A Brief Statistical Primer

Basic distributions. The central limit theorem. Sampling, estimation and hypothesis testing.
IC Production Requirements

- IC design must be modified for manufacturability.
- Incoming material and utilities must be qualified.
- Equipment must be qualified for volume production.
- Process must be debugged and transferred to high volume fab line.
- Yield detractors must be identified and eliminated.
- Efficient test procedures must be established.

Issues that SPC & DOE deal with

SPC helps remove the causes of problems and makes the process more stable over time.

DOE allows the exploration of the process settings and helps find the optimum operating conditions.
Quality is fitness for use.

There are two kinds of Quality:
  a. Quality of design.
  b. Quality of conformance.

**DOE** and **SQC** deal with quality of conformance.

**SPC** is one of the tools used to implement SQC.

Statistics is the art of making inferences about the whole by observing a sample.

---

The Evolution of Manufacturing Science

1. Invention of machine tools. English system (1800).
   *mechanical* - *accuracy*

2. Interchangeable components. American system (1850).
   *manufacturing* - *repeatability*

   *industrial* - *reproducibility*

   *quality* - *stability*

   *system* - *adaptability*

   *knowledge* - *versatility*
Outline

- Discrete Distributions
- Continuous Distributions (Normal)
- Central Limit Theorem
- Sampling
- Estimation
- Hypothesis Testing

Presenting and Summarizing Data

The histogram

The sample average

\[ \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \]

The sample variance

\[ s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} \]
The Distribution of x

Discrete Distributions

Continuous Distributions

A discrete distribution: the binomial

The probability distributions. (a) Discrete case. (b) Continuous case.

P\{D = x\} = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, ..., n

mean = np, variance = np(1-p)

where \( \binom{n}{x} = \frac{n!}{x!(n-x)!} \cdot n! \equiv 1 \cdot 2 \cdot ... \cdot (n-1) \cdot n \)

the sample fraction \( \hat{p} = \frac{D}{n} \) mean = p, variance = \( \frac{p(1-p)}{n} \)
Example: Using the Binomial Distribution

The probability of a broken wafer in a loading operation is 0.1%. After loading 100 wafers, what is the probability that:

- All wafers are OK?
- 1 wafer broke?
- 5 wafers broke?
- More than 5 wafers broke?

Another discrete distribution: the Poisson

\[ p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \]

\[ x = 0, 1, \ldots \]

where \( x! = 1 \cdot 2 \cdot 3 \cdots (x-1) \cdot x \)

mean: \( \mu = \lambda \)

variance: \( \sigma^2 = \lambda \)
A continuous distribution: the Normal

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}, \quad -\infty < x < \infty \quad x \sim N(\mu, \sigma^2) \]

\[ P\{x < a\} = F(a) = \int_{-\infty}^{a} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} dx \]

\[ z = \frac{x - \mu}{\sigma}, \quad P\{x < a\} = P\left\{ z < \frac{a-\mu}{\sigma} \right\} = \Phi\left( \frac{a-\mu}{\sigma} \right) \]

Useful Definitions on the Normal Distribution

\( z \sim N(0, 1) \) This is the Standard Normal Distribution

for any Normal distribution, \( z \) signifies the “number of sigmas away from the mean”.

\( z_\alpha \) is the number that “cuts” an upper tail with area \( \alpha \)

Notice that \( z_\alpha = -z_{1-\alpha} \)
The Normal Distribution

The Normal Law of Error stands out in the experience of mankind as one of the broadest generalizations of natural philosophy. It serves as the guiding instrument in researches in the physical and social sciences and in medicine, agriculture, and engineering. It is an indispensable tool for the analysis and the interpretation of the basic data obtained by observation and experiment.

Example: Table Lookup for Normal Distribution

The wafer to wafer thickness of a poly layer is distributed normally around 500nm with a $\sigma$ of 20nm:

$$P_{\text{th}} \sim N(500\text{nm}, 400\text{nm}^2)$$

What is the probability that a given wafer will have polysilicon thicker than 510nm?

... thinner than 480nm?

... between 490 and 515nm?
The Additivity of Variance

If \( y = a_1x_1 + a_2x_2 + \ldots + a_nx_n \) then \( \mu_y = a_1\mu_1 + \ldots + a_n\mu_n \)
and
\( \sigma^2_y = a_1^2\sigma^2_1 + \ldots + a_n^2\sigma^2_n \)

This applies under the assumption that the parameters \( x \) are independent.

