Lecture 7. Linear Regression.

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 and standard deviation σ. Test if the value of the standard deviation is $\sigma = 2.2$ with a level of significance $\alpha = 5\%$.

$H_0: \sigma = 2.2 \quad (\sigma_0 = 2.2)$

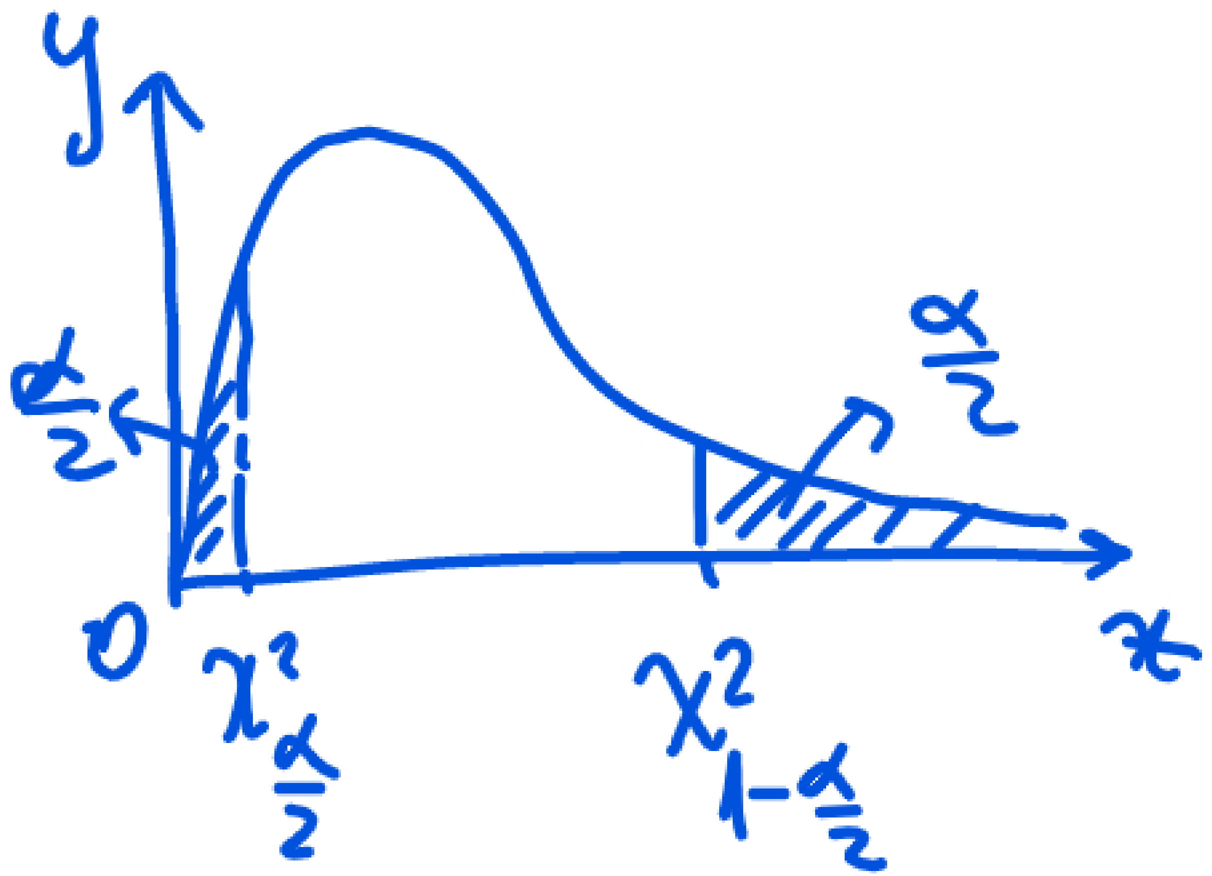
$H_a: \sigma \neq 2.2 \Rightarrow R = (0, \chi^2_{\frac{\alpha}{2}}) \cup (\chi^2_{1-\frac{\alpha}{2}}, \infty)$

$\alpha = 5\% = 0.05$

$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{5 \cdot 6.23}{2.2^2} = 6.44$

