Problem 20.

The daily milk production is described by the variable X of normal distribution, which has expected value 22,1 and variance $1{,}5^2$.

Then $\frac{X-22{,}1}{1{,}5}$ follows the standard normal distribution (here we „standardized" the variable X)

We have to determine P(23 < X < 25):

P(23 < X < 25)=P(0,6 < $\frac{X-22{,}1}{1{,}5} < \frac{5{,}8}{3}$) = Φ($\frac{5{,}8}{3}$) - Φ(0,6).

Problem 21.

X is a variable of normal distribution with expected value 110 and variance 100. By standardizing it we get that $\frac{X-110}{10}$ followsdistandard normal distribution.

We have to determine P(X>120).

P(X>120)=1- P(X<120) = 1-P($\frac{X-110}{10}$ < 1) = 1 – Φ(1).

Problem 27.

P(X=x_i) = $\binom{m}{x_i} \cdot p^{x_i} \cdot (1-p)^{x_i}$.

The likelihood function:

$\prod_{i=1}^{n} \binom{m}{x_i} \cdot p^{x_i} \cdot (1-p)^{x_i}$, where p lies in [0,1].

The log-likelihood-function:

$\ln\left[\prod_{i=1}^{n} \binom{m}{x_i} \cdot p^{x_i} \cdot (1-p)^{m-x_i}\right] =$

$= \sum_{i=1}^{n} ln\left(\binom{m}{x_i}\right) + \sum_{i=1}^{n} ln\,(p^{x_i}) + \sum_{i=1}^{n} ln\,(1-p)^{m-x_i} =$

$= \sum_{i=1}^{n} ln\left(\binom{m}{x_i}\right) + \sum_{i=1}^{n} x_i \cdot ln\,p + \sum_{i=1}^{n} (m-x_i) \cdot ln\,(1-p) =$

The derivative of that must be zero:

$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(m-x_i)}{1-p} = 0.$

Thus $\frac{\sum_{i=1}^{n} x_i}{p} = \frac{\sum_{i=1}^{n}(m-x_i)}{1-p}$, so

$(1-p)\cdot\sum_{i=1}^{n} x_i = p\cdot\sum_{i=1}^{n}(m-x_i).$

We get that $\sum_{i=1}^{n} x_i$=pmn, thus p=$\frac{\sum_{i=1}^{n} x_i}{mn}$ is a potential extremum point (using the theorem of Weierstrass it is easy to see that it must be the maximum point).

Problem 28.

Let $x_1, x_2, ..., x_n$ be a sample set with independent elements from an exponentila distribution having parameter λ.

The log-likelihood-function is

$\ln\prod_{i=1}^{n}\left(\lambda e^{-\lambda x_i}\right) = \ln(\lambda^n) + \sum_{i=1}^{n} ln\left(e^{-\lambda x_i}\right)=$

$= \ln(\lambda^n) - \sum_{i=1}^{n} \lambda x_i = n\cdot\ln\lambda - \lambda\cdot\sum_{i=1}^{n} x_i$

The derivative (by λ) must be zero:

$0 = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$, thus $\frac{n}{\lambda} = \sum_{i=1}^{n} x_i$.

By arranging that we get that

$\lambda=\frac{n}{\sum_{i=1}^{n} x_i}.$

It is easy to check that this is a maximum point.

Problem 29.

H_0: m ≤ 15, H_1: m > 15

We will use the t-test.

The sample mean is 13,475; and we will estimate the standard deviation by

$$\sqrt{\frac{(14{,}8-13{,}475)^2+(12{,}2-13{,}475)^2+(16{,}8-13{,}475)^2+(11{,}1-13{,}475)^2}{3}}$$

(this is an unbiased estimation).

The critical range consists of the numbers greater than

$t_{3,95/100} = t_{3,5/100}$.

$$t=\frac{\frac{13{,}475-15}{2}}{\sqrt{\frac{(14{,}8-13{,}475)^2+(12{,}2-13{,}475)^2+(16{,}8-13{,}475)^2+(11{,}1-13{,}475)^2}{3}}}=$$

$$=\frac{(13{,}475-15)\cdot\sqrt{\frac{(14{,}8-13{,}475)^2+(12{,}2-13{,}475)^2+(16{,}8-13{,}475)^2+(11{,}1-13{,}475)^2}{3}}}{2}$$

= -1,97257733134...

This is greater than $t_{3,5/100}$-nél, thus we reject the zero hypothesis.

Problem 30.

The standard deviations are unknown, our samples are from two independent normal distributions. This means that we must first determine by F-test whether we can assume that the standard deviations of the two normal distributions are the same.

If this is assumed, a two-sample t-test is used, if discarded, a Welch test is used. The confidence level is 1-α=0,95.

