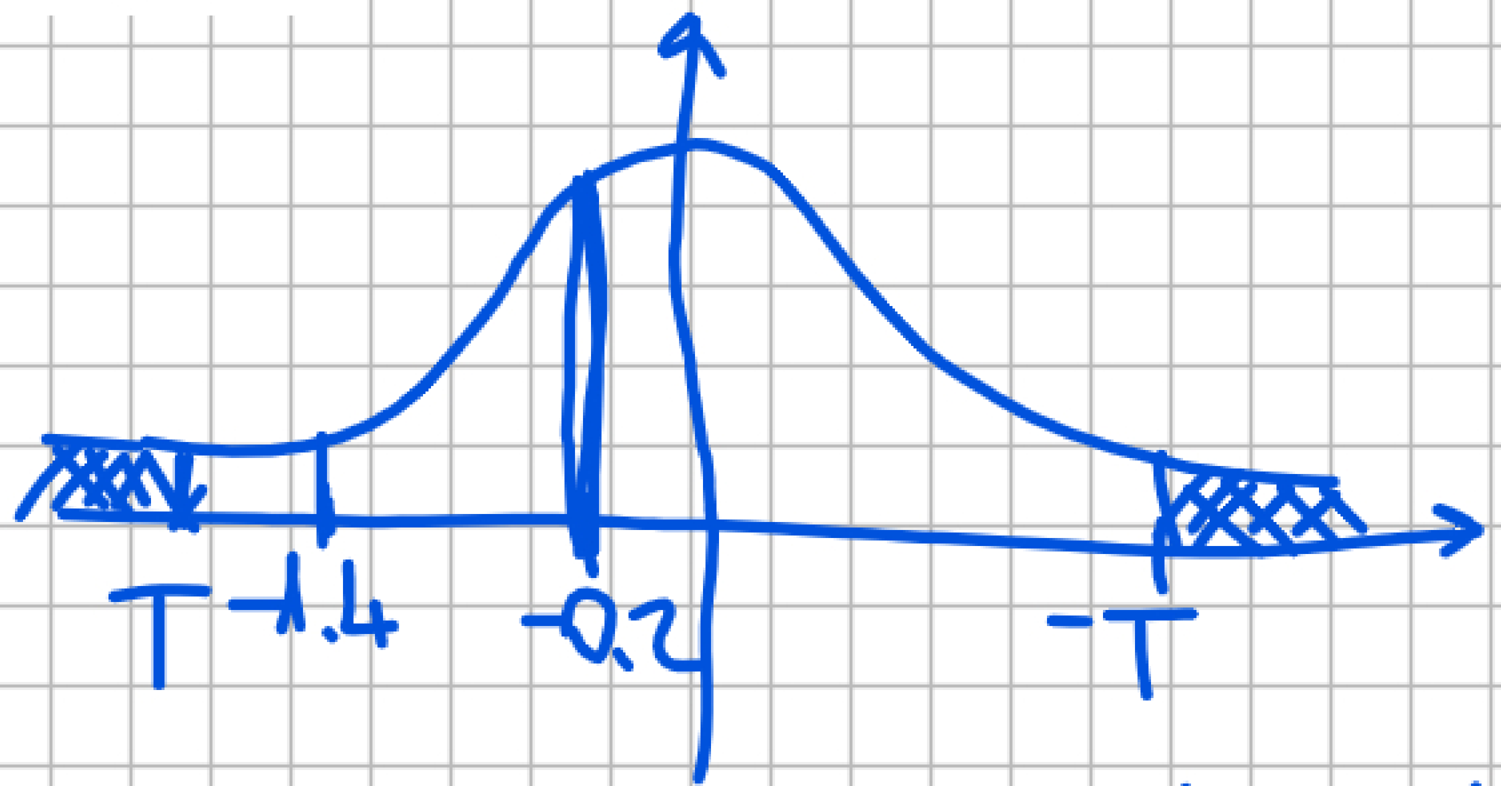
$H_a: \mu_1 \neq \mu_2$ (two-sided test)

$\alpha = 5\% = 0.05$

Acceptance region $A = [-1.4, -0.2]$

$\Rightarrow R = \mathbb{R} \setminus A = (-\infty, -1.4) \cup (-0.2, \infty)$

$T = \dfrac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$ - test statistic



two-sample t-test with unequal variances

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

p-value $= 2 \cdot pt(T, 25) = 0.011$

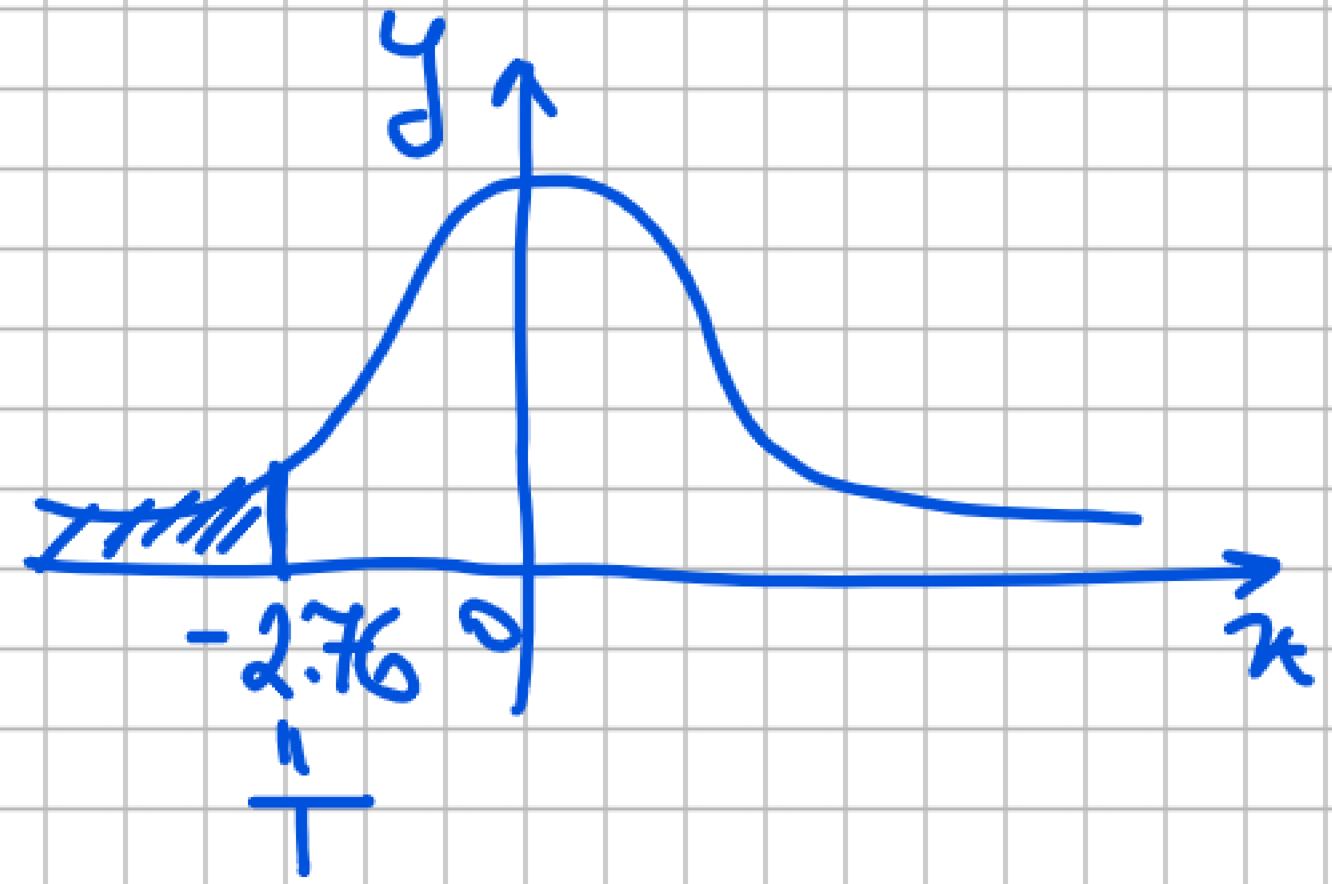
p-value $< \alpha \Rightarrow$ we reject H_0

c) $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 < \mu_2$

$T = \dfrac{\bar{X}-\bar{Y}}{\sqrt{\dfrac{S_x^2}{m}+\dfrac{S_y^2}{m}}} \sim$ Student t distribution with ν df

$\nu = 25$



$T = -2.76$

p-value $= pt(-2.76, 25) = 0.005$

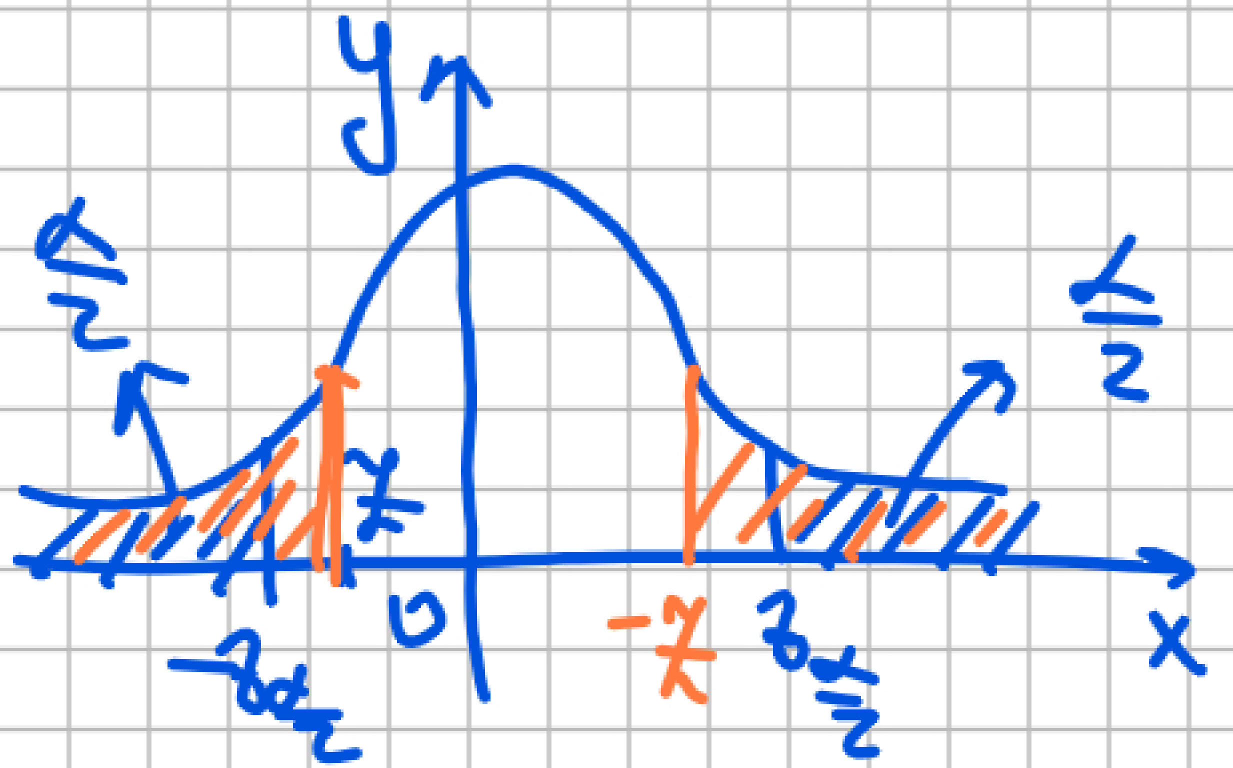
p-value $< \alpha \Rightarrow$ we reject H_0

(server A is faster at the 5% level of significance)

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?



b) $H_0: p_1 = p_2$

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

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

$\hat{p}$ - pooled proportion

$n = 250$

$m = 300$

$\hat{p}_1 = \frac{10}{250} = \frac{1}{25} = 0.04$

$\hat{p}_2 = \frac{18}{300} = \frac{6}{100} = 0.06$

proportions of defective items in the 2 samples

$$\hat{p} = \frac{10+18}{250+300} = \frac{28}{550} = 0.05$$

$\alpha = 0.02$

$$z = \frac{0.04 - 0.06}{\sqrt{0.05 \cdot 0.95\left(\frac{1}{250} + \frac{1}{300}\right)}} = -1.07$$