For example, consider the thickness variance of a layer defined by two consecutive growths:

\[
\begin{align*}
\mu_t &= \mu_{g1} + \mu_{g2} \\
\sigma_t^2 &= \sigma_{g1}^2 + \sigma_{g2}^2
\end{align*}
\]

or by a growth step followed by an etch step:

\[
\begin{align*}
\mu_t &= \mu_{g} - \mu_e \\
\sigma_t^2 &= \sigma_{g}^2 + \sigma_{e}^2
\end{align*}
\]

Example: How to combine consecutive steps

The thickness of a SiO\(_2\) layer is distributed normally around 600nm with a \( \sigma \) of 20nm:

\[ \text{GOx}\sim N(600\text{nm}, 400\text{nm}^2) \]

During a polysilicon removal step with limited selectivity, some of the oxide is removed. The removed oxide is:

\[ \text{ROx}\sim N(50\text{nm}, 25\text{nm}^2) \]

What is the probability that the final oxide thickness is between 540 and 560nm?
The Correlation Coefficient

The correlation coefficient $\rho$, is a statistical moment that gives a measure of *linear dependence* between two random variables. It is estimated by:

$$ r = \frac{s_{xy}}{s_x s_y} $$

where $s_x$ and $s_y$ are the square roots of the estimates of the variance of $x$ and $y$, while $s_{xy}$ is an estimate of the covariance of the two variables and is estimated by:

$$ s_{xy} = \sum_{i=1}^{n} \frac{|x_i - \bar{x}| |y_i - \bar{y}|}{n-1} $$

**Correlation Example**

A plot showing the correlation of two variables $X_1$ and $X_2$ over a range of $n$ values from 0 to 100.
Using the Correlation Coefficient

In fact, the only way to stop variances from accumulating is by making sure that the parameters have the appropriate correlation $\rho$.

$$\mu_z = a \mu_x + b \mu_y$$

$$\sigma^2_z = a^2 \sigma^2_x + b^2 \sigma^2_y + 2ab \rho \sigma_x \sigma_y$$

For example, two consecutive growths can be controlled if we set $\rho = -1$

$$\mu_t = \mu_{g1} + \mu_{g2}$$

$$\sigma^2_t = \sigma_{g1}^2 + \sigma_{g2}^2 - 2 \sigma_{g1} \sigma_{g2}$$

and a growth followed by an etch can be controlled by setting $\rho = 1$

$$\mu_t = \mu_g - \mu_e$$

$$\sigma^2_t = \sigma_g^2 + \sigma_e^2 - 2 \sigma_g \sigma_e$$
The Central Limit Theorem:

The sum of independent random variables tends to have a normal distribution.

Uniformly distributed number | Sum of 5 unif. distr. numbers:

Statistical Moments and their Estimators

The mean (μ), the sigma (σ), the correlation (ρ), etc. are symbolized with Greek characters and they refer to “true”, but hidden values which we cannot know exactly.

These are the moments of a population.

The average, the standard deviation, the correlation coefficient, etc. are represented by Latin characters and they refer to values, which we can calculate from the data.

These values are used to estimate the true (but impossible to know exactly) moments. They are estimators.
Sampling and Estimation

Sampling: the act of making inferences about populations.

Random sampling: when each observation is identically and independently distributed.

Statistic: a function of sample data containing no unknowns. (e.g. average, median, standard deviation, etc.)

A statistic is a random variable. Its distribution is a sampling distribution.

Example: Estimating the mean of a normal dist.

The thickness of a poly layer is distributed normally around 500nm with a σ of 10nm:

\[ \text{Pth} \sim \text{N}(500 \text{nm}, 100 \text{m}^2) \]

We randomly select 50 wafers, measure the poly thickness and calculate the average of the fifty readings:

\[ \bar{\text{Pth}} = \frac{1}{50} \sum_{i=1}^{50} \text{Pth}_i \]

What is the distribution of \( \bar{\text{Pth}} \)?

What is the probability that \( \bar{\text{Pth}} \) will be between 495 and 505nm?
Example: The Mean of a Normal Dist. (cont)

Now that we understand how the average is distributed, let's use it to estimate the mean (assuming that we know that $\sigma$ is 10nm).

If the 50 measurements yielded an average of 503nm, what can we say about the unknown mean?

What is the estimated value of the unknown mean?

What is the probability that the unknown mean is between 500 and 506nm?

The chi-square distribution

The **chi-square** distribution with $n$ degrees of freedom is defined as:

$$x_1^2 + x_2^2 + \ldots + x_n^2 \sim \chi^2_n$$

**Usage:** To find the distribution of the standard deviation when the mean is known. Also to define the student-t distribution (more on this later).