Test 2

1. Sales of icecream in 20 consecutive days at an icecream shop are: 62 62 70 47 57 65 47 55 44 65 57 53 68 48 41 68 59 55 40 42. Compute:
- the five-point summary
- the variance and standard deviation
- IQR; are there any outliers?

Plot the boxplot using the five-point summary.
Find the 95% CI for the mean number of icecreams sold this summer (population mean) and the 99% CI for the standard deviation of the sales.

2. T-test for the population mean.

3. Chi-square test for the variance/standard deviation.

4. Linear regression model.

1. Five point summary: min, Q_1, Me, Q_3, max

40, 41, 42, 44, $\underbrace{47, 47}$, 48, 53, 55, $\underbrace{55, | 57}$, 57, 59, 62, $\underbrace{62, 65}$, 65, 68, 68, 70

min = 40

max = 70

$Me = \frac{55+57}{2} = \frac{112}{2} = 56$

$Q_1 = \frac{47+47}{2} = 47$

$Q_3 = \frac{62+65}{2} = \frac{127}{2} = 63.5$

Variance: $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$n = 20$

$\bar{x}$ – sample mean

$s^2 = 93.46$

Standard deviation: $s = \sqrt{s^2} = \sqrt{93.46} = 9.67$

$IQR = Q_3 - Q_1 = 63.5 - 47 = 16.5$

Outliers: $< Q_1 - 1.5\,IQR$ or $> Q_3 + 1.5\,IQR$

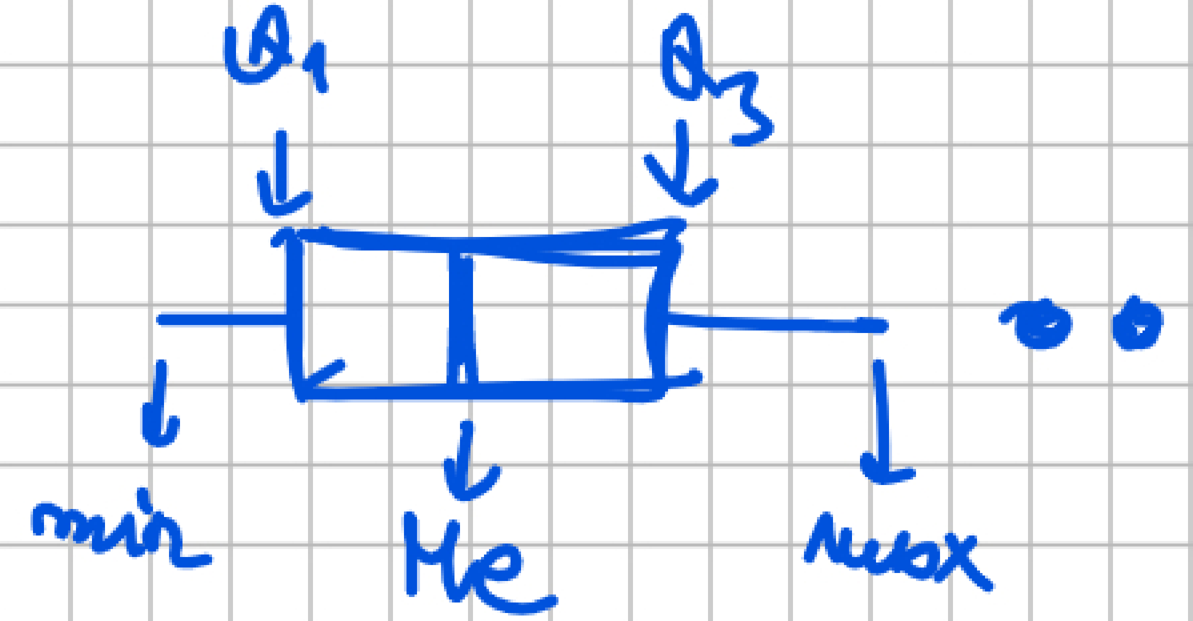
$Q_1 - 1.5\,IQR = 47 - 1.5 \cdot 16.5 = 22.25$

