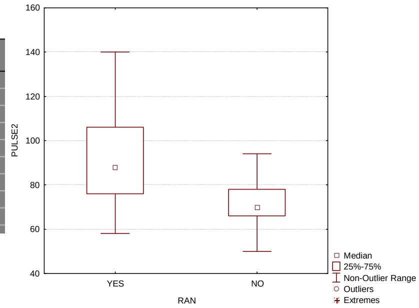
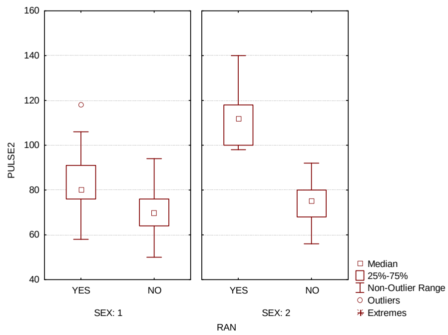
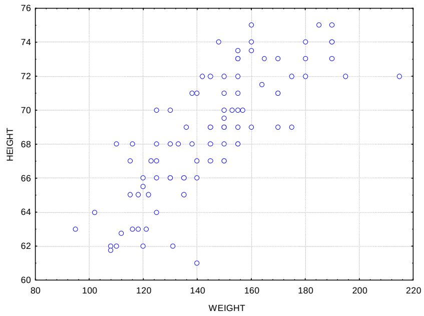
	1 PULSE1	2 PULSE2	3 RAN	4 SMOKES	5 SEX	6 HEIGHT	7 WEIGHT	8 ACTIVITY
1	64	88	YES	NO	1	66.0	140	2
2	58	70	YES	NO	1	72.0	145	2
3	62	76	YES	YES	1	73.5	160	3
4	66	78	YES	YES	1	73.0	190	1
5	64	80	YES	NO	1	69.0	155	2
6	74	84	YES	NO	1	73.0	165	1
7	84	84	YES	NO	1	72.0	150	3
8	68	72	YES	NO	1	74.0	190	2
9	62	75	YES	NO	1	72.0	195	2
10	76	118	YES	NO	1	71.0	[illegible]	[illegible]







1

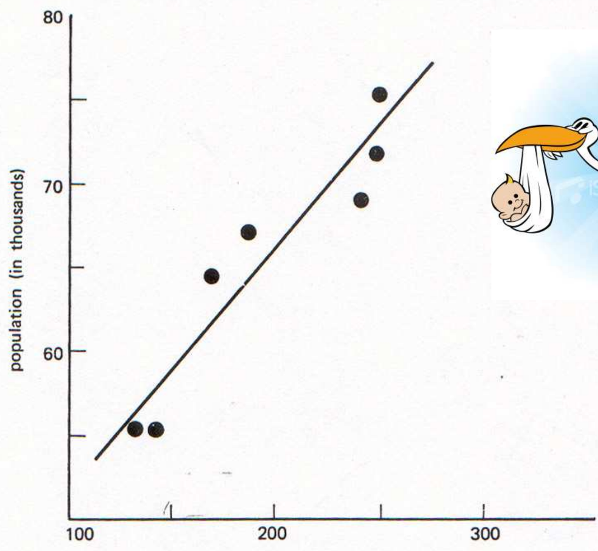
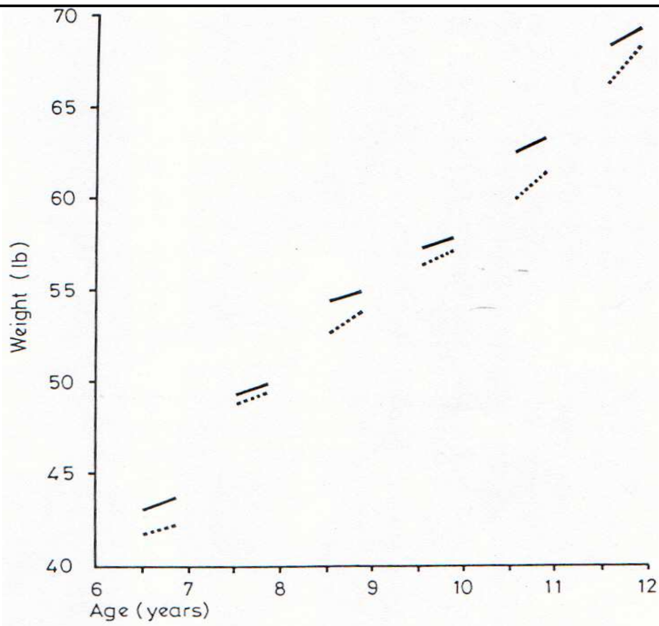


FIGURE 1.4. A plot of the population of Oldenburg at the end of each year against the number of storks observed in that year, 1930–1936.