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

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

$Z = -1.07 \Rightarrow Z \notin R \Rightarrow$ we cannot reject H_0

At the 2% level of significance there isn't a significant difference in the quality of the two lots.

p-value = 2 * pnorm (Z) ≈ $\underline{0.28}$

p-value > α $\Rightarrow$ we cannot reject H_0

9.18. Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

8.1. The numbers of blocked intrusion attempts on each day during the first two weeks of the month were

56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58

After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.

b) $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

two-sample ~~t~~ test $\left\langle \begin{array}{l} \text{equal variances} \\ \text{unequal variances} \end{array} \right.$

If we assume equal variances:

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

S_p – pooled standard deviation

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

• t.test(x, y, alternative = "greater", var.equal = T)

p-value = 0.00054

$\alpha = 0.05$

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

(At the 5% level of significance there is enough evidence to claim that there is a significant reduction in the rate of intrusion attempts)

If we assume the variances are not equal.

t.test(x,y, alternative="greater")

p-value = 0.0055

p-value < α => H_0 is rejected

Confidence intervals for the population mean:

9.7. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

$(1-\alpha)\cdot 100\%$ confidence interval for the population mean μ: $\left[\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right] = \left[37.7 - 1.66 \cdot \frac{9.2}{\sqrt{100}}, 37.7 + 1.66 \cdot \frac{9.2}{\sqrt{100}}\right]$

$= [36.17, 39.23]$

$\bar{X} = 37.7$

$n = 100$

$S = 9.2$

$\alpha = 1 - 0.9 = 0.1$

$-t_{\frac{\alpha}{2}} = qt(\alpha/2, n-1) = -1.66 \Rightarrow t_{\frac{\alpha}{2}} = 1.66$

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

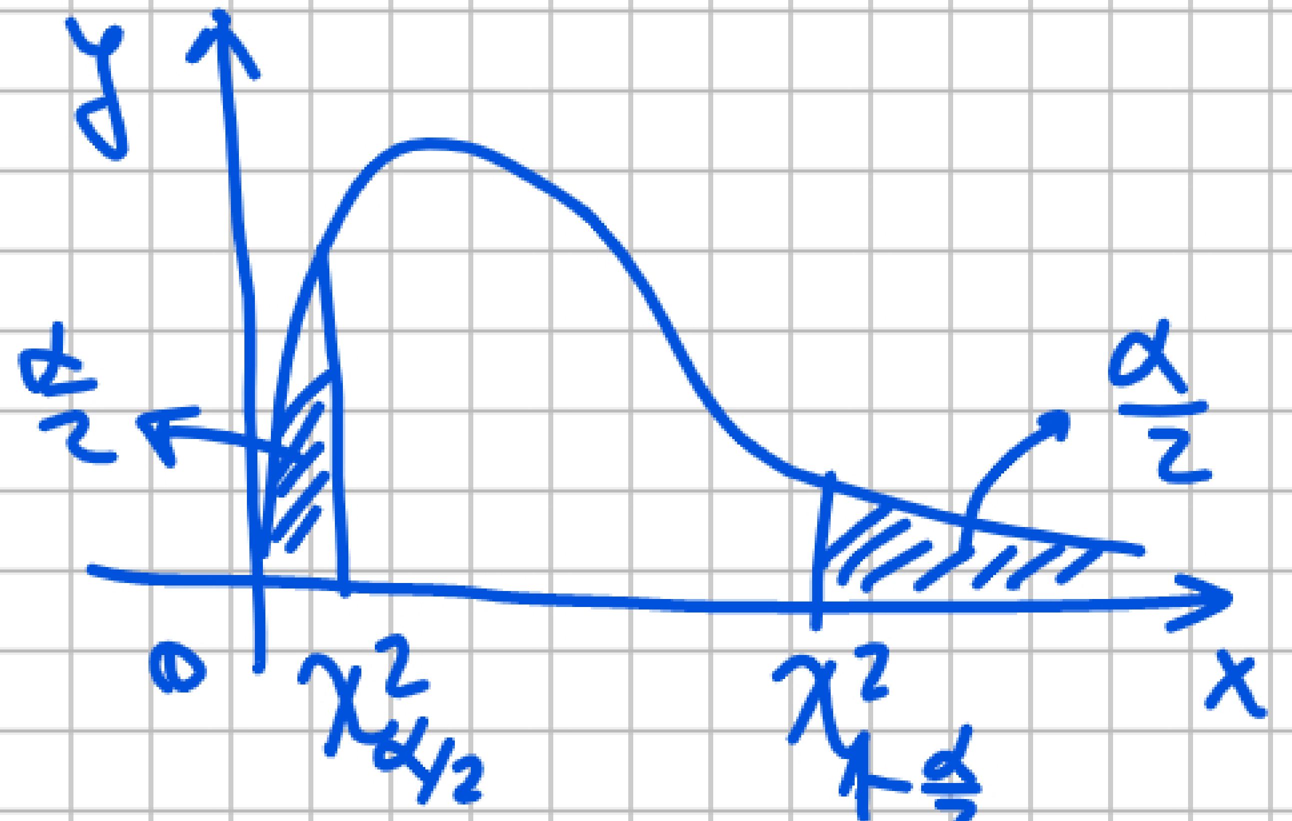
is collected from a Normal distribution with mean μ. Construct the 95% confidence intervals for the population variance σ^2 and standard deviation σ.

$(1-\alpha)100\%$ confidence interval of the population variance σ^2:

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}\right]$$

$(1-\alpha)100\%$ CI for the population std. deviation σ:

$$\left[\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}}}\right]$$



s^2 – sample variance

$n=6$, $\alpha=0.05$

$\chi^2_{\frac{\alpha}{2}} = qchisq(\alpha/2, n-1) = 0.83$

$\chi^2_{1-\frac{\alpha}{2}} = qchisq(1-\alpha/2, n-1) = 12.83$

$s^2 = 6.23$

95% ci for σ^2: $\left[\frac{5\cdot 6.23}{12.83}, \frac{5\cdot 6.23}{0.83}\right] = [2.43, 37.53]$

95% ci for σ^2: $[\sqrt{2.43}, \sqrt{37.53}] = [1.56, 6.13]$

sgr. 2

9.7. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(s)

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?

$n = 100$, $\bar{x} = 37.7$

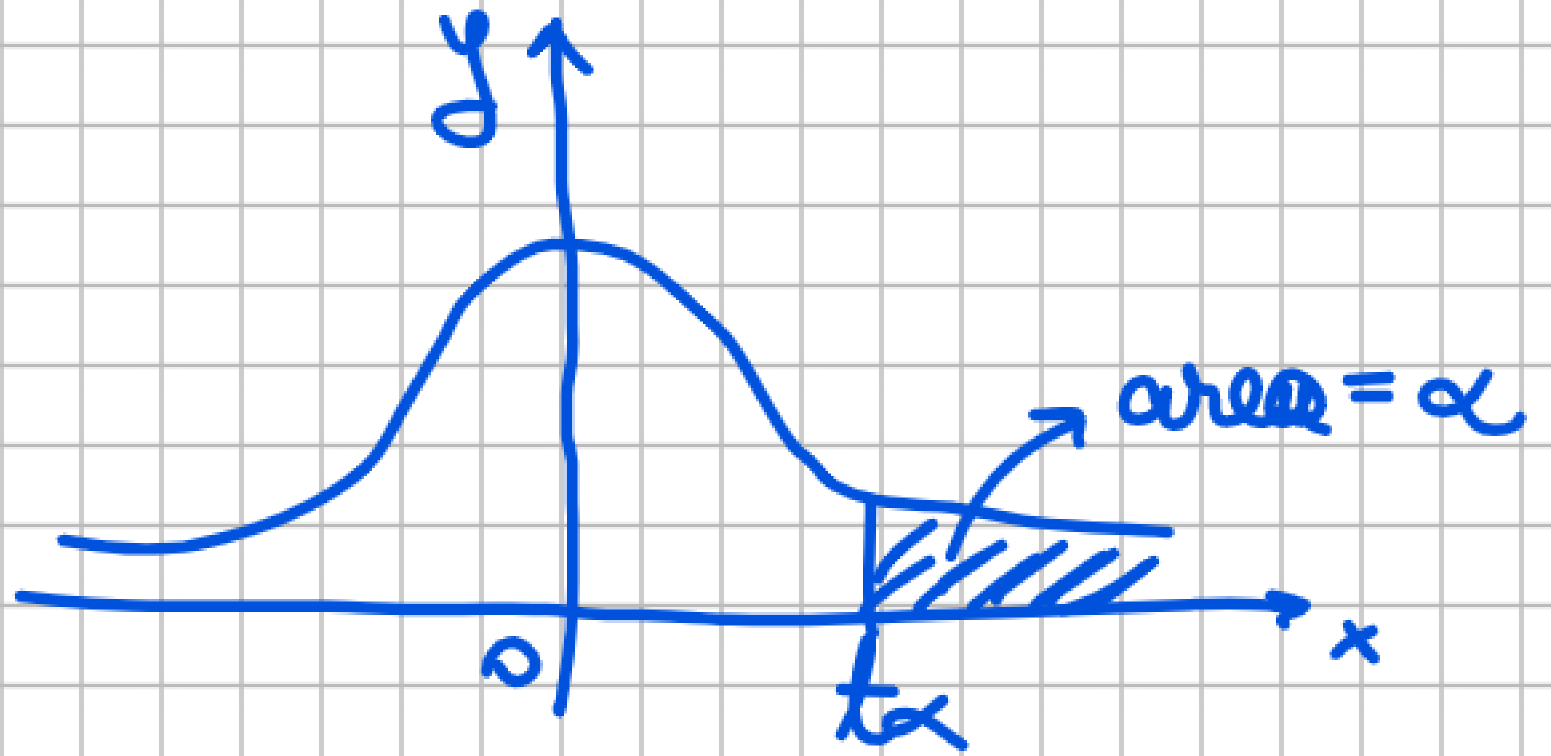
$s = 9.2$ (sample mean)

b) $\alpha = 1\% = 0.01$

$H_0: \mu = 35$ ($\mu_0 = 35$)

$H_a: \mu > 35 \Rightarrow R = (t_\alpha, \infty)$

σ-unknown $\Rightarrow$ t-test



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

$T = \dfrac{37.7 - 35}{9.2/\sqrt{100}} = \dfrac{2.7 \cdot 10}{9.2} = \dfrac{27}{9.2} = 2.93$

$t_\alpha = qt(1-\alpha, n-1) = 2.36$

$\Rightarrow R = (2.36, \infty)$

$T = 2.93 \in R \Rightarrow H_0$ is rejected

(At the 1% level of significance,
the data provides significant
evidence that the mean nr. of
concurrent users is greater than 35.)

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?

