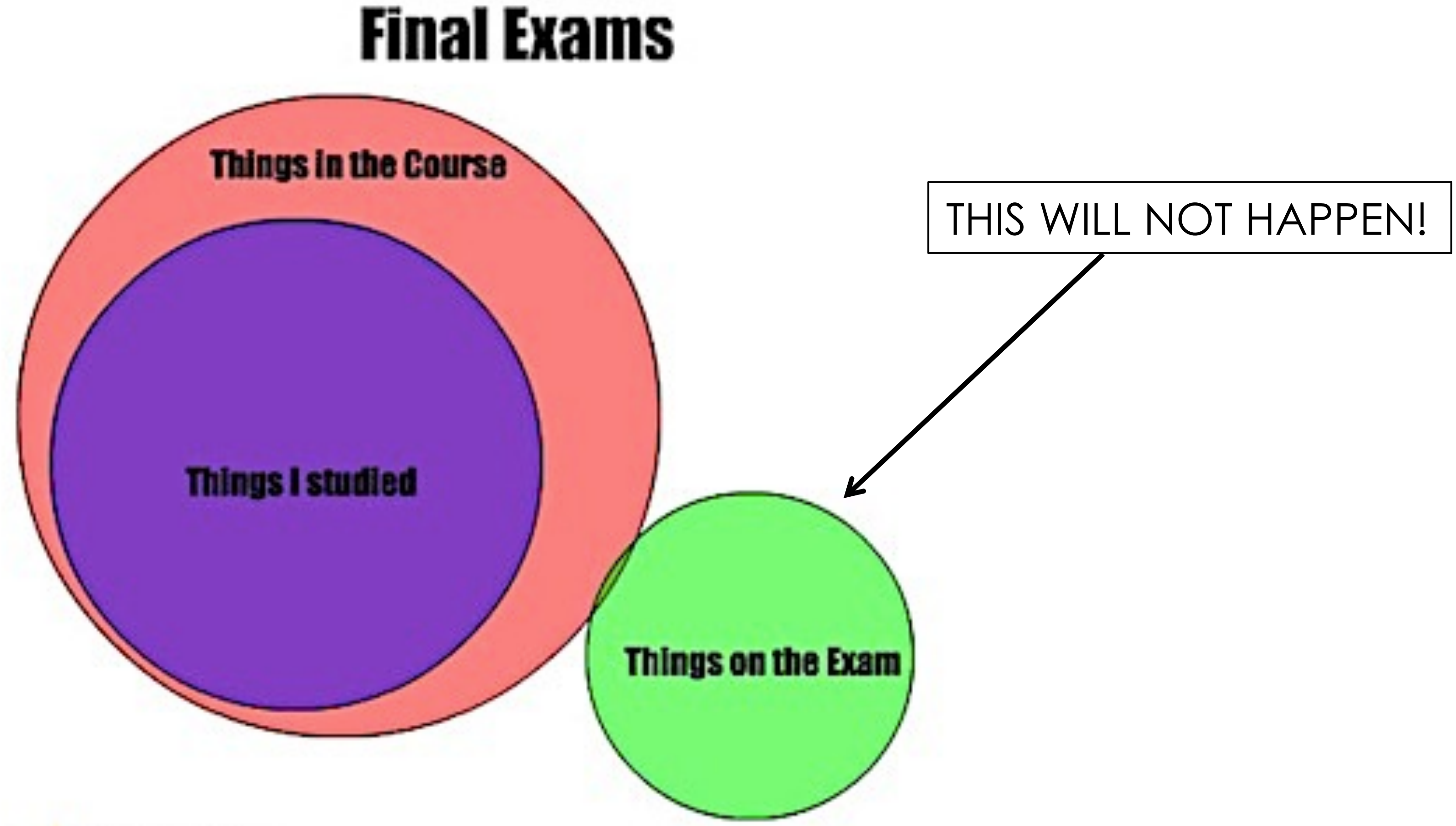
Our Goal Today!



