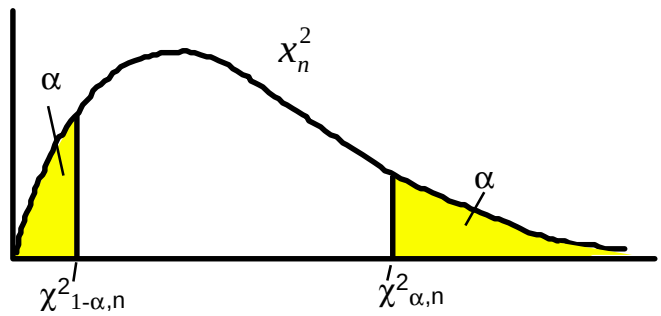
Useful Definitions around the distribution



Statistical tables can give us the value of the random number that "cuts" an α upper tail of the distribution.

The student-t distribution

The *student-t* distribution with k degrees of freedom is defined as:

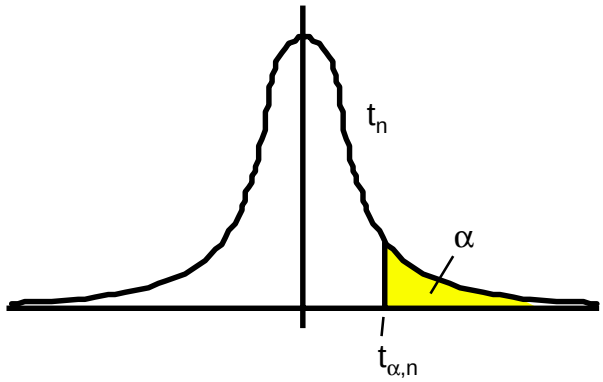
if z ~ N(0,1) then

$$\frac{z}{\sqrt{y/k}} \sim t_k$$

$$y \sim \chi^2_k$$

<u>Usage</u>: Find the distribution of the average when σ is unknown.

Useful Definitions about the student-t



Again, $t_{\alpha,n}$ signifies the value of the random number that cuts an a upper tail with area "α".
Notice that $t_{\alpha,n} = - t_{1-\alpha,n}$.
Also notice that $t_{\alpha,n}$ is a bit larger than z_α.

The F distribution

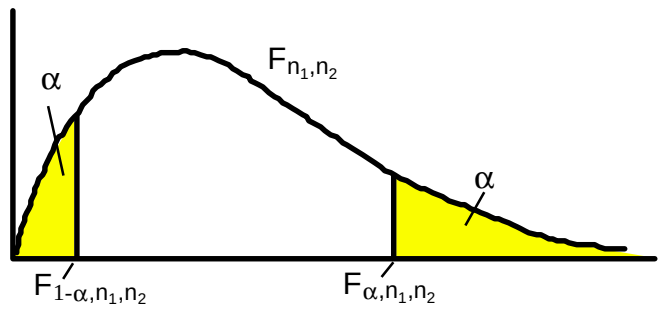
The F distribution with n_1, n_2 degrees of freedom is defined as:

$$\frac{y_1/n_1}{y_2/n_2} \sim F_{n_1,n_2}$$

$$y_1 \sim \chi^2_{n_1} \qquad y_2 \sim \chi^2_{n_2}$$

<u>Usage</u>: to compare the spread of two populations.

Useful Definitions around the F distribution



Statistical tables can give us the value of the random number that "cuts" an α upper tail of the F distribution.

It so happens that $F_{1-\alpha,n_1,n_2} = \dfrac{1}{F_{\alpha,n_1,n_2}}$

Sampling from a Normal Distribution

The statistics of the average:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \ \sigma \text{ is known}$$

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}, \ \sigma \text{ is } \textit{not} \text{ known}$$

The statistics of the standard deviation:

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

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} \sim F_{n_1-1,n_2-1}$$

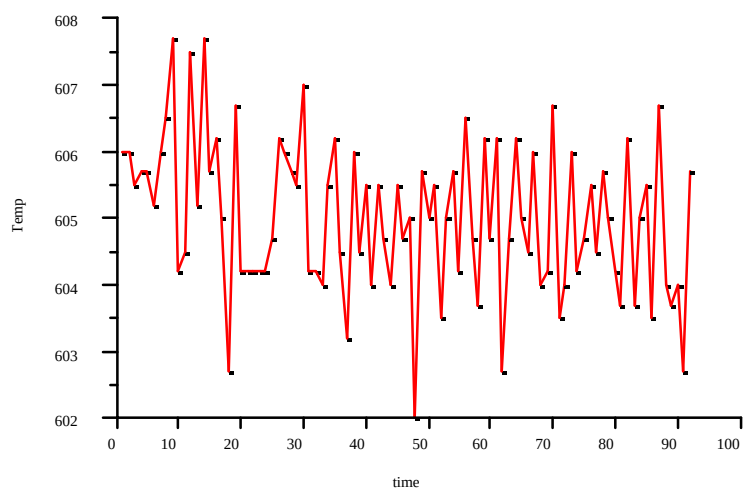
Estimation: The art of guessing at unknowns

Point estimation must be unbiased and must have minimum variance.