(b) $H_0: p_1 = p_2$

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

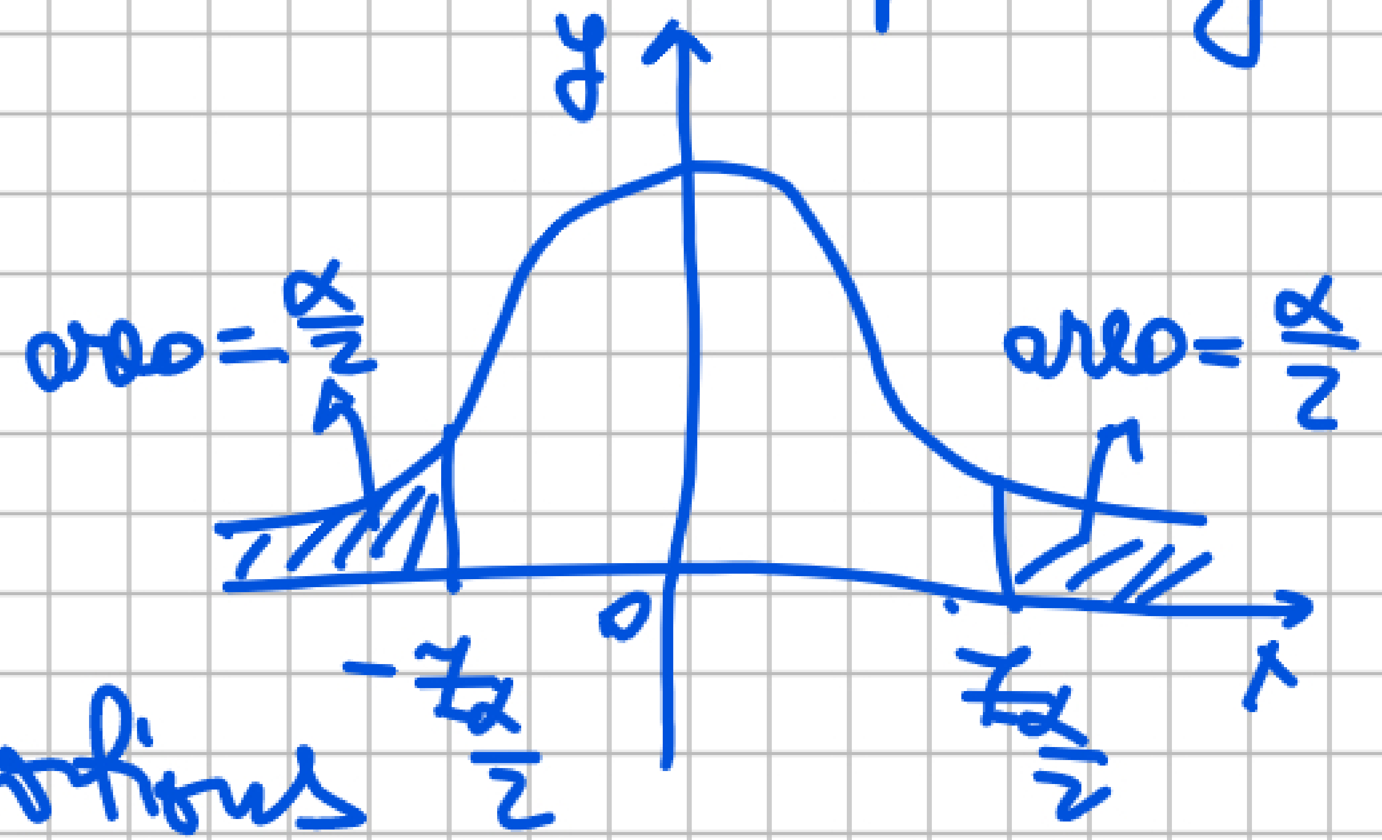
p_1, p_2 – proportion (percentage) of defective items in lot A and B, respectively

$n = 250$, $m = 300$

$\hat{p}_1 = \frac{10}{250} = \frac{1}{25} = 0.04$

$\hat{p}_2 = \frac{18}{300} = \frac{6}{100} = 0.06$

$\hat{p}_1, \hat{p}_2$ – sample proportions



$\alpha = 0.02$

Z-test

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

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

$$\hat{p} = \frac{10+18}{250+300} = \frac{28}{550} = 0.05$$

$$Z = \frac{0.04 - 0.06}{\sqrt{0.05 \cdot 0.95\left(\frac{1}{250} + \frac{1}{300}\right)}} = -1.07$$

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

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

$Z = -1.07$

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

(There isn't enough evidence to suggest a difference in the quality of the 2 lots, at the 2% level of significance)

9.18. Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

8.1. The numbers of blocked intrusion attempts on each day during the first two weeks of the month were

$$56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58$$

After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

$$53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.$$

$H_0: \mu_1 = \mu_2$

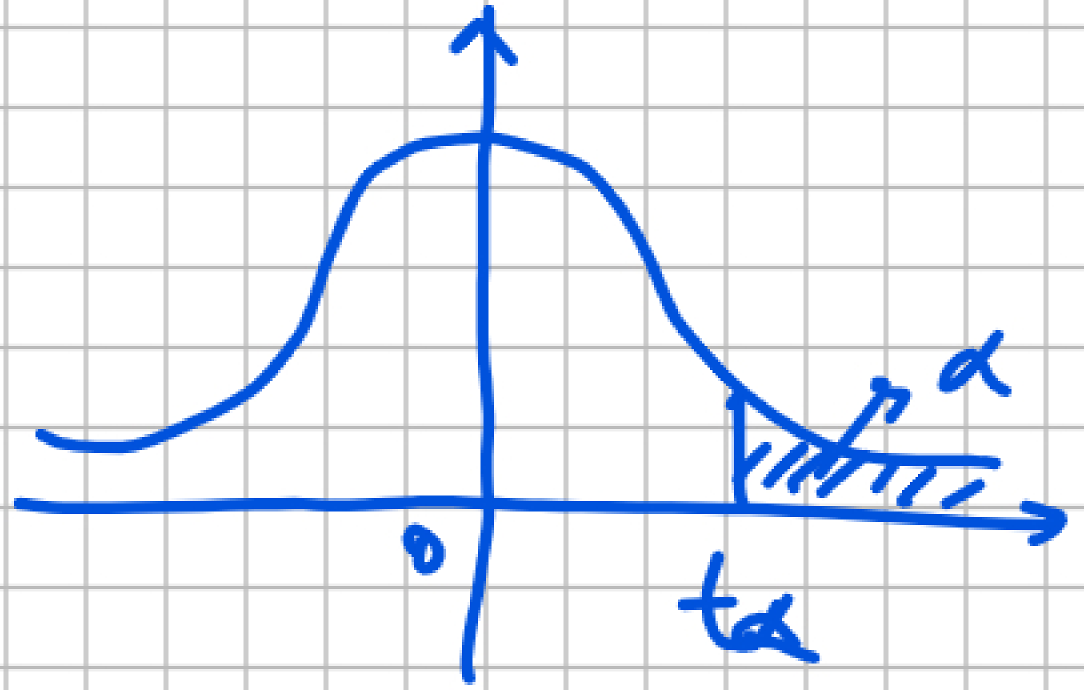
$H_a: \mu_1 > \mu_2 \Rightarrow R = (t_\alpha, \infty)$

2-sample t-test < equal variances / unequal variances

If we assume equal variances:

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

$t_\alpha = qt(1-\alpha, n+m-2)$, $\alpha = 0.05$



$t_\alpha = 1.69$

$R = (1.69, \infty)$

$\bar{X} = 50$, $\bar{Y} = 40.2$, $s_p = \cancel{9.13}\ 7.82$

$$T = \frac{50 - 40.2}{\cancel{9.13}\ 7.82 \sqrt{\frac{1}{14} + \frac{1}{20}}} = 3.08$$

$\Rightarrow T \in R \Rightarrow H_0$ is rejected

There is a signif. reduction in intrusion attempts at $\alpha = 0.05$ level of signif.

t.test(x, y, alternative = "greater", var.equal = T)

p-value = 0.00054 < α $\Rightarrow$ H_0 is rejected

If we don't assume equal variances:

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{m}}} \sim \text{Student } t \text{ distribution with } \nu \text{ df}$$

$\nu \Rightarrow$ slides/textbook

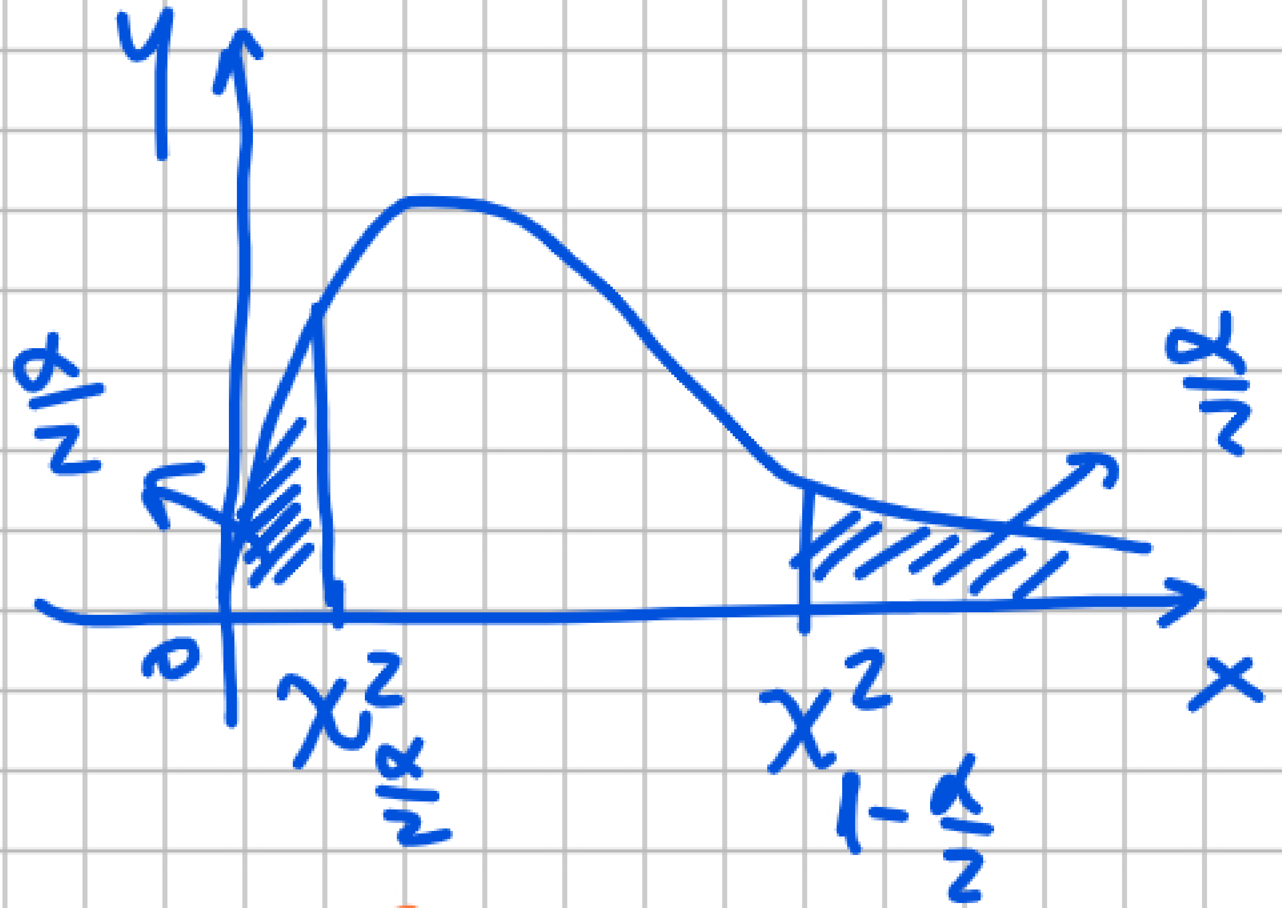
$\longrightarrow$ t.test(x, y, alternative = 'greater', var.equal = F)

p-value = 0.00055 < α $\Rightarrow$ H_0 is rejected

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

is collected from a Normal distribution with mean μ. Construct the 95% confidence intervals for the population variance σ^2 and standard deviation σ.



$(1-\alpha)100\%$ confidence interval for the

- population variance (σ^2) is $\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right] = \left[\frac{5\cdot 6.23}{12.83}, \frac{5\cdot 6.23}{0.83}\right] = [2.43, 37.53]$

- population std. deviation (σ) is $\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}}\right] = [1.56, 6.13]$

s^2 - sample variance

$n = 6$

$s^2 = 6.23$

$\chi^2_{\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1) = 0.83$, $\chi^2_{1-\frac{\alpha}{2}} = \text{qchisq}(1-\alpha/2, n-1) = 12.83$

$\alpha = 5\% = 0.05$

9.14. Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

Example 9.21 (Comparison of two servers). An account on server A is more expensive than an account on server B. However, server A is faster. To see if it's optimal to go with the faster but more expensive server, a manager needs to know how much faster it is. A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results,

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

$[-1.4, -0.2]$. → 95% ci for the difference of population means

(a) $H_0: \mu_1 = \mu_2$ $(\mu_1 - \mu_2 = \underline{0})$

$H_a: \mu_1 \neq \mu_2$

$\alpha = 5\%$

Acceptance region $A = [-1{,}4; -0{,}2]$

$\Rightarrow R = \mathbb{R} \setminus A = (-\infty, -1.4) \cup (-0.2, \infty)$

If $D = 0$ belongs to R $\Rightarrow$ H_0 is rejected

$0 \in R \Rightarrow H_0$ is rejected

p-value → compute T

sgr. 5

9.18. Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

8.1. The numbers of blocked intrusion attempts on each day during the first two weeks of the month were

56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58

After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.

