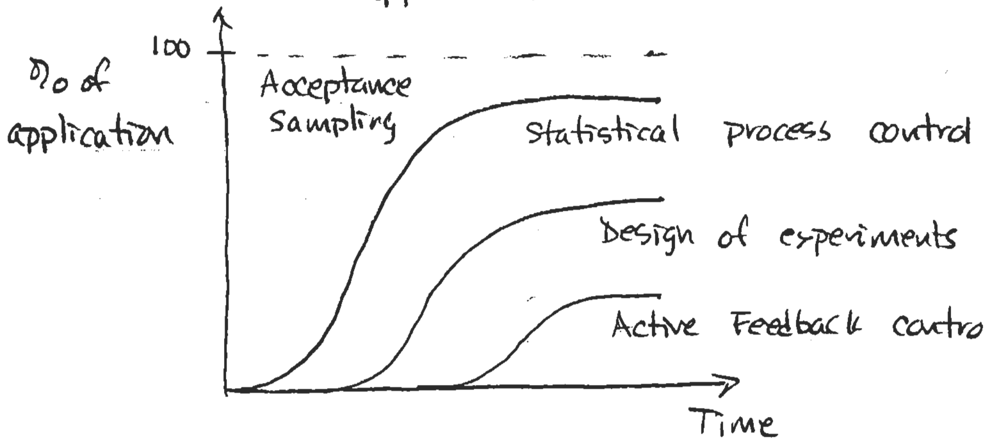
6.780

Spring 2003
D. Boning

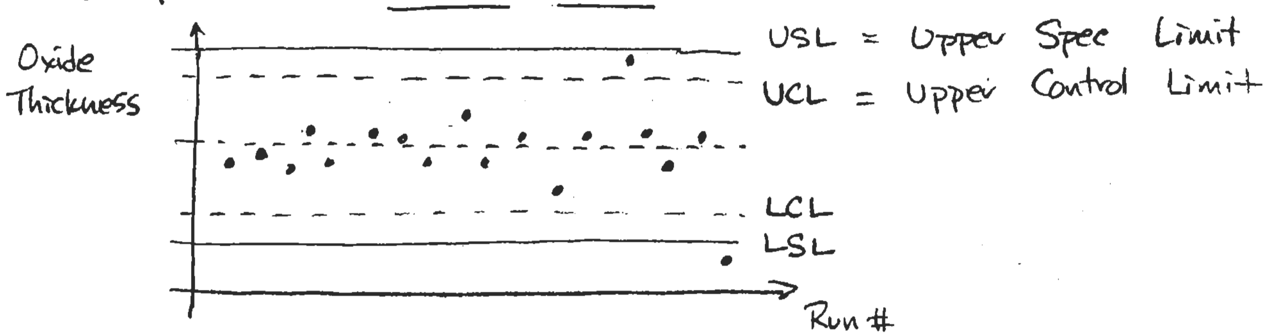
STATISTICAL PROCESS CONTROL

• Phases of method application:



SPC: Systematic approach to identify unusual behavior, identify, and then eliminate assignable causes to achieve minimum natural or inherent variation

• Example of a control chart



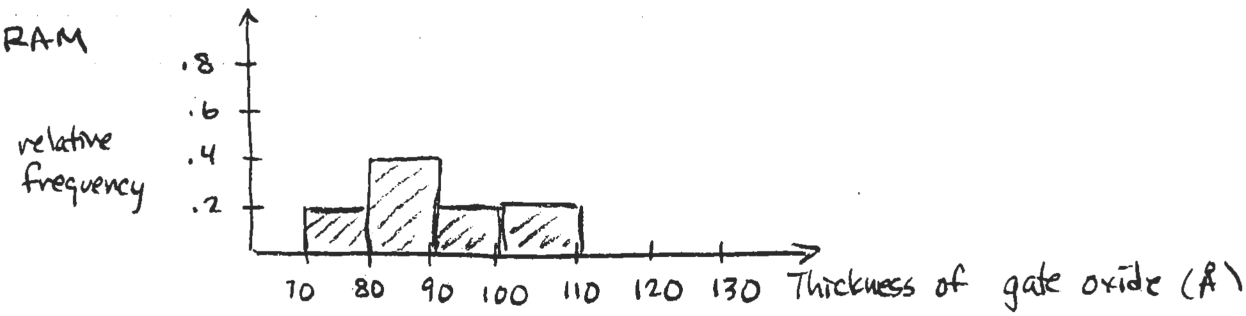
ISSUES:
- underlying statistical model?
- design of chart, Pr (false alarm)
  Pr (miss deviation)
- types of charts