$Q_3 + 1.5\,IQR = 63.5 + 1.5 \cdot 16.5 = 88.25$

$\Rightarrow$ no outliers

Boxplot

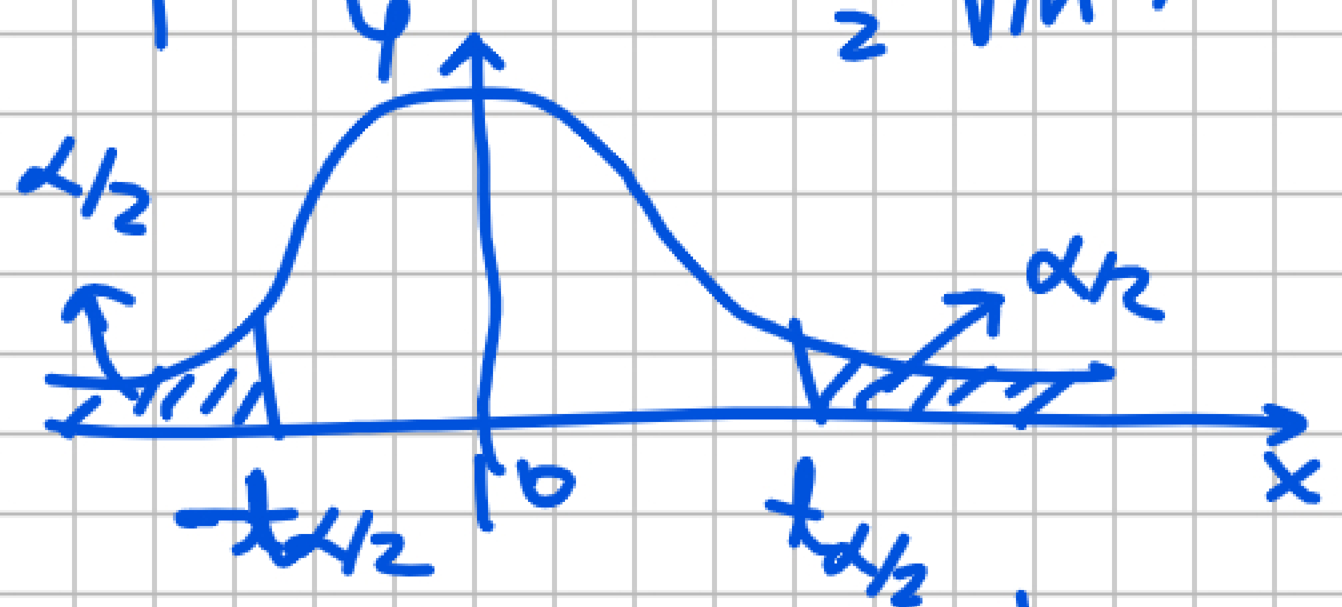
5-point summary: 40, 47, 56, 63.5, 70



95% CI for the population mean (μ): $\left[\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}\right] =$

s - sample std. deviation

n - sample size



$= [50.73; 59.77]$

$-t_{\frac{\alpha}{2}} = qt(\alpha/2, n-1) = -2.09 \Rightarrow t_{\alpha/2} = 2.09$, $\bar{x} = 55.25$ | $55.25 \pm 2.09 \cdot \frac{9.67}{\sqrt{20}}$

