ESD.86
Exam #2 Review

Dan Frey
Associate Professor of Mechanical Engineering and Engineering Systems

Some Study Suggestions

- Review all the lecture notes
  - If there is a concept test, know the answer and WHY it's the right answer and WHY other answers are wrong
  - If there is a word you don't understand, look it up, talk to colleagues...
- Review the last two homeworks
  - For concepts, not details
- Review the reading assignments (April and May)
  - For concepts, not details
  - If there is a graph or table, make sure you can describe how to use it and the procedure by which it was made

Suggested Resources
(I would not emphasize this as much as the lecture notes and problem sets)

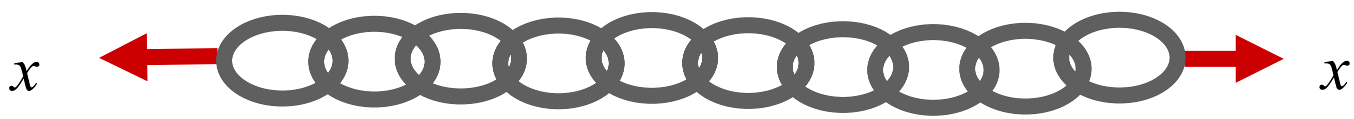
Wu and Hamada. *Experiments: Planning, Analysis and Parameter Design Optimization*. Chapters 4 and 5 are relevant. The "practical summaries" are good condensations of the main points. Many of the exercises are close to what you might find on the test. Many are not e.g. "prove that ..." or "find the defining words and resolution..."

***Problem Solvers: Statistics* (Research & Education Association) Solving lots of problems is good practice. There are books filled with problems and solutions. A large fraction of these involve lots of number crunching, these are fine to review but that's not the sort of question I intend to give. Many are conceptual or involve modest computations or estimation. Those are closer in style to what you can expect. There are many topics we didn't cover so you don't have to study them.**

Weibull's Derivation

Call P_n the probability that a chain will fail under a load of x

Let's define a cdf for *each* link meaning the link <u>will fail</u> at a load X less than or equal to x as $P(X \leq x) = F(x)$



If the chain <u>does not fail</u>, it's because all n links <u>did not fail</u>

If the n link strengths are probabilistically independent

$$1 - P_n = (1 - P)^n$$

Weibull, W., 1951, "A Statistical Distribution Function of Wide Applicability," *J. of Appl. Mech*.

Some Terms Related to Estimation

- Consistent – for any c $\lim_{n\to\infty} P\left(\left|\hat{\theta}-\theta\right| \geq c\right) = 0$ MLEs are
- Unbiased – $E(\hat{\theta}) = \theta$ MLEs are not always
- Minimum variance

$$\mathrm{var}(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

MLEs are pretty close

Complex Distributions

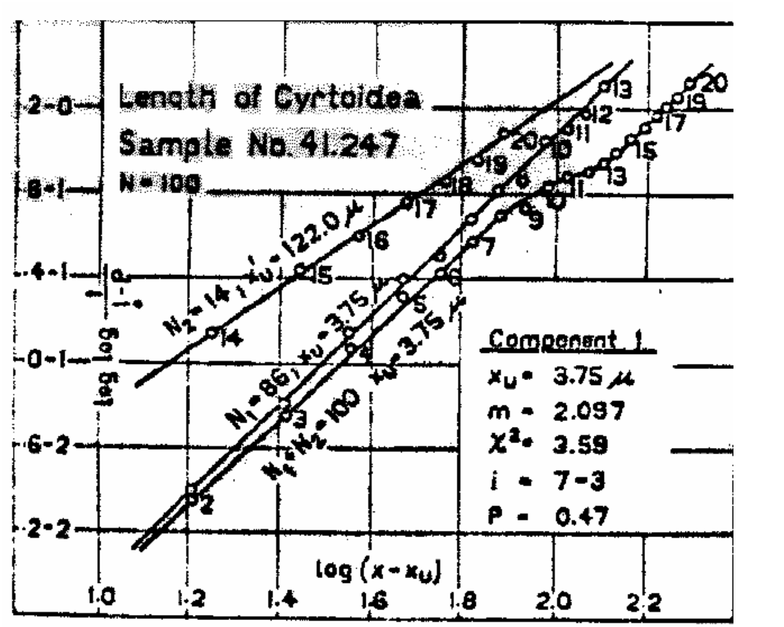


FIG. 4 LENGTH OF CYRTOIDEAE

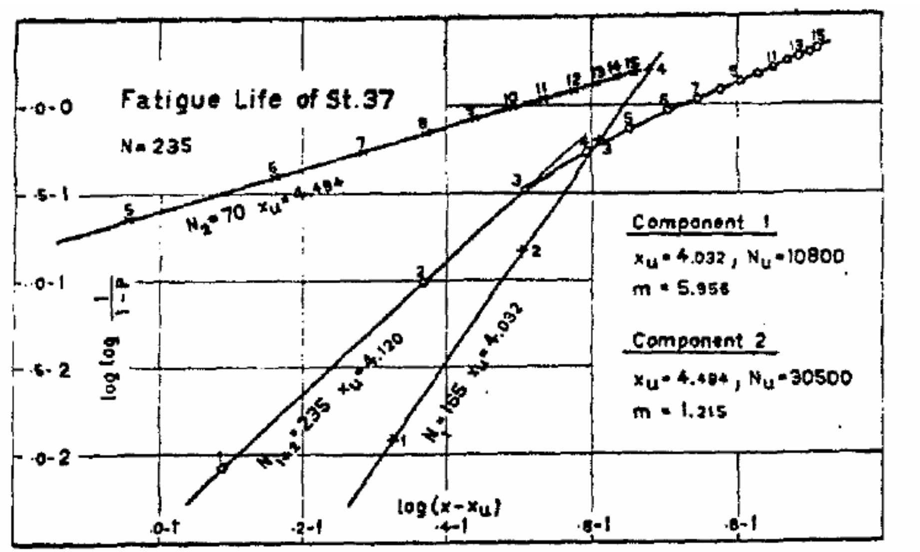


FIG. 6 FATIGUE LIFE OF ST-37 STEEL

Weibull, W., 1951,"A Statistical Distribution Function of Wide Applicability," *J. of Appl. Mech*.

Looking for Further Evidence of Two Populations

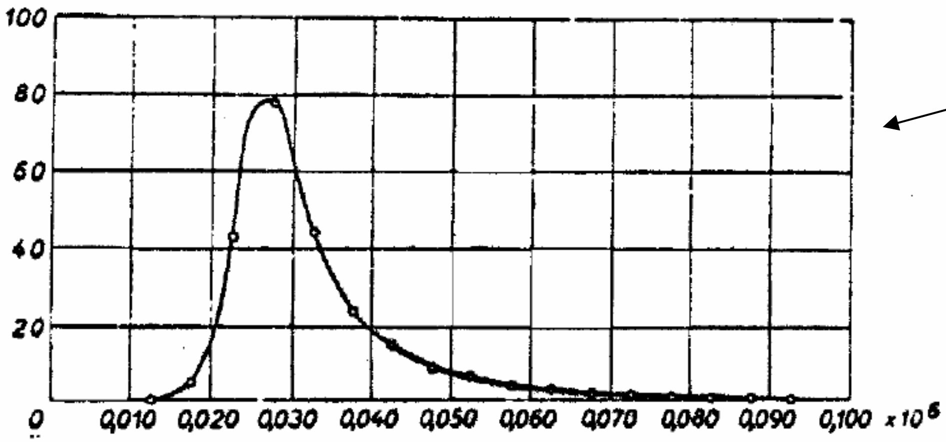


FIG. 5 FREQUENCY CURVE OF FATIGUE LIFE OF ST-37 STEEL
(Number of specimens versus number of stress cycles.)

No evidence of bi-modality in fatigue data

Clear evidence of bi-modality in strength data

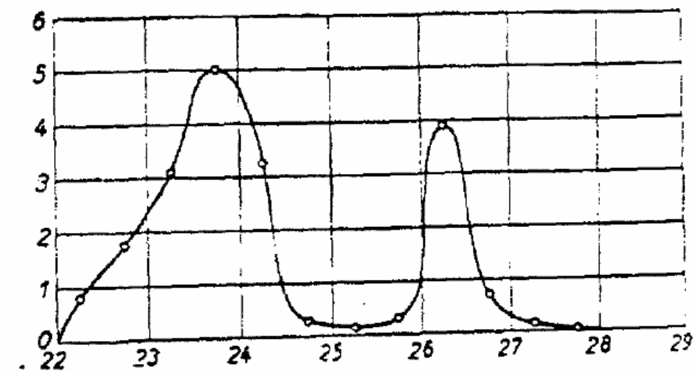


FIG. 7 FREQUENCY CURVE OF YIELD STRENGTH OF ST-37 STEEL
(Number of specimens versus yield strength in kg/mm².)