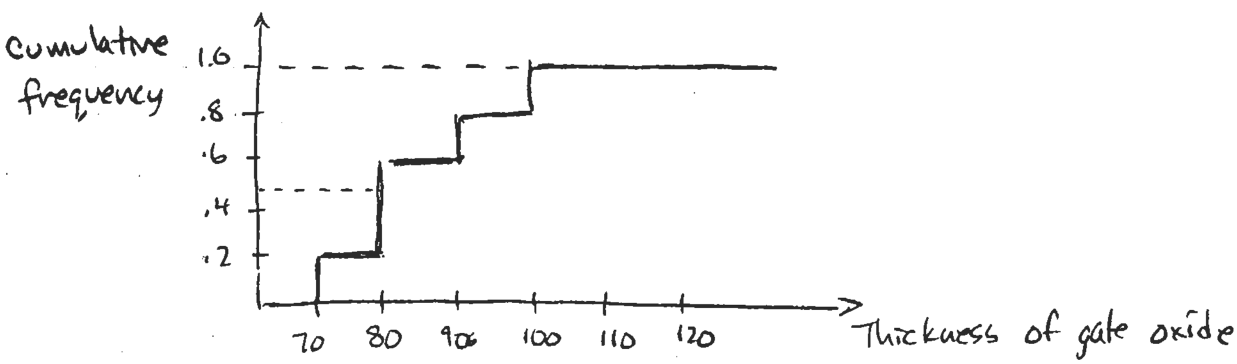
STATISTICS REVIEW

Descriptive Statistics

Focuses on the summarization and exposition (tabulation, grouping, graphical representation) of observed data and the derivation of numerical characteristics, such as measures of location, dispersion, shape.

HISTOGRAM





MEAN: measure of location $\bar{z} = \sum_{i=70}^{110} \phi_i z_i = 84$

MODE: value that occurs most frequently 80–90Å in above

MEDIAN: value in the "middle" of data 80 Å above
(# observations below = # above")

VARIANCE: $\sigma^2 = \sum_{i=70}^{110} (z_i - \bar{z})^2 \phi_i = 104$

STD. DEV: $\sigma = \sqrt{\sigma^2} = 10.2$ Å

Probabilistic Basis for Statistics

Drawback to descriptive statistics: study of observed data enables us to draw conclusions which relate only to the data we have.

Probability Theory

- Classical Definition: Prob. (event A) $P_A = \frac{N_A}{N}$

  in an experiment w/. N mutually exclusive and equally likely outcomes, N_A of which result in event A.

- Frequency Approach: $\lim_{n \to \infty} \left(\frac{n_A}{n} \right) = P_A$

- Subjective Approach: Degrees of belief

- Axiomatic Approach: Probability as set function on a field $\mathcal{F}$ which is a set of subsets of elementary event space S;

  that satisfies $P(A) \geq 0 \quad \forall A \in \mathcal{F}$

  $P(S) = 1$

  $P\left(\bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i)$ if A_i mutually exclusive