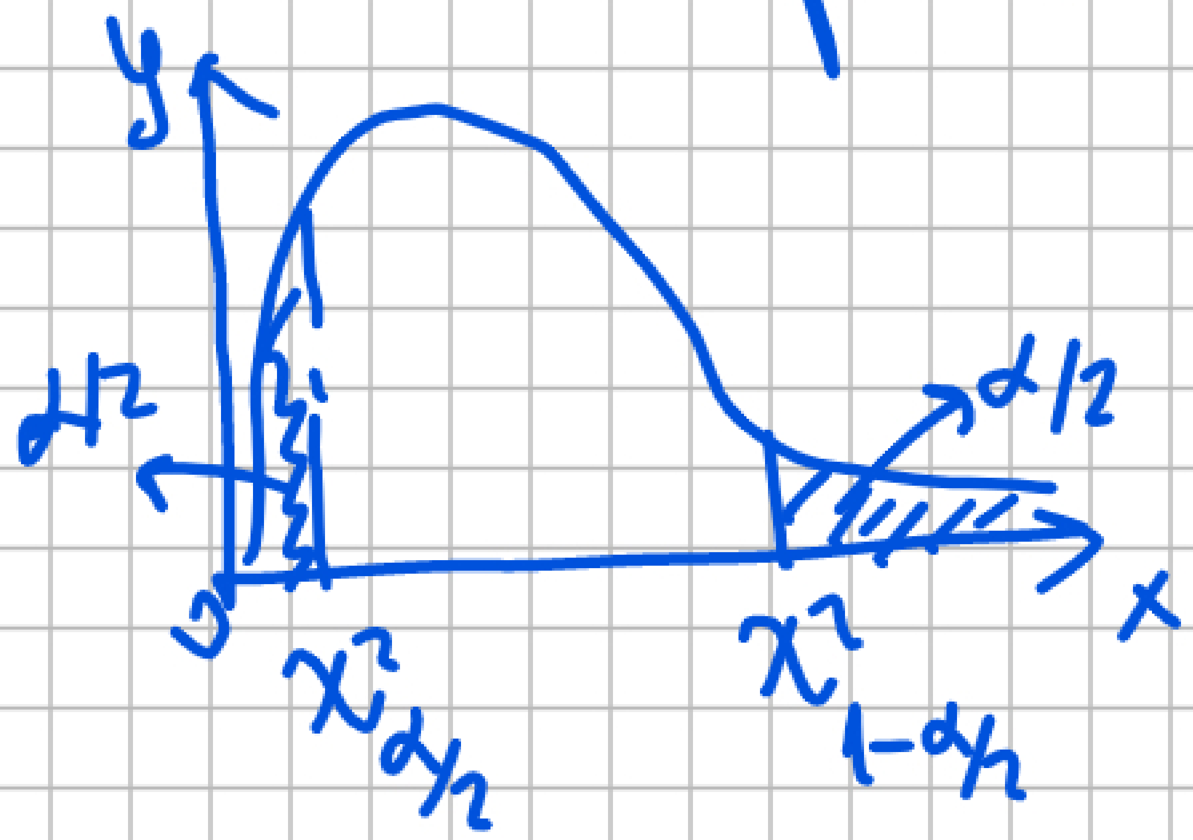
$\alpha = 0.05$, $n = 20$

99% ci for the population std. deviation (σ): $\left[\sqrt{\dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}, \sqrt{\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}}}\right] =$

$= \left[\sqrt{46.03}, \sqrt{259.61}\right] =$

$= \underline{[6.78, 16.11]}$

s^2 - sample variance



$\chi^2_{\alpha/2} = \text{qchisq}(\alpha/2, n-1) = 6.84$

$\chi^2_{1-\alpha/2} = \text{qchisq}(1-\alpha/2, n-1) = 38.58$

$\dfrac{19 \cdot 93.46}{38.58} = 46.03$

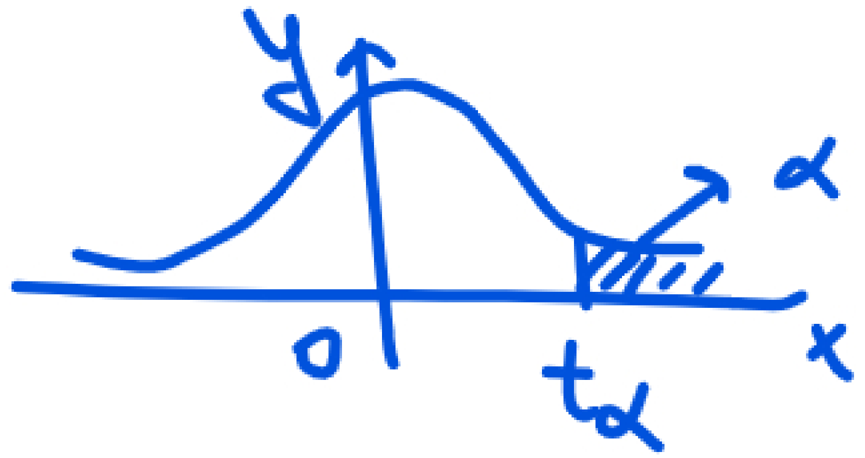
$\dfrac{19 \cdot 93.46}{6.84} = 259.61$

$\alpha = 0.01$

$s^2 = 93.46$

$n = 20$

2. The shop owner wonders if the mean icecream sales is 50 or larger. Does the sample show significant evidence to support this claim at the 5% level of significance?



$H_0: \mu = 50 \quad (\mu_0 = 50)$

$H_a: \mu > 50 \Rightarrow R = (t_\alpha, \infty)$

one sample t-test for the population mean μ · t.test( )

$T = \dfrac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} \sim$ Student t distribution with n-1 df