Reliability Terminology

- Reliability function *R*(*t*) -- The probability that a product will continue to meet its specifications over a time interval
- Mean Time to Failure *MTTF* -- The average time *T* before a unit fails
- Instantaneous failure rate $\lambda(t)$

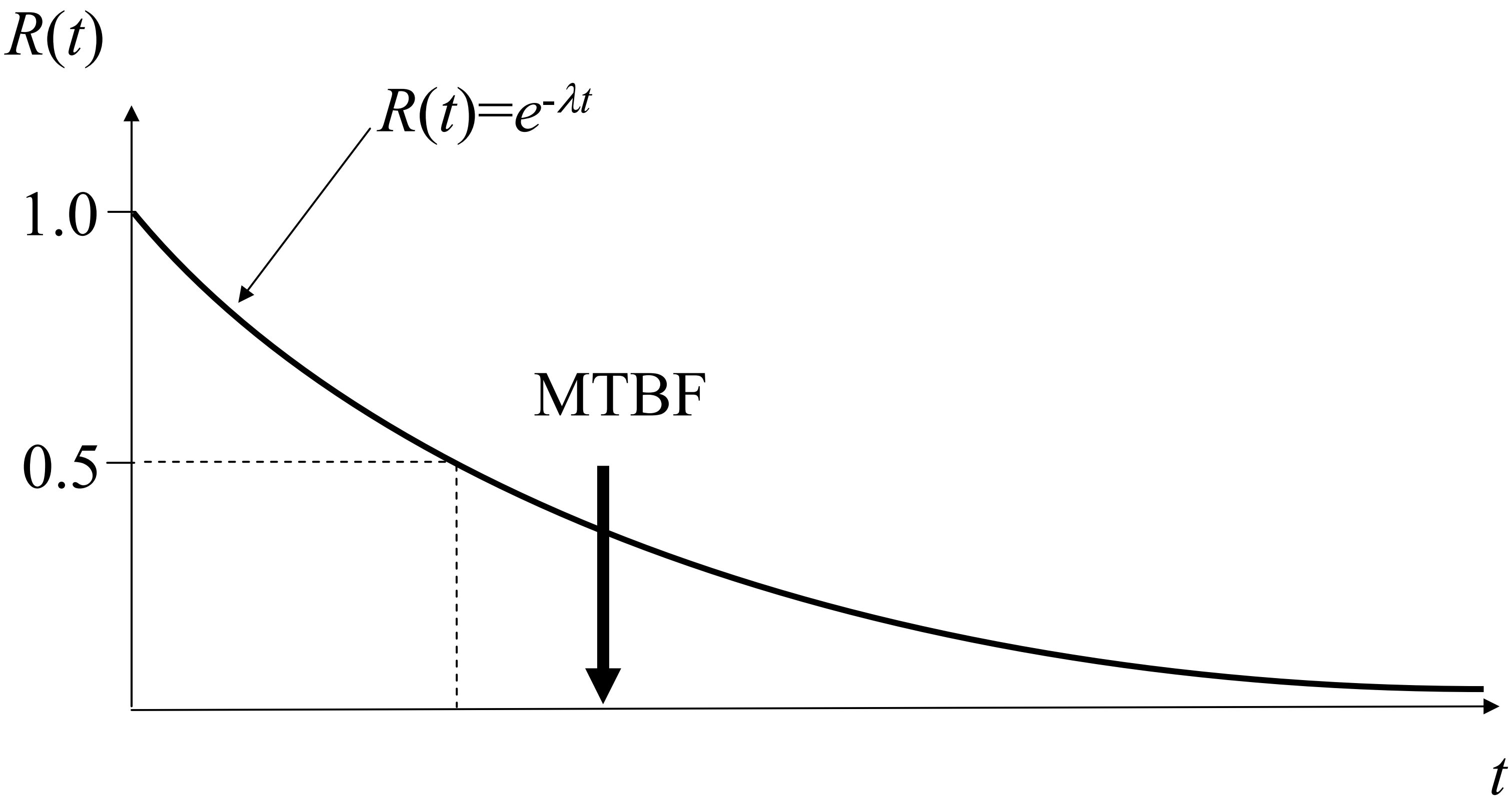
$$MTTF = \int_0^\infty R(t)\,dt$$

$$\lambda(t) = \Pr(\text{System survives to } t + dt \mid \text{System survives to } t)$$

$$R(t) = e^{-\int_0^t \lambda(\xi)\,d\xi}$$

Constant Failure Rates

"When the system operating time is the MTBF, the reliability is 37%"
- Blanchard and Fabrycky



Fisher's Null Hypothesis Testing

1. Set up a statistical null hypothesis. The null need not be a nil hypothesis (i.e., zero difference).
2. Report the exact level of significance ... Do not use a conventional 5% level, and do not talk about accepting or rejecting hypotheses.
3. Use this procedure only if you know very little about the problem at hand.

Gigernezer's Quiz

Suppose you have a treatment that you suspect may alter performance on a certain task. You compare the means of your control and experimental groups (say 20 subjects in each sample). Further, suppose you use a simple independent means t-test and your result is significant (t = 2.7, $d.f.$ = 18, p = 0.01). Please mark each of the statements below as “true” or “false.” ...

1. You have absolutely disproved the null hypothesis
2. You have found the probability of the null hypothesis being true.
3. You have absolutely proved your experimental hypothesis (that there is a difference between the population means).
4. You can deduce the probability of the experimental hypothesis being true.
5. You know, if you decide to reject the null hypothesis, the probability that you are making the wrong decision.
6. You have a reliable experimental finding in the sense that if, hypothetically, the experiment were repeated a great number of times, you would obtain a significant result on 99% of occasions.

Concept Question

- This Matlab code repreatedly generates and tests simulated "data"
- 20 "subjects" in the control and treatment groups
- Both normally distributed with the same mean
- How often will the *t*-test reject H_0 (α=0.01)?

```
for i=1:1000
  control=random('Normal',0,1,1,20);
  trt=random('Normal',0,1,1,20);
  reject_null(i) = ttest2(control,trt,0.01);
end
mean(reject_null)
```

1) ~99% of the time
2) ~1% of the time
3) ~50% of the time
4) None of the above

Concept Question

- This Matlab code repreatedly generates and tests simulated "data"
- 20 "subjects" in the control and treatment groups
- Both normally distributed with the different means
- How often will the *t*-test reject H_0 (α=0.01)?

```
for i=1:1000
   control=random('Normal',0,1,1,200);
   trt=       random('Normal',1,1,1,200);
   reject_null(i) = ttest2(control,trt,0.01);
end
mean(reject_null)
```

1) ~99% of the time
2) ~1% of the time
3) ~50% of the time
4) None of the above

Concept Question

- How do “effect” and "alpha" affect the rate at which the *t*-test rejects H_0?

```
effect=1;alpha=0.01;
for i=1:1000
   control=random('Normal',0,1,1,20);
   trt=        random('Normal',effect,1,1,20);
   reject_null(i) = ttest2(control,trt,alpha);
end
mean(reject_null)
```

a) ↑ effect, ↑ rejects
b) ↑ effect, ↓ rejects
c) ↑ alpha, ↑ rejects
d) ↑ alpha, ↓ rejects

1) a & c
2) a & d
3) b & c
4) b & d

NP Framework and Two Types of Error

- Set a critical value c of a test statistic T or else set the desired confidence level or "size" α
- Observe data X
- Reject H_1 if the test statistic $T(X) \geq c$
- Probability of Type I Error – The probability of $T(X) < c \mid H_1$
  - (i.e. the probability of rejecting H_1 given H_1 is true
- Probability of Type II Error – The probability of $T(X) \geq c \mid H_2$
  - (i.e. the probability of not rejecting H_1 given H_2 is true
- The *power* of a test is 1 - probability of Type II Error
- In the N-P framework, power is <u>maximized</u> subject to Type I error being <u>set</u> to a fixed critical value c or of α

or other confidence region (e.g. for "two-tailed" tests)

Measures of Central Tendency

- Arithmetic mean $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
  - an unbiased estimate of $$\mu = E(x) = \int_S x f_x(x)\,\mathrm{d}x$$
- Median $$= \begin{cases} X_{\frac{n+1}{2}} \text{ if } n \text{ is odd} \\ \frac{1}{2}\left(X_{n/2} + X_{1+n/2} \right) \text{if } n \text{ is even} \end{cases}$$
- Mode – The most frequently observed value in the sample

Confidence Intervals

- Assuming a given distribution and a sample size n and a given value of the parameter θ the 95% confidence interval from U to V is s.t. the estimate of the parameter $\hat{\theta}$

$$\Pr(U < \hat{\theta} < V \middle| \theta) = 95\%$$

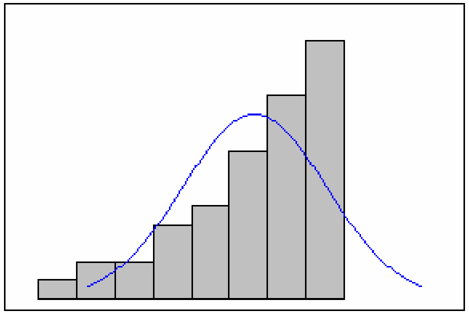
- The confidence interval depends on the confidence level, the sample variance, and the sample size

Measures of Dispersion

- Population Variance $$VAR(X) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$
- Sample variance $$S^2 = \frac{1}{n-1}\sum_{i-1}^{n}(X_i - \overline{X})^2$$
  - an unbiased estimate of $$\sigma^2 = E((x - E(x))^2)$$
- *n*th central moment $$E((x - E(x))^n)$$
- *n*th moment about m $$E((x - m)^n)$$

