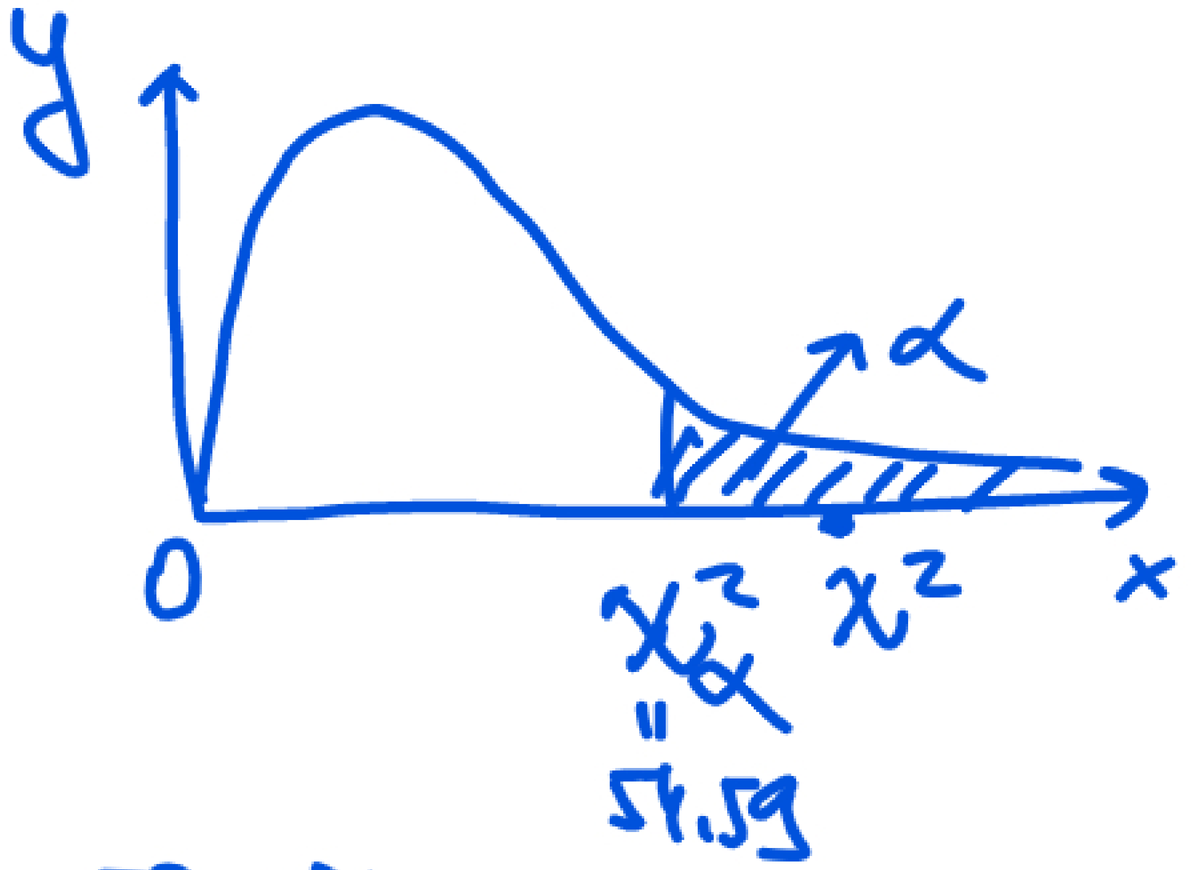
Example 8. Installation of a certain hardware takes random time with a standard deviation of 5 minutes. A manager questions this assumption. Her pilot sample of 40 installation times has a sample standard deviation of $s = 6.2$ min, and she says that it is significantly different from the assumed value of $\sigma = 5$ min. Do you agree with the manager? Conduct the suitable test of a standard deviation.



$H_0: \sigma = 5 \quad (\sigma_0 = 5)$

$H_a: \sigma \neq 5 \Rightarrow R = (\chi^2_\alpha, \infty)$ (*)

$s = 6.2$

$n = 40$

$\alpha = 0.05$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ has a χ^2 distribution with $n-1$ df

$\chi^2 = \frac{39 \cdot 6.2^2}{25} = 59.97$

$\chi^2_\alpha = \text{qchisq}(1-\alpha, n-1) = 54.59$

$R = (54.59, \infty)$

$\chi^2 \in R \Rightarrow H_0$ is rejected

At the 5% level of significance, the population std. deviation σ is significantly

(*) different from the assumed value of 5 min.

9.23. Anthony says to Eric that he is a stronger student because his average grade for the first six quizzes is higher. However, Eric replies that he is more stable because the variance of his grades is lower. The actual scores of the two friends (presumably, independent and normally distributed) are in the table.

	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6
Anthony	85	92	97	65	75	96
Eric	81	79	76	84	83	77

(a) Is there significant evidence to support Anthony's claim? State H_0 and H_A. Test equality of variances and choose a suitable two-sample t-test. Then conduct the test and state conclusions.

(b) Is there significant evidence to support Eric's claim? State H_0 and H_A and conduct the test.

For each test, use the 5% level of significance.

b) $H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2 \Rightarrow R = (F_\alpha, \infty)$

$\alpha = 0.05$

$S_x^2 = 162.8$

$S_y^2 = 10.4$

$$F = \frac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2} = \frac{162.8}{10.4} = 15.65$$

$F_\alpha = qf(1-\alpha, n-1, m-1) = 5.05$

$n = m = 6$

$R = (5.05, \infty)$

$F \in \mathcal{R} \Rightarrow H_0$ is rejected

At the 5% level of significance there is enough evidence to support Eric's claim.

p-value $= 0.0045 < \alpha = 0.05 \Rightarrow H_0$ is rejected

var.test( )

11.12. Masha weighed 7.5 lbs when she was born. Then her weight increased according to the table.

Age (months)	0	2	4	6	8	10	12	14	16	18	20	22
Weight (lbs)	7.5	10.3	12.7	14.9	16.8	18.5	19.9	21.3	22.5	23.6	24.5	25.2

(a) Construct a time plot of Masha's weight. What regression model seems appropriate for these data? linear → plot( )

(b) Fit a quadratic model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. Why does the estimate of β_2 appear negative? → lm( )

(c) What additional portion of the total variation is explained by the quadratic term? Would you keep it in the model?

(d) Predict Masha's weight at 24 months using the quadratic model and using the linear model. Which prediction is more reasonable? linear → predict( )

(e) Can one test significance of the quadratic term by an F-test? Why or why not?

$y = 9.46 + 0.79 \cdot x$

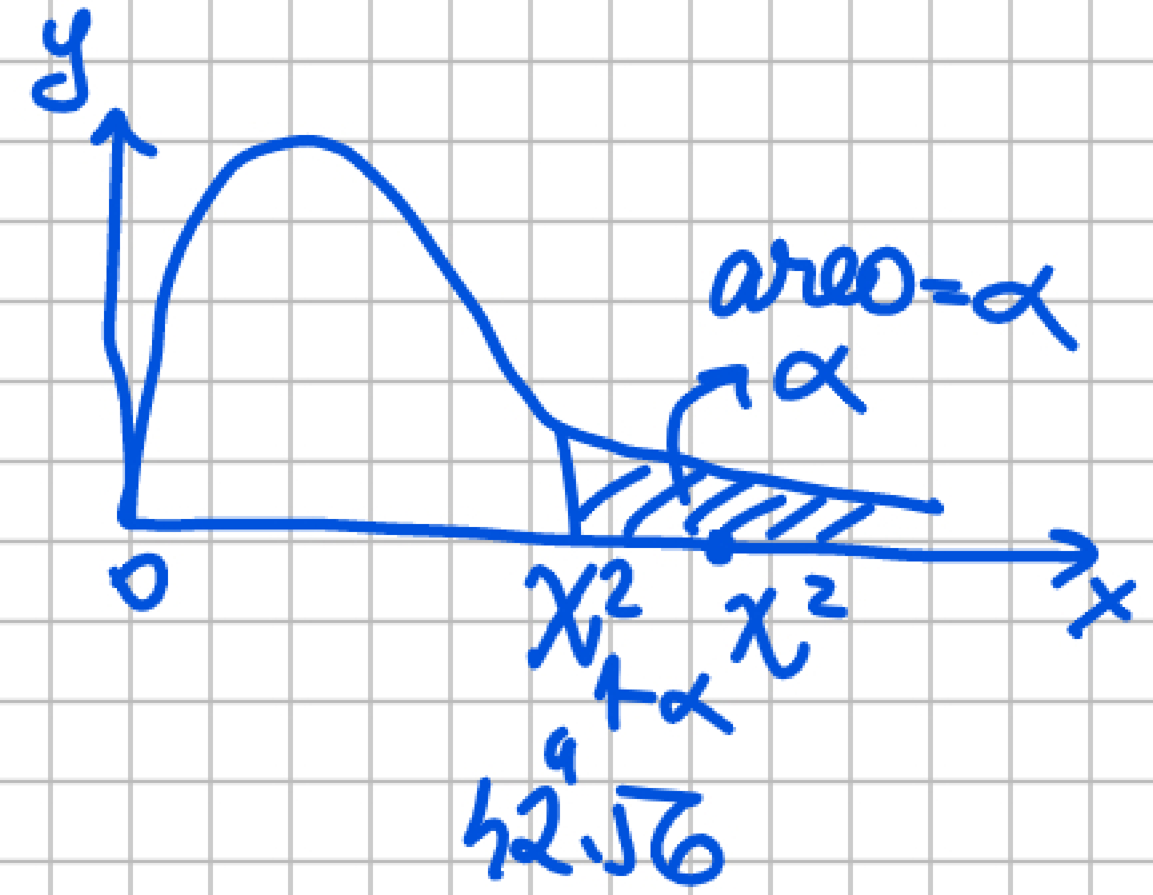
↑ weight ↑ age

(d) age = 24 ($x = 24$) ⇒ weight $= 9.46 + 0.79 \cdot 24 = 28.42$

sgr. 2

Exercise 103. A production process is considered to be out of control if the produced parts have a mean length different from 27.5 millimeters or a standard deviation that is greater than 0.5 millimeter. A sample of 30 parts yields a sample mean of 27.63 millimeters and a sample standard deviation of 0.87 millimeters.

a. At the 0.05 level of significance, does this sample indicate that the process should be adjusted in order to correct the standard deviation of the product?