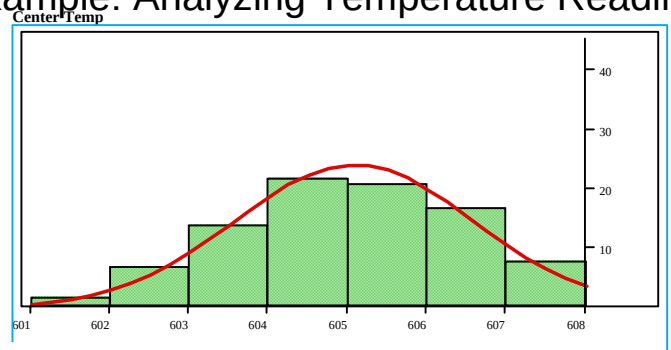
Interval estimation yields bounds that contain the actual value with a given certainty.

(For interval estimation we need the sampling distribution.)

Example - Analyze LPCVD Temp. Readings



Example: Analyzing Temperature Readings



Moments

Mean	605.0791
Std Dev	1.5058
Std Err Mean	0.1579
upper 95% Mean	605.3927
lower 95% Mean	604.7655
N	91.0000

Example: The 95% confidence intervals on furnace temperature statistics.

Mean:

$$\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$$

or, at the 5% level, $\mu = 605.1 \pm 0.313$

Variance:

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2},\, n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},\, n-1}}$$

or, at the 5% level, $1.31 < \sigma < 1.76$

Example: Use the tables to find the 1% confidence intervals for μ and σ, when:

Average: 605.1μm
Standard Deviation: 1.51μm
N=91
(The distribution is assumed to be normal.)

Hypothesis Testing

Hypothesis testing is a selection between:

Ho (null hypothesis)

and

H1 (alternative hypothesis).

α = P{type I error} = P{reject Ho / Ho is true}

β = P{type II error} = P{accept Ho / Ho is false}

Power = 1 - β = P{reject Ho / Ho is false}

<u>Example</u>:

Ho : the means of two populations are equal.

H1 : the means of two populations are different.

Comparing Population Mean to a Constant

(to simplify this, we assume that the variances are known)

Since the arithmetic average is distributed as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

given a constant μ_o , we have:

$$\bar{x} - \mu_o \sim N\left(\mu - \mu_o, \frac{\sigma^2}{n}\right)$$

and under the H_o that $\mu = \mu_o$, we have:

$$\bar{x} - \mu_o \sim N\left(0, \frac{\sigma^2}{n}\right)$$

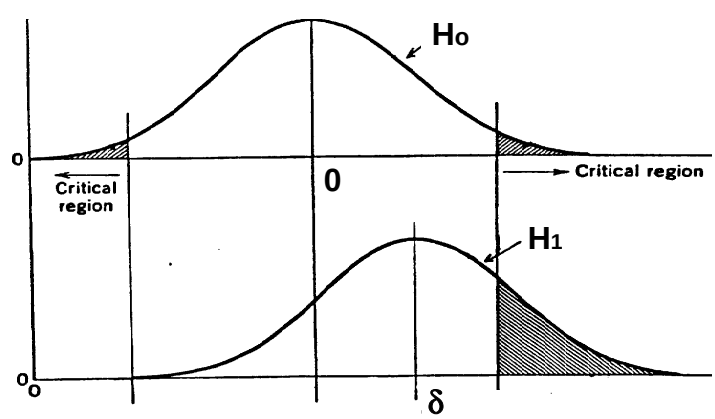
Under this assumption, we can find the probability that $\bar{x} - \mu_o$ will differ from zero by a certain amount...

Comparing Mean to a Constant (cont)

If the observed value of $\bar{x} - \mu_o$ is so far from zero that its probability under H_0 would be small, the H_0 should be rejected.

H_o is rejected at a *level of significance* "α" when $\bar{x} - \mu_o$ falls in a predesignated area of the H_0 distribution that is selected to correspond to a small probability "α".

Graphically:



Example:
Is the mean of the temperature 605 degrees?

Data File: tyl16stats L1

Single Sample...

Variable:	center	Population
Mean:	6.0510e+2	6.0500e+2
Std. Deviation:	1.5108e+0	
Observations:	92	

t-statistic:	6.3486e-1	Hypothesis:
Degrees of Freedom:	91	Ho: µ1 = µ2
Significance:	5.27e-1	Ha: µ1 ° µ2

Use the student-t tables to test the hypothesis:

H_o: The mean temperature is 605 degrees.

H_1: The mean temperature is *not* 605 degrees.

Step 1: Collect your data.

Step 2: Assume that H_o is true.

Step 3: Calculate the appropriate test statistic.

Step 4: Place the statistic against its assumed distribution and see how well it fits.

Step 5: Accept or reject H_o based on Step 4.

More Tests

Simple tests can also be devised to compare the mean value to a standard when the sigma is unknown.

Simple tests are also in use to compare two (or more) population means with sigmas known, unknown, identical or not.

A test based on the F distribution is available to compare sigmas.

Tests are also available to compare correlation coefficients, etc.

These tests are implicitly or explicitly used in experimental design when we compare the results of various treatments.

More Tests to Compare Populations

Compare means if the two sigmas are the same...

$$s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

$$\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Compare means if the two sigmas are different...

$$n = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1+1}+\frac{(s_2^2/n_2)^2}{n_2+1}}$$

$$\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} \sim t_n$$

Summary

Distributions are used to describe properties of populations.

Moments are used to describe aspects of distributions.

Estimators are used to approximate the values of moments.

Moments can be estimated as numbers and/or as intervals.

Populations can be compared by building and testing various Hypotheses about their moments.

What does all this have to do with the manufacturing of ICs?