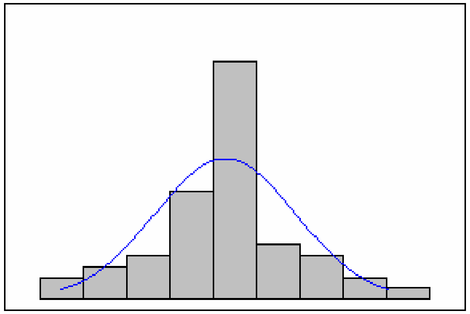
Skewness and Kurtosis

- Skewness $E((x-E(x))^3)$



positively skewed distribution

- Kurtosis $E((x-E(x))^4)$



positive kurtosis

Correlation Coefficient

- Sample

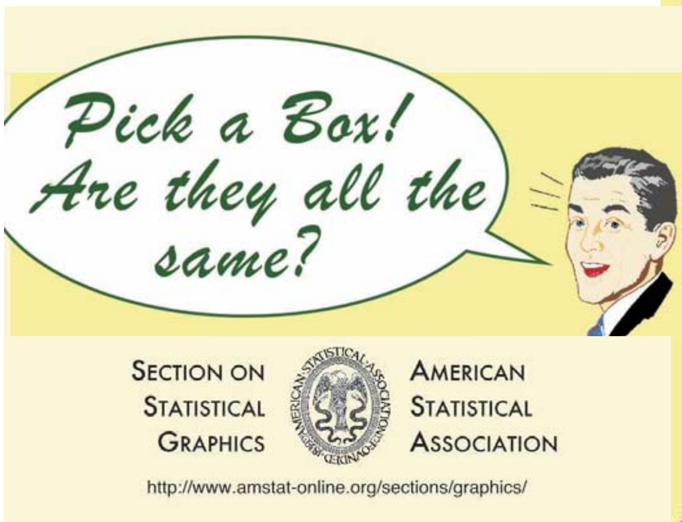
$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\left(n-1\right)S_X S_Y}$$

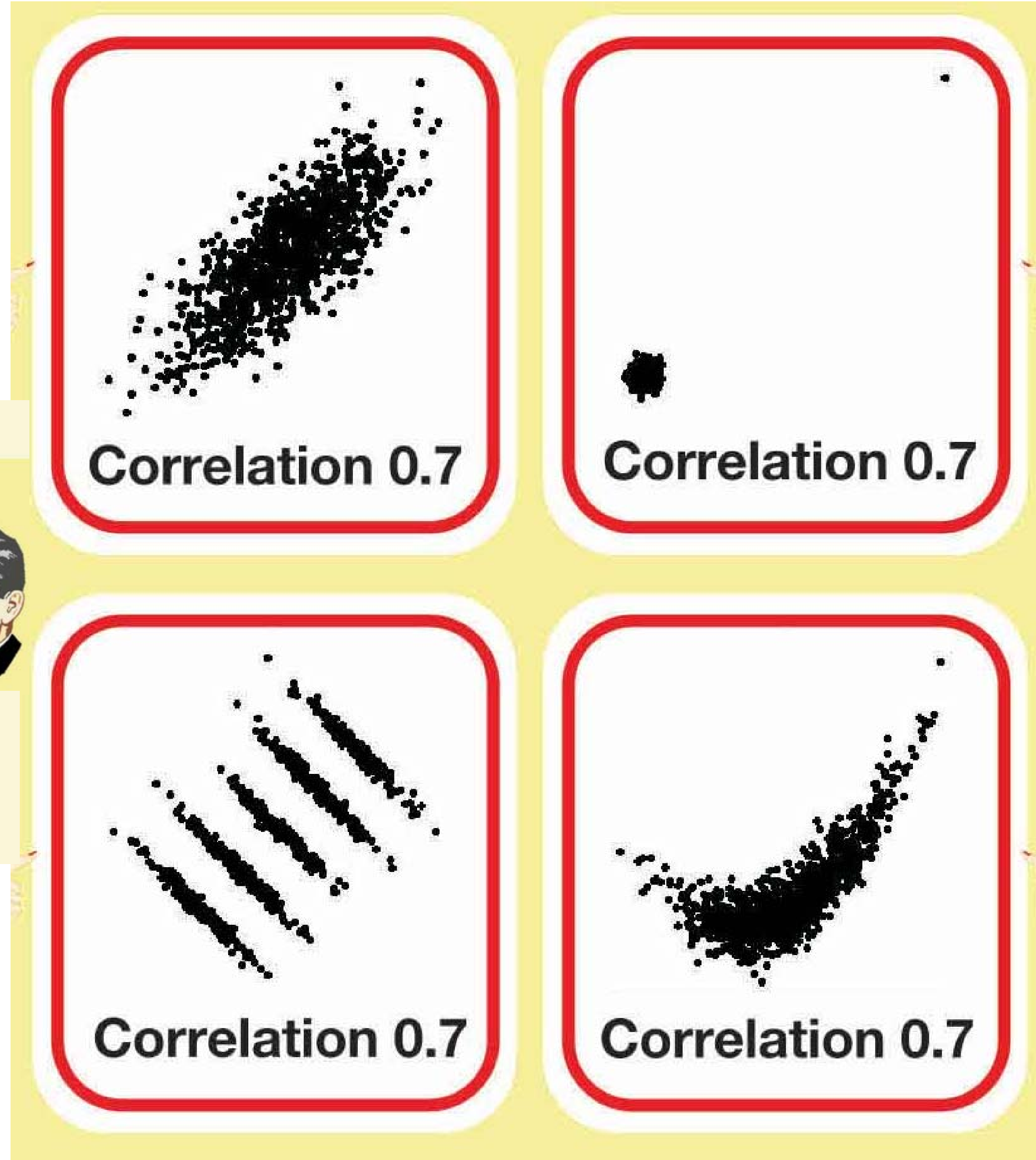
$$S_X{}^2 = \frac{1}{n-1}\sum_{i-1}^{n}(X_i - \overline{X})^2$$

- Which is an estimate of

$$\frac{E((x - E(x))(y - E(y)))}{\sigma_x \sigma_y}$$

But What Does it Mean?





Courtesy of American Statistical Association. Used with permission.

What is Linear Regression?

1. Form a probabilistic model

random — theoretical parameters — independent variable — random $E(\varepsilon)=0$

$$Y = \alpha + \beta x + \varepsilon$$

2. Get a sample of data in pairs (X_i, Y_i), $i=1\ldots n$
3. Estimate the parameters of the model from the data

The Method of Least Squares

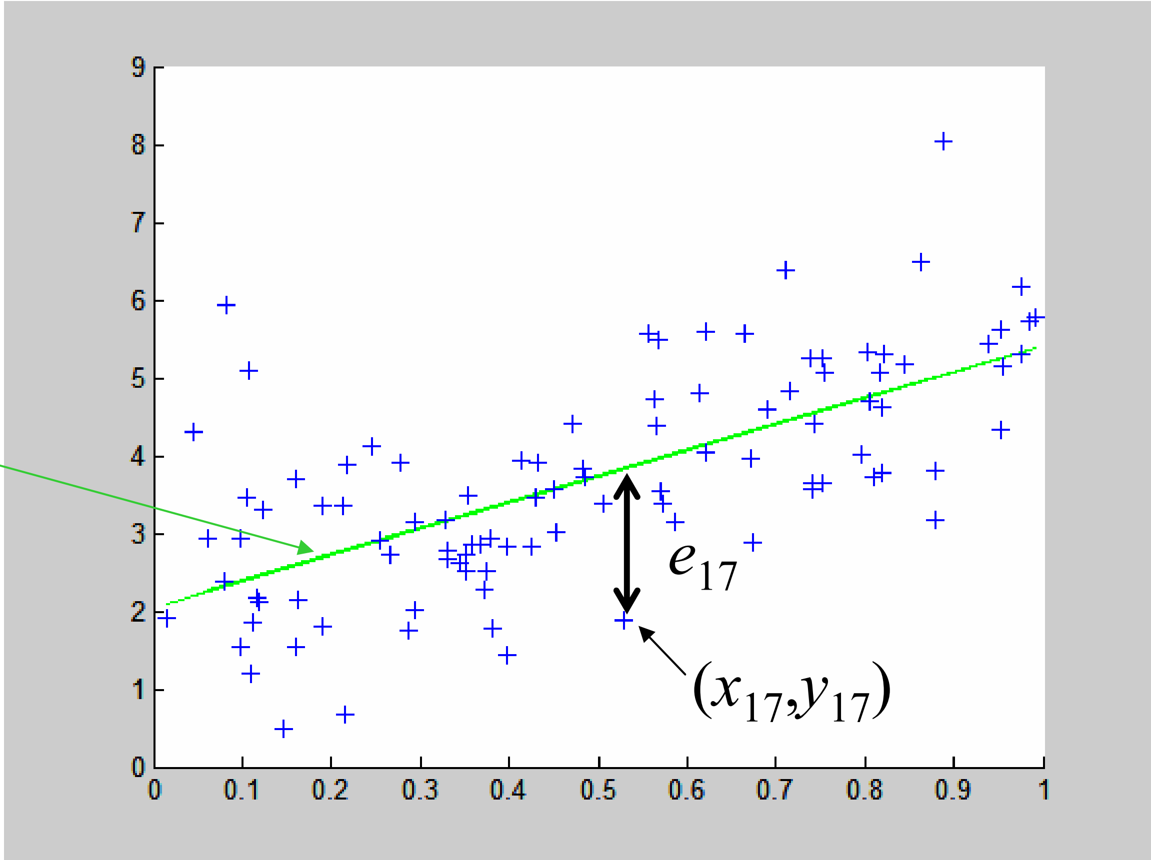
Given a set of n data points (x,y) pairs