$H_0: \sigma = 0.5 \quad (\sigma_0 = 0.5)$

$H_a: \sigma > 0.5 \Rightarrow R = (\chi^2_{1-\alpha}, \infty)$

$s = 0.87, \quad n = 30$

$\alpha = 0.05$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{29 \cdot 0.87^2}{0.5^2} = 87.8 \rightarrow$ has a χ^2 distribution with n-1 df

$\chi^2_{1-\alpha} = \text{qchisq}(1-\alpha, n-1) = 42.56 \Rightarrow R = (42.56, \infty)$

$= \text{qchisq}(\alpha, n-1, \text{lower.tail} = F)$

$\chi^2 \in R \Rightarrow H_0$ is rejected

At the 0.05 level of significance, the sample indicates that the process should be adjusted!

9.23. Anthony says to Eric that he is a stronger student because his average grade for the first six quizzes is higher. However, Eric replies that he is more stable because the variance of his grades is lower. The actual scores of the two friends (presumably, independent and normally distributed) are in the table.

	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6
Anthony	85	92	97	65	75	96
Eric	81	79	76	84	83	77

(a) Is there significant evidence to support Anthony's claim? State H_0 and H_A. Test equality of variances and choose a suitable two-sample t-test. Then conduct the test and state conclusions.

(b) Is there significant evidence to support Eric's claim? State H_0 and H_A and conduct the test.

For each test, use the 5% level of significance.

(b) $H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2 \Rightarrow R = (F_{1-\alpha}, \infty)$

$\alpha = 5\% = 0.05$

$F = \dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}$ has a F distribution with $n-1, m-1$ degrees of freedom

$F = \dfrac{S_x^2}{S_y^2} = \dfrac{162.8}{10.4} = 15.65$

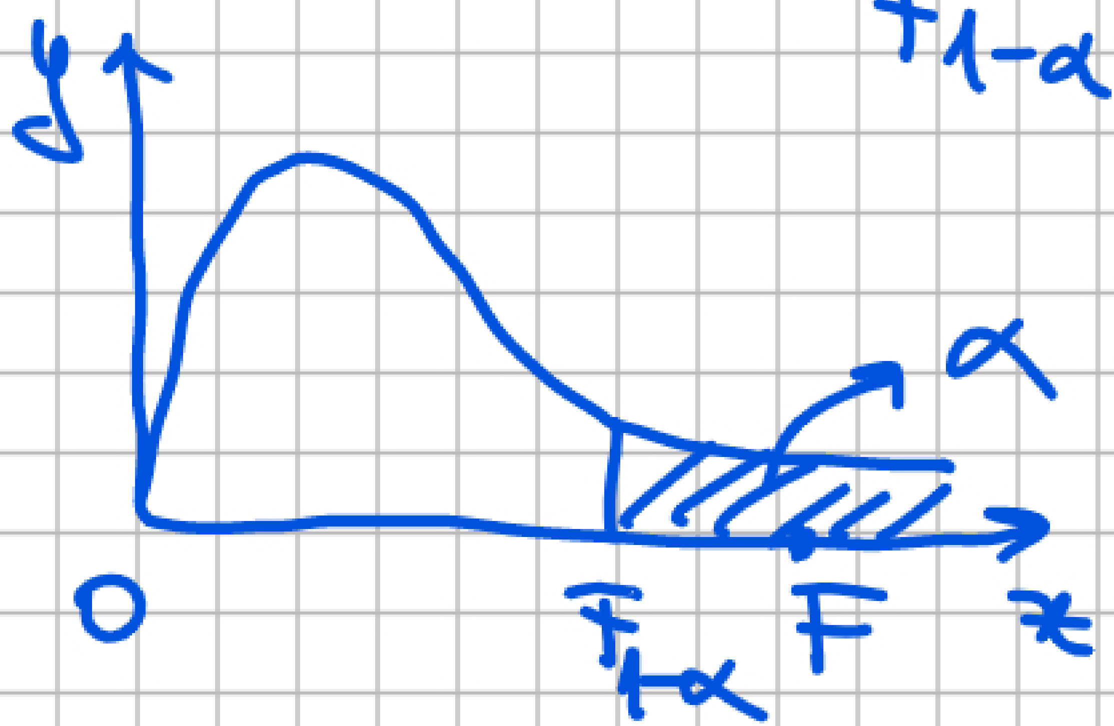
S_x^2, S_y^2 – sample variance

$S_x^2 = 162.8$

$S_y^2 = 10.4$

$F_{1-\alpha} = qf(1-\alpha, n-1, m-1) = 5.05$

$n = m$



$\Rightarrow R = (5.05, \infty)$

$F \in R \Rightarrow H_0$ is rejected

There is significant evidence to support Eric's claim, at 5% level of signif.

p-value $= 0.0045 < \alpha = 0.05 \Rightarrow H_0$ is rejected

R: var.test( )

Example 1. **(World population).** According to the International Data Base of the U.S. Census Bureau, population of the world grows according to Table 11.1. How can we use these data to predict the world population in years 2015 and 2020?

Year	Population mln. people	Year	Population mln. people	Year	Population mln. people
1950	2558	1975	4089	2000	6090
1955	2782	1980	4451	2005	6474
1960	3043	1985	4855	2010	6864
1965	3350	1990	5287	2015	?
1970	3712	1995	5700	2020	?

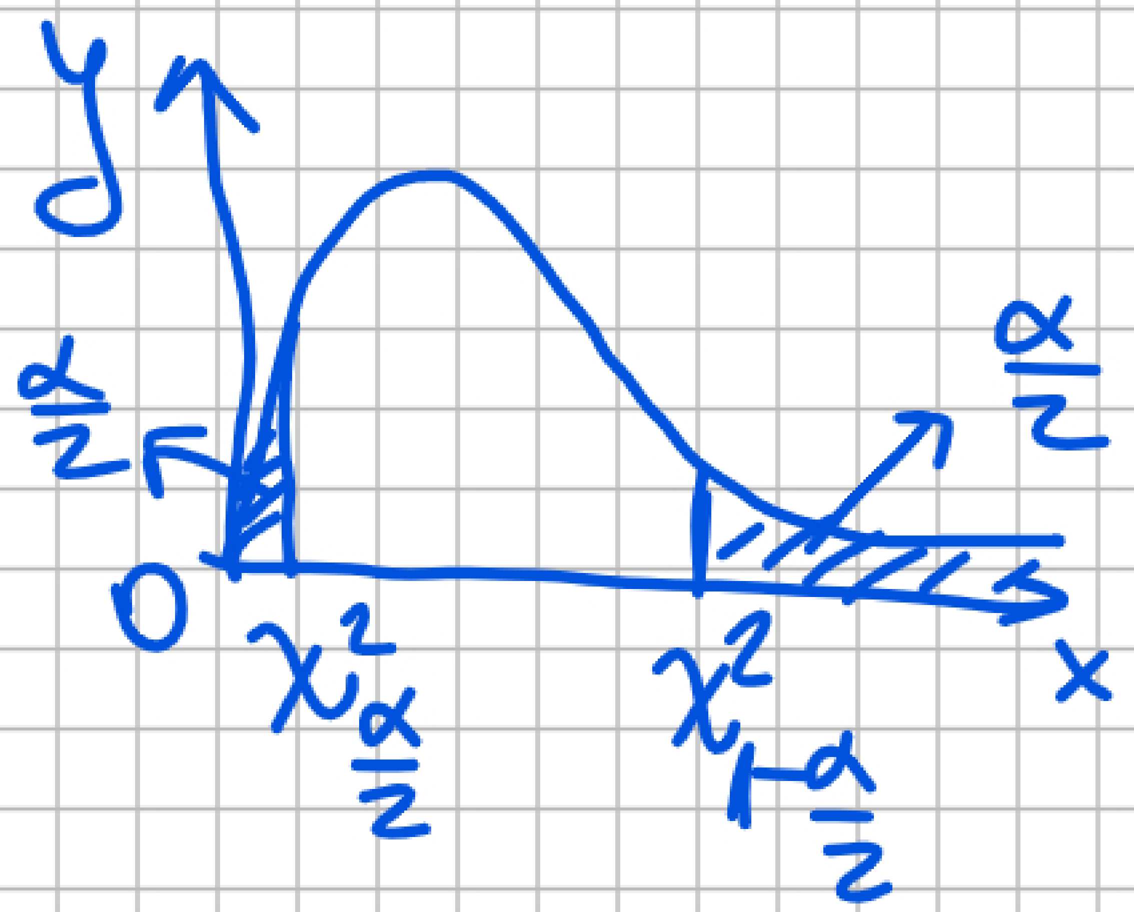
- define 2 vectors in R for the year and the population
- do a scatterplot : plot( )
- do a regression model : lm( )