Final Exam Logistics

- **Date:** Friday, June 6 – Sunday, June 8
- **Time:** Any 8 hour period from 8:00AM on Friday to 8:00PM on Sunday
- **Place:** Will be available electronically
- **Open book exam**: Allow the use notes, lectures, recitations, homework
- **ABSOLUTELY NO COLLABORATION! Violation of this will result in a failing grade for the course.**
- Formula Sheet provided on Stellar

Outline

- ***Topic 1 :*** *Probability Theory*
- ***Topic 2 :*** *Decision Analysis*
- **Topic 3 :** Discrete Random Variables
- **Topic 4 :** Continuous Random Variables
- **Topic 5 :** Covariance and Correlation
- **Topic 6 :** Regression
- **Topic 7 :** Linear Optimization
- **Topic 8 :** Discrete Optimization
- **Topic 9 :** Nonlinear Optimization

Random Variables

- Random variable: a function that assigns a numerical value to every possible outcome of an experiment

Example: Coin toss

If Heads: X=1 If Tails: X=0

Probability Laws

- What rules does this function follow?
  1. The probability of each outcome must be non-negative ($p \geq 0$) The *largest* possible probability is 1: this tells us that we are 100% certain that this outcome will occur
  2. MOST IMPORTANTLY: the sum of the probabilities of **all possible** outcomes **is always = 1**
     - If all outcomes are: $x_1, x_2, x_3, \ldots, x_n$
     - Then the sum: $P(X = x_1) + P(X = x_2) + \ldots + P(X = x_n) =$

       $$= p_1 + p_2 + \ldots + p_n = 1$$
  3. Mutually exclusive events: If you have a set of non-overlapping events, then the probability of this set is just the sum of the individual event probabilities

Mean (Expected Value)

- Mean: the "**weighted**" average outcome
  - Formula:

$$\boxed{E[X] = \mu_x = \sum_{i=1}^{n} x_i \cdot P(X = x_i) = \sum_{i=1}^{n} x_i p_i}$$

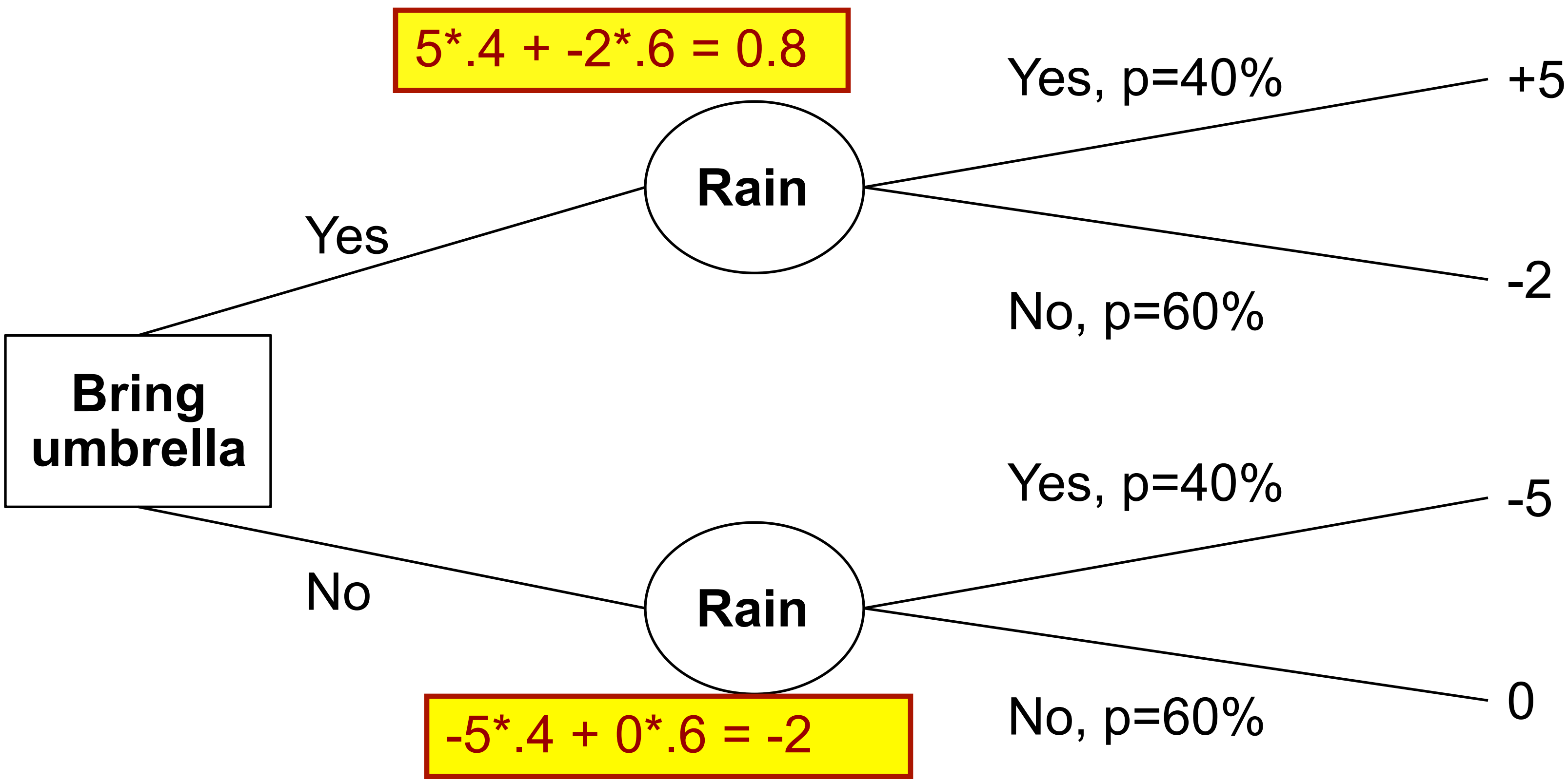
- Average: the center of a set of numbers (all equally likely)
  - Formula:

$$\boxed{A = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

- <u>THEY ARE ONLY EQUAL IF ALL PROBABILITIES p_i ARE THE SAME!!</u>

Decision Trees

Should I bring an umbrella today in case it rains?



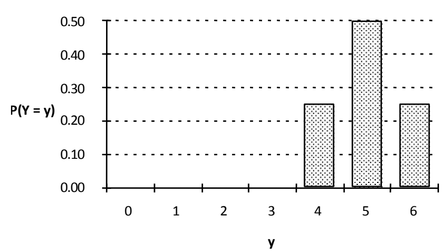
Since 0.8 > -2, I should bring an umbrella!!!

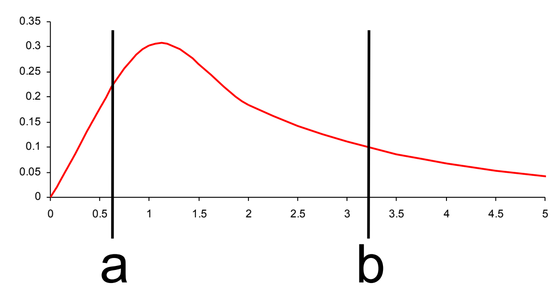
Outline

- **Topic 1 :** Probability Theory
- **Topic 2 :** Decision Analysis
- ***Topic 3 :*** *Discrete Random Variables*
- ***Topic 4 :*** *Continuous Random Variables*
- ***Topic 5 :*** *Covariance and Correlation*
- **Topic 6 :** Regression
- **Topic 7 :** Linear Optimization
- **Topic 8 :** Discrete Optimization
- **Topic 9 :** Nonlinear Optimization

RVs: Discrete v. Continuous

	Discrete	Continuous
Examples	# from rolling a die # of stock prices which will go up tomorrow	Distance die rolls Amount stock prices go up tomorrow
Values	$x_1, x_2, \ldots, x_n$	range
Probabilities	$p_1, p_2, \ldots, p_n$ $P(X = x_i) = p_i$ $\sum p_i = 1$	$P(b \le X \le a)$, $P(b \le X)$, $P(X \ge a)$, etc.
Graphic	Histogram: $P(X = x_i) = p_i$	Probability Density Function: $P(b \le X \le a)$ = area under curve between a and b





RVs: Important Summary Statistics

	Meaning	Equation (Discrete)
Mean	μ Expected value	$E(X) = \mu_X = \sum_i p_i * x_i$
Variance	σ^2 Measure of spread around mean (units^2)	$Var(X) = \sigma_X^2 = \sum_i p_i * (x_i - \mu_x)^2$
Standard Deviation	σ Measure of spread around mean (units)	$\sigma_X = \sqrt{Var(X)}$
Coefficient of variation	**CV** Unitless measure of spread	$CV_X = \sigma_X / \mu_X$

RVs: Binomial Distribution (Discrete)

- There were three T/F questions on the Econ Exam, you have a 70% chance of getting each of them right, what is the probability you get all three correct? What is the mean number of questions you get right? The variance?

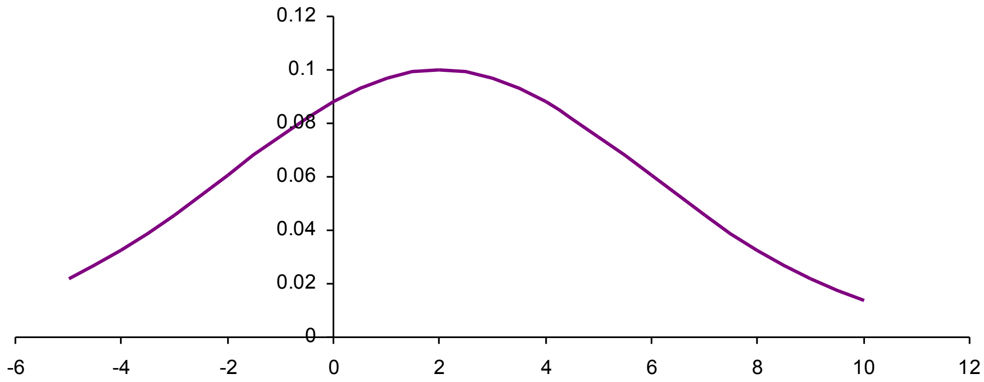
- **Binomial!**
  - n independent trials, each with a success rate of p
    X ~ the total number of successes
    $\mu = np$, $\sigma^2 = np*(1-p)$