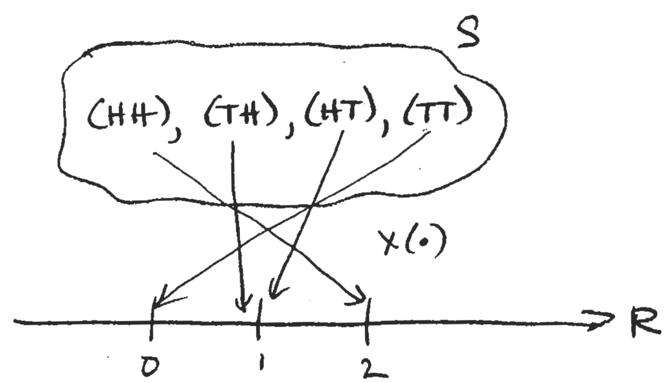
  A probability space includes a probability function

A random variable is a function $X(\cdot)$ that maps the sample space onto the real line $\mathbb{R}$

$X(\cdot): S \to R_X$

Example: two coin flips

$X(\cdot)$: # heads



Importance: lets us relate observations to some underlying probability model and thus make inferences about that process.

DISCRETE DISTRIBUTIONS

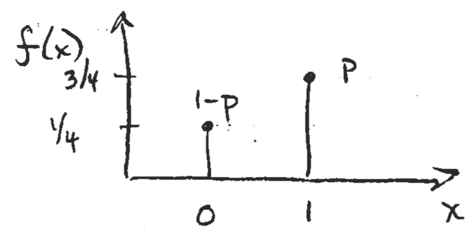
Bernoulli Distribution

A Bernoulli trial: experiment with two outcomes

$$\Pr(\text{success}) = \Pr(1)$$
$$\Pr(\text{failure}) = \Pr(0)$$

Probability mass function (pmf)

$$f(x,p) = \begin{cases} p, & x=1 \\ 1-p, & x=0 \end{cases}$$



Binomial Distribution

Most important discrete distribution: repeated random Bernoulli trials: x = # of successes in n trials

$$f(x,p,n) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,2,\ldots n$$

where $\binom{n}{x}$ "n choose x" $= \dfrac{n!}{x!\,(n-x!)}$

$$n! = 1\cdot 2\cdot 3 \cdots (n-1)\cdot n$$

$x \sim B(n,p)$

$\sim$ "distributed as"

De Moivre & Laplace: approximation for large n

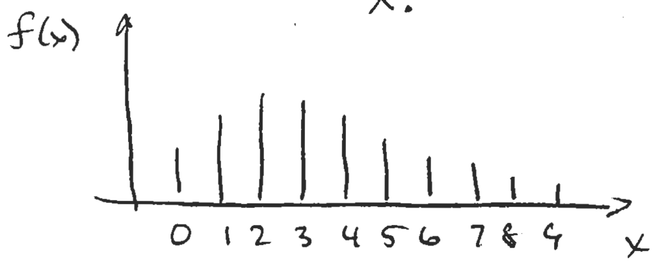
$$\binom{n}{x} p^x (1-p)^{n-x} \simeq \frac{1}{\sqrt{2\pi\, np\,(1-p)}}\, e^{-\frac{1}{2}z^2}$$

where $z = \dfrac{x - np}{\sqrt{np\,(1-p)}}$ -- pdf of the normal!

Poisson Distribution

Another important distribution; approximation to the binomial when n is large and p is small:

$$f(x, \lambda) = \frac{e^{-\lambda} \lambda^{x}}{\lambda!}, \quad \lambda = 0, 1, 2 \ldots$$



Examples:

- # misprints on page(s) of a book
- # transistors fail on 1st day of operation

$$\lambda \doteq n \cdot p$$

Continuous Distributions

- Distribution Function or cumulative density function (cdf): $F(x)$

vs. Probability Density Function (pdf): $f(x)$

$$\underset{\text{cdf}}{F(x)} = \int_{-\infty}^{x} \underset{\text{pdf}}{f(u)}\, du \qquad \forall x \in \mathbb{R}$$

$$f(x) = \frac{dF(x)}{dx} = \lim_{h \to 0} \Pr\left(\frac{x-h < x < x+h}{h}\right), \quad h > 0$$