Y= population
X= year

ex. 11.2 page 400 in the textbook → linear regression

sgr. 5

Example 8. Installation of a certain hardware takes random time with a standard deviation of 5 minutes. A manager questions this assumption. Her pilot sample of 40 installation times has a sample standard deviation of $s = 6.2$ min, and she says that it is significantly different from the assumed value of $\sigma = 5$ min. Do you agree with the manager? Conduct the suitable test of a standard deviation.



$H_0: \sigma = 5$

$H_a: \sigma \neq 5$. $\Rightarrow R = (0, \chi^2_{\frac{\alpha}{2}}) \cup (\chi^2_{1-\frac{\alpha}{2}}, \infty)$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(40-1)\cdot 6.2^2}{5^2} = 59.97$ - has a χ^2 distribution with $n-1$ df

$\chi^2_{\frac{\alpha}{2}} = \text{qchisq}(\alpha/2, n-1) = \text{qchisq}(0.05/2, 39) = 23.65$

$\alpha = 5\% = 0.05$ $\chi^2_{1-\frac{\alpha}{2}} = \text{qchisq}(1-\alpha/2, n-1) = 58.12$

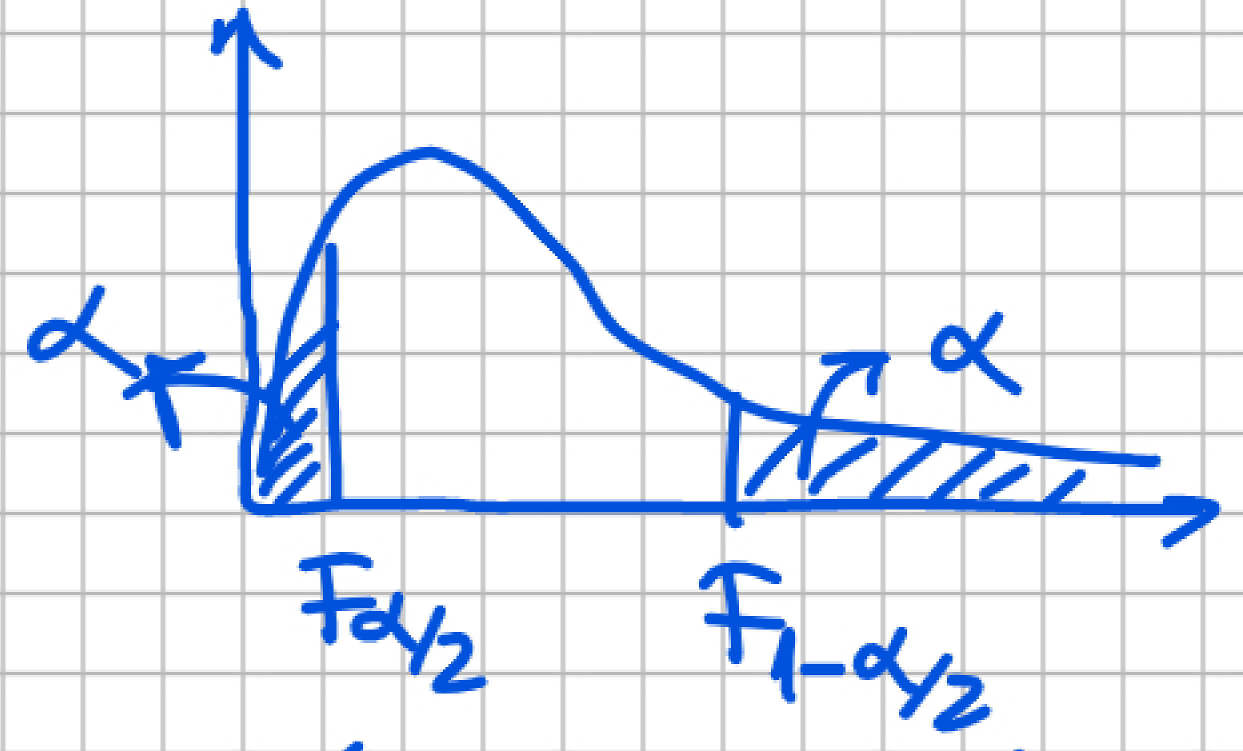
$\Rightarrow R = (0, 23.65) \cup (58.12, \infty)$

$\chi^2 = 59.97 \in R \Rightarrow H_0$ is rejected

At the 5% level of significance there is significant evidence to suggest that the std. deviation is different from 5 min.

An insurance company sells health insurance and motor insurance policies. Premiums are paid by customers for these policies. The CEO of the insurance company wonders if premiums paid by either of insurance segments (health insurance and motor insurance) are more variable as compared to another. He finds the following data for premiums paid:

	A	B	C	D
1		**Health Insurance**	**Motor Insurance**	
2	**Variance**	$200	$50	
3	**Sample Size**	11	51	
4				



$\alpha = 5\% = 0.05$

$H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 \neq \sigma_2^2 \Rightarrow R = (0, F_{\alpha/2}) \cup (F_{1-\alpha/2}, \infty)$

qf( )

$F = \dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}$ has a F distribution with $n-1, m-1$ df

n, m – sample sizes

S_x^2, S_y^2 – sample variance (s_1^2, s_2^2)

$F = \dfrac{S_x^2}{S_y^2} = \dfrac{200}{50} = 4$

$F_{\alpha/2} = qf(\alpha/2, n-1, m-1) = 0.31$

n, m – sample sizes

$F_{1-\alpha/2} = qf(1-\alpha/2, n-1, m-1) = 2.32$

$R = (0, 0.31) \cup (2.32, \infty)$

$F = 4 \Rightarrow F \in R \Rightarrow H_0$ is rejected

At the 5% level of significance there is significant evidence to suggest that either insurance segments are more variable compared to the other.

11.12. Masha weighed 7.5 lbs when she was born. Then her weight increased according to the table.

Age (months)	0	2	4	6	8	10	12	14	16	18	20	22
Weight (lbs)	7.5	10.3	12.7	14.9	16.8	18.5	19.9	21.3	22.5	23.6	24.5	25.2

(a) Construct a time plot of Masha's weight. What regression model seems appropriate for these data? linear → scatterplot : plot()

(b) Fit a quadratic model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. Why does the estimate of β_2 appear negative? → lm(y ~ x)

summary(lm(y ~ x))

(c) What additional portion of the total variation is explained by the quadratic term? Would you keep it in the model?

(d) Predict Masha's weight at 24 months using the quadratic model and using the linear model. Which prediction is more reasonable?

(e) Can one test significance of the quadratic term by an F-test? Why or why not?

Linear regression model : $y = +9.45 + 0.79 \cdot x$

(y = weight, x = age)

age = 24 ⇒ weight = 9.45 + 0.79 · 24 = **28.41**

sgr. 6

Math instructors are not only interested in how their students do on exams, on average, but how the exam scores vary. To many instructors, the variance (or standard deviation) may be more important than the average.