Lanarkshire milk experiment[1]: comparison of control group (————) and milk feeding group (- - - - -) showing mean weight at beginning and end of study for each yearly age group.

3

Gas bottle problem

There are 2.5 million gas bottles in Hungary. They want to fill more gas in them. The volume of the bottle limits the maximum ammount of gas. (i.e. the smaller the bottle the higher the pressure in it – for a given ammount of gas)

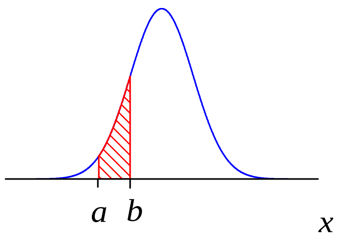
What is the lot and what is the sample?

Sample	conclusion	Lot
Known	⟶	Unknown
Histogram		Density function

Random variable

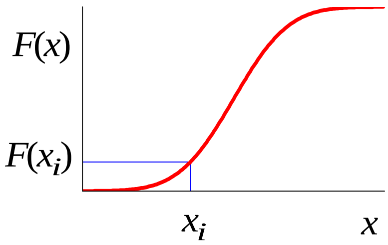
4

Continuous random variable



$$P(a < x \le b) = \int_a^b f(x)dx$$

density function



$$F(x_i) = P(x \le x_i) = \int_{-\infty}^{x_i} f(x)dx$$

distribution function

Parameter and statistic

- expected value:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

- variance

$$Var(x) = \int_{-\infty}^{\infty} [x - E(x)]^2 f(x)dx$$

- sample mean:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- sample variance

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

The expected value and variance of the function of random variables

$$E(cx)=cE(x)$$

The expected value of the volume of gas bottles is 25.8 dm^3. What is the expected value in cm^3?

$$Var(cx)=c^2Var(x)$$

The variance of the volume of gas bottles is 0.25 $(dm^3)^2$. What is the variance in $(cm^3)^2$?

7

Independent measurements

The data in the samples are the results of measurements. These are subject to error.

Type of measurement errors:
- systematic
- random

Why is it important to do independent repetitions?

Two measurements are independent if their errors are independent.

8

The expected value and variance of the function of random variables

If $x_1, x_2, \ldots, x_n$ are independent:

$$E(x_1 + x_2 + \ldots + x_n) = E(x_1) + E(x_2) + \ldots + E(x_n)$$

$$Var(x_1 + x_2 + \ldots + x_n) = Var(x_1) + Var(x_2) + \ldots + Var(x_n)$$

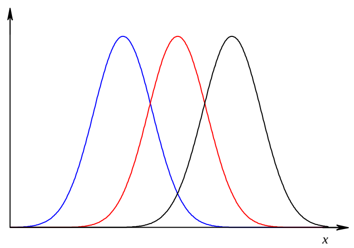
If $x_1, x_2, \ldots, x_n$ are independent and has the same distribution, with E(x) expected value and Var(x) variance.

$$E(x_1 + x_2 + \ldots + x_n) = nE(x) \qquad Var(x_1 + x_2 + \ldots + x_n) = nVar(x)$$

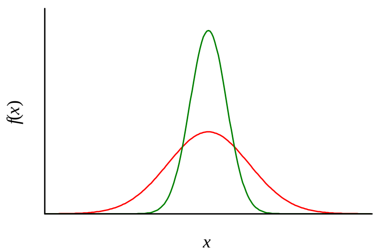
The most important continuous distribution: Gauss (normal) distribution

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Two parameters: μ and σ^2



μ is different



σ is different

Expected value and variance:

$$E(x) = \mu \qquad Var(x) = \sigma^2$$

Short notation:

$$N(\mu, \sigma^2) \quad \text{e.g.} \quad N(0,1)$$

Standardisation:

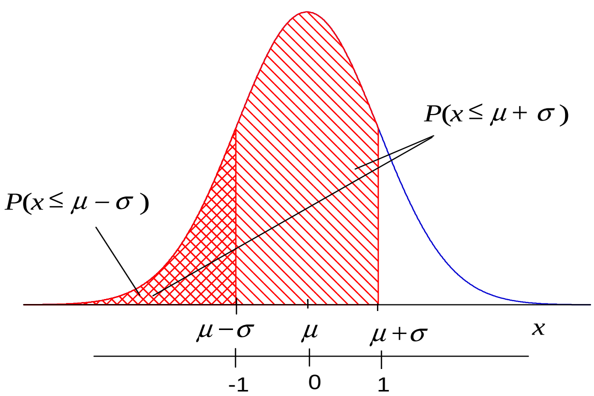
$$z = \frac{x - \mu}{\sigma} \qquad f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

$$\mu = E(z) = 0 \qquad \sigma^2 = Var(z) = 1$$

11

What is the probability of finding the x Gauss d. random variable in the $(\mu-\sigma, \mu+\sigma)$ range?

$$P(\mu - \sigma < x \le \mu + \sigma) = F(\mu + \sigma) - F(\mu - \sigma)$$



12

$$z_{lower} = \frac{\mu - \sigma - \mu}{\sigma} = -1 \qquad z_{upper} = \frac{\mu + \sigma - \mu}{\sigma} = 1$$

Width of the interval	$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
P	0.68268	0.9545	0.9973

The variance of a measurement is σ^2 =0.25g^2. The measurement is unbiased. We measure a 10 g etalon weight. In which range will be the measured weight with 95% probability?

This is the question:

$$P\left(x_{lower} < x \leq x_{upper}\right) = 0.95$$

This is what we know from the distribution function:

$$P\left(-z_{\alpha/2} < z \leq z_{\alpha/2}\right) = 0.95$$

Connection: $z = \frac{x - \mu}{\sigma}$

$$P\left(-z_{\alpha/2} < \frac{x-\mu}{\sigma} \leq z_{\alpha/2}\right) = 0.95$$

$$P\left(\underbrace{\mu - z_{\alpha/2}\sigma}_{x_{lower}} < x \leq \underbrace{\mu + z_{\alpha/2}\sigma}_{x_{upper}}\right) = 1 - \alpha$$

$$P\left(10 - 1.96 \cdot 0.5 < x \leq 10 + 1.96 \cdot 0.5\right) = 0.95$$

The volume of gas bottles are normally distributed with 25.8 dm^3 expected value and 0.0625 $(dm^3)^2$ variance. What is that minimum volume, that 99.5% of the bottles exceed?

How many percent of the bottles will be in the 25.8±0.3 dm^3 interval?

In what interval will be 99% of the bottles?

15

Homework

The variance of a measurement is $\sigma^2 = 4g^2$. The bias of the measurement is 2 g. We measure an object, its weight is 200 g.

1. In which range will be the outcome of the measurement with 99% probability?
2. What is the probability that the measured weight is above 205g?
3. What is the probability that the measured weight is below 200g?
4. What is that maximum value, that the measured weight will not achieve with 90% probability?

16

The sample mean

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum x_i$$

$$E(\bar{x}) = \frac{1}{n}[nE(x)] = E(x) = \mu$$

$$\sigma_{\bar{x}}^2 = Var(\bar{x}) = \frac{Var(x)}{n} = \frac{\sigma_x^2}{n}$$

17

Central limit theorem

The mean of sample elements taken from any distribution approximately follows Gauss d. around the expected value of the original d. with variance σ^2/n.

$$\bar{x} \sim N(\mu, \sigma^2/N)$$

Sum as well $\sum_{i=1}^{N} x_i \sim N(N\mu, N\sigma^2)$

Based on the Central Limit Theorem: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

18

Calculate the 95% probability interval for the mean of a n = 5 sample taken from a population of μ=10 and σ^2=0.25 !