NOTE: $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

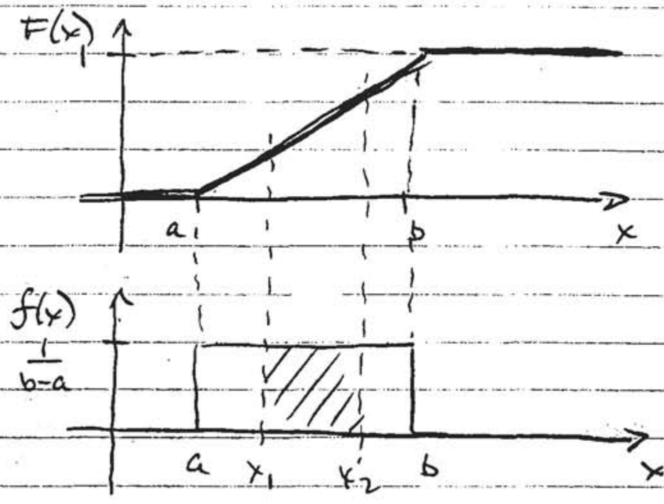
$$\int_{-\infty}^{\infty} f(u)\, du = 1$$

$$\Pr(a < x < b) = \int_{a}^{b} f(x)\, dx = F(b) - F(a)$$

Uniform Distribution

$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x < b \\ 1 & x \ge b \end{cases}$$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x < b \\ 0 & \text{otherwise} \end{cases}$$

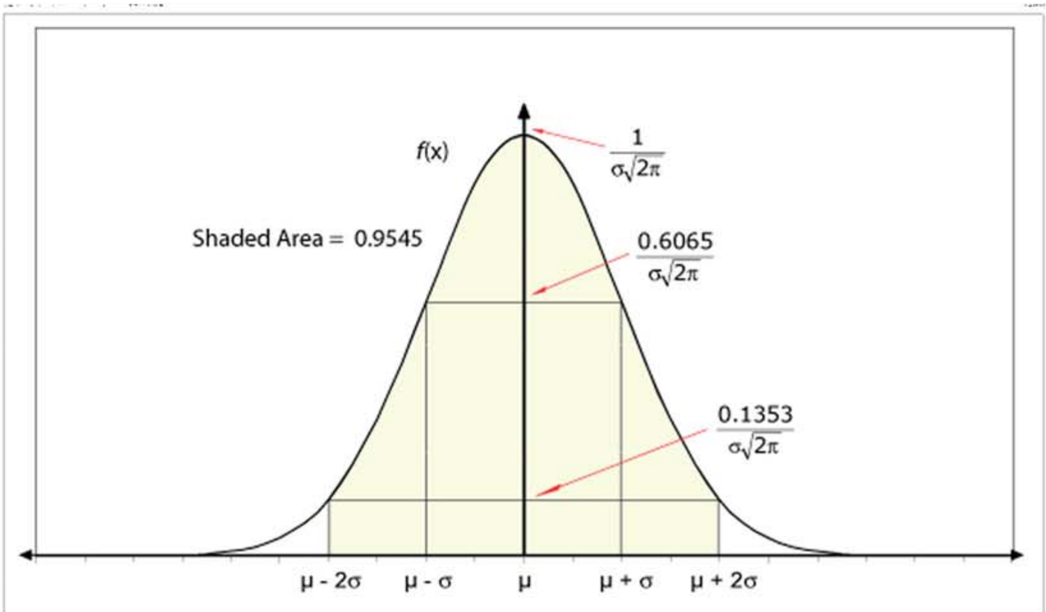


$X \sim U(a, b)$

E.g. $\quad Pr(x_1 < x < x_2) = \displaystyle\int_{x_1}^{x_2} \frac{1}{b-a}\,dx = \frac{x_2 - x_1}{b-a}$

Normal Distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^{1/2}} \qquad X \sim N(\mu, \sigma^2)$$