There exists a unique line $\hat{y} = a + bx$

that minimizes the *residual sum of squares*

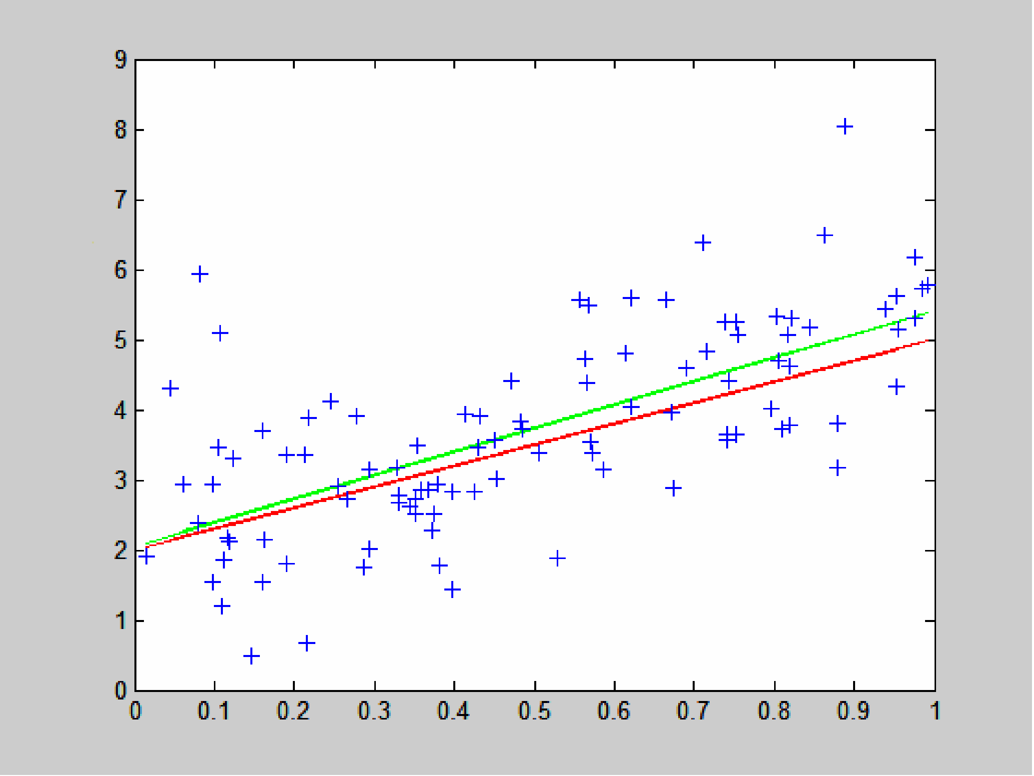
$$\sum_{i=1}^{n} e_i^{\ 2} \qquad e_i = y_i - \hat{y}_i$$

$$s_e^{\ 2} = \frac{1}{n-2}\sum_{i=1}^{n} e_i^{\ 2}$$



Matlab Code for Regression

```
p = polyfit(x,Y,1)
y_hat=polyval(p,x);
plot(x,y_hat,'-','Color', 'g')
```

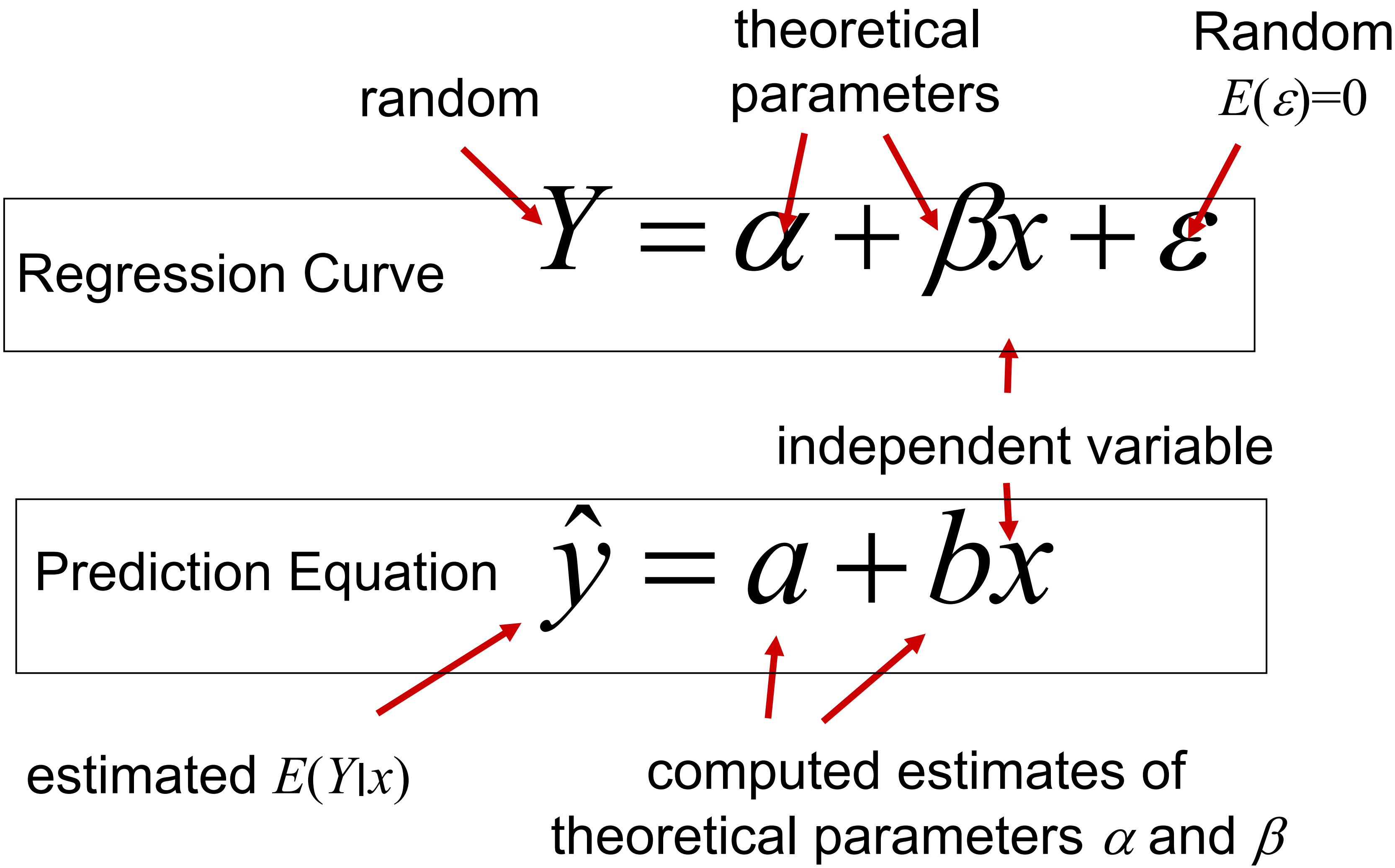


Concept Question

You are seeking to calibrate a load cell. You wish to determine the regression line relating voltage (in Volts) to force (in Newtons). What are the units of a, b, S_{xx}, and S_{xy} respectively?

1) N, N, N, and N
2) V, V, V^2, and V^2
3) V, V/N, N^2, and VN
4) V/N, N, VN, and V^2
5) None of the variables have units

Regression Curve vs Prediction Equation



Evaporation vs Air Velocity Hypothesis Tests

Air vel (cm/sec)	Evap coeff. (mm^2/sec)
20	0.18
60	0.37
100	0.35
140	0.78
180	0.56
220	0.75
260	1.18
300	1.36
340	1.17
380	1.65

Air vel (cm	Evap coeff. (mm2/sec)
20	0.18
60	0.37
100	0.35
140	0.78
180	0.56
220	0.75
260	1.18
300	1.36
340	1.17

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.934165
R Square	0.872665
Adjusted R Square	0.854474
Standard Error	0.159551
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regression	1	1.221227	1.221227	47.97306	0.000226
Residual	7	0.178196	0.025457		
Total	8	1.399422			

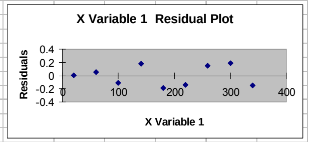
	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	0.102444	0.106865	0.958637	0.369673	-0.15025	0.355139	-0.15025	0.355139
X Variable 1	0.003567	0.000515	6.926259	0.000226	0.002349	0.004784	0.002349	0.004784

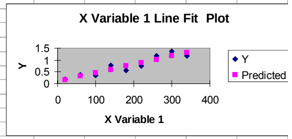
RESIDUAL OUTPUT

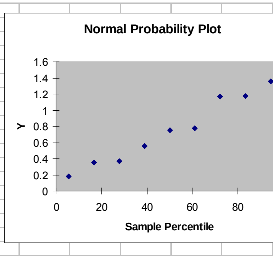
Observation	*Predicted Y*	*Residuals*	*dard Residuals*
1	0.173778	0.006222	0.041691
2	0.316444	0.053556	0.35884
3	0.459111	-0.10911	-0.73108
4	0.601778	0.178222	1.194149
5	0.744444	-0.18444	-1.23584
6	0.887111	-0.13711	-0.91869
7	1.029778	0.150222	1.006539
8	1.172444	0.187556	1.256685
9	1.315111	-0.14511	-0.97229

PROBABILITY OUTPUT

Percentile	*Y*
5.555556	0.18
16.66667	0.35
27.77778	0.37
38.88889	0.56
50	0.75
61.11111	0.78
72.22222	1.17
83.33333	1.18
94.44444	1.36