This is the question:

$$P\left(\overline{x}_{lower} < \overline{x} \le \overline{x}_{upper}\right) = 0.95$$

This is what we know from the distribution function:

$$P\left(-z_{\alpha/2} < z \le z_{\alpha/2}\right) = 0.95$$

Connection: $z = \dfrac{\overline{x} - \mu}{\sigma/\sqrt{n}}$

$$P\left(-z_{\alpha/2} < \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 0.95$$

$$P\Big(\underbrace{\mu - z_{\alpha/2}\,\sigma/\sqrt{n}}_{\overline{x}_{lower}} < \overline{x} \le \underbrace{\mu + z_{\alpha/2}\,\sigma/\sqrt{n}}_{\overline{x}_{upper}}\Big) = 1 - \alpha$$

$$P\left(10 - 1.96 \cdot 0.5/\sqrt{5} < \overline{x} \le 10 + 1.96 \cdot 0.5/\sqrt{5}\right) = 0.95$$

The mean of five measurements is 10, The variance of the measurements is known form previous data: σ^2 =0.25. In what range can be the true expected value of the measurements with 95% probability?

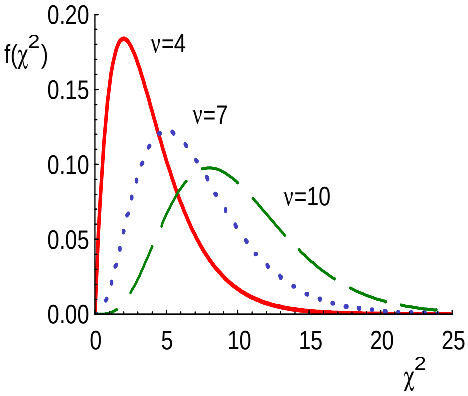
(Give a 95% confidence interval for the expected value !)

The way of thinking is the same as it was on the previous slide, but the inequality is rearranged so, that the expected value (mu) remains in the middle:

$$P\left(\overline{x} - z_{\alpha/2}\,\sigma/\sqrt{n} < \mu \le \overline{x} + z_{\alpha/2}\,\sigma/\sqrt{n}\right) = 1 - \alpha$$

$$P\left(10 - 1.96 \cdot 0.5/\sqrt{5} < \mu \le 10 + 1.96 \cdot 0.5/\sqrt{5}\right) = 0.95$$

χ^2- (chi-square) distribution

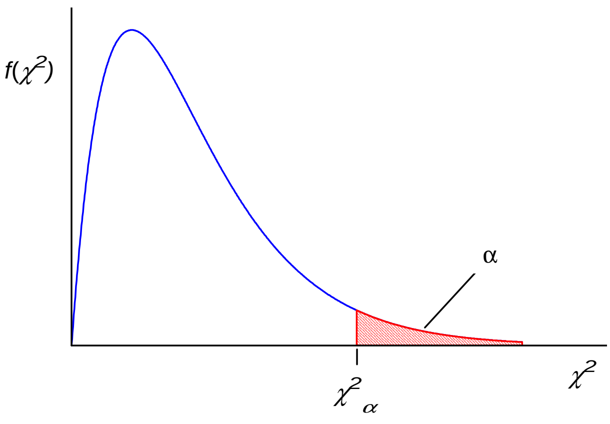


$$\chi^2 = \sum_{i=1}^{n} z_i^2$$

$$E\left(\chi^2\right) = \nu = n-1$$

$$Var\left(\chi^2\right) = 2\nu = 2n-2$$

ν is the degrees of freedom



Distribution of the variance (squared standard deviation) of a sample taken from a normally distributed population

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

$$\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \chi^2\sigma^2, \qquad \nu = n-1$$

$$s^2 = \frac{\chi^2\sigma^2}{\nu}$$

A sample of 8 elements are taken from a normal distribution having σ^2 = 0.08. Calculate the interval in which the s^2 is found at 95% probability.

This is the question:

$$P\left(s^2_{\text{lower}} < s^2 \le s^2_{\text{upper}}\right) = 0.95$$

This is what we know from the distribution function:

$$P\left(\chi^2_{\text{lower}} < \chi^2 \le \chi^2_{\text{upper}}\right) = 0.95$$