- $Pr(x \le a) = F(a) = \displaystyle\int_{-\infty}^{a} f(x)\,dx$

- Symmetric around μ

Standardize: $z = \dfrac{x-\mu}{\sigma}$ $\quad f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-1/2\, z^2}$

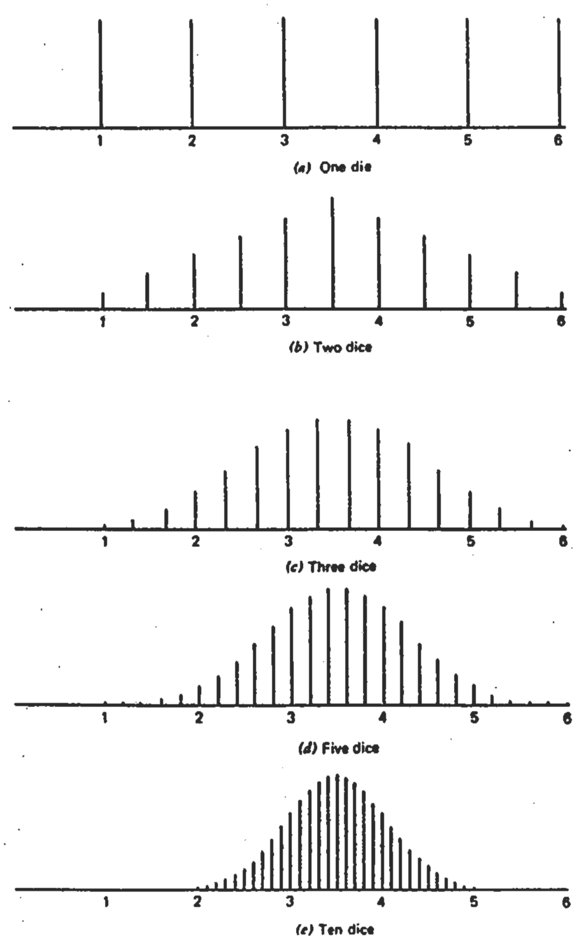
"standard normal distribution"

$Z \sim N(0,1)$

$$Pr(x \le a) = Pr\left(z \le \frac{a-\mu}{\sigma}\right) \triangleq \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$\Rightarrow$ use std. normal and cumulative normal statistical tables

Importance of Normal Distribution: <u>CENTRAL LIMIT THEOREM</u>



Distribution of average scores from throwing various numbers of dice.

Numerical Characteristics of Random Variables

Expectation

$$E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx \qquad \text{for continuous r.v.'s}$$

$$= \sum_i x_i\, f_i(x) \qquad \text{for discrete r.v.'s}$$

(a) $X \sim B(1, p)$ Bernoulli

x:	0	1
$f(x)$:	$1-p$	p

$$\Rightarrow E(x) = 0 \cdot (1-p) + 1 \cdot p = p$$

(b) $X \sim U(a,b)$ Uniform

$$\Rightarrow E(x) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_a^b x \left(\frac{1}{b-a}\right) dx$$

$$= \frac{1}{2}\left(\frac{1}{b-a}\right) x^2 \Big|_a^b = \frac{a+b}{2}$$

(c) $X \sim N(\mu, \sigma^2)$

$$\Rightarrow E(x) = \int_{-\infty}^{\infty} x \left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]\right) dx \; ; \quad z = \frac{x-\mu}{\sigma}$$

$$= \int_{-\infty}^{\infty} \frac{(\sigma z + \mu)}{\sqrt{2\pi}}\, e^{-1/2 z^2} =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{z\, e^{-1/2 z^2}}_{\text{odd fct } z} dz + \mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-1/2 z^2} dz$$

$$= 0 + \mu \cdot 1 = \mu$$

Properties

$E(c) = c$ if c is a constant

$E(ax_1 + bx_2) = a\,E(x_1) + b\,E(x_2)$ if x_1 & x_2 are independent

(d) $X \sim B(n, p)$ Binomial = sum of n indep. Bernoulli trials

$$\Rightarrow E(x) = n \cdot p$$

Variance

$$\mathrm{Var}(x) = E\left([x - E(x)]^2\right) = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)\, dx \qquad \text{continu.}$$