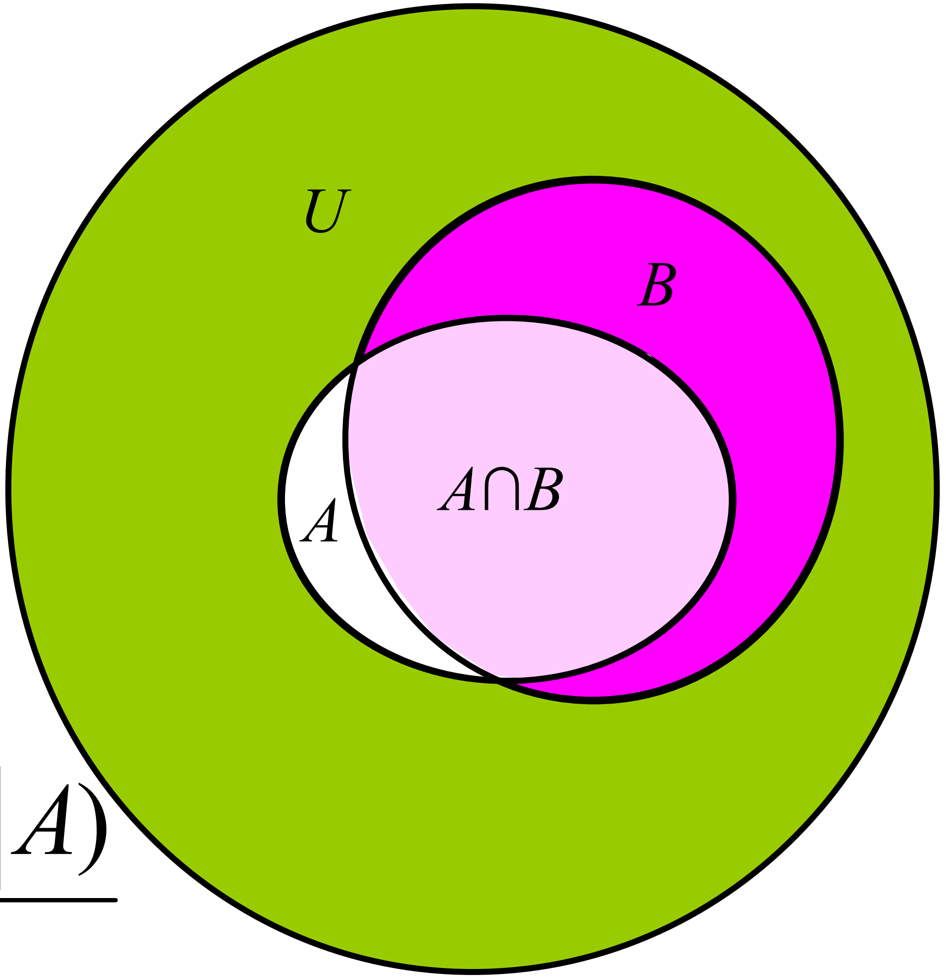
Bayes' Theorem

$$\Pr(A|B) \equiv \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(B|A) \equiv \frac{\Pr(A \cap B)}{\Pr(A)}$$

with a bit of algebra

$$\Pr(A|B) = \frac{\Pr(A)\Pr(B|A)}{\Pr(B)}$$



False Discovery Rates

Image removed due to copyright restrictions.
Source: Figure 2 in Efron, Bradley. "Modern Science and the Bayesian-Frequentist Controversy."
http://www-stat.stanford.edu/~brad/papers/NEW-ModSci_2005.pdf

Single Factor Experiments

- A single **experimental factor** is varied
- The parameter takes on various **levels**

Cotton weight percentage	Observations				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

$a=5$ replicates

Each cell is a y_{ij}

Each row is a treatment i

↑ experimental factor

Fiber strength in lb/in^2

Breakdown of Sum Squares

"Grand Total Sum of Squares"

$$GTSS = \sum_{i=1}^{a}\sum_{j=1}^{n} y_{ij}^{2}$$

"Total Sum of Squares"

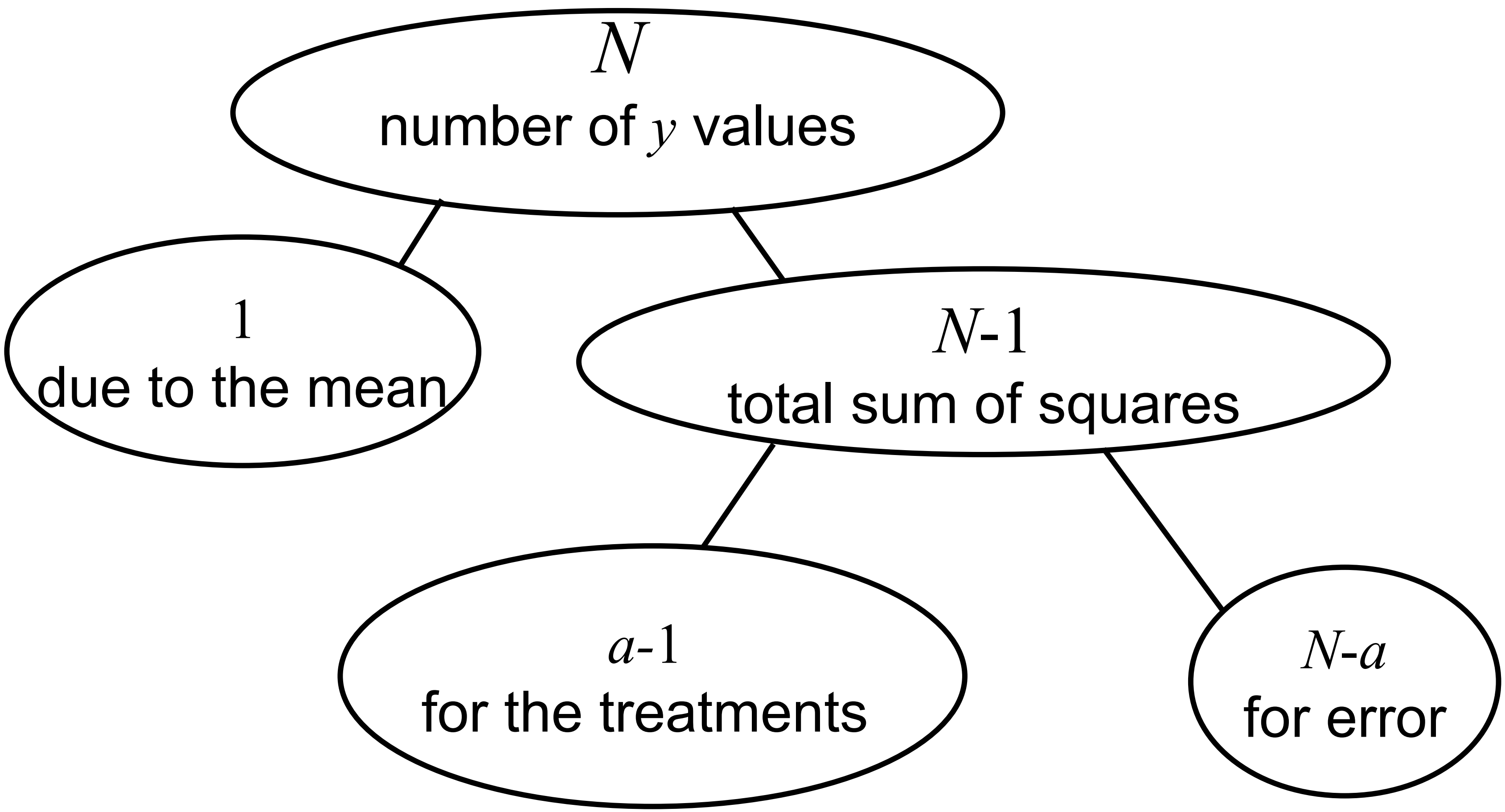
$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^{2}$$

SS due to mean $= n\bar{y}_{..}^{2}$

$$SS_{Treatments} = n\sum_{i=1}^{a} (\bar{y}_{i.} - \bar{y}_{..})^{2}$$

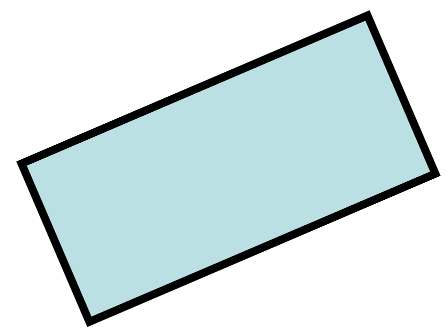
$$SS_E$$

Breakdown of DOF

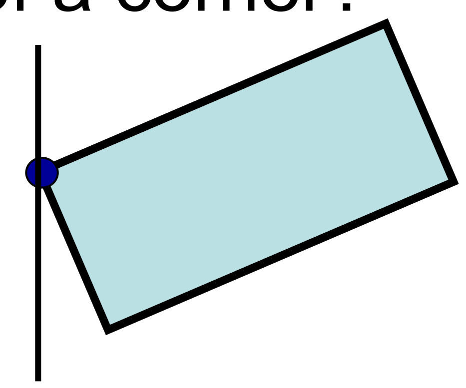


What is a "Degree of Freedom"?

- How many scalar values are needed to unambiguously describe the state of this object?



- What if I were to fix the x position of a corner?



What is a "Degree of Freedom"?

- How many scalar values are needed to unambiguously describe the outcome o this experiment?

Cotton weight percentage	Observations				
	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

- What if I were to tell you $\bar{y}_{..}$?
- What if I were to tell you $\bar{y}_{i.}$ $i = 1...4$?