$T = \dfrac{55.25 - 50}{\frac{9.67}{\sqrt{80}}} = 2.46$

$t_\alpha = qt(1-\alpha, n-1) = 1.73 \Rightarrow R = (1.73, \infty)$

$\alpha = 0.05$

$T \in R \Rightarrow H_0$ is rejected

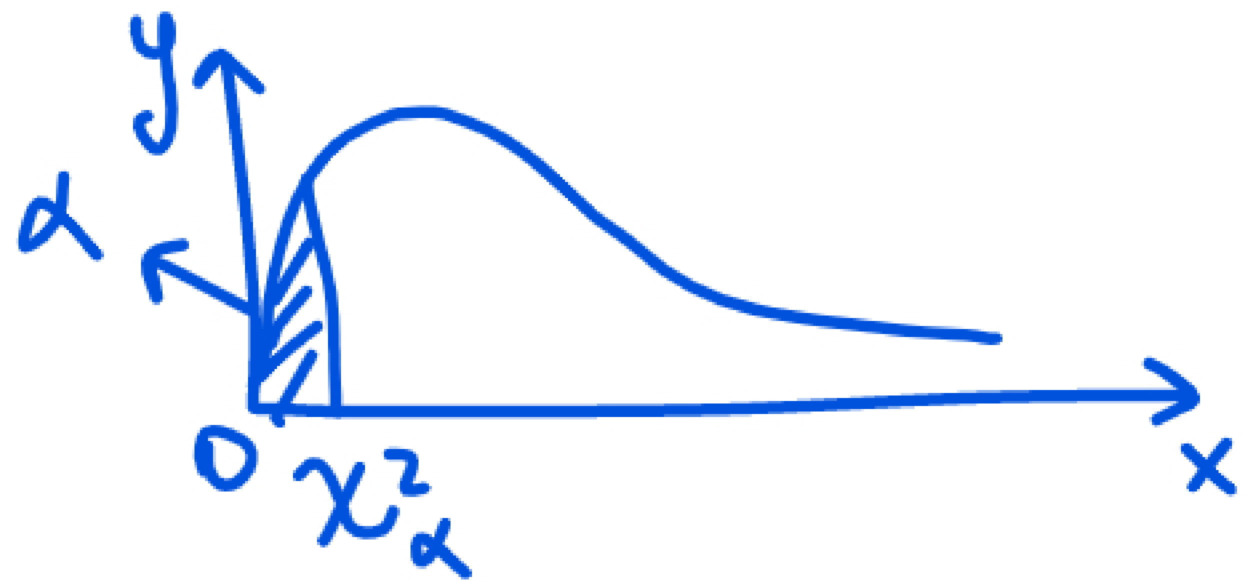
The sample shows significant evidence to support the claim that the mean sales is larger than 50 at 5% level of signif.

3. Test the hypothesis that the standard deviation of the icecream sales is 11 or smaller, at the 1% level of significance.

$H_0: \sigma = 11 \quad (\sigma_0 = 11)$

$H_a: \sigma < 11 \Rightarrow R = (0, \chi^2_\alpha)$

$\alpha = 0.01$



varTest() → "EnvStats" package

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

s^2 - sample variance

n - sample size

$\chi^2 = \frac{19 \cdot 93.46}{11^2} = 14.68$

$\chi^2_\alpha = qchisq(\alpha, n-1) = 7.63 \Rightarrow R = (0, 7.63)$

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

The sample doesn't show signif. evidence to support the claim that σ is < 11 at 1% level of signif.

- Z-test
  - pop. mean (μ) $\longrightarrow$ σ is known
  - pop. proportion (p)
  - comparison of 2 pop. proportions
- t-test
  - pop. mean (μ) $\longrightarrow$ σ is unknow
  - comparison of 2 pop. means
    - $\sigma_1 = \sigma_2$
    - $\sigma_1 \neq \sigma_2$
- χ^2 test of variance / std. deviation
- F test for comparison of 2 variances / std. dev.

4. At a local food stand icecream and soda sales are measured in 10 consecutive days in July:

icecream sales (x): 18.4 15.8 11.8 13.9 12.7 11.8 13.0 12.4 10.6 13.2

soda sales (y): 11.6 10.4 8.5 9.6 8.8 8.5 9.0 8.7 7.7 9.0

Construct a scatterplot of the data and say what trend can you see.
Fit a linear regression model for the dataset. What is the slope and intercept? Perform the t-test for them.
Compute the correlation coefficient for the icecream and soda sales.
Predict soda sales for icecream sales of 15 units.

icecream = c(...)
soda = c(...)

lm(soda ~ icecream)

intercept = 2.620, slope = 0.491

$y = 2.620 + 0.491 \cdot x$

p-value for intercept $= 7.11 \cdot 10^{-7} < \alpha = 0.05$
p-value for slope $= 4.71 \cdot 10^{-10} < \alpha = 0.05$

$\Rightarrow$ they pass the t-test (we reject $H_0: \beta_0 = 0$, $H_0: \beta_1 = 0$)

$y = \beta_0 + \beta_1 \cdot x$

cor(icecream, soda) = 0.99

$x = 15 \Rightarrow y = 2.620 + 0.491 = 9.985$