$H_0: \mu_1 = \mu_2$

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

$n = 14, m = 20, \alpha = 0.05$

σ_1, σ_2 – unknown $\Rightarrow$ two-sample t-test

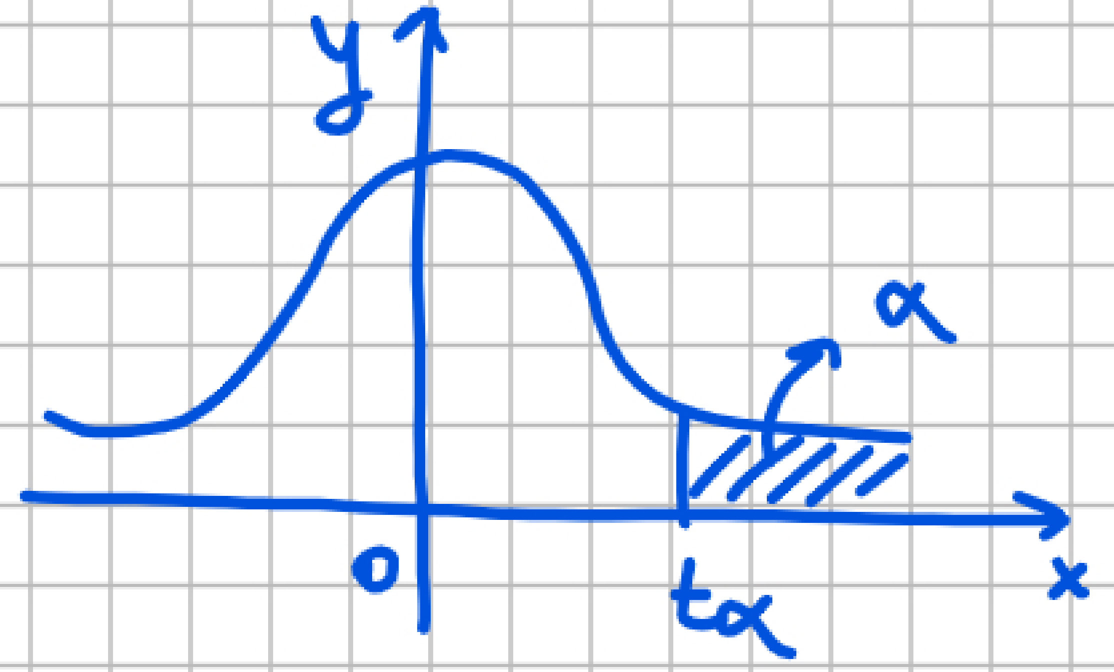
equal var. unequal var.

If we assume equal variances:

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

$S_p^2 = ((n-1)S_x^2 + (m-1)S_y^2) / (n+m-2)$

$T = 3.59$



$t_\alpha = qt(1-\alpha, n+m-2) = 1.72$

$R = (1.72, \infty)$

$T \in R \Rightarrow H_0$ is rejected

There is a significant reduction in the rate of intrusion attempts, at the 5% level of significance.

If we assume unequal variances:

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

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

$\nu = 29$

$S_x = 7.62$

$S_y = 7.96$

$T = 3.62$

$R = (1.72, \infty)$

$T \in R \Rightarrow H_0$ is reject

t.test(x,y, alternative = "greater", var.equal = T) → p-value = 0.00054 < α
F → p-value = 0.00055 < α
H_0 is rejected

9.14. Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

Example 9.21 (Comparison of two servers). An account on server A is more expensive than an account on server B. However, server A is faster. To see if it's optimal to go with the faster but more expensive server, a manager needs to know how much faster it is. A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results,

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

c) $H_0: \mu_A = \mu_B \quad (\mu_1 = \mu_2)$

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

σ_A, σ_B (population std. deviations) are unknown

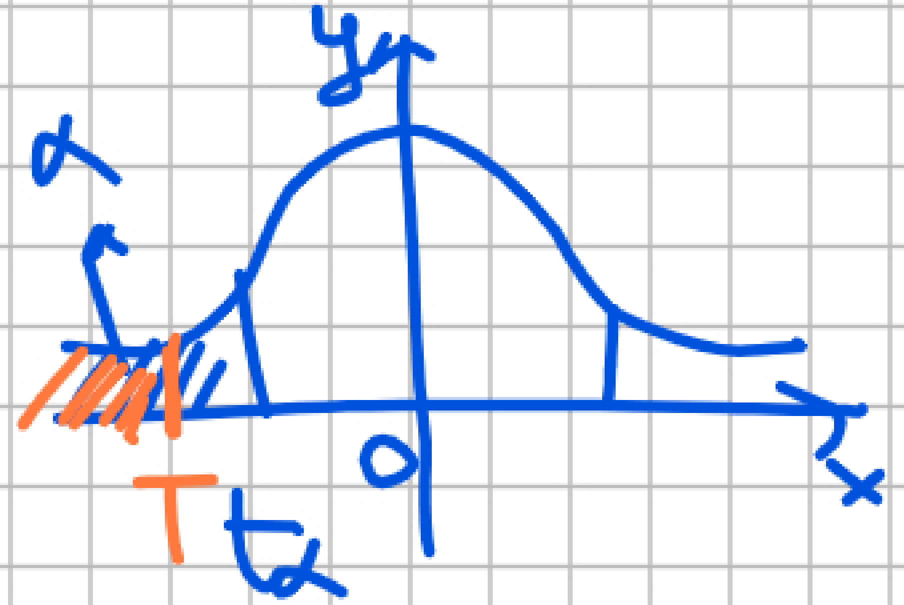
$\Rightarrow$ 2-sample t-test, unequal variance

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_A^2}{n} + \frac{S_B^2}{m}}} \sim \text{Student } t \text{ distribution with } \nu \text{ df}$$

$\nu = 25$

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

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



$T = -2.76 \in R \Rightarrow H_0$ is rejected

p-value $= pt(T, \nu) = 0.0053 < \alpha \Rightarrow H_0$ is rejected

$\alpha = 0.05$

$R = (-\infty, \overset{z_\alpha}{t_\alpha}) \Rightarrow$ p-value $= pt(T, \nu)$ / $pnorm(z)$

$R = (\underset{z_\alpha}{t_\alpha}, \infty) \Rightarrow$ p-value $= 1 - pt(T, \nu)$ / $1 - pnorm(z)$

$R = (-\infty, \underset{-z_{\frac{\alpha}{2}}}{-t_{\frac{\alpha}{2}}}) \cup (\underset{z_{\frac{\alpha}{2}}}{t_{\frac{\alpha}{2}}}, \infty) \Rightarrow$ p-value $= 2 * pt(T, \nu)$ / $2 * pnorm(z)$

Example 4. (Unauthorized use of a computer account). If an unauthorized person accesses a computer account with the correct username and password (stolen or cracked), can this intrusion be detected? The following times between keystrokes (sec) were recorded when a user typed the username and password:

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

Construct a 99% confidence interval for the mean time between keystrokes assuming Normal distribution of these times.

$(1-\alpha)\cdot 100\%$ CI for the population mean: $\left[\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$

99% CI for μ:

$\alpha = 1\% = 0.01$

$\bar{x} = 0.29$, $t_{\frac{\alpha}{2}} = qt(\alpha/2, n-1) = -2.90$

$s = 0.07$, $t_{\frac{\alpha}{2}} = 2.90$

$n = 18$

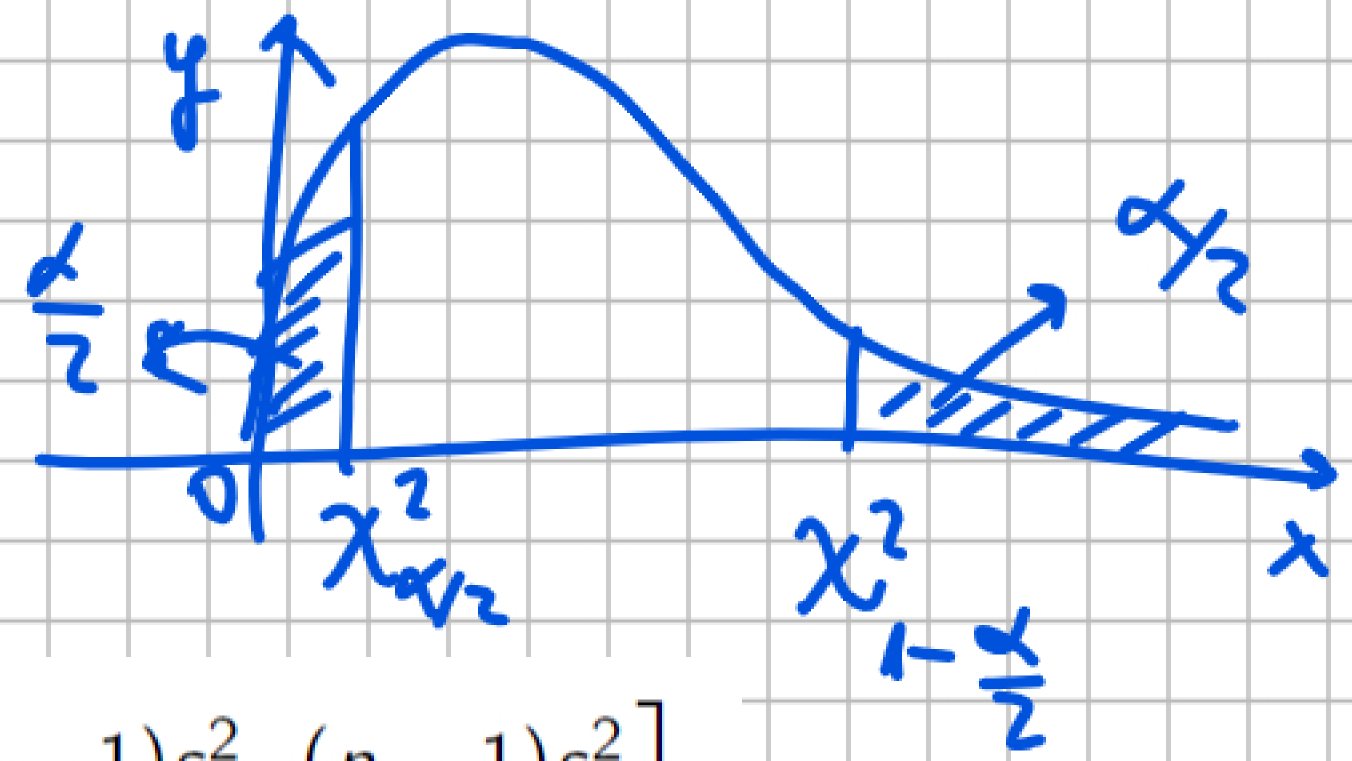
$= \left[0.29 - 2.90 \cdot \frac{0.07}{\sqrt{18}}, 0.29 + 2.90 \cdot \frac{0.07}{\sqrt{18}}\right]$

$= [0.24, 0.34]$

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

is collected from a Normal distribution with mean μ. Construct the 95% confidence intervals for the population variance σ^2 and standard deviation σ.



$(1-\alpha)100\%$ CI for the population variance σ^2: $\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right]$

— " — std. deviation σ: $\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}}\right]$

$\alpha = 5\% = 0.05$

s^2 – sample variance

$\chi^2_{\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1)$ degrees of freedom

$\chi^2_{1-\frac{\alpha}{2}} = \text{qchisq}(1-\alpha/2, n-1)$, continue at home (full solution at sgl.3 for ex.)

sgr. 6

9.7. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?