Adding h.o.t. to the Model Equation

Each row of **X** is paired with an observation

You can add interactions

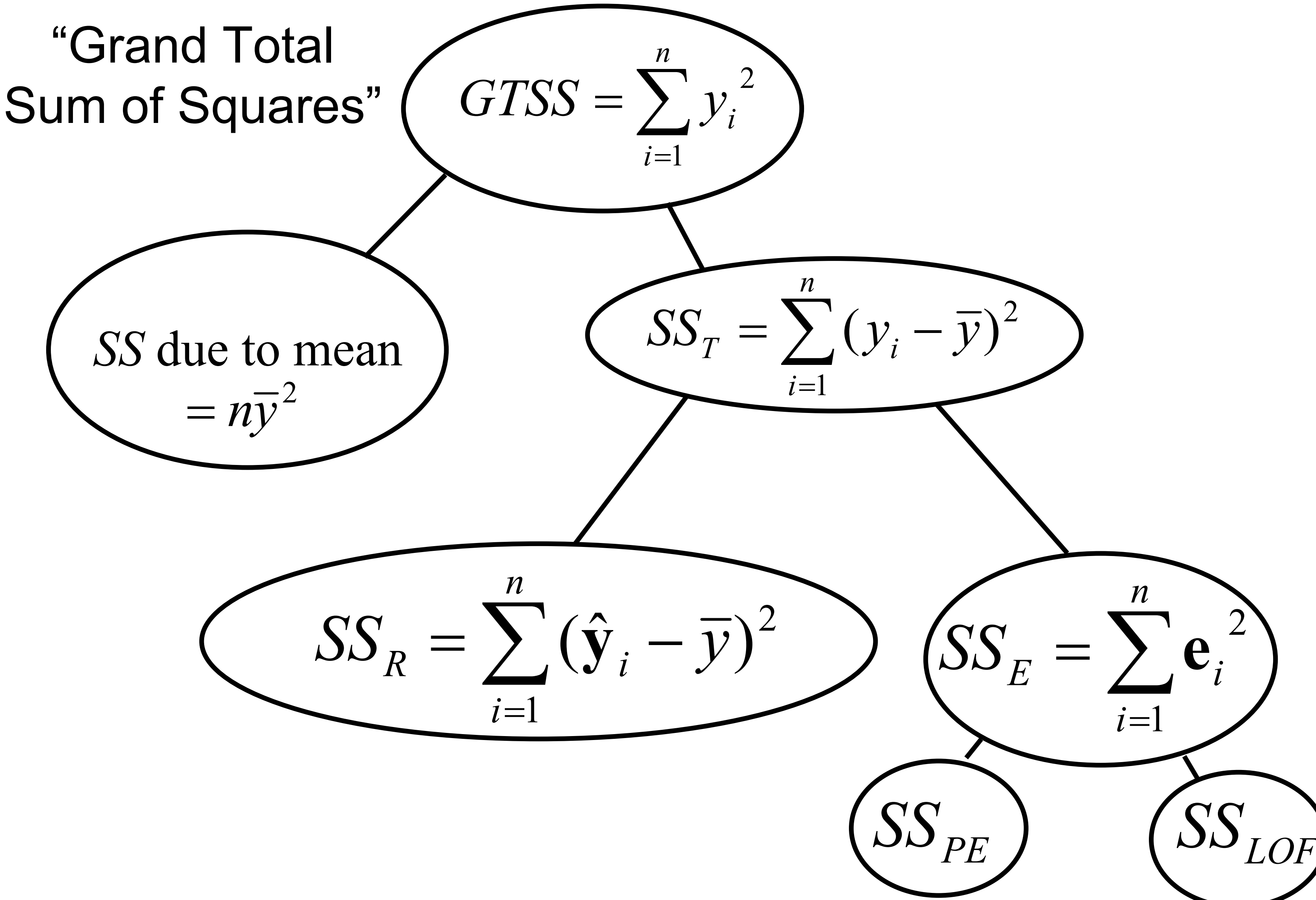
You can add curvature

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{11}x_{12} & x_{11}^{\;2} \\ 1 & x_{21} & x_{22} & x_{21}x_{22} & x_{21}^{\;2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & x_{n1}x_{n2} & x_{n1}^{\;2} \end{bmatrix}$$

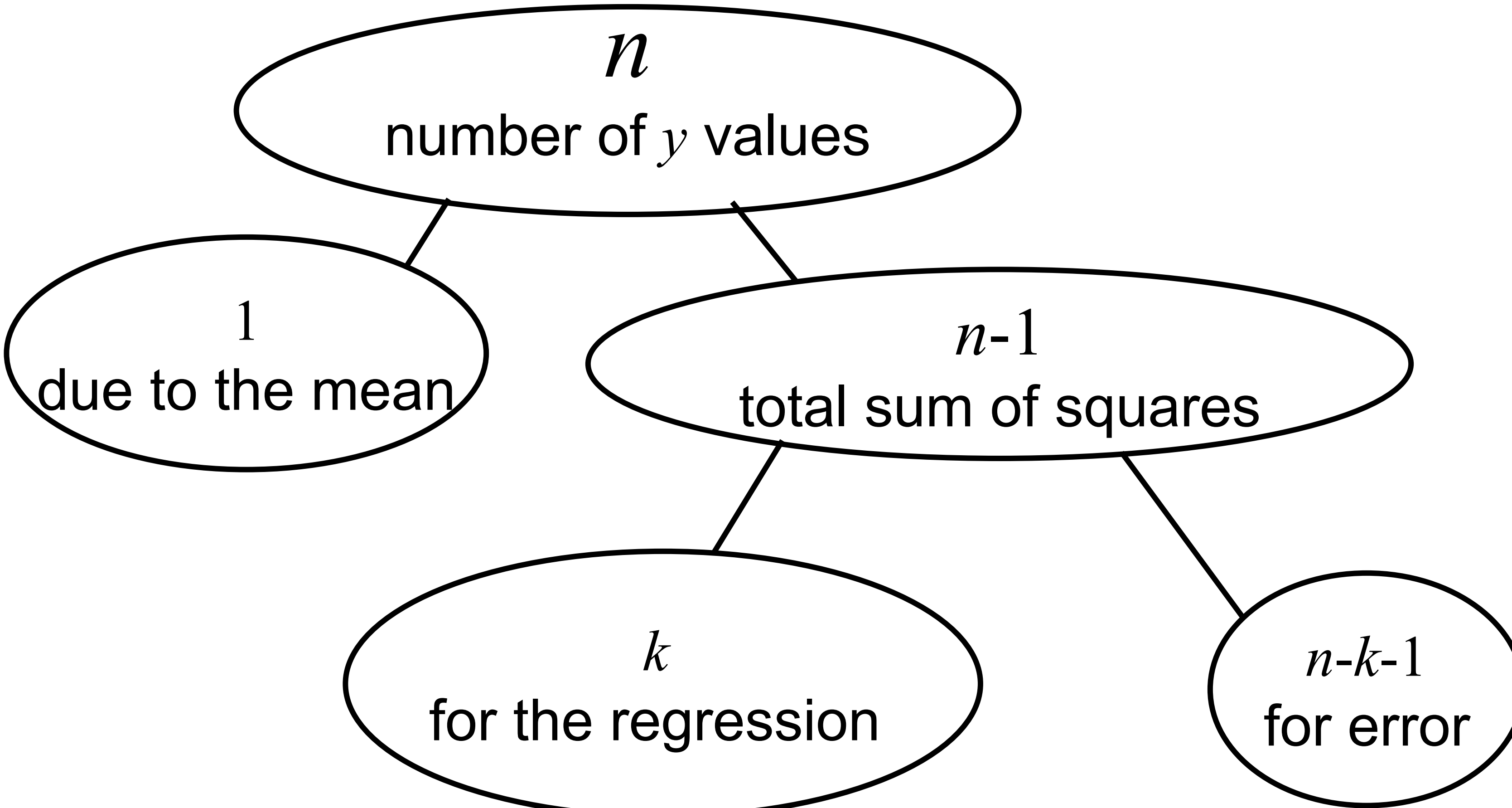
There are n observations of the response

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_{12} & \beta_{11} \end{bmatrix}$$

Breakdown of Sum Squares



Breakdown of DOF



Estimation of the Error Variance σ^2

$\boldsymbol{\varepsilon} \sim N(0, \sigma)$

Remember the the model equation $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