$$= \sum_i (x_i - E(x_i))^2 f_i(x) \qquad \text{discrete}$$

(a) $x \sim B(1, p)$ Bernoulli $\quad E(x) = p$

$$(x_0 - \bar{x})^2 pr(x=0) + (x_1 - \bar{x})^2 p_r(x=1)$$

$$\mathrm{Var}(x) = (0 - p)^2 (1 - p) + (1 - p)^2 p = p(1-p)$$

(b) $x \sim U(a, b)$ Uniform

$$\mathrm{Var}(x) = \frac{(b-a)^2}{12}$$

(c) $x \sim N(\mu, \sigma^2)$ Normal

$$\mathrm{Var}(x) = \sigma^2$$

Properties

$$\mathrm{Var}(x) = E(x^2) - [E(x)]^2 \quad \text{where } E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)\, dx$$

$\mathrm{Var}(c) = 0$ for c constant

$\mathrm{Var}(ax) = a^2\, \mathrm{Var}(x)$ for a constant

Generally, if x_i are independent

$$y = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

$$\mu_y = E(y) = a_1 \mu_1 + a_2 \mu_2 + \dots + a_n \mu_n$$

$$\sigma_y^2 = \mathrm{Var}(y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2$$

NOTE: $\sigma_y \neq a_1 \sigma_1 + a_2 \sigma_2$

Covariance

$$\sigma_{xy}^2 \triangleq \mathrm{Cov}(X,Y) = E[(X-E(X))(Y-E(Y))]$$
$$= E(XY) - E(X)E(Y)$$

If X, Y are independent, then $E(XY) = E(X)E(Y)$
$\Rightarrow \mathrm{Cov}(X,Y) = 0$

Correlation Coefficient

$$\rho_{xy} \triangleq \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Properties $\mathrm{Corr}(X_1, X_2) = 0$ for X_1 and X_2 indep. r.v.'s

$-1 \le \mathrm{Corr}(X_1, X_2) \le 1$

$\mathrm{Corr}(X_1, X_2) = 1$ for $X_1 = a + bX_2$, a, b real const.

Example: bivariate normal random vector

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

$$f(x_1,x_2) = \frac{(1-\rho^2)^{-1/2}}{2\pi\sigma_1\sigma_2} \exp\left\{ \frac{-1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

$\Rightarrow \mathrm{Cov}(X_1, X_2) = \sigma_1\sigma_2\rho$

Here, $\rho = 0$ implies independence, i.e. $f(x_1,x_2) = f(x_1)\cdot f(x_2)$
This is NOT generally true (i.e. for other distributions)

	MOMENTS of the POPULATION	SAMPLE STATISTICS	
Mean:	$\mu = \mu_x = E(x)$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$	average or "sample mean"
Variance:	$\sigma^2 = \sigma^2_{xx} = E(x-\mu_x)^2$	$s^2 = s^2_{xx} = s^2_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$	
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$	
Covariance	$\sigma^2_{xy} = E[(x-\mu_x)(y-\mu_y)] = E(xy) - E(x)E(Y)$	$s^2_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$	
Correlation Coefficient	$\rho_{xy} = \frac{\sigma^2_{xy}}{\sigma_x\sigma_y} = \frac{Cov(xY)}{\sqrt{Var(x)Var(Y)}}$	$r_{xy} = \frac{s_{xy}}{s_x s_y}$	

SAMPLING and ESTIMATION

- Sampling: act of making inferences about populations
- Random Sampling: when each observation is identically and independently distributed (i.i.d)
- Statistic: a function of sample data containing no unknowns (e.g. average, median, std.dev. etc.)
- A statistic is a random variable, which itself has a sampling distribution

Sampling and Estimation, an example

- Suppose we know the thickness of a field oxide is normally distributed:

$$T \sim N(\mu, \sigma^2)$$

We sample 50 random wafers and compute the mean oxide thickness:

$$\bar{T} = \frac{1}{50} \sum_{i=1}^{50} T_i$$

Key questions: (1) what is distribution of $\bar{T}$?
(2) what is $Pr(a \le \bar{T} \le b)$?