- Remember: $$P(X = x) = \frac{n!}{x!(n-x)!} * p^x (1-p)^{n-x}$$

- So $$P(X = 3) = \frac{3!}{3!(3-3)!} * .7^3 (1-.7)^{3-3} = .7^3 = .34$$

- And $\mu = 3*.7 = 2.1$, $\sigma^2 = 3*.7*.3 = .62$

RVs: Normal Distribution (Continuous)



Bell-shaped curve

Computing probabilities with the Normal distribution:

You want : P(a ≤ X ≤ b) where X is N(μ,σ)

1. Define : $\boxed{Z = \dfrac{X-\mu}{\sigma}}$: **Z is N(0,1)**

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$$

2. Use the standard normal probability table (**Z table**)

Multiple Random Variables

	Meaning	Equation
Covariance	Measures the extent to which two variables vary together (units^2)	$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$ *Discrete:* $= \sum_{i,j} P(X = x_i; Y = y_j)(x_i - \mu_X)(y_j - \mu_Y)$
Correlation	Standardized Covariance, [-1,1]	$\mathrm{CORR(X, Y)} = \frac{\mathrm{COV(X, Y)}}{\sigma_X \sigma_Y}$

2005 Exam: Problem 1 (d)

d) If X has mean 1, standard deviation 2 and Y has mean 1, standard deviation 4, then the standard deviation of Z=X+Y cannot exceed 6.

$\mathrm{Var}(Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2*\sigma_X*\sigma_Y*\mathrm{CORR}(X,Y)$

Maximum value achieved when CORR(X,Y) = 1

$\mathrm{Var}(Z) = 36$, $\sigma_Z = 6$

Sums of i.i.d RVs: Central Limit Theorem

$X_1, X_2, \ldots, X_n$ **independent identically distributed** random variables:

$$E[X_i] = \mu, \mathrm{Var}(X_i) = \sigma^2$$

- For n>30, $S_n = X_1 + X_2 + \ldots + X_n$ is approximately normal with mean n.μ and variance n.σ^2 (standard deviation $\sqrt{n}.\sigma$)
- For n>30, $M_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ is approximately normal with mean μ and variance σ^2/n (standard deviation $\sigma/\sqrt{n}$)

- The probability distribution of X_i does not matter;
- n does not have to be very large (30 is good enough);
- CLT requires only 2 pieces of information: the mean and SD of X_i

Silverware Example

- SilverwareInc has 3 products: Forks, Knives, and Spoons. They have kept track of sales, but only in aggregate form. They have found that the mean of yearly Silverware revenues is \$3M, and the standard deviation of the yearly revenues is \$1M. Revenues are Normally Distributed

- What can you say about the distribution of the yearly revenues?
  - $W = F+K+S \sim N(\mu=3,\ \sigma=1)$, in \$M

Silverware Example

- SilverwareInc is selling world-class Forks, Knives, and Spoons. The mean and standard deviation of the number of daily sets sold are recorded in the table below, together with the price.

	Forks	Knives	Spoons
Mean	15.6	17.2	22.4
St Dev	8.3	7.2	8.5
Price ($)	10.0	10.0	11.0

- What can you say about the distribution of yearly sales of forks?
  - $F \sim N(\mu=365*15.6,\ \sigma=8.3*\sqrt{365})$

Silverware Example

- What can you say about the distribution of yearly sales of forks?
  - F ~ N(μ=365*15.6, σ=8.3*√365)

- What is the probability that SilverwareInc' s revenues from forks is greater than 6,000 units this year?

$$P(W >= 6000) = P(\frac{W-\mu}{\sigma} >= \frac{6000-\mu}{\sigma})$$

$$= P(Z >= \frac{6000-5694}{158.57}) = P(Z >= 1.92)$$

$$= 1 - P(Z <= 1.92)$$

Outline

- **Topic 1 :** Probability Theory
- **Topic 2 :** Decision Analysis
- **Topic 3 :** Discrete Random Variables
- **Topic 4 :** Continuous Random Variables
- **Topic 5 :** Covariance and Correlation
- ***Topic 6 :*** *Regression*
- **Topic 7 :** Linear Optimization
- **Topic 8 :** Discrete Optimization
- **Topic 9 :** Nonlinear Optimization