If assumptions of the model equation hold, then

$$E\left(SS_E/(n-k-1)\right) = \sigma^2$$

So an unbiased estimate of σ^2 is

$$\hat{\sigma}^2 = SS_E/(n-k-1)$$

Test for Significance Individual Coefficients

The hypotheses are $H_0 : \beta_j = 0$

$H_1 : \beta_j \neq 0$

The test statistic is

$$t_0 = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 C_{jj}}}$$

$$C = \left(\mathbf{X}^T \mathbf{X}\right)^{-1}$$

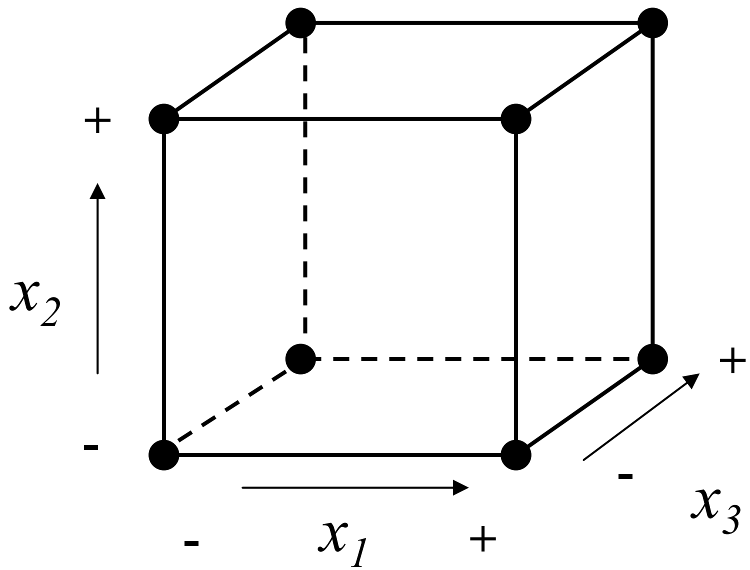
Standard error

$$\sqrt{\hat{\sigma}^2 C_{jj}}$$

Reject H_0 if $|t_0| > t_{\alpha/2, n-k-1}$

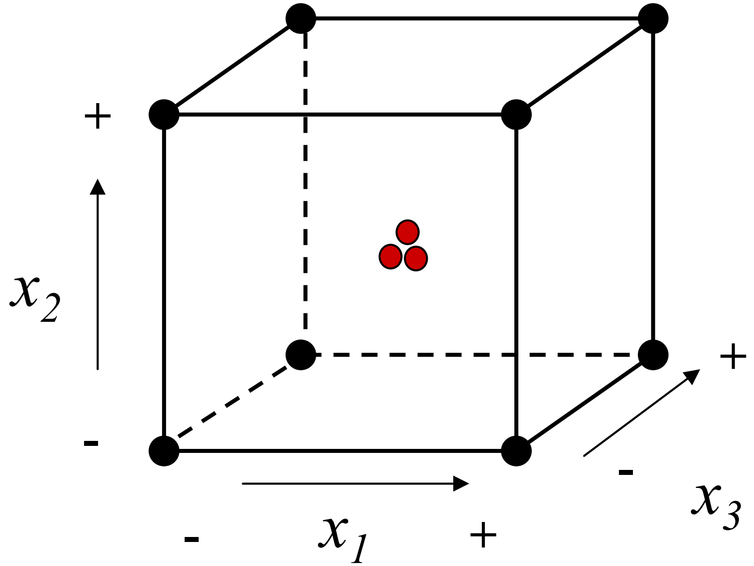
Factorial Experiments

Cuboidal Representation



Exhaustive search of the space of discrete 2-level factors is the full factorial 2^3 experimental design

Adding Center Points



Center points allow an experimenter to check for curvature and, if replicated, allow for an estimate of pure experimental error

Concept Test

- You perform a linear regression of 100 data points (n=100). There are two independent variables x_1 and x_2. The regression R^2 is 0.72. Both β_1 and β_2 pass a t test for significance. You decide to add the interaction x_1x_2 to the model. Select all the things that cannot happen:

1) Absolute value of β_1 decreases

2) β_1 changes sign

3) R^2 decreases

4) β_1 fails the t test for significance

Screening Design

Image removed due to copyright restrictions.
TABLE 2: Design I: Layout and Data for 2_{IV}^{8-4} Screening Design in Box and Liu, 1999.

- What is the objective of screening?
- What is special about this matrix of 1s and -1s?

Analysis of Variance

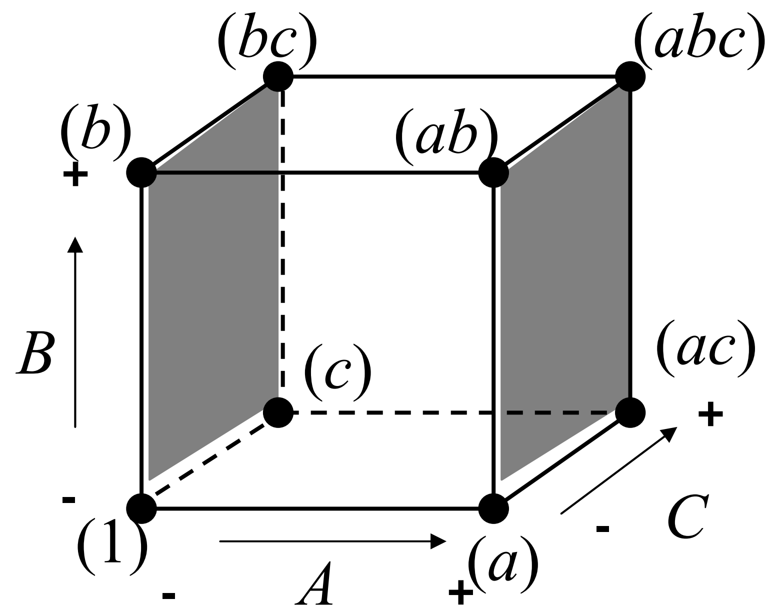
Image removed due to copyright restrictions.
TABLE 10: Design III: Analysis of Variance for Completed Composite Design in Box and Liu, 1999.

- What would you conclude about lack of fit?
- What is being used as the denominator of F?

Concept Question

Say the independent experimental error of observations (a), (ab), et cetera is σ_{ε}.

We define the main effect estimate A to be

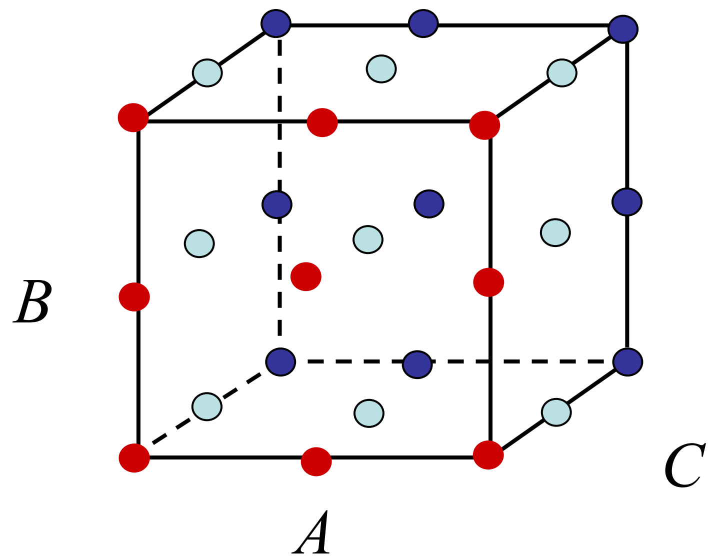


$$A \equiv \frac{1}{4}\left[(abc)+(ab)+(ac)+(a)-(b)-(c)-(bc)-(1)\right]$$

What is the standard deviation of the main effect estimate A?

1) $\sigma_A = \frac{1}{2}\sqrt{2}\sigma_{\varepsilon}$ 2) $\sigma_A = \frac{1}{4}\sigma_{\varepsilon}$ 3) $\sigma_A = \sqrt{8}\sigma_{\varepsilon}$ 4) $\sigma_A = \sigma_{\varepsilon}$

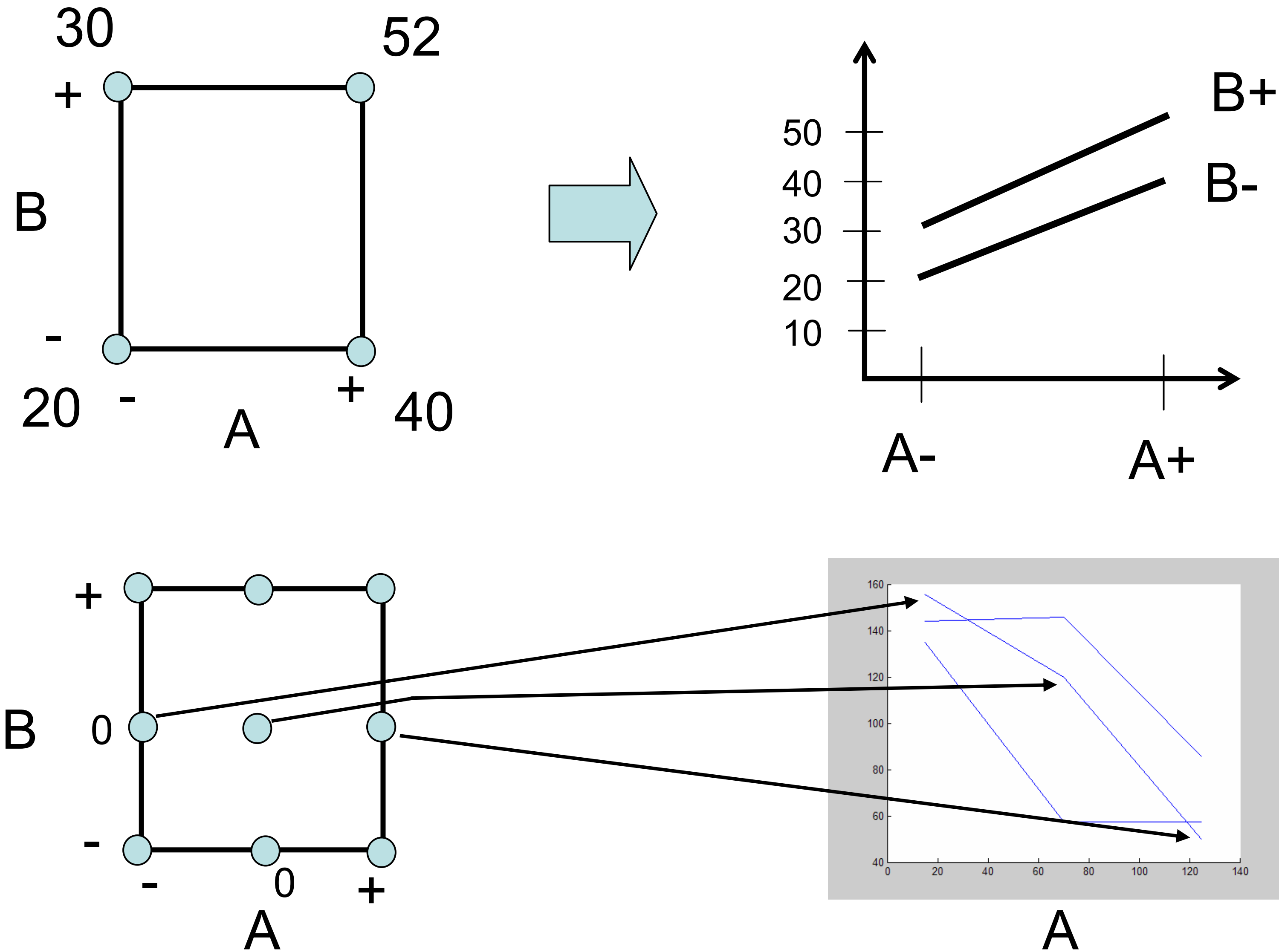
Three Level Factors



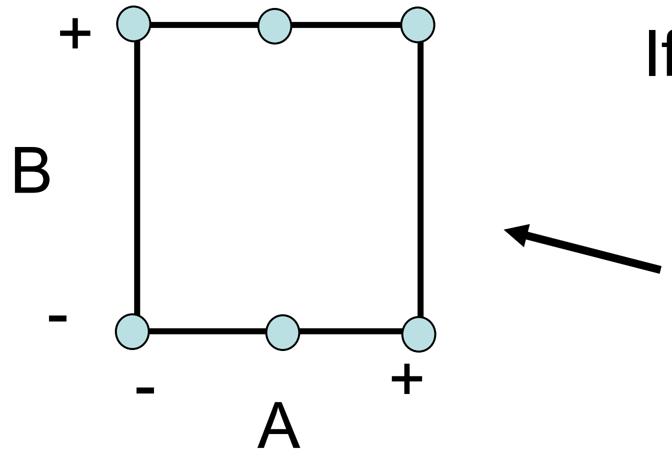
8 vertices +
12 edges +
6 faces +
1 center =
27 points