$H_0: \mu = 35 \quad (\mu_0 = 35)$

$H_a: \mu > 35. \Rightarrow R = (t_\alpha, \infty)$

σ - is unknown (std. deviation of the population)

$\rightarrow$ t-test

$\bar{X} = 37.7$

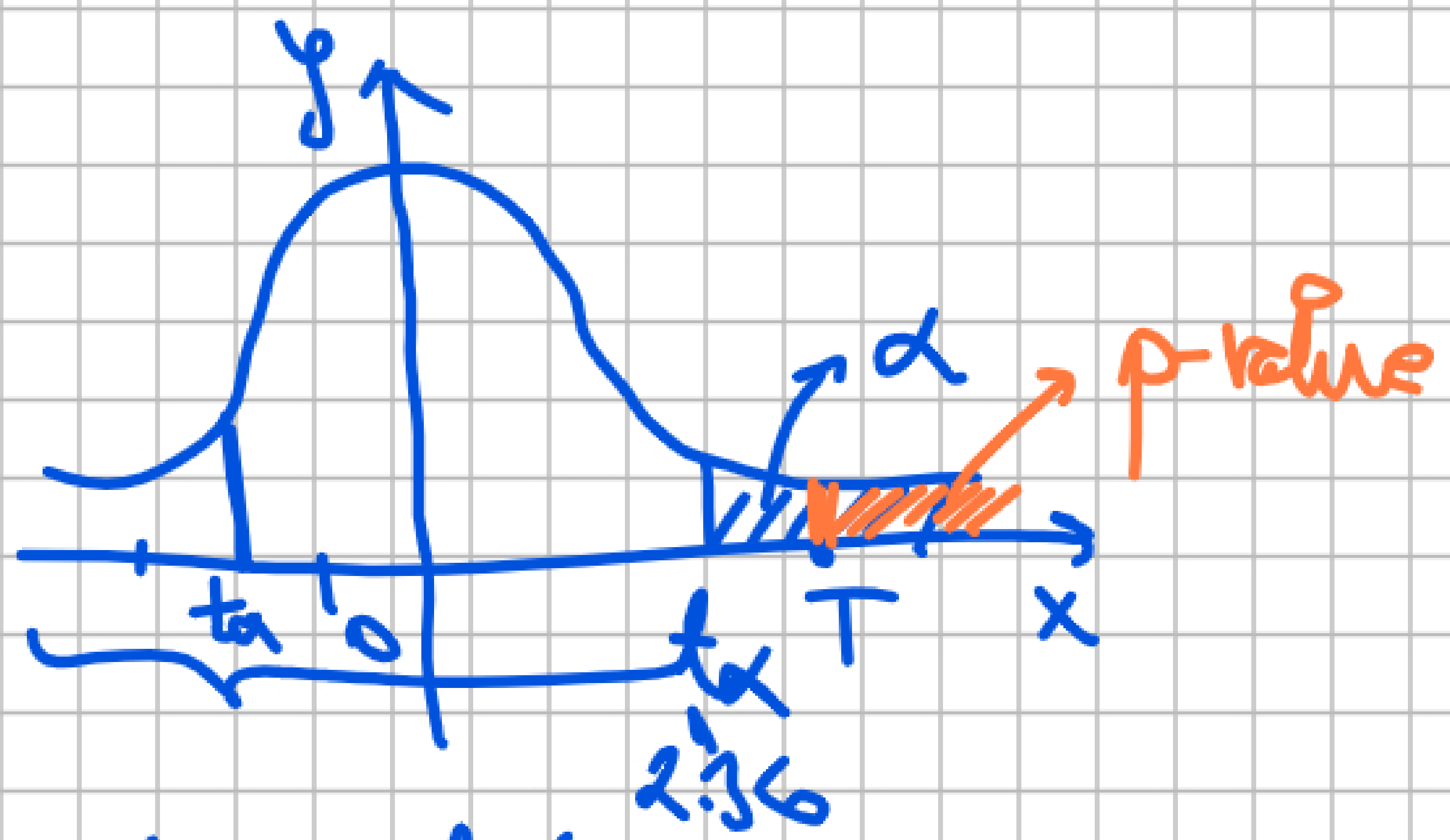
$n = 100$

$s = 9.2$ (std. deviation of the sample)

$\alpha = 1\% = 0.01$

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

$T = \dfrac{37.7 - 35}{\frac{9.2}{\sqrt{100}}} = \dfrac{2.7 \cdot 10}{9.2} = \dfrac{27}{9.2} = 2.93$



$t_\alpha = qt(1-\alpha, n-1) = 2.36$

$R = (2.36, \infty)$

$T = 2.93 \Rightarrow T \in R \Rightarrow H_0$ is reject

The data provide significant evidence to say that the mean nr. of concurrent users is greater than 35, at the 1% level of significance.

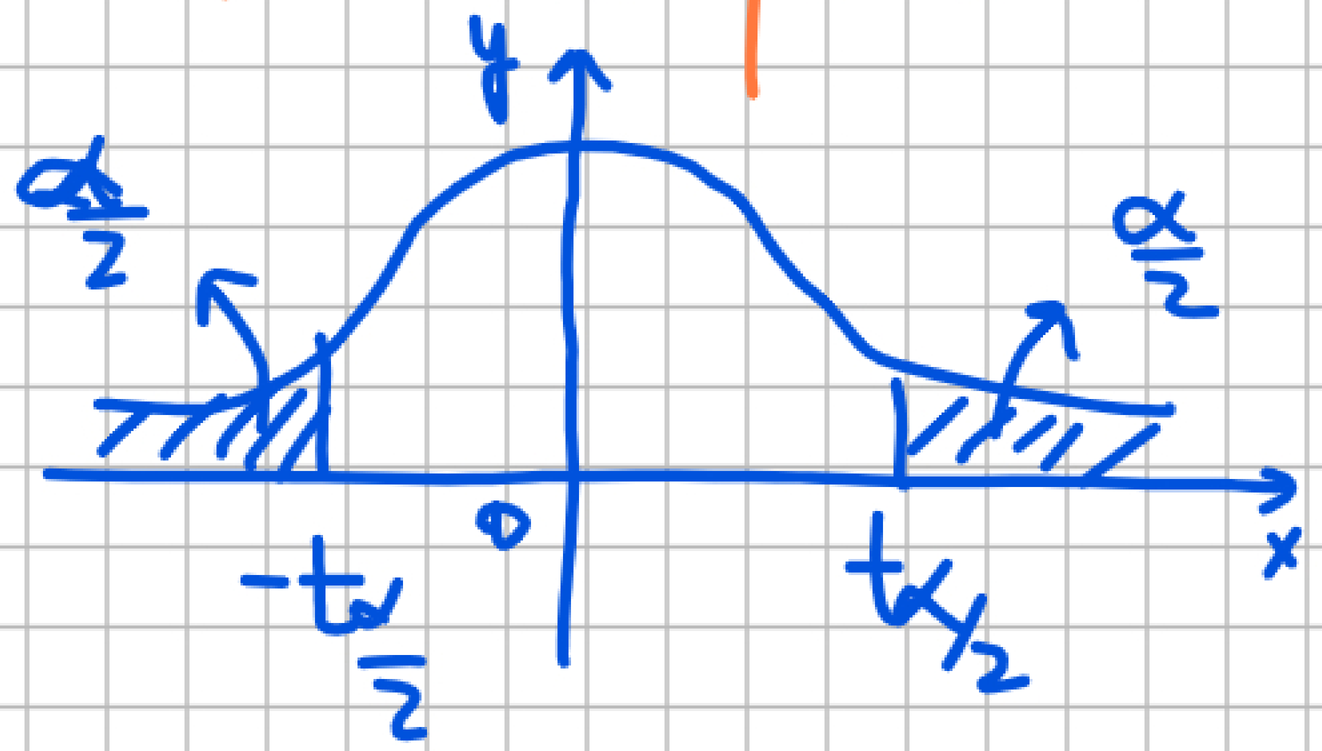
- p-value $= 1 - pt(T, n-1) = pt(T, n-1, \text{lower.tail}=F)$

p-value $= 0.002 < \alpha = 0.01 \Rightarrow H_0$ is rejected

- $H_a: \mu < \mu_0 \Rightarrow$ p-value $= pt(T, n-1)$
- $H_a: \mu \neq \mu_0 \Rightarrow$ p-value $= 2 * pt(T, n-1)$ or $2 * pt(T, n-1, \text{lower.tail}=F)$

a) $(1-\alpha)100\%$ confidence interval for the population mean μ is:

$$\left[\bar{X} - t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}} \cdot \frac{S}{\sqrt{n}}\right]$$



$\bar{X} = 37.7$, $S = 9.2$, $n = 100$

- $t_{\frac{\alpha}{2}} = qt(\alpha/2, n-1) = -1.66$

$t_{\alpha/2} = 2.63$

$\alpha = 0.1$

90% CI for μ: $\left[37.7 - 2.63 \cdot \frac{9.2}{\sqrt{100}}, 37.7 + 2.63 \cdot \frac{9.2}{\sqrt{100}}\right] = [36.17, 39.23]$

If $\mu_0 \in CI \Rightarrow H_0$ is not rejected

$\mu_0 \notin CI \Rightarrow H_0$ is rejected

9.14. Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

Example 9.21 (COMPARISON OF TWO SERVERS). An account on server A is more expensive than an account on server B. However, server A is faster. To see if it's optimal to go with the faster but more expensive server, a manager needs to know how much faster it is. A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results,

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

c) $H_0: \mu_1 = \mu_2$

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

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

$S_x = 0.6$, $S_y = 1.2$ (sample std. deviations)

$n = 30$, $m = 20$ (sample sizes)

$\alpha = 5\% = 0.05$

σ_1, σ_2 – unknown (pop. std. deviations)

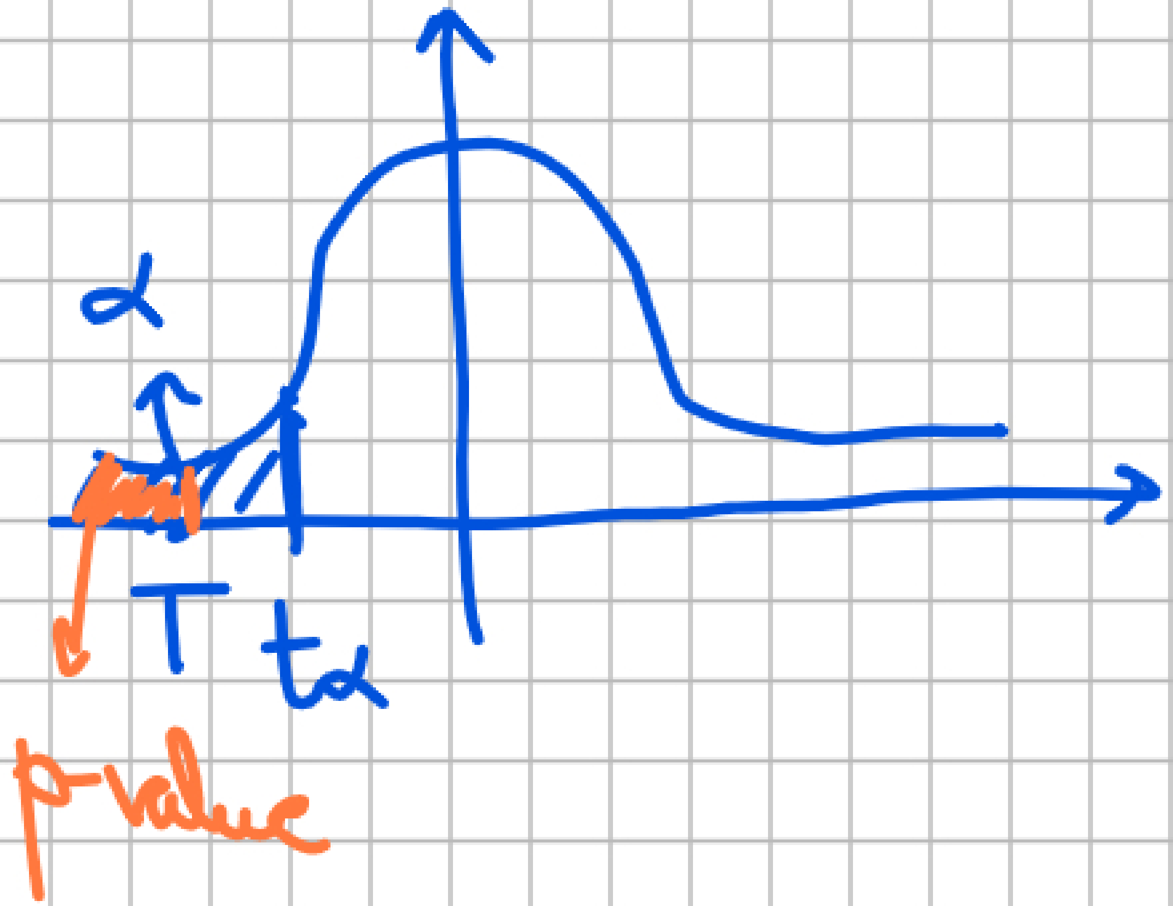
$\Rightarrow$ two-sample t-test

$S_x = \frac{1}{2} S_y \Rightarrow$ we assume σ_1, σ_2 are unequal

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \sim \text{Student } t \text{ distribution with } \nu \text{ df}$$

$\nu = 25$

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



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

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

$T = -2.76 \Rightarrow T \in R \Rightarrow H_0$ is reject

There is significant evidence to say that server A is faster at the 5% level of significance.

p-value = area from $-\infty$ to T under the PDF

$= pt(T, \nu) = 0.005$

p-value $< \alpha = 0.05 \Rightarrow H_0$ is rejected

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?