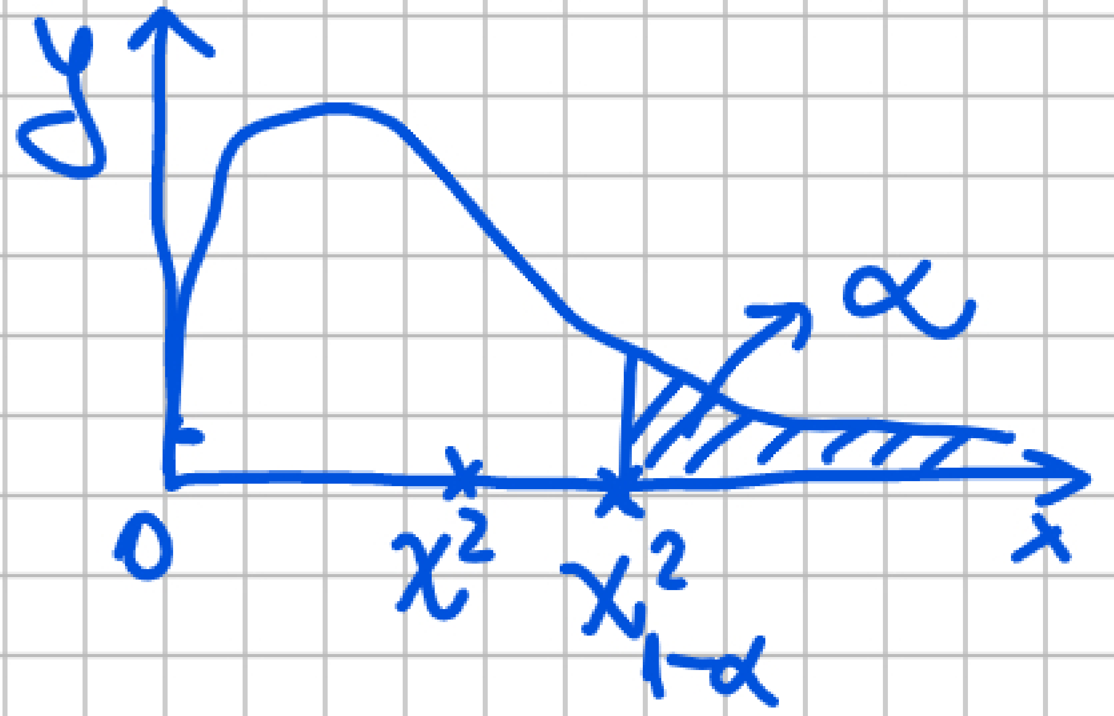
Suppose a math instructor believes that the standard deviation for his final exam is five points. One of his best students thinks otherwise. The student claims that the standard deviation is more than five points. If the student were to conduct a hypothesis test, what would the null and alternative hypotheses be?

$H_0: \sigma^2 = 5^2$

$H_a: \sigma^2 > 5^2 \Rightarrow R = (\chi^2_{1-\alpha}, \infty)$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ has a χ^2 distribution with n-1 df

$s^2 = 5.7^2$



qchisq(1-α, n-1) = qchisq(α, n-1, lower.tail=F) = 30.14

$\alpha = 5\% = 0.05$, $n = 20$ $\Rightarrow R = (30.14, \infty)$

$\chi^2 = \frac{19 \cdot 5.7^2}{25} = 24.69$

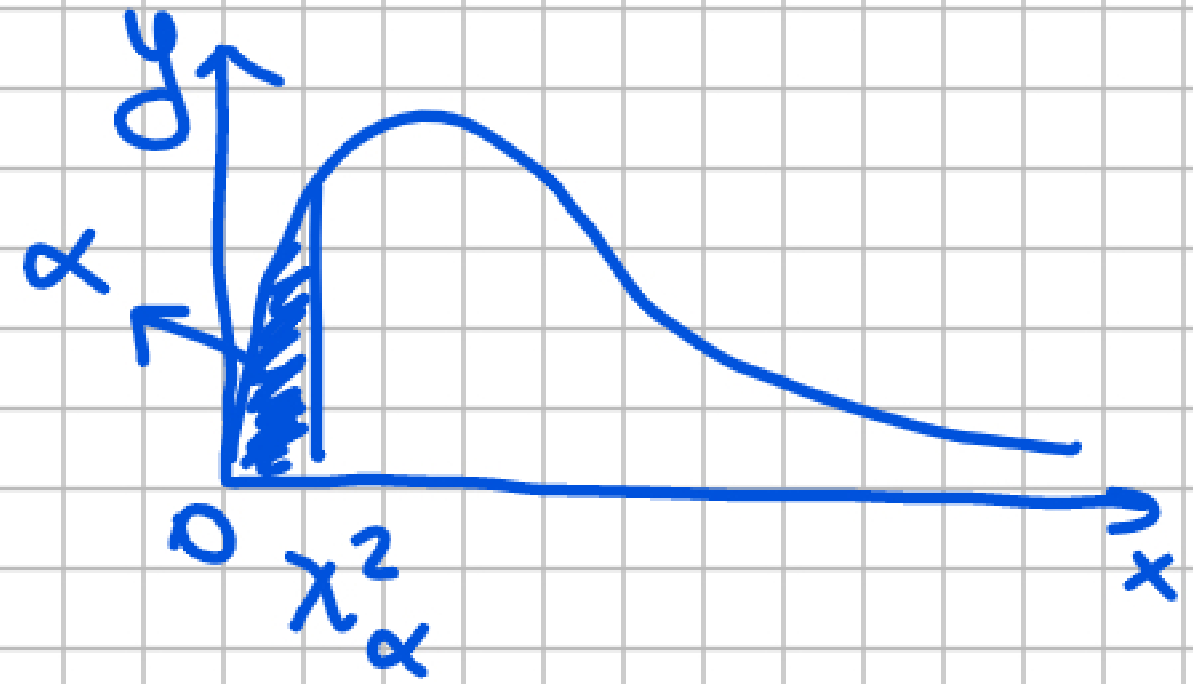
$\chi^2 \notin R \Rightarrow H_0$ is not rejected

At the 5% level of significance, there isn't signif. evidence to say the std. dev. is larger than 5

A statistician wishes to test the claim that the standard deviation of the weights of firemen is less than 25 pounds. She selected a random sample of 20 firemen and found $s = 23.2$ pounds. Assuming that the weights of firemen are normally distributed, test the claim of the statistician at the 0.05 level of significance.

$H_0: \sigma^2 = 25^2$

$H_a: \sigma^2 < 25^2 \Rightarrow R = (0, \chi^2_\alpha)$



$s = 23.2$, $n = 20$

$\alpha = 0.05$

$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2} = 16.36$

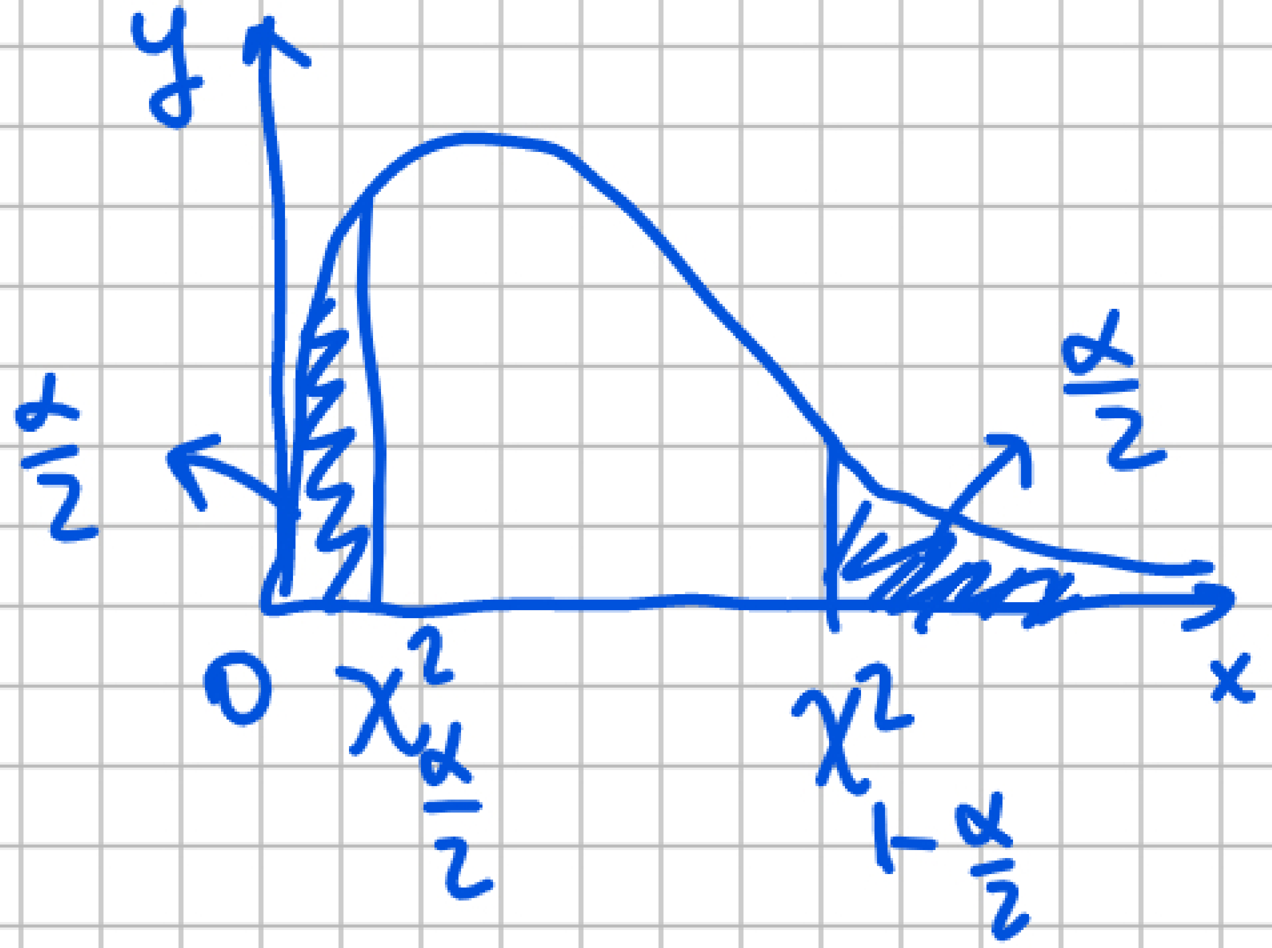
$\chi^2_\alpha = \text{qchisq}(0.05, n-1) = 10.1$