Linear Regression

- Predict a dependent variable **Y** from a set of independent variables (predictors) $\mathbf{X_i}$

- Assume Y is a "straight-line" function of each of the X_i variables, when all the others are fixed.
  - The slopes of each X_i's relationship with Y are coefficients β_i

- **Goal:** Pick coefficients to minimize errors (sum of squared residuals)

$$\hat{y}_i = b_0 + b_1 x_{1i} + \ldots + b_k x_{ki}$$

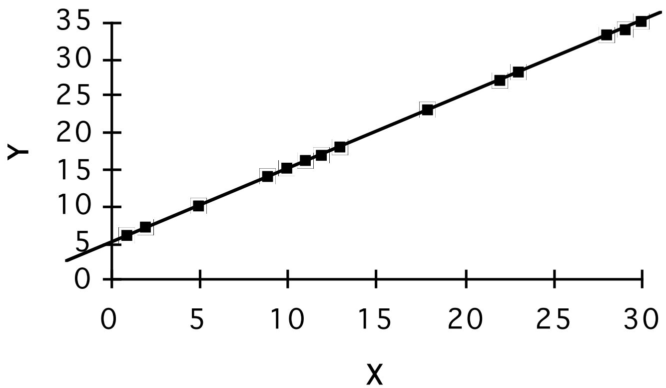
- The steps in which we accomplish this goal are as follows:
  - Calculate the model parameters that "best agree" with the available data (β_k).
  - Analyze the model and estimates and verify that they are reasonable.
  - Potentially revise initial model based on this analysis, and repeat.

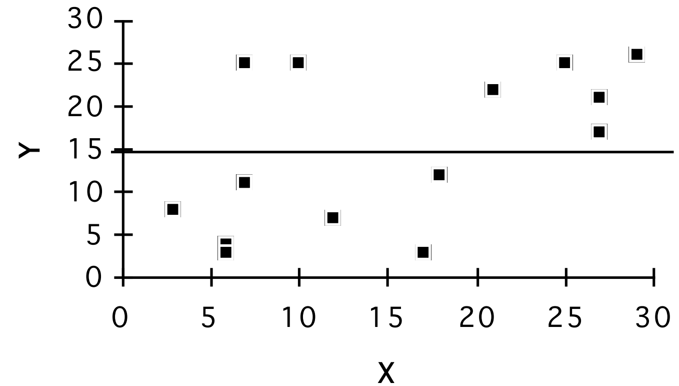
Linear Regression

- **Residuals** (errors), defined as: $e_i = y_i - \hat{y}_i$
- So we're minimizing SSE $$SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
- **R^2**: Coefficient of determination

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

R^2 takes values between 0 and 1:





R^2: is the explained variation (by the model) divided by the total variation = the percent of variation that can be explained by the regression equation

Linear Regression

Confidence Intervals

- A confidence interval tells us a range of values in which we expect an UNKNOWN parameter to lie (in our case, each of the β_k).

- This intervals depends on the data sample (x_{k1}, x_{k2}, …, x_{km})!!!!! It depends on the mean and variance of this data set. If we had different data, we would get a different interval!

- A 95% confidence interval DOES NOT tell us that we have 95% likelihood of correctly estimating β_k.

- A 95% confidence interval DOES tell us that if we constructed this interval from 100 different data samples of size m, then 95 of these intervals would correctly contain the true (unknown) value of β_k.

Linear Regression

Regression model is good?

- R^2 closer to 1. (Although good value depends on situation.)
- Estimated coefficients make sense
    - Sign? Magnitude?
- Confidence Interval
    - If zero does not lie in the confidence interval**:** we are confident at the α% level that β_m is different from 0
    - If zero lies in the confidence interval**:** we should be skeptical that Y depends linearly on x_m and we might want to eliminate x_m from the model.

Checklist for Evaluating a Linear Regression Model

- **Linearity:** Construct a scatter-plot for each explanatory variable check for linearity.
- **Signs of Regression Coefficients:** Check to see that the signs make intuitive sense.
- **Significance tests:** check if the regression coefficients are significantly different from zero (not included in the confidence interval).
- **R^2:** Check if the value of R^2 is reasonably high.
- **Normality:** Check that the residuals are approximately Normally distributed by constructing a histogram of residuals.
- **Heteroscedasticity:** Do error terms have constant standard deviation? Plot the residuals with the observed values of each of the explanatory variables.
- **Multicollinearity:** Are two explanatory variables correlated?