$H_0: p_1 = p_2$

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

$\Rightarrow$ Z-test, p_1, p_2 – population proportions

$n = 250$, $m = 300$, $\alpha = 0.02$

$\hat{p}_1 = \frac{10}{250} = \frac{1}{25} = 0.04$

$\hat{p}_2 = \frac{18}{300} = \frac{6}{100} = 0.06$

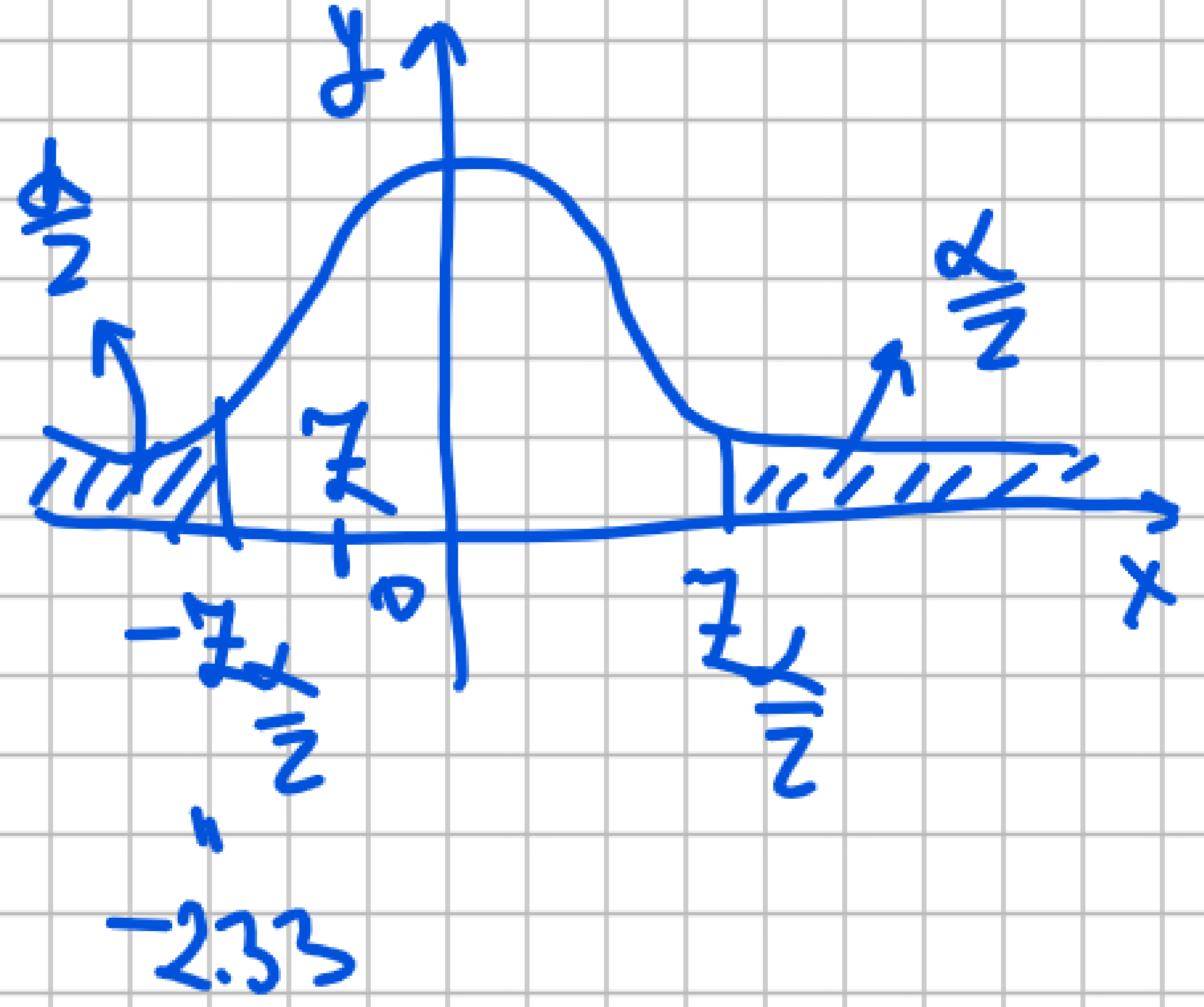
sample proportions

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

$\hat{p}$ – pooled sample proportion

$$\hat{p} = \frac{10+18}{250+300} = \frac{28}{550} = 0.05$$

$$Z = \frac{0.04 - 0.06}{\sqrt{0.05 \cdot 0.95 \cdot \left(\frac{1}{250} + \frac{1}{300}\right)}} = -1.07$$



$$-Z_{\frac{\alpha}{2}} = qnorm(\alpha/2) = -2.33 \Rightarrow Z_{\frac{\alpha}{2}} = 2.33$$

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

$Z = -1.07 \notin R \Rightarrow H_0$ is not reject

At the 2% level of significance the data does not provide suff. evidence to say the quality of the two lots is different

9.18. Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

8.1. The numbers of blocked intrusion attempts on each day during the first two weeks of the month were

$$56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58$$

After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

$$53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.$$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

$\Rightarrow$ two-sample t-test (σ_1, σ_2 - unknown)

• if we assume equal variance : t.test(x, y, alternative = "greater", var.equal = T)

p-value = 0.00054 < α = 0.05

$\Rightarrow H_0$ is rejected

- if we don't assume equal variances:

  t.test(x, y, alternative = "greater", var.equal = F)

  p-value = 0.0055 < α = 0.05 => H_0 is rejected

Example 7. A sample of 6 measurements

$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

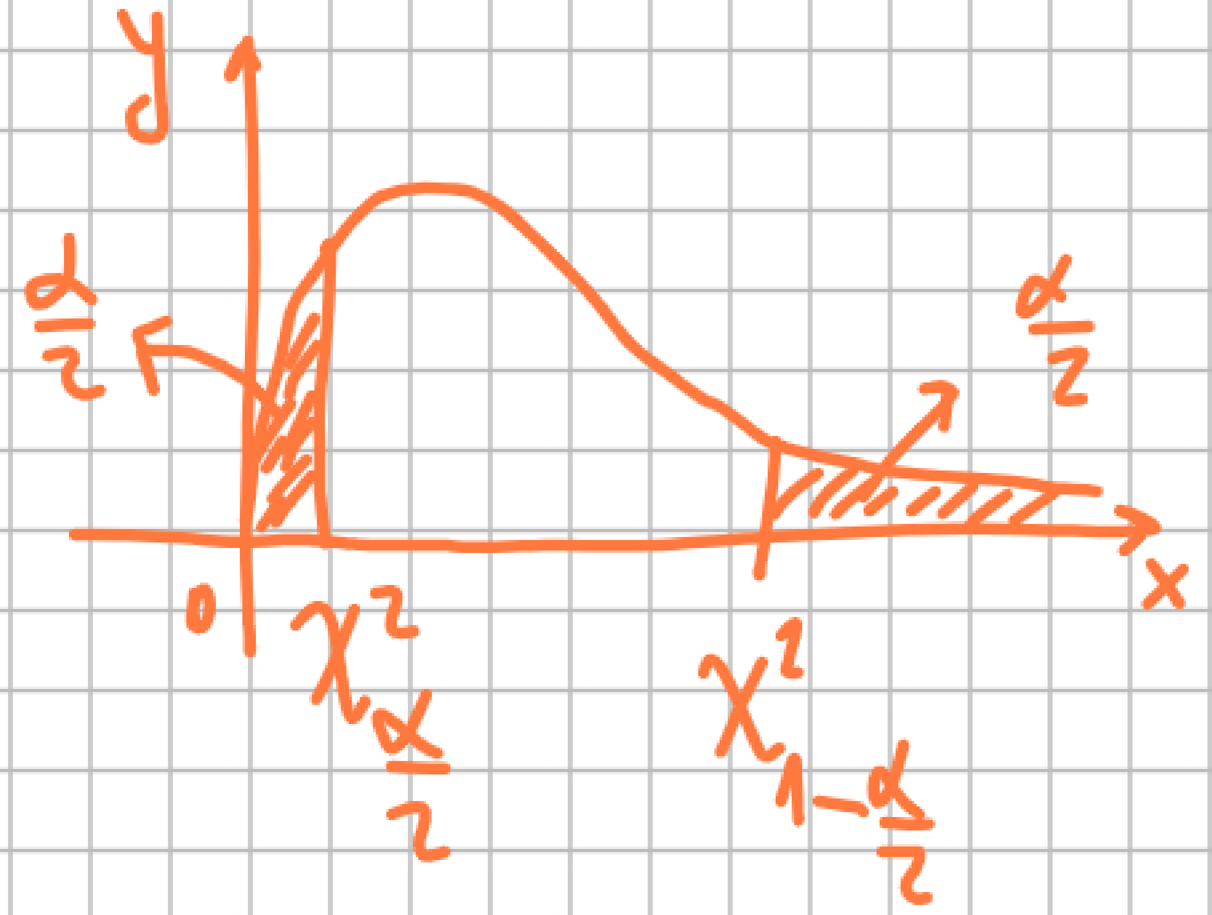
is collected from a Normal distribution with mean μ. Construct the 95% confidence intervals for the population variance σ^2 and standard deviation σ.

$(1-\alpha)100\%$ CI for the population variance σ^2: $\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right]$

— " — — " — std. deviation σ: $\left[\sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}}\right]$

n - sample size

s^2 - sample variance



$$\chi^2_{\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1)$$

$$\chi^2_{1-\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1, \text{lower.tail}=F)$$

95% ci for σ^2: $\left[\frac{5\cdot 6.23}{12.83}, \frac{5\cdot 6.23}{0.83}\right] = [2.43, 37.53]$

$\chi^2_{\frac{\alpha}{2}} = 0.83$, $\chi^2_{1-\frac{\alpha}{2}} = 12.83$

$s^2 = 6.23$

95% ci for σ: $[\sqrt{2.43}, \sqrt{37.53}] = [1.56, 6.13]$

sgr. 1

9.7. In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation $\sigma = 9.2$.

(a) Construct a 90% confidence interval for the expectation of the number of concurrent users.

(b) At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?