$R = (0, 10.1)$

$\chi^2 \notin R \Rightarrow H_0$ is not rejected

At the 5% level of significance, the sample doesn't show sufficient evidence to support the claim that the std. dev. is less than 25.

$$H_0: \sigma^2 \neq \sigma_0^2 \Rightarrow R = \left(0, \chi^2_{\frac{\alpha}{2}}\right) \cup \left(\chi^2_{1-\frac{\alpha}{2}}, \infty\right)$$

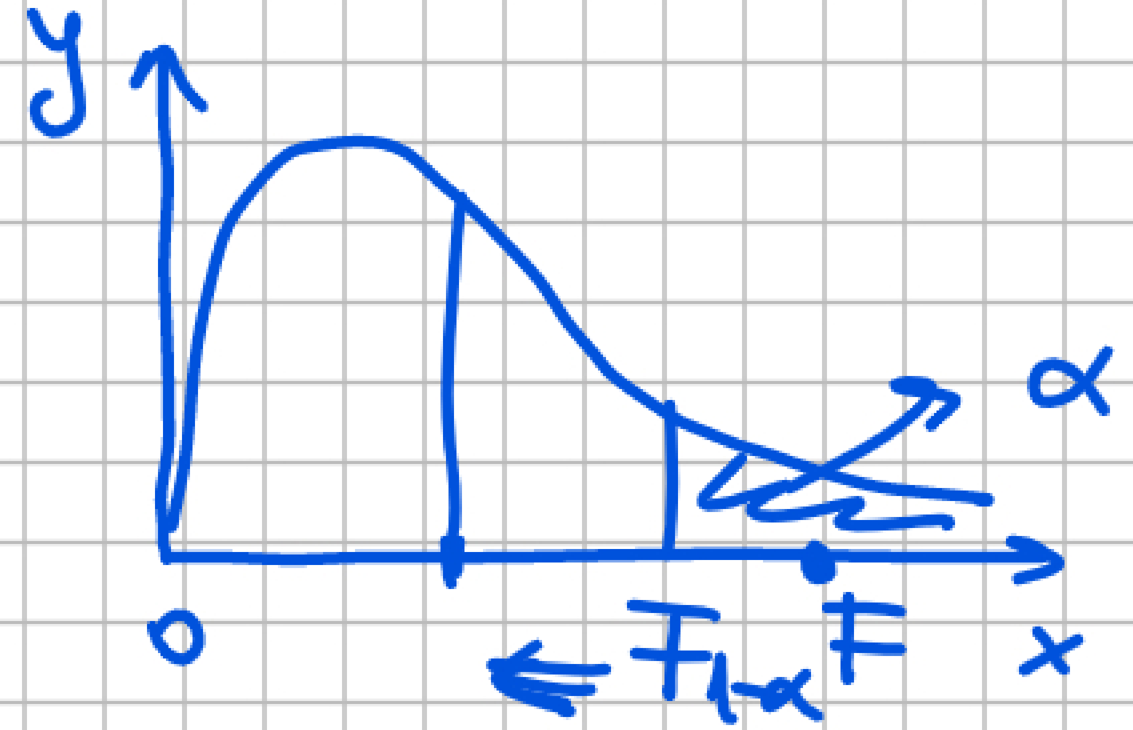


A researcher wanted to see if women varied more than men in weight. Nine women and sixteen men were weighed. The variance for the women was 525 and the variance for the men was 142. What can be concluded at the 0.05 level of significance?

$H_0: \sigma_1^2 = \sigma_2^2$

$H_a: \sigma_1^2 > \sigma_2^2 \Rightarrow R = (F_{1-\alpha}, \infty)$

$F = \dfrac{S_X^2}{S_Y^2}$ has a F distribution with $n-1, m-1$ df



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

$\alpha = 0.05$

$qf(1-\alpha, n-1, m-1) = 2.64 \Rightarrow R = (2.64, \infty)$

$n = 9,\ m = 16$

$F = \dfrac{525}{142} = 3.70$

$F \in R \Rightarrow H_0$ is rejected

At the 5% level of significance the samples show sufficient evidence to say women varied more than men in weight.

11.12. Masha weighed 7.5 lbs when she was born. Then her weight increased according to the table.

x Age (months)	0	2	4	6	8	10	12	14	16	18	20	22
y Weight (lbs)	7.5	10.3	12.7	14.9	16.8	18.5	19.9	21.3	22.5	23.6	24.5	25.2

(a) Construct a time plot of Masha's weight. What regression model seems appropriate for these data?

(b) Fit a ~~quadratic~~ linear model $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$. Why does the estimate of β_2 appear negative?

(c) What additional portion of the total variation is explained by the quadratic term? Would you keep it in the model?

(d) Predict Masha's weight at 24 months using the ~~quadratic model and using the~~ linear model. Which prediction is more reasonable?

(e) Can one test significance of the quadratic term by an F-test? Why or why not?

→ plot()

$y = \beta_0 + \beta_1 x$; lm()

model=lm(y ~ x)

summary(model)

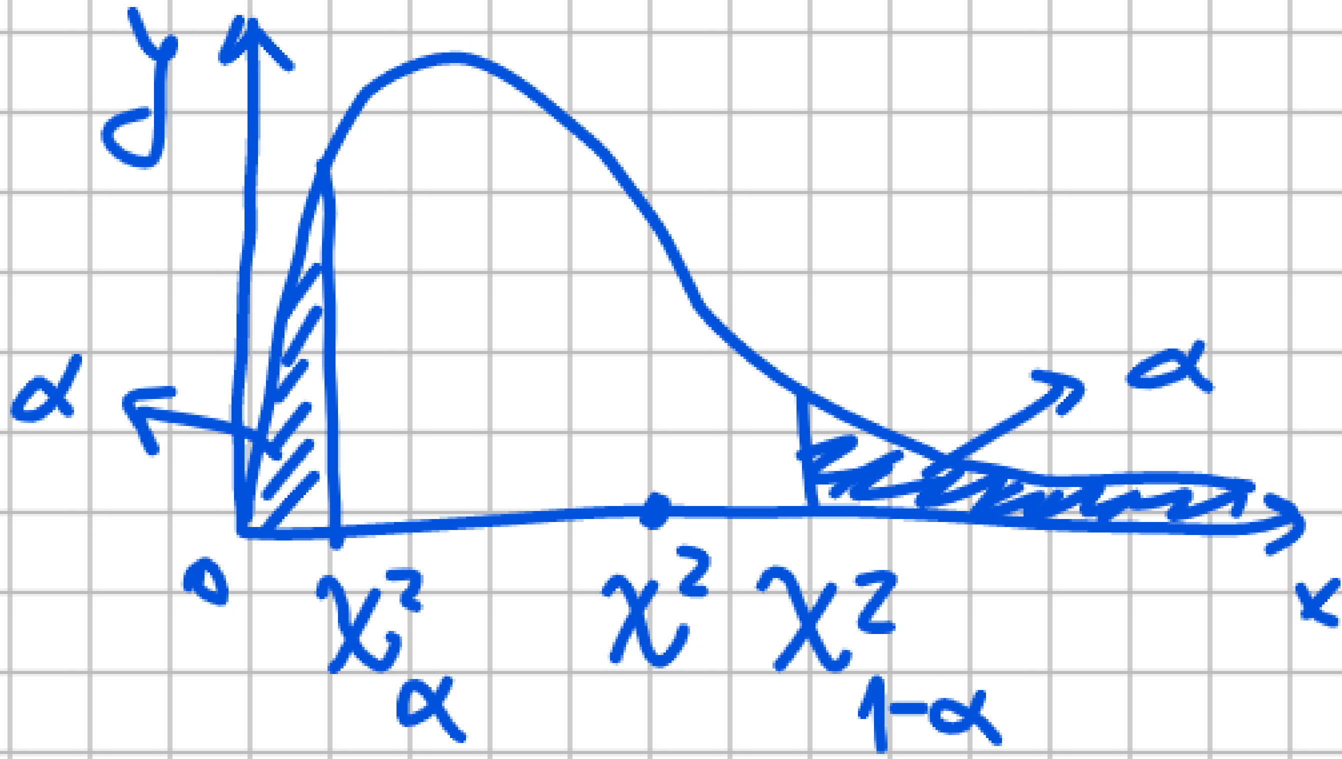
(d) $x = 24 \Rightarrow y = ?$

$y = 9.46 + 0.79 \cdot x$

$x = 24 \Rightarrow y = 9.46 + 0.79 \cdot 24 = 28.42$

An engineer is investing the amount of standard deviation in the time it takes a 3D printer to make a particular part. The engineer believes that the standard deviation in the time it takes to make the part is more than 2. Test this at $\alpha = 0.01$ level of significance, using 11 sample times taken while the printer was making these parts. The sample standard deviation is 2.3.