(1) Distribution of $\bar{T}$? $\quad \bar{T} \sim N(\mu, \sigma^2/n)$

Why? $\quad \bar{T} = \frac{T_1}{n} + \frac{T_2}{n} + \dots \frac{T_n}{n} \quad$ where $T_i \sim N(\mu, \sigma^2)$

$$E(\bar{T}) = n \cdot \frac{1}{n} E(T_i) = \mu$$

$$Var(\bar{T}) = \frac{1}{n^2} Var(T_1) + \frac{1}{n^2} Var(T_2) + \dots \frac{1}{n^2} Var(T_n)$$
$$= \frac{1}{n} \sigma^2$$

This is a very <u>general</u> result, and depends only on IID and <u>not</u> specifics of the distribution

Returning to key questions: EXAMPLE

- Know $T_{ox} \sim N(\mu, \sigma^2)$

μ: unknown

σ^2: known, $\sigma = 0.02\,\mu m$, $\sigma^2 = 0.0004\,\mu m^2$

Take $n = 50$ measurements, observe $\bar{T} = 1.22\,\mu m$

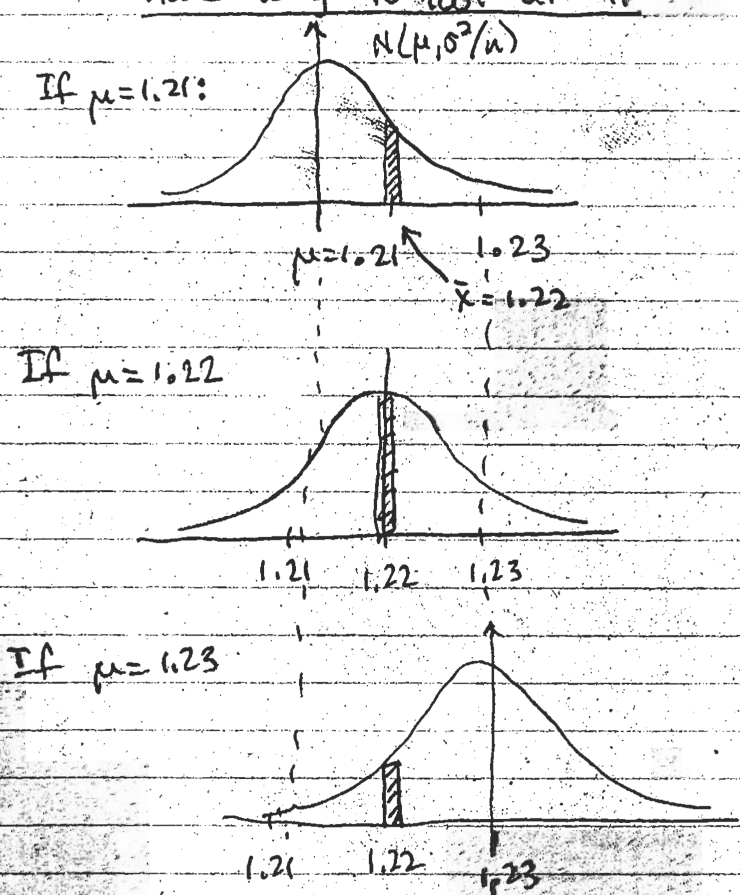
What can we infer about the unknown mean?