$H_0: \mu = 35$ $(\mu_0 = 35)$

$H_a: \mu > 35 \Rightarrow R = (t_\alpha, \infty)$

$s = 9.2$ (sample std. deviation), σ is unknown (population std. deviation)

$\Downarrow$

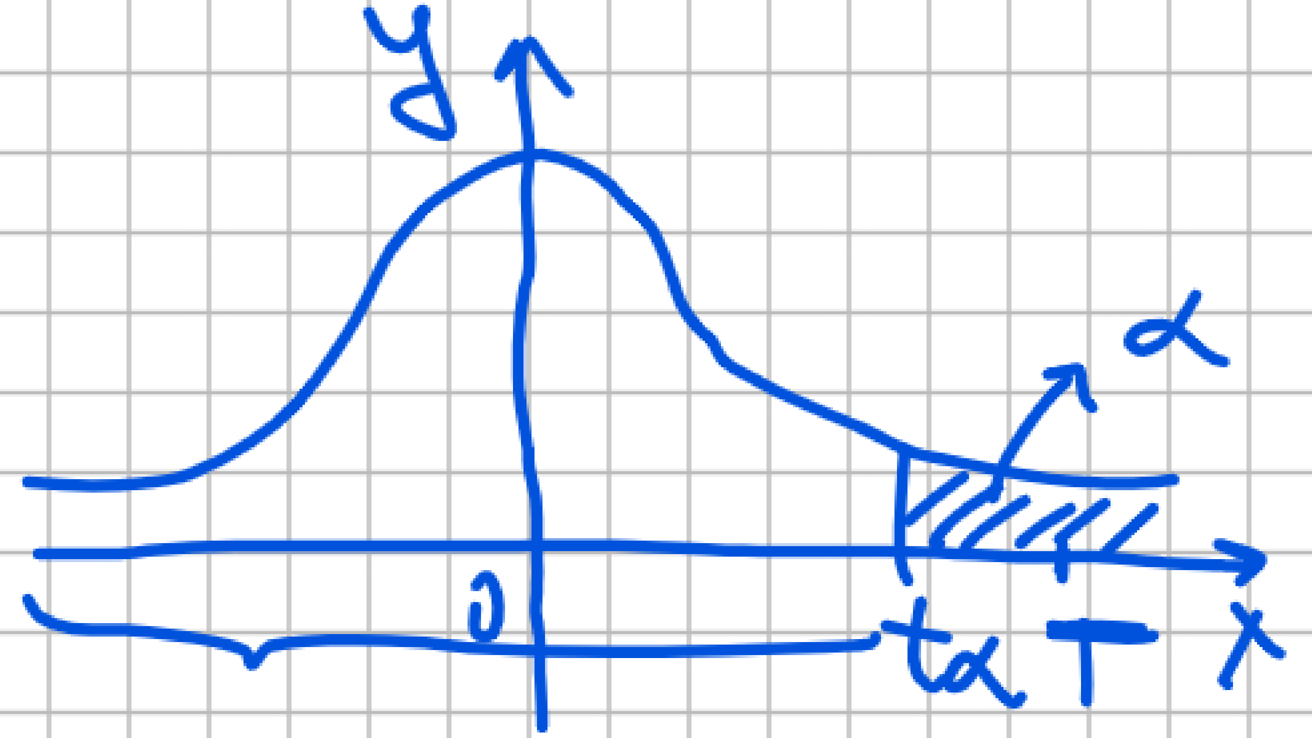
t-test

$\bar{X} = 37.7$

$n = 100$

$\alpha = 1\% = 0.01$

$$T = \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \sim \text{Student } t \text{ distribution with } n-1 \text{ degrees of freedom}$$



$$T = \frac{37.7-35}{\frac{9.2}{\sqrt{100}}} = \frac{2.7\cdot 10}{9.2} = \frac{27}{9.2} = 2.93$$

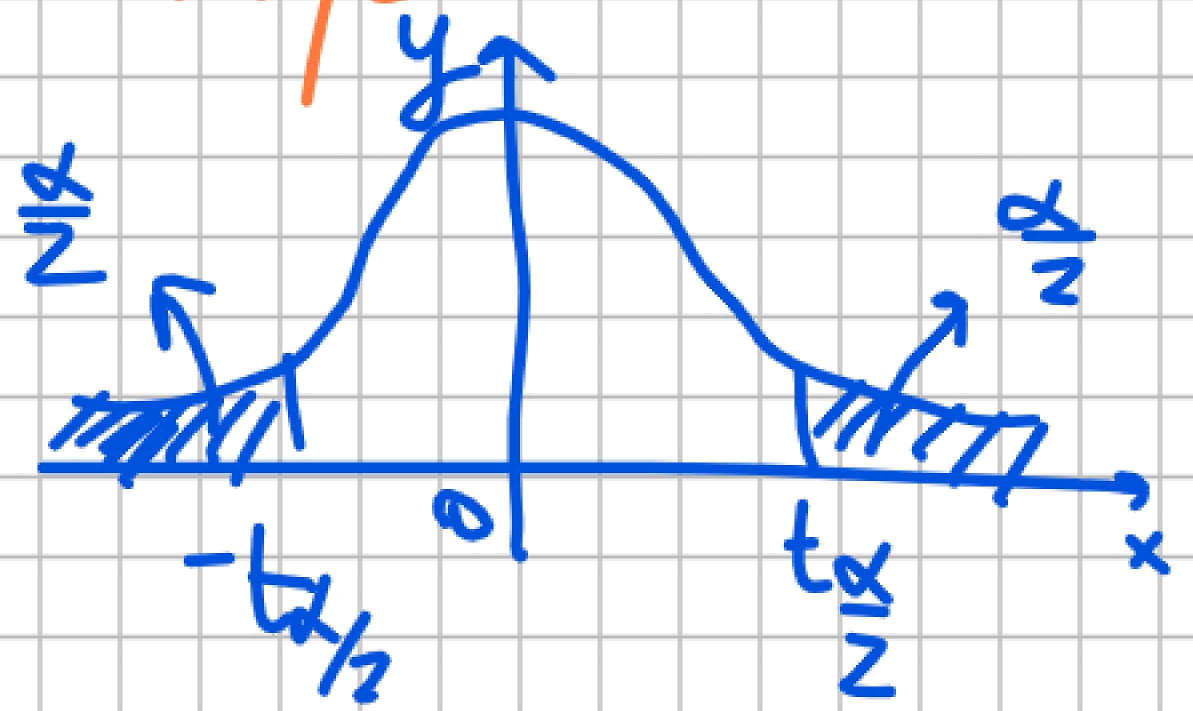
$$t_\alpha = qt(1-\alpha, n-1) = 2.36 \Rightarrow R = (2.36, \infty)$$

$T \in R \Rightarrow H_0$ is rejected

At the 1% significance level, the sample provides significant evidence to say the mean nr. of concurrent users is greater than 35.

a) $\underbrace{(1-\alpha)100\%}_{90\%}$ confidence interval for the population mean μ:

$$\left[\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right]$$



$$t_{\frac{\alpha}{2}} = qt(1-\alpha/2, n-1) = 1.66$$

$\alpha = 0.1$

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 37.7 - 1.66 \cdot \frac{9.2}{\sqrt{100}} = 36.17$$

$$\bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} = 37.7 + 1.66 \cdot \frac{9.2}{\sqrt{100}} = 39.23$$

90% ci for μ: $[36.17, 39.23]$

9.14. Is there significant difference in speed between the two servers in Example 9.21 on p. 263?

(a) Use the confidence interval in Example 9.21 to conduct a two-sided test at the 5% level of significance.

(b) Compute a P-value of the two-sided test in (a).

(c) Is server A really faster? How strong is the evidence? Formulate the suitable hypothesis and alternative and compute the corresponding P-value.

Example 9.21 (COMPARISON OF TWO SERVERS). An account on server A is more expensive than an account on server B. However, server A is faster. To see if it's optimal to go with the faster but more expensive server, a manager needs to know how much faster it is. A certain computer algorithm is executed 30 times on server A and 20 times on server B with the following results,

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

c) $H_0: \mu_1 = \mu_2$

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

σ_1, σ_2 - unknown (pop. std. deviations)

$\Rightarrow$ two-sample t-test.

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim \text{Student } t \text{ distribution with } \nu \text{ degrees of freedom}$$

(we assume unequal variance $\sigma_1^2 \neq \sigma_2^2$)

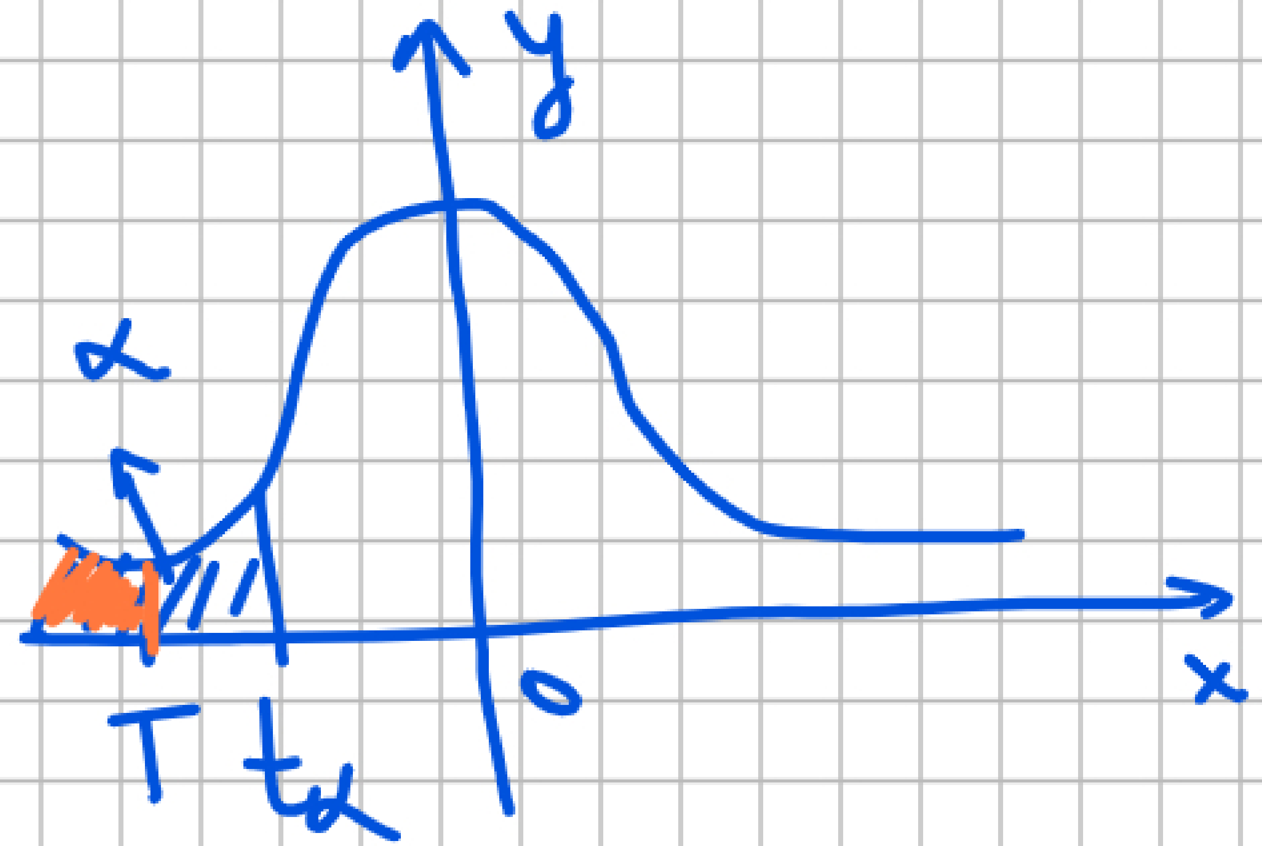
$\nu = 25$

$\bar{x} = 6.7$, $\bar{y} = 7.5$

$s_x = 0.6$, $s_y = 1.2$

$n = 30$, $m = 20$, $\alpha = 5\% = 0.05$

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



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

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

$T = -2.76 \Rightarrow T \in R \Rightarrow H_0$ is reject

At the 5% level of significance, we have enough evidence to say server A is faster.