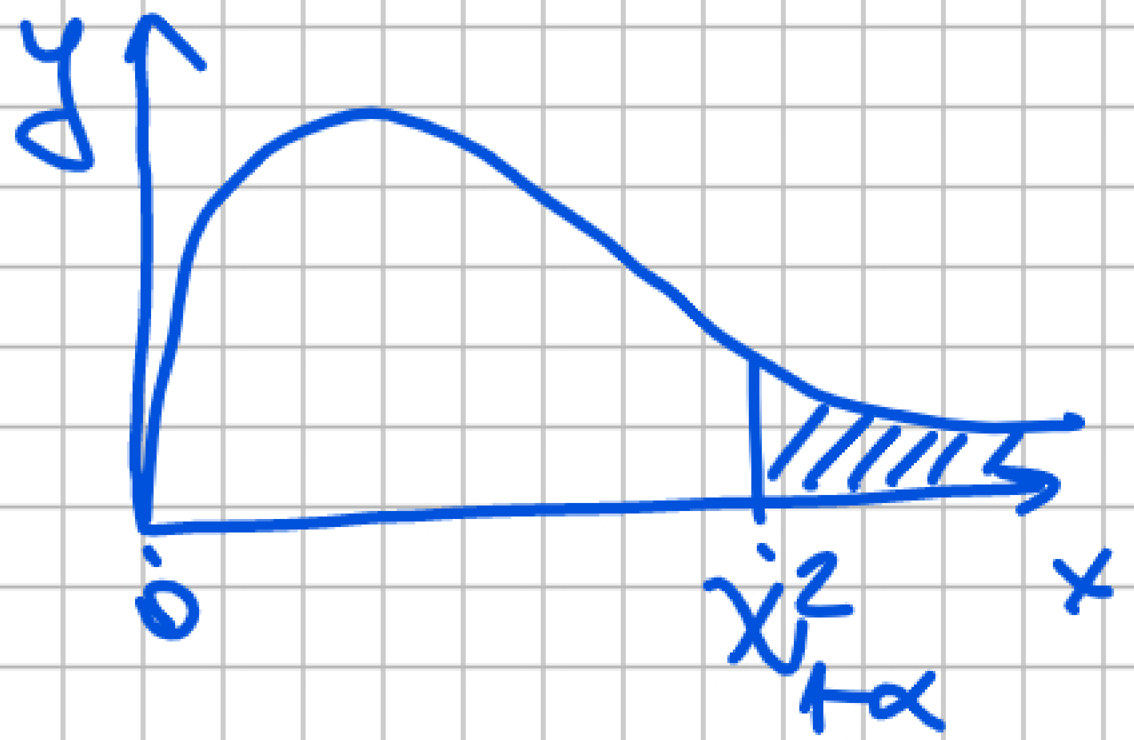
sgr. 1



$H_0: \sigma = 2$

$H_a: \sigma > 2 \Rightarrow R = (\chi^2_{1-\alpha}, \infty)$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{10 \cdot 2.3^2}{2^2} = 13.23$



qchisq(α / $1-\alpha$, $n-1$)

n = sample size

$s = 2.3$

$n = 11$

$\alpha = 0.01$

$\chi^2_{1-\alpha}$ = qchisq($1-\alpha$, 10) = 23.21

$\rightarrow R = (23.21, \infty)$

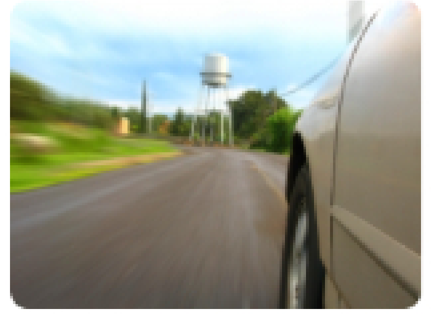
$\chi^2 \notin R \Rightarrow H_0$ is not rejected

At the 1% level of significance there isn't enough evidence to suggest that the std. deviation of the population is more than 2.

A cigarette manufacturer wishes to test the claim that the variance of nicotine content of its cigarettes is 0.644. Nicotine content is measured in milligrams and is assumed normally distributed. A sample of 20 cigarettes has a standard deviation of 1.00 milligram. At $\alpha = 0.01$, is there enough evidence to reject the manufacturer's claim?

→ homework

A psychologist was interested in exploring whether or not male and female college students have different driving behaviors. The particular statistical question she framed was as follows:



> Is the mean fastest speed driven by male college students different than the mean fastest speed driven by female college students?

The psychologist conducted a survey of a random $n = 34$ male college students and a random $m = 29$ female college students. Here is a descriptive summary of the results of her survey:

Males (X)	**Females** (Y)
$n = 34$	$m = 29$
$\overline{x} = 105.5$	$\overline{y} = 90.9$
$s_x = 20.1$	$s_y = 12.2$

- Is there sufficient evidence at the $\alpha = 0.05$ level to conclude that the variance of the fastest speed driven by male college students differs from the variance of the fastest speed driven by female college students?

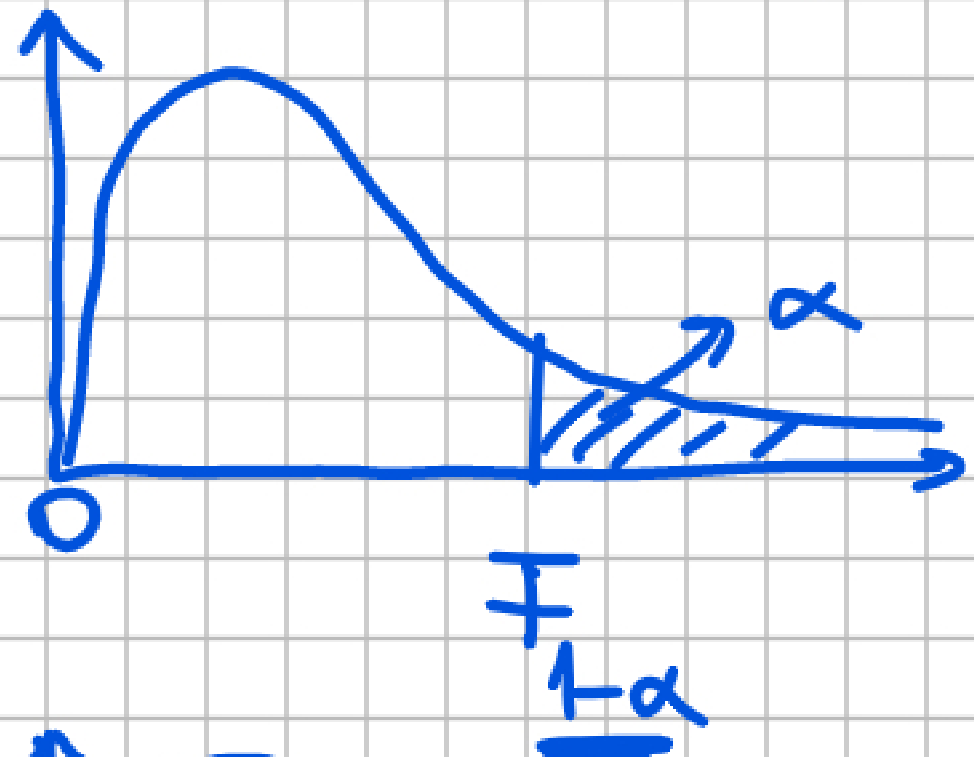
$H_0: \sigma_1^2 = \sigma_2^2$

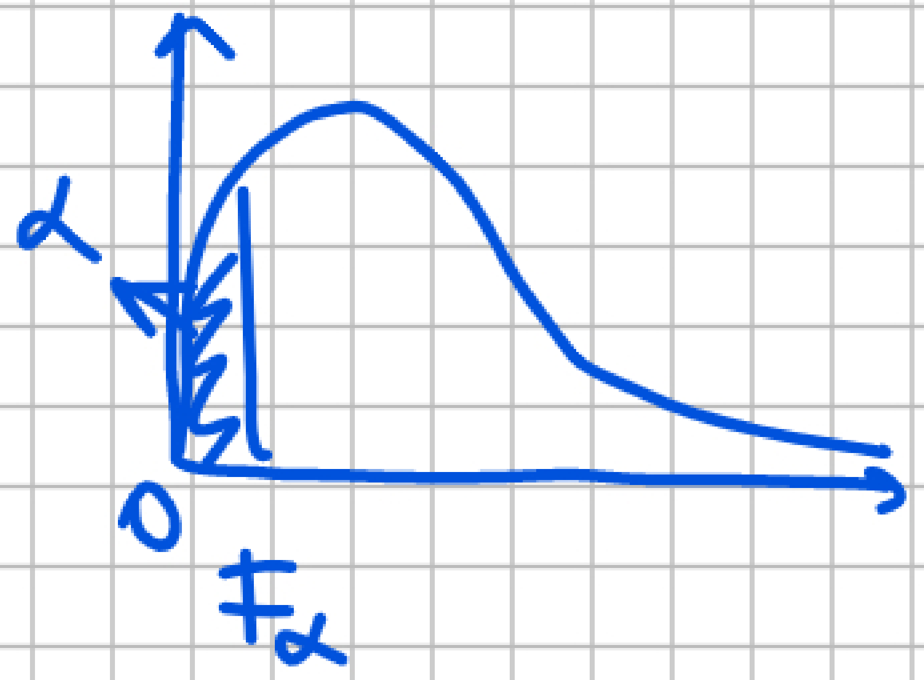
$H_a: \sigma_1^2 \neq \sigma_2^2 \Rightarrow R = (0, F_{\frac{\alpha}{2}}) \cup (F_{1-\frac{\alpha}{2}}, \infty)$