3^3 Design

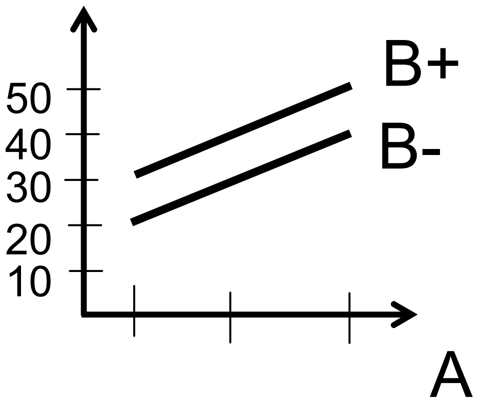
Factor Effect Plots



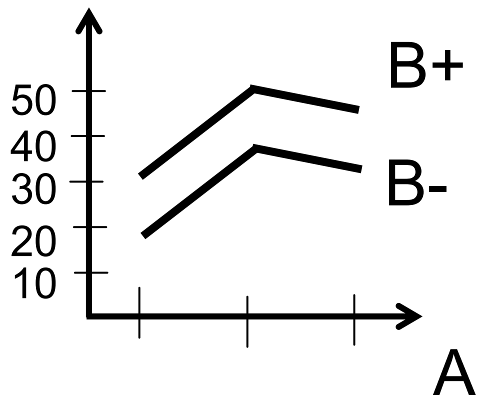
Concept Test



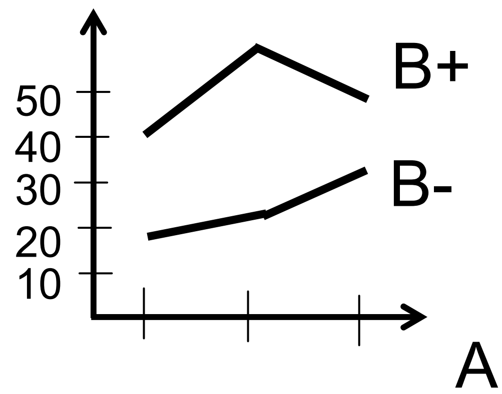
If there are no interactions in this system, then the factor effect plot from this design could look like:



1



2



3

Hold up all cards that apply.

Concept Test

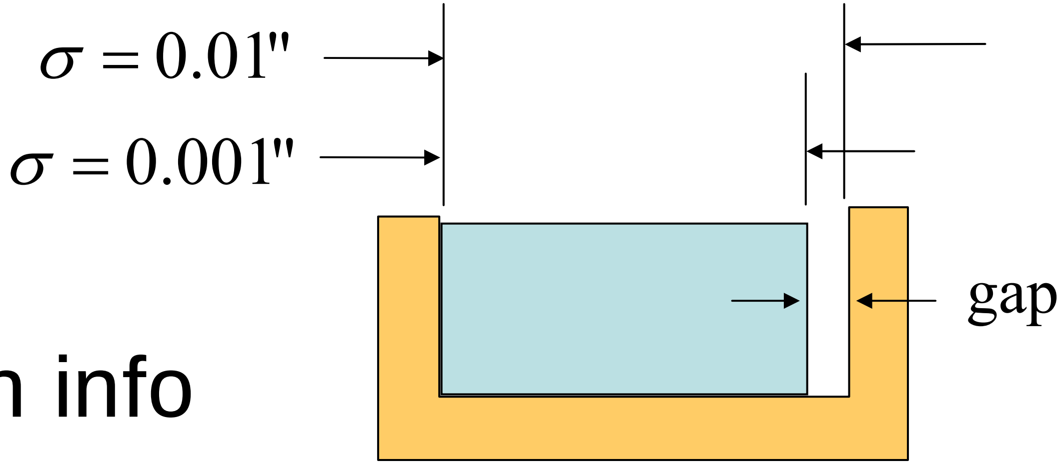
- A bracket holds a component as shown. The dimensions are strongly correlated random variables with standard deviations as noted. Approximately what is the standard deviation of the gap?

A) 0.011”

B) 0.01”

C) 0.009”

D) not enough info



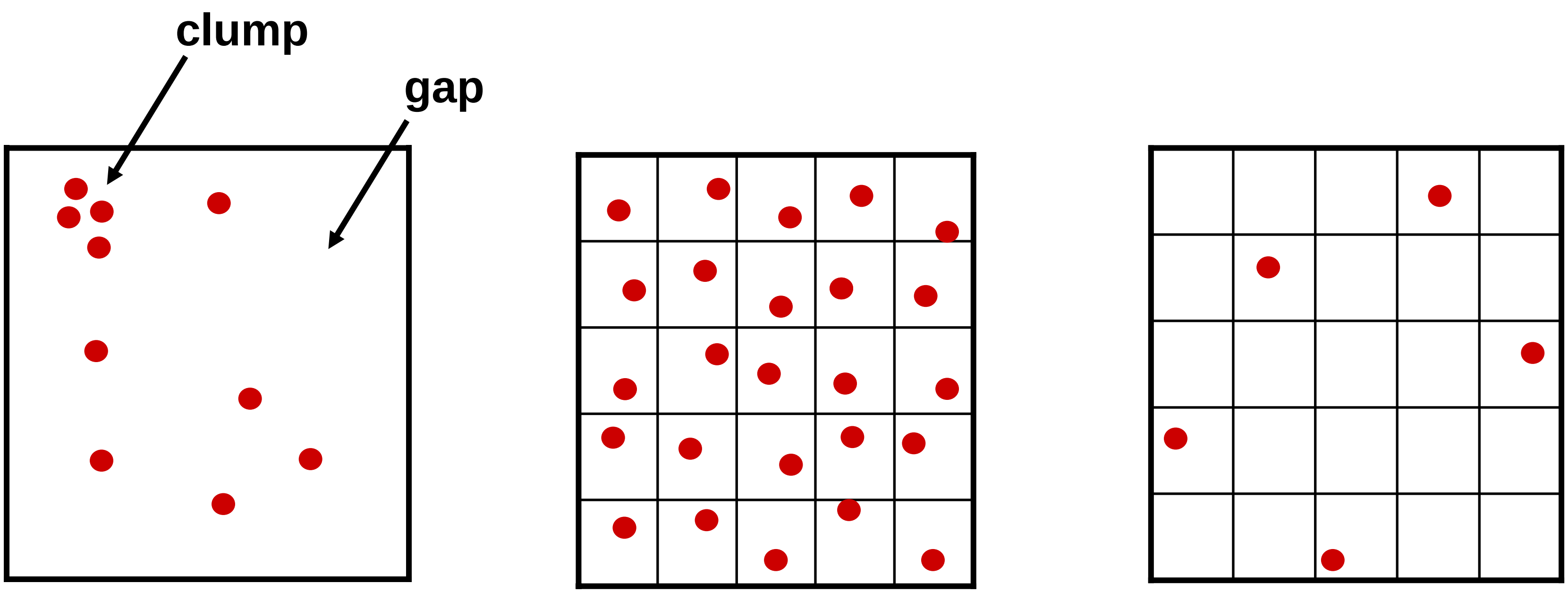
Monte Carlo Simulations
What are They Good at?

$$\text{Accuracy} \propto 1/\sqrt{N} \qquad N \equiv \#\text{Trials}$$

- Above formulae apply regardless of dimension
- So, Monte Carlo is good for:
  - Rough approximations <u>or</u>
  - Simulations that run quickly
  - <u>Even if</u> the system has many random variables

Fishman, George S., 1996, *Monte Carlo: Concepts, Algorithms, and Application*s, Springer.

Sampling Techniques for Computer Experiments



Random Sampling

Stratified Sampling

Latin Hypercube Sampling

Errors in Scientific Software

- Experiment T1
  - Statically measured errors in code
  - Cases drawn from many industries
  - ~10 serious faults per 1000 lines of commercially available code
- Experiment T2
  - Several independent implementations of the same code on the same input data
  - One application studied in depth (seismic data processing)
  - Agreement of 1 or 2 significant figures on average

Hatton, Les, 1997, "The T Experiments: Errors in Scientific Software", *IEEE Computational Science and Engineering*.

Next Steps

- Monday 7 May – Frey at NSF
- Wednesday 9 May – Exam #2
- Wed and Fri, May 14 and 16
  - Final project presentations