$s^2 = 6.23$

$n = 6$



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

χ^2 - has a χ^2 distribution with $n-1$ degrees of freedom

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

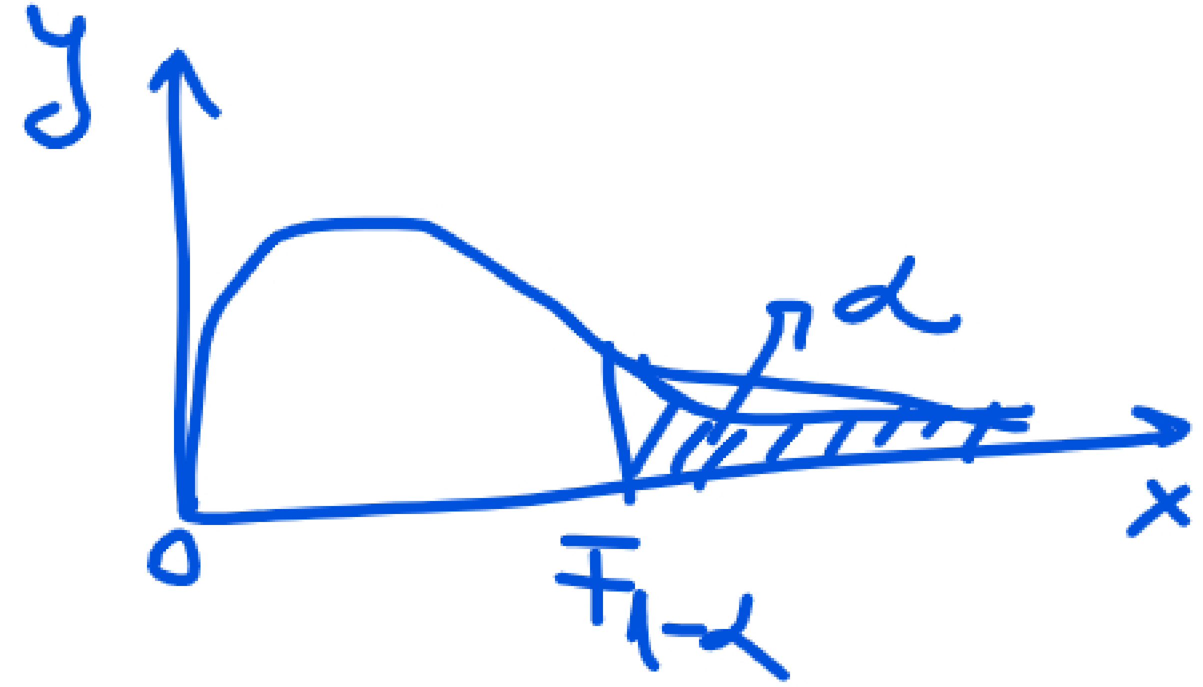
$\Rightarrow R = (0, 0.83) \cup (12.83, \infty)$

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

p-value $= 0.53 > \alpha = 0.05 \Rightarrow H_0$ is rejected

Ex.8 $\rightarrow$ homework

Example 9. A data channel has the average speed of 180 Megabytes per second. A hardware upgrade is supposed to improve stability of the data transfer while maintaining the same average speed. Stable data transfer rate implies low standard deviation. After the upgrade, the instantaneous speed of data transfer, measured at 16 random instants, yields a standard deviation of 14 Mbps. Records show that the standard deviation was 22 Mbps before the upgrade, based on 27 measurements at random times. Test whether the stability wasimproved with a significance level $\alpha = 5\%$.



$H_0: \sigma_1 = \sigma_2$

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

$n = 27, m = 16$

$S_x = 22, S_y = 14$

$\alpha = 5\%$

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

$F = \dfrac{S_x^2}{S_y^2} = \dfrac{22^2}{14^2} = 2.47$

$F_{1-\alpha} = qf(1-\alpha, n-1, m-1) = 2.27$

$R = (2.27, \infty)$

$F \in R \Rightarrow H_0$ is rejected

The stability was improved, at the 5% level of significance.

var.test(x, y, ...)

Example 3. (Efficiency of computer programs). A computer manager needs to know how efficiency of her new computer program depends on the size of incoming data. Efficiency will be measured by the number of processed requests per hour. Applying the program to data sets of different sizes, she gets the following results:

Data size (gigabytes), x	6	7	7	8	10	10	15
Processed requests, y	40	55	50	41	17	26	16

a) Estimate the regression line. b) Compute the ANOVA table and estimate the variance of the model. c) Perform the t test for the slope. d) Perform the F test for the model.

f) summary( )

a) $y = \underbrace{72.29}_{\text{intercept}} - \underbrace{4.14}_{\text{slope}} \cdot x$ lm( )

d) p-value $= 0.03 < 0.05 = \alpha \Rightarrow H_0$ is reject

$H_0: \beta_1 = 0$

$H_a: \beta_1 \neq 0$