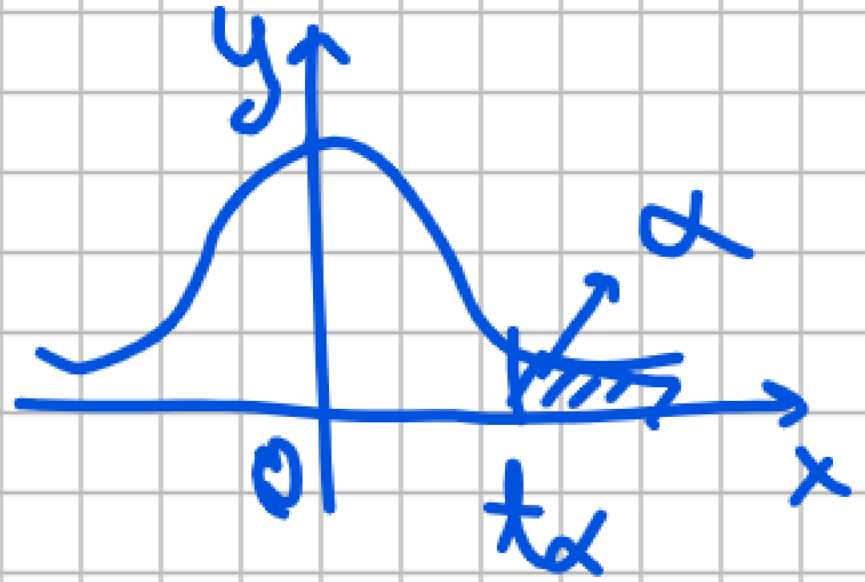
p-value = area from $-\infty$ to T under the PDF

p-value $= pt(T, \nu) = 0.005 < 0.05 = \alpha \Rightarrow H_0$ is rejected

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

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

$R=(t_\alpha,\infty) \Rightarrow$ pvalue $= 1-pt(T, df) = pt(T, df, lower.tail=F)$



$R=(-\infty,-t_{\frac{\alpha}{2}})\cup(t_{\frac{\alpha}{2}},\infty) \Rightarrow$ p-value $= 2*pt(T, df)$ or

$2*pt(T, df, lower.tail=F)$

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?

$\hat{p}_1, \hat{p}_2$ - sample proportions

$\hat{p}_1 = \frac{10}{250} = \frac{1}{25} = 0.04$, $\hat{p}_2 = \frac{18}{300} = \frac{6}{100} = 0.06$

$n=250$, $m=300$, $\alpha=0.02$

$H_0: p_1 = p_2$

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

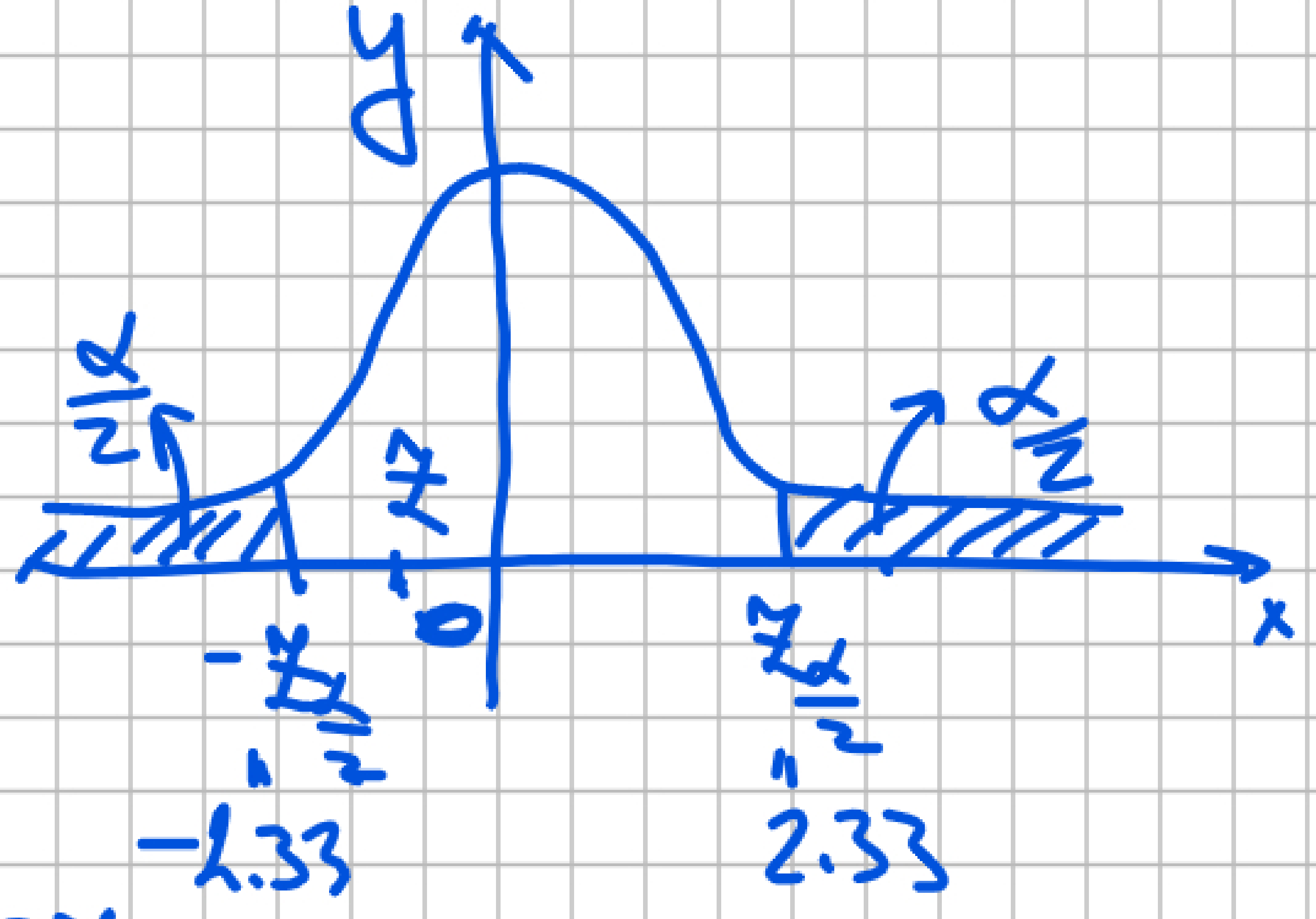
p_1, p_2 - population proportions of defective items

Z-test

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

$\hat{p}$ - pooled proportion

$$\hat{p} = \frac{10+18}{250+300} = \frac{28}{550} = 0.05$$



$$z = \frac{0.04-0.06}{\sqrt{0.05 \cdot 0.95 \cdot \left(\frac{1}{30}+\frac{1}{20}\right)}} = -1.07$$

$$-z_{\frac{\alpha}{2}} = qnorm(\alpha/2) = -2.33 \Rightarrow z_{\frac{\alpha}{2}} = 2.33$$

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

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

At the 2% level of significance there isn't a significant difference in the quality of the 2 lots

9.18. Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

(a) Construct a 95% confidence interval for the difference between the average number of intrusion attempts per day before and after the change of firewall settings (assume equal variances).

(b) Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusion attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?

8.1. The numbers of blocked intrusion attempts on each day during the first two weeks of the month were

56, 47, 49, 37, 38, 60, 50, 43, 43, 59, 50, 56, 54, 58

After the change of firewall settings, the numbers of blocked intrusions during the next 20 days were

53, 21, 32, 49, 45, 38, 44, 33, 32, 43, 53, 46, 36, 48, 39, 35, 37, 36, 39, 45.

$H_0: \mu_1 = \mu_2$

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

σ_1, σ_2 - unknown $\Rightarrow$ two-sample t-test

• If we assume equal variances ($\sigma_1^2 = \sigma_2^2$)

• t.test(x, y, alternative="greater", var.equal = T)

p-value $= 0.00054 < 0.05 = \alpha$

$\Rightarrow H_0$ is rejected

We can claim a signif. reduction in the rate of intrusion attempts, at 2% level signif.

• If we don't assume equal variances

t.test(x, y, alternative='greater')

p-val $= 0.00055 < \alpha \Rightarrow H_0$ is rejected

Example 7. A sample of 6 measurements

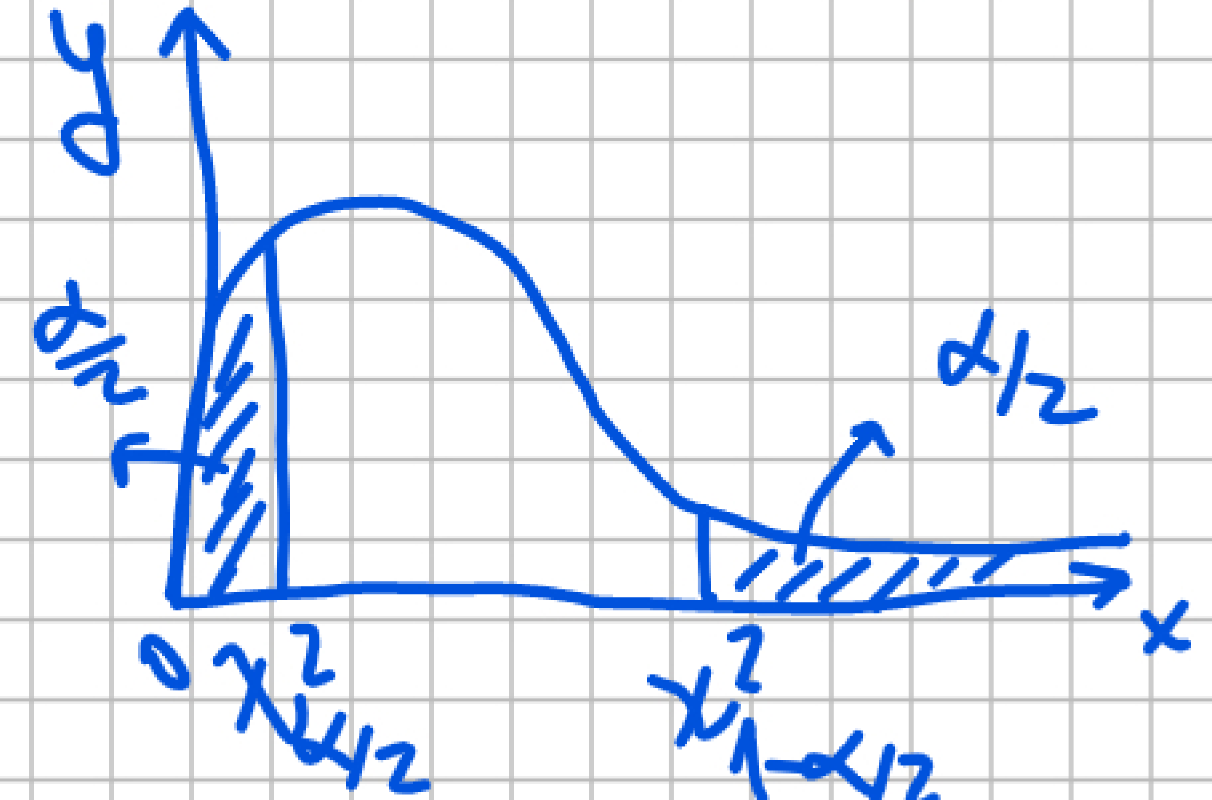
$$2.5, 7.4, 8.0, 4.5, 7.4, 9.2$$

is collected from a Normal distribution with mean μ. Construct the 95% confidence intervals for the population variance σ^2 and standard deviation σ.

$(1-\alpha)100\%$ ci for the population variance (σ^2): $\left[\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}\right]$

n - sample size, s^2 - sample variance

$(1-\alpha)100\%$ ci for the pop. std. dev. σ: $\left[\sqrt{\dfrac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}}, \sqrt{\dfrac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}}\right]$



$\chi^2_{\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1)$

$\chi^2_{1-\frac{\alpha}{2}} = \text{qchisq}(1-\alpha/2, n-1)$

95% ci for σ^2: $\left[\dfrac{5\cdot 6.23}{12.83}, \dfrac{5\cdot 6.23}{0.83}\right] = [2.43, 37.53]$

$n=6$, $s^2 = 6.23$

$\chi^2_{\frac{\alpha}{2}} = 0.83$, $\chi^2_{1-\frac{\alpha}{2}} = 12.83$

95% ci for σ: $\left[\sqrt{2.43}, \sqrt{37.53}\right] = [1.56, 6.13]$