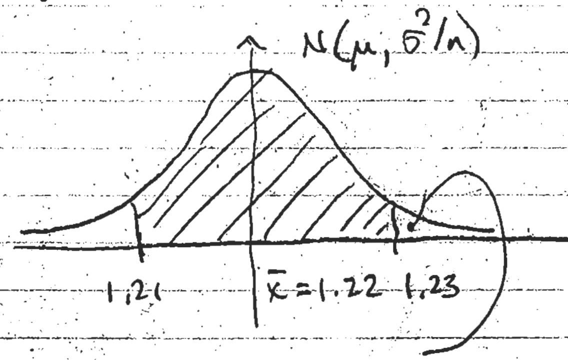
① $\hat{\mu}$ (estimate of μ)? $\Rightarrow$ $\hat{\mu} = \bar{x} = 1.22\,\mu m$

② Probability true mean is between $1.21\,\mu m$ & $1.23\,\mu m$?
$Pr(1.21 \le \mu \le 1.23\,\mu m)$

$\Rightarrow$ Use sampling distribution: $\bar{T} \sim N(\mu, \sigma^2/n)$

- Hard way to look at it vs. Easy way to look at it





$$1 - \Phi\left(\frac{1.23 - 1.22}{0.02/\sqrt{50}}\right)$$
$$= 1 - \Phi(3.54)$$
$$= 1 - 0.99980 = 0.0002$$

So $Pr(1.21 \le \mu \le 1.23)$
$= 1 - 2(0.0002) = 0.9996$

or with 99.96% CONFIDENCE we know mean is between $1.21\,\mu m$ and $1.23\,\mu m$!

$n\uparrow$, confidence in $\hat{\mu}\uparrow$

KEY DISTRIBUTIONS ARISING IN SAMPLING

Chi-Square Distribution

- If $x_i \sim N(0,1)$ for $i = 1, 2, \ldots, n$
  and $y = x_1^2 + x_2^2 + \ldots x_n^2$

  Then $Y \sim \chi^2_n$ - Chi-square distribution with n degrees of freedom

- $$f(\chi^2, n) = \frac{1}{2^{n/2}\,\Gamma(n/2)} (\chi^2)^{n/2-1} e^{-\chi^2/2}, \quad \chi^2 > 0$$ TABLES!

- Typical Use: Find distribution of variance when mean known

- Example: Suppose we know $x_i \sim N(\mu, \sigma^2)$

  (1) $\bar{x} = \frac{1}{n}\sum_i x_i \sim N(\mu, \sigma^2/n)$

  (2) standardize it! $y = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{x}-\mu}{\sigma}\right) \sim N(0,1)$

  (3) square it $y^2 \sim \chi^2_1$ by definition

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

  $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ // ... why n-1 degrees of freedom?

  $$\sum_i \left(\frac{x_i-\mu}{\sigma}\right)^2 = \left[n\left(\frac{\bar{x}-\mu}{\sigma}\right)^2 + \frac{(n-1)s^2}{\sigma^2}\right] \sim \chi^2_n$$

  lose one d.o.f. here.

Student - t Distribution

- If $z \sim N(0,1)$ then $\dfrac{z}{\sqrt{y/k}} \sim t_k$ - student t with k degrees of freedom

  if $y \sim \chi^2_k$

- Typical Use: Find distribution of average when σ is NOT known

- For $k \to \infty$, $t_k \to N(0,1)$

- Consider $x_i \sim N(\mu, \sigma^2)$

$$\frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{\left(\dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)}{s/\sigma} \sim \frac{N(0,1)}{\sqrt{\frac{1}{n-1}\cdot \chi^2_{n-1}}} \sim t_{n-1}$$

F Distribution

- If $y_1 \sim \chi^2_u$, $y_2 \sim \chi^2_v$

  Then $Y = \dfrac{y_1/u}{y_2/v} \sim F_{u,v}$ - F distribution with u, v degrees of freedom

- Typical Use: Compare the spread of two populations

- Example: $X \sim N(\mu_x, \sigma_x^2)$ from which sample $x_1, x_2 \ldots x_n$
  $Y \sim N(\mu_y, \sigma_y^2)$ " $y_1, y_2 \ldots y_m$

$$\frac{s_x^2/\sigma_x^2}{s_y^2/\sigma_y^2} \sim F_{n-1,m-1}$$