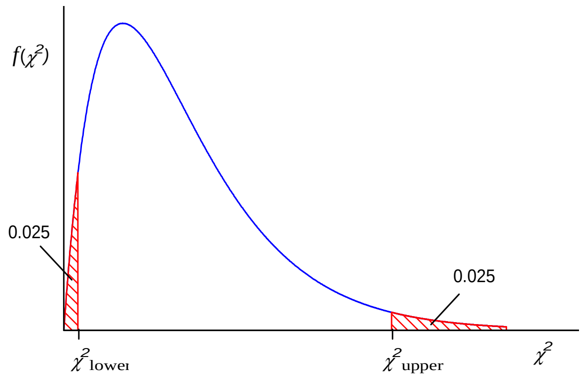
Connection: $s^2 = \frac{\chi^2\sigma^2}{\nu} \rightarrow \frac{s^2\nu}{\sigma^2} = \chi^2$

$$P\left(\chi^2_{\text{lower}} < \frac{s^2\nu}{\sigma^2} \le \chi^2_{\text{upper}}\right) = 0.95$$

s^2_{lower} $\quad$ s^2_{upper}

$$P\left(\frac{\chi^2_{\text{lower}}\sigma^2}{\nu} < s^2 \le \frac{\chi^2_{\text{upper}}\sigma^2}{\nu}\right) = 0.95$$

Critical χ^2 values

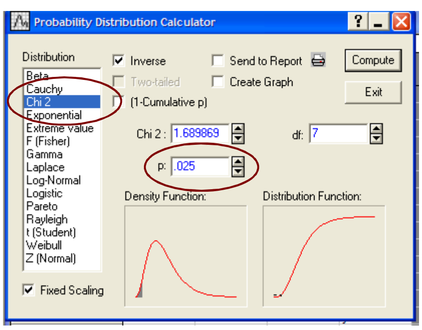


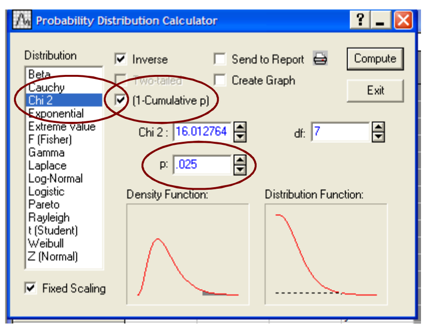
$$\chi^2_{\text{lower}} = 1.69$$

$$\chi^2_{\text{upper}} = 16.0$$

$$\nu = 7$$

Statistics > Probability Calculator > Distributions...





$$P\left(\frac{\chi^2_{\text{lower}}\sigma^2}{\nu} < s^2 \le \frac{\chi^2_{\text{upper}}\sigma^2}{\nu}\right) = 0.95$$

$$= P\left(\frac{1.69 \cdot 0.08}{7} < s^2 \le \frac{16.0 \cdot 0.08}{7}\right) = P(0.0193 < s^2 \le 0.183) = 0.95$$

A sample of 8 elements are taken from a normal distribution. The sample variance is $s^2 = 0.08$. Calculate the interval in which the the variance (σ^2) is found at 95% probability!

(Give a 95% confidence interval for σ^2!)

$$P\left(\frac{s^2\nu}{\chi^2_{\text{lower}}} > \sigma^2 \ge \frac{s^2\nu}{\chi^2_{\text{upper}}}\right) = 0.95$$

! Inequality sign is reversed !

27

Confidence interval

This interval contains the parameter (e.g. μ, σ…) with 1-α probability.

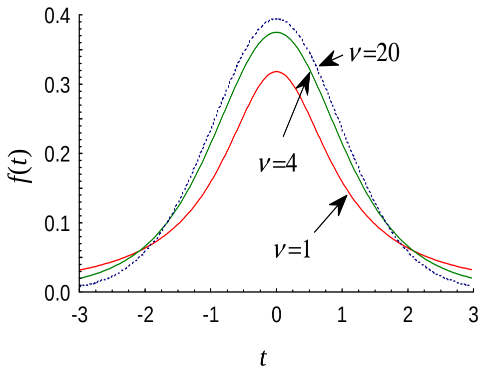
Interpreting

If the 95% confidence interval is calculated for the expected value from 100 different sample, than approximately 95 interval contains the true expected value out of the 100.