Signs: if regression coeffs have "wrong" sign or we find high R^2 but one or more of the regression coeffs is not significantly different from 0. Check correlation table. Remove one of the variables from the model if there is high correlation.

An Ice Cream Example

The fat content in a gallon of chocolate ice cream is believed to depend on **Cream, Chocolate and Sugar** according to:

Fat =A + B*Cream + C*Chocolate + D*Sugar

A multiple regression was run on data from **20 different batches** of chocolate ice cream:

R Square: 0.8433

Standard Error: 13.73

Observations: 20

Degrees of Freedom: 16

	Coefficients	**Standard Error**	**Lower 95%**	**Upper 95%**
Intercept	-8.94	19.95	-51.24	33.35
Cream (ounces)	0.93	0.12	0.67	1.18
Choc. (ounces)	2.07	0.60	0.80	3.34
Sugar (ounces)	2.47	1.33	- 0.34	5.29

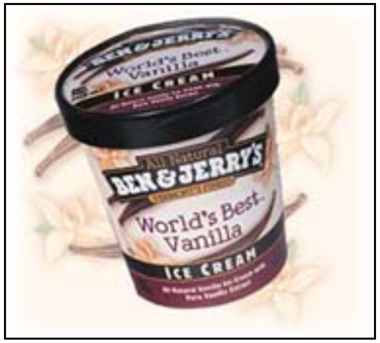
An Ice Cream Example

Correlation between different variables:

	Fat (gm)	Cream (ounces)	Choc. (ounces)	Sugar (ounces)
Fat (gm)	1			
Cream (ounces)	0.769	1		
Choc. (ounces)	0.486	0.025	1	
Sugar (ounces)	0.280	-0.099	0.409	1

- **Check the 95% CI for Choc. coefficient**

- **Critique model**



- **Critique model**

An Ice Cream Example

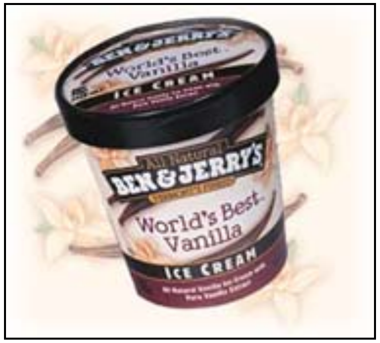
Signs of Regression Coefficients

	Coefficients	Standard Error	Lower 95%	Upper 95%
Intercept	-8.94	19.95	-51.24	33.35
Cream (ounces)	0.93	0.12	0.67	1.18
Choc. (ounces)	2.07	0.60	0.80	3.34
Sugar (ounces)	2.47	1.33	- 0.34	5.29

The coefficients for Cream, Choc and Sugar appear to make sense.

Significance test:

0 is in the confidence interval for Sugar coeff.
so Sugar should be excluded from the regression.



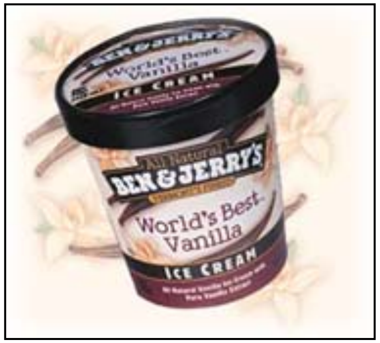
An Ice Cream Example

R^2 - The value for R^2 is 0.8433 which indicates that the model has a high level of prediction.

Multicollinearity:

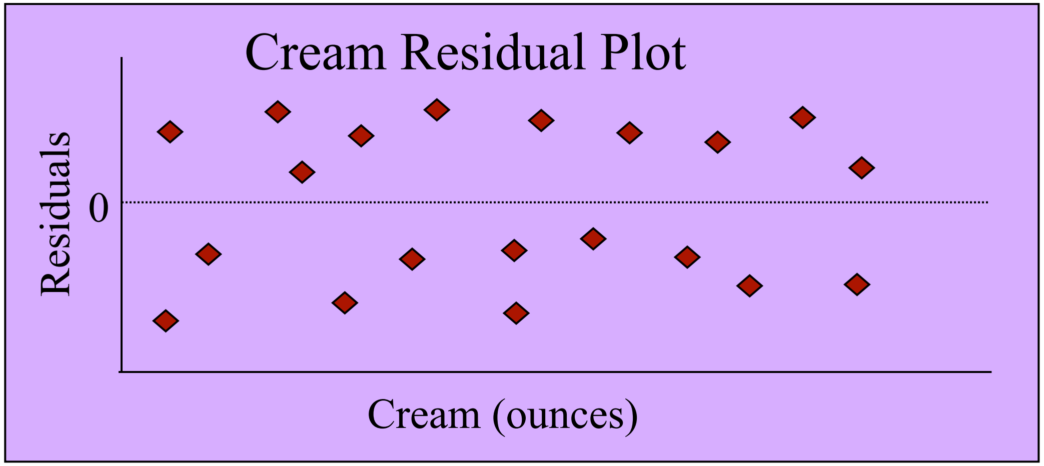
	Fat (gm)	Cream (ounces)	Choc. (ounces)	Sugar (ounces)
Fat (gm)	1			
Cream (ounces)	0.769	1		
Choc. (ounces)	0.486	0.025	1	
Sugar (ounces)	0.280	-0.099	0.409	1

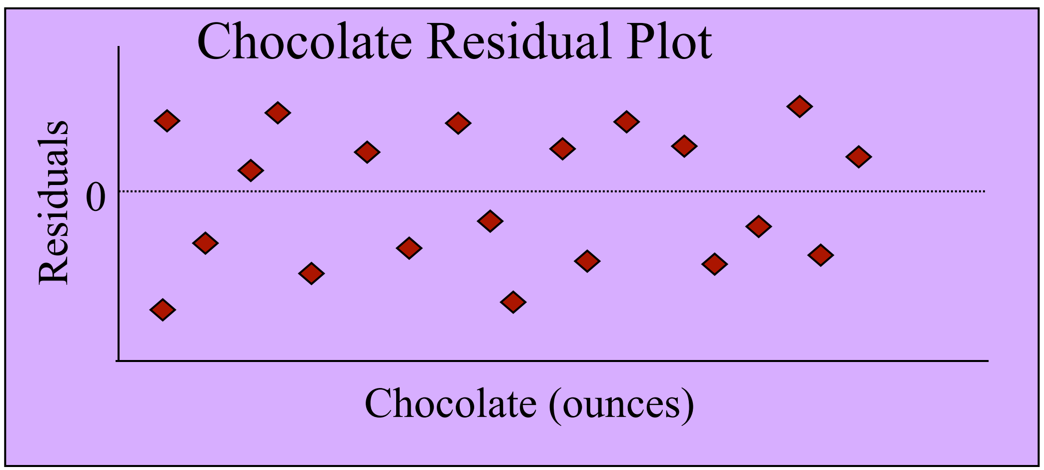
There is a high correlation between chocolate and sugar (>0.4) hence we should eliminate one of these variables - sugar because it is less correlated with fat.



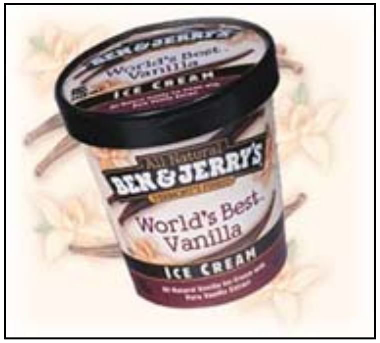
Heteroscedasticity:

An Ice Cream Example



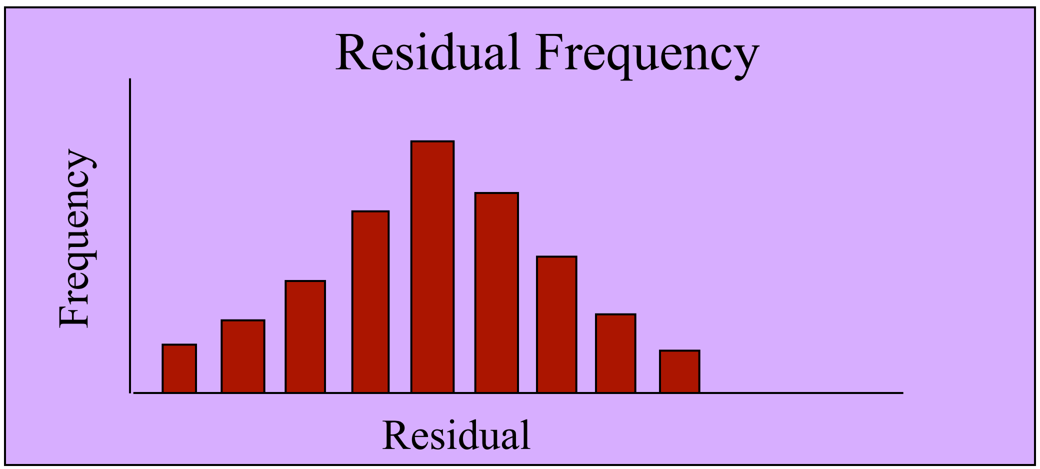


There appears to be no heteroscedasticity



An Ice Cream Example

Residual Distribution:



The residuals appear to be normally distributed

Outline

- **Topic 1 :** Probability Theory
- **Topic 2 :** Decision Analysis
- **Topic 3 :** Discrete Random Variables
- **Topic 4 :** Continuous Random Variables
- **Topic 5 :** Covariance and Correlation
- **Topic 6 :** Regression
- ***Topic 7 :*** *Linear Optimization*
- **Topic 8 :** Discrete Optimization
- **Topic 9 :** Nonlinear Optimization

Optimization terminology

Decision Variable: Describes a decision that needs to be made, e.g. how many items to produce.

Objective Function: An expression (in terms of the variables) that needs to be minimized or maximized.

Constraint: An expression that restricts the values of the variables.

Steps in formulation

1. Define the decision variables.

2. Write the objective as a function of these vars.
   Determine whether max or min.

3. Write the constraints as functions of these vars.
   Either $\leq, \geq, =$.

4. Determine the variable restrictions,
   e.g. non-negative, integer.
   Be careful of units!

About Shadow Prices

- **Associated with each constraint.**
- **Shadow price = 0 for non-binding constraints.**
- **The *shadow price* is the change in the objective value per unit change in the right hand side, given all other data remain the same.**
- **Allowable range (decrease/increase)**
  - If r.h.s changes within range: Shadow price tells us rate of change in the optimal objective function value; we know exactly the new objective value without re-solving the problem.
  - If r.h.s changes outside range: We cannot determine the rate of change in the optimal objective value anymore; we need to solve the optimization pb again!

Avoid frequent mistakes!

- Forgetting the **non-negativity** restrictions
- Confusing **Maximizing with Minimizing**
- **Inconsistent** and/or **incorrect** units
- Reversing the **signs** of the constraints
- Wrong interpretation of the **shadow prices**.
- Change in R.H.S **outside** the allowable range

2007 Exam – Problem 3

Sloan Alloy Corp. manufactures four different types of alloys for aircraft construction, denoted A, B, C and D, from three basic metals: aluminum, copper and magnesium. The profit margins for each of the alloys are \$105, \$45, \$95, and \$140 per ton, for alloys A, B, C, and D, respectively. The proportion of metals in the alloys is shown in the following table.

	A	B	C	D
Aluminum	0.3	0.4	0.1	0.15
Copper	0.3	0.1	0.25	0.4
Magnesium	0.4	0.5	0.65	0.45

For example, one ton of alloy A consists of 0.3 tons of aluminum, 0.3 tons of copper and 0.4 tons of magnesium. The monthly maximum supplies of aluminum, copper and magnesium are 600, 400 and 800 tons per month, respectively.

2007 Exam – Problem 3

Each type of alloy requires a different number of machine hours for production. Alloys A, B, C, and D require 4, 5, 3, and 6 hours/ton, respectively.

	A	B	C	D
Machine hours per ton	4	5	3	6

The total number of machine hours available per month is 8,000 hours. The company can sell all alloy that it produces. The company would like to determine how many tons of each alloy to produce to maximize the profits per month.

2007 Exam – Problem 3

A linear optimization model formulation and solution to this problem is as follows:

MAX: $105\,A + 45\,B + 95\,C + 140\,D$

SUBJECT TO:

Aluminum Supply	$0.3\,A + 0.4\,B + 0.1\,C + 0.15\,D \leq 600$
Copper Supply	$0.3\,A + 0.1\,B + 0.25\,C + 0.4\,D \leq 400$
Magnesium Supply	$0.4\,A + 0.5\,B + 0.65\,C + 0.45\,D \leq 800$
Machine hours	$4\,A + 5\,B + 3\,C + 6\,D \leq 8000$
Nonnegativity	$A, B, C, D \geq 0$

2007 Exam – Problem 3 (a)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(a) (4 points) What is the optimal solution? What is the optimal profit per month?

Optimal solution: 0 tons of A, 632.4 of B, 284.6 of C and 664 of D.
Optimal profit: 105 * 0 + 45 * 632.4 + 95 * 284.6 + 140 * 664 = $148,455/mth

2007 Exam – Problem 3 (b)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(b) (2 points) Which are the binding constraints?

Binding constraints: Supply of copper and magnesium and the available machine time