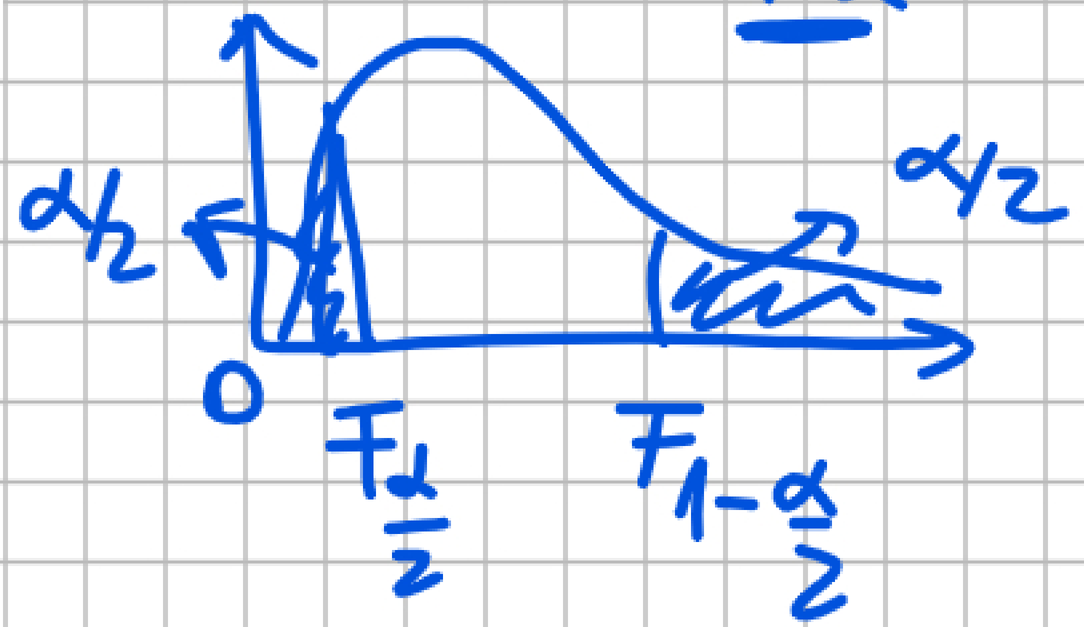
$F = \frac{s_x^2}{s_y^2}$ has a F distribution with $n-1$, $m-1$ df

n, m - sample sizes, $\alpha = 0.05$

s_x^2, s_y^2 - sample variances







qf(— , n-1, m-1)

$F = 2.71$

$F_{\alpha/2} = qf(\alpha/2, n-1, m-1) = 0.48$

$F_{1-\alpha/2} = qf(1-\alpha/2, n-1, m-1) = 2.08$

$\Rightarrow R = (0, 0.48) \cup (2.08, \infty)$

$F = 2.71$

$F \in R \Rightarrow H_0$ is rejected

At the 5% level of significance

we can conclude that the variance of the fastest speed driven by male college students is different from the fastest speed driven by female college students.

Example 1. (World population). According to the International Data Base of the U.S. Census Bureau, population of the world grows according to Table 11.1. How can we use these data to predict the world population in years 2015 and 2020?

Year	Population mln. people	Year	Population mln. people	Year	Population mln. people
1950	2558	1975	4089	2000	6090
1955	2782	1980	4451	2005	6474
1960	3043	1985	4855	2010	6864
1965	3350	1990	5287	2015	?
1970	3712	1995	5700	2020	?

TABLE 11.1: *Population of the world, 1950–2020.*

$y = \beta_0 + \beta_1 x$

y → pop. ; x → year

$x = 2015$

$x = 2020$

- define in R two vectors : year, population
  seq( )
- do a scatterplot of the data : plot( )
- find the linear regression model
  lm(y ~ x)

$y = -142201.26 + 74.12 \cdot x$

$x = 2015 \Rightarrow y = 7150.54$

$x = 2020 \Rightarrow y = 7521.14$

cor(year, pop)

0.998

sgr. 4

With individual lines at its various windows, a post office finds that the standard deviation for normally distributed waiting times for customers on Friday afternoon is 7.2 minutes. The post office experiments with a single, main waiting line and finds that for a random sample of 25 customers, the waiting times for customers have a standard deviation of 3.5 minutes.

With a significance level of 5%, test the claim that **a single line causes lower variation among waiting times (shorter waiting times) for customers.**

$H_0: \sigma^2 = 7.2^2 \quad (\sigma_0^2 = 7.2^2)$

$H_a: \sigma^2 < 7.2^2 \Rightarrow R = (0, \chi^2_\alpha)$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$ – has a χ^2 distribution with $n-1$ df

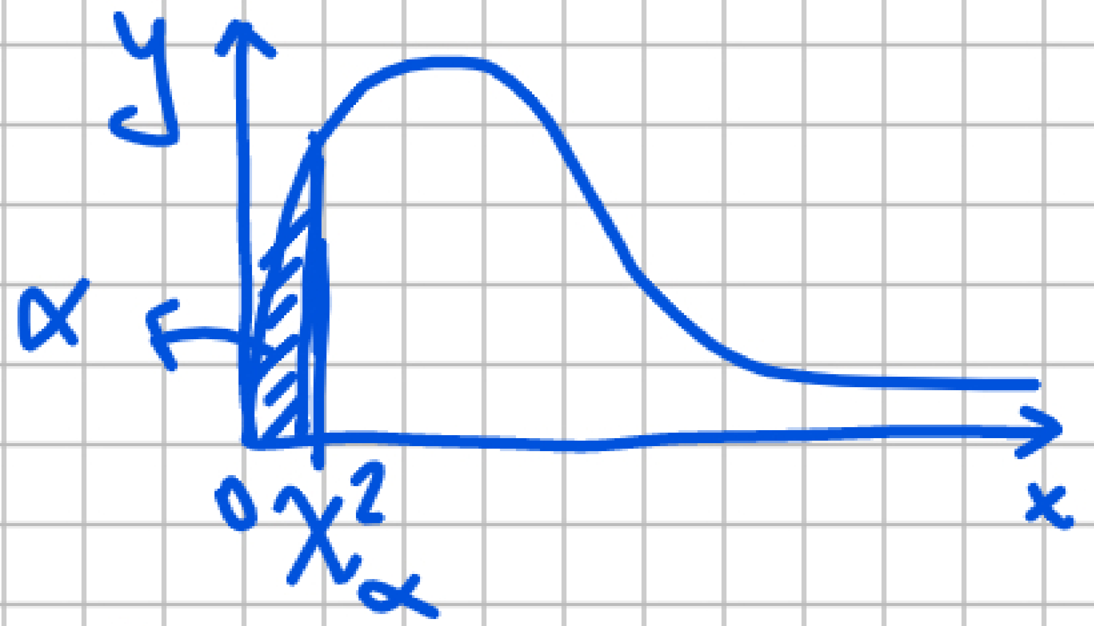
$n = 25, \ s^2 = 3.5^2$

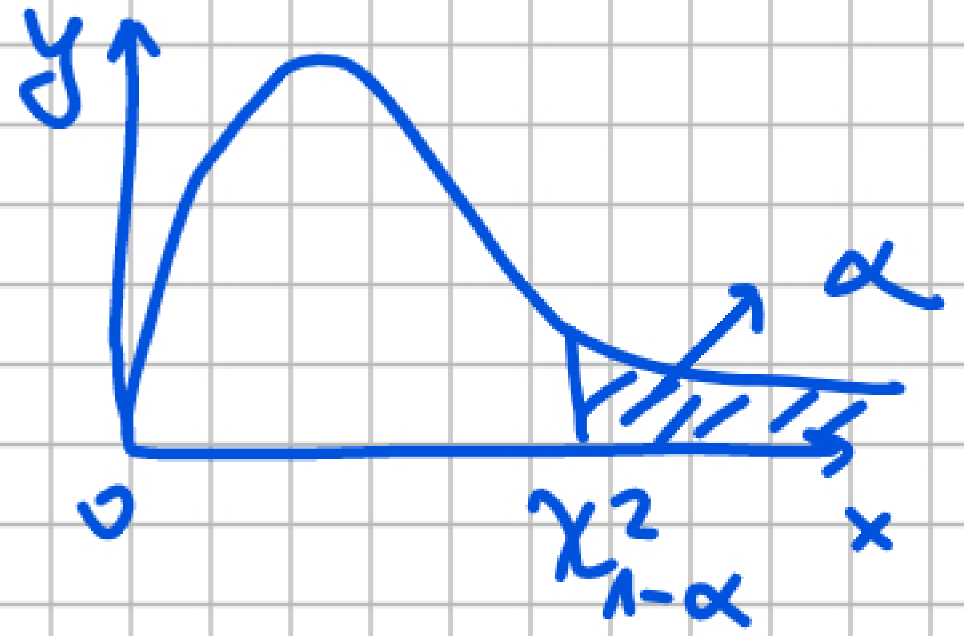
$\chi^2_\alpha = \text{qchisq}(\alpha, n-1) = 13.85$