The confidence interval refers to the PARAMETER.
(Not to x, or s…)

28

Student's *t* distribution

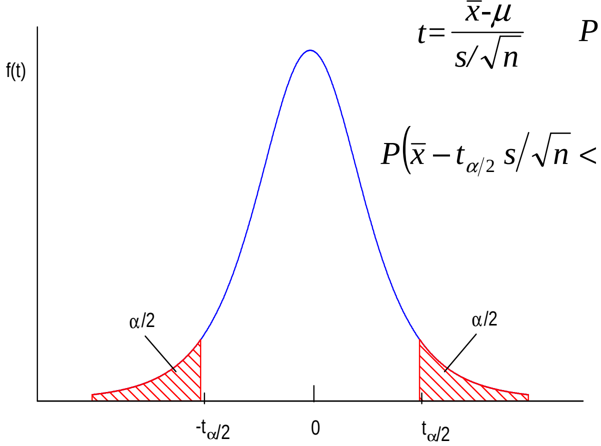


$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

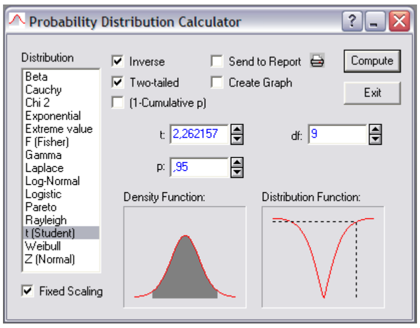
$$E(t) = 0$$

Results of 10 measurements:
24.46; 23.93; 25.79; 25.17; 23.82; 25.39; 26.54; 23.85; 24.19; 25.50.
Give the 95% confidence interval for the true (expected) value!



$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \qquad P\left(-t_{\alpha/2} < t \le t_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(\bar{x} - t_{\alpha/2}\, s/\sqrt{n} < \mu \le \bar{x} + t_{\alpha/2}\, s/\sqrt{n}\right) = 1 - \alpha$$



F distribution

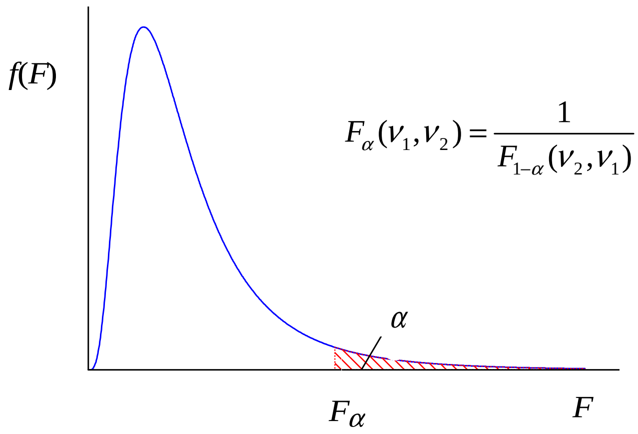
two parameters:
ν_1 is the degrees of freedom for the numerator,
ν_2 is for the denominator

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

if $\sigma_1^2 = \sigma_2^2$

$$F = s_1^2 / s_2^2$$

Critical values for the *F* distribution



$$F_{\alpha}(\nu_1,\nu_2)=\frac{1}{F_{1-\alpha}(\nu_2,\nu_1)}$$

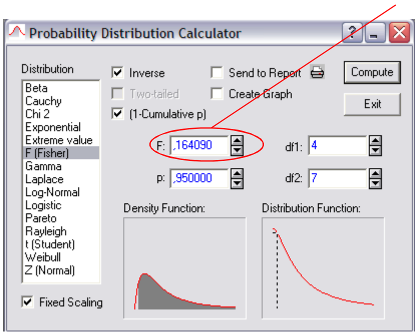
33

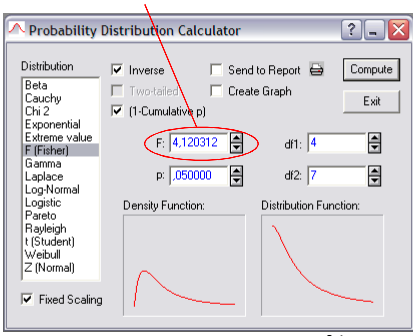
Two sets of measurements (4 and 7 repetitions) were performed using the same method and device.

Give the 90% probability interval for the ratio of sample variances (squared standard deviations).

The population variances are equal: $\sigma_1^2=\sigma_2^2$

$$P\left(F_{\text{lower}}<s_1^2/s_2^2\le F_{\text{upper}}\right)=P\left(F_{0.95}<s_1^2/s_2^2\le F_{0.05}\right)=0.90$$





34