The sample mean in factory A:

$$\frac{11{,}9+12{,}1+12{,}8+12{,}2+12{,}5+11{,}9+12{,}5+11{,}8+12{,}4+12{,}9}{10} = 12{,}3$$

The sample mean in factory B:

$$\frac{12{,}1+12{,}0+12{,}9+12{,}2+12{,}7+12{,}6+12{,}6+12{,}8+12{,}0+13{,}1}{10} = 12{,}5.$$

Unbiased estimation for the variance in Factory A:

$$\frac{(0{,}4)^2+(0{,}2)^2+(0{,}5)^2+(0{,}1)^2+(0{,}2)^2+(0{,}4)^2+(0{,}2)^2+(0{,}5)^2+(0{,}1)^2+(0{,}6)^2}{9} = \frac{132}{900}$$

Unbiased estimation for the variance in Factory B:

$$\frac{(0{,}4)^2+(0{,}5)^2+(0{,}4)^2+(0{,}3)^2+(0{,}2)^2+(0{,}1)^2+(0{,}1)^2+(0{,}3)^2+(0{,}5)^2+(0{,}6)^2}{9} = \frac{142}{900}.$$

If the standard deviations of the two normal distributions can be assumed to be the same, then the confidence interval for the quotient of the larger and smaller empirical standard deviations is (-$F_{10-1,10-1;\,1-\frac{\alpha}{2}}$, $F_{10-1,10-1;\,1-\frac{\alpha}{2}}$).

The value $\frac{\frac{142}{900}}{\frac{132}{900}} = \frac{142}{132}$ lies here thus we can assume that the variances are the same.

So we can use the two sample t-test.

Let m_1 be the expected value in factory A, and let m_2 be the expected value in factory B.

Zero hypothesis: $m_1 \geq m_2$.

The value of the test statistics:

$$t = \sqrt{\frac{10\cdot 10}{10+10}} \cdot \frac{12{,}3-12{,}5}{\sqrt{\frac{\frac{132}{100}+\frac{142}{100}}{10+10-2}}} = \sqrt{5}\cdot(-0{,}2)\cdot\sqrt{18}\cdot\frac{10}{\sqrt{274}} = \sqrt{\frac{90}{274}}\cdot(-2) = -\sqrt{\frac{360}{274}} = -1{,}146...$$

This is greater than $t_{18,\alpha}$. So we can accept our zero hypothesis.

Problem 31.

The servers are independent, and the standard deviations are equal. We assume that our samples are taken from normal distributions.

We will use the two sample Z-test.

We have to decide whether we can assume that the expected values of the running times are equal.

Using our standard notations we get

$$u = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{\sigma_1{}^2}{n}+\frac{\sigma_2{}^2}{m}}} = \frac{6{,}7-7{,}2}{\sqrt{\frac{0{,}25}{30}+\frac{0{,}25}{20}}} = \frac{-0{,}5}{0{,}5\cdot\sqrt{\frac{1}{30}+\frac{1}{20}}} = -\frac{1}{\sqrt{\frac{1}{12}}} = -\sqrt{12}.$$

The confidence level is 1-α=0,95.

$-\sqrt{12}$ does not lie in the interval $\left(-u_{0,975}, u_{0,975}\right)$ thus we reject our zero hypothesis.

Problem 32.

The two sample sets are not independent, we will apply a t-test for the difference set, which is the following

0,5; -0,4; 0,6; 0,4; -0,2; 1,0; 0,2; -0,1; 0,3; 0,7.

Our zero hypothesis will be that the toxic gas emission remained unchanged.

We have to check whether the value of the test statistics lies on the interval $(-t_{9,1-\frac{\alpha}{2}}, t_{9,1-\frac{\alpha}{2}})$.

Using the formula (where according to our zero hypothesis $m_0 = 0$):

$$t = \frac{\bar{X}-m_0}{S_n*}\cdot\sqrt{n} = \frac{\frac{0{,}5-0{,}4+0{,}6+0{,}4-0{,}2+1{,}0+0{,}2-0{,}1+0{,}3+0{,}7}{10}}{\sqrt{\frac{(0{,}2)^2+(0{,}7)^2+(0{,}1)^2+(0{,}1)^2+(0{,}5)^2+(0{,}7)^2+(0{,}1)^2+(0{,}4)^2+(0)^2+(0{,}4)^2}{9}}}\cdot\sqrt{10} =$$

$$= \frac{0{,}3}{\sqrt{\frac{162}{900}}}\cdot\sqrt{10} = \frac{0{,}3}{\sqrt{162}}\cdot 30\cdot\sqrt{10} = 2{,}236...$$

This value lies on the interval (-2,262; 2,262) thus we accept the zero hypothesis.