$1-\alpha$

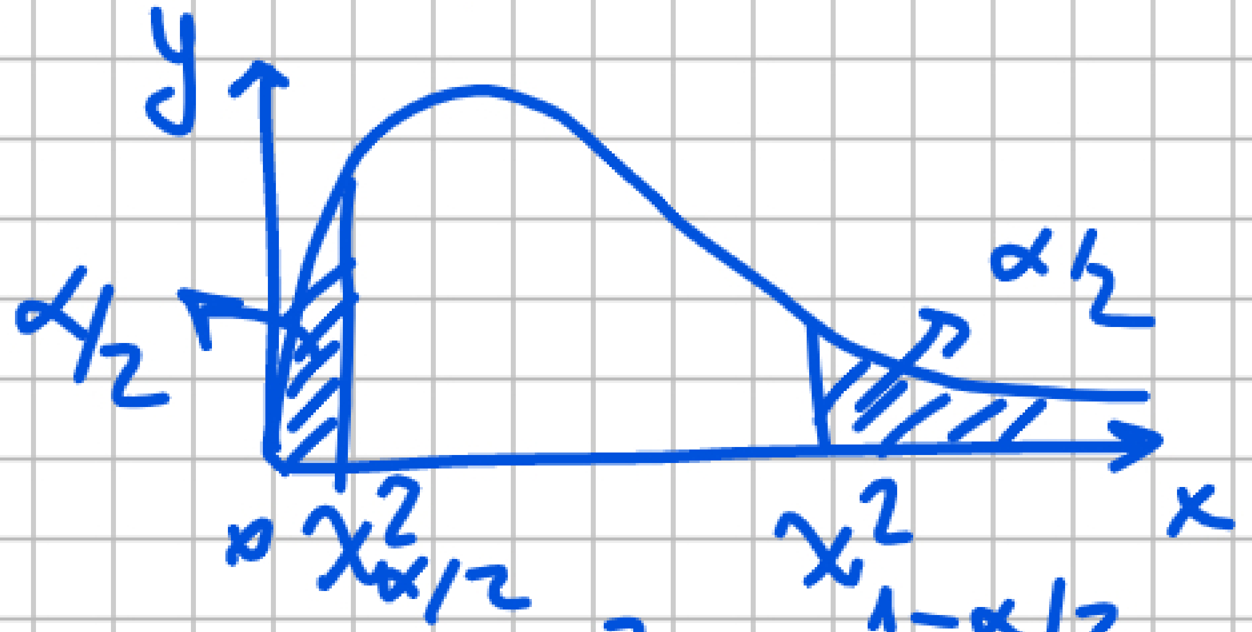
$\alpha/2$

$1-\alpha/2$





$\alpha = 0.05$



$\chi^2 = \frac{24 \cdot 3.5^2}{7.2^2} = 5.67$

$R = (0, 13.85)$

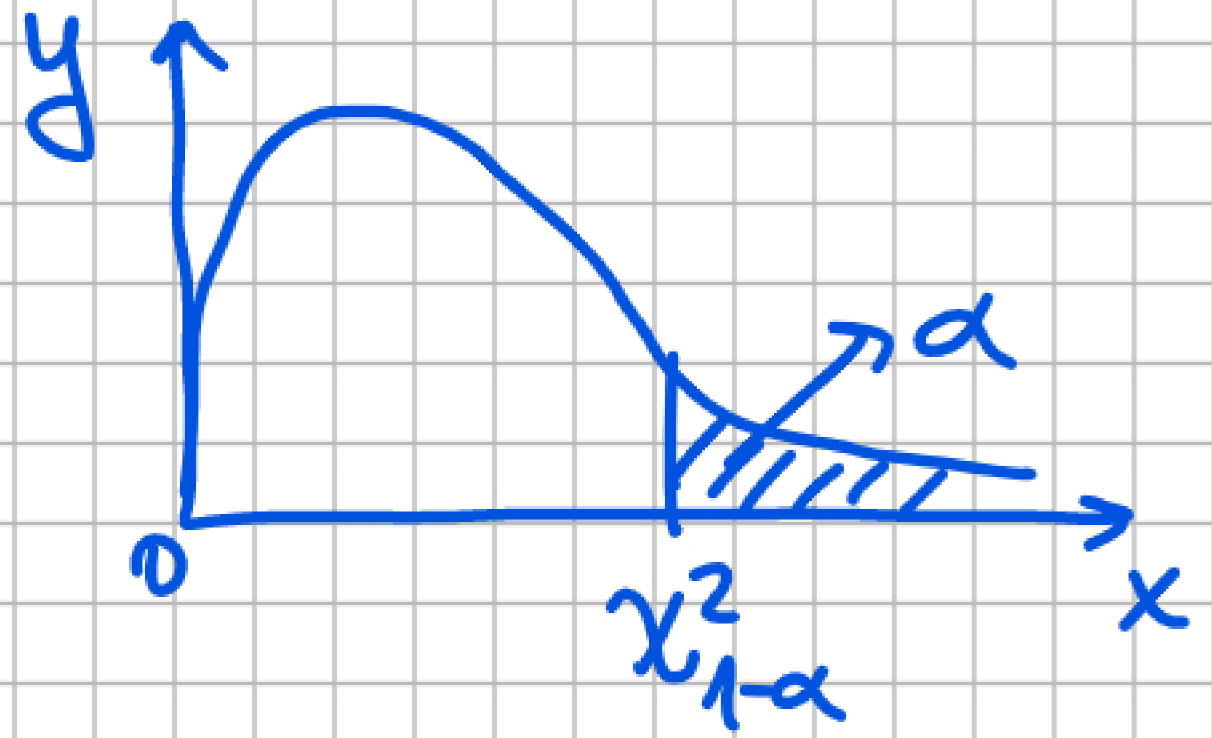
$\chi_i^2 = 5.67 \rightarrow \chi^2 \in R \Rightarrow H_0$ is rejected

At the 5% level of significance, the sample shows signif. evidence to say that a single line causes lower variation among waiting times.

The FCC conducts broadband speed tests to measure how much data per second passes between a consumer's computer and the internet. As of August of 2012, the standard deviation of Internet speeds across Internet Service Providers (ISPs) was 12.2 percent. Suppose a sample of 15 ISPs is taken, and the standard deviation is 13.2. An analyst claims that the standard deviation of speeds is more than what was reported. State the null and alternative hypotheses, compute the degrees of freedom, the test statistic and draw a conclusion. Test at the 1% significance level.

$H_0: \sigma^2 = 12.2^2 \quad (\sigma_0^2 = 12.2^2)$

$H_a: \sigma^2 > 12.2^2 \Rightarrow R = (\chi^2_{1-\alpha}, \infty)$



$\int_{-\infty\,(0)}^{\infty} f(x)\,dx = 1$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{14 \cdot 13.2^2}{12.2^2} = 16.38$

$P(\chi^2 \le \chi^2_{1-\alpha}) = 1-\alpha$

$n = 15, \; s^2 = 13.2^2$

$\chi^2_{1-\alpha} = \text{qchisq}(1-\alpha, n-1) = \text{qchisq}(\alpha, n-1, \text{lower.tail}=F) = 29.14$

$\text{pchisq}(\chi^2_{1-\alpha}) = 1-\alpha$

$\alpha = 1\% = 0.01 \quad \Rightarrow R = (29.14, \infty)$

$\chi^2 \notin R \Rightarrow H_0$ is not rejected

At the 1% level of sig., we cannot support the claim that std. dev. is more than it was reported.

Exercise 107. For a sample of 10 students, the following bivariate data represents the distance and the duration of their travel to school.

x	1	3	5	5	7	7	8	10	10	12
y	5	10	15	20	15	25	20	25	35	35

1. Determine the scatter diagram and the correlation coefficient of the sample.
2. Does this sample show sufficient evidence for the positive linear correlation of the distance and the duration of travel in the case of all students ?
3. Find the equation of the regression line.
4. Does the slope b_1 of the regression line show sufficient evidence to claim that $\beta_1 > 0$ at a significance level $\alpha = 0.05$?

1. scatterplot → plot(x,y)
   x,y → seq()
   cor(x,y)
2. cor.test(x,y)
3. lm(y~x) → model
   $y = \beta_0 + \beta_1 \cdot x$
4. summary(model)
5. predict the duration of travel to school for a distance of 14.
   $x = 14 \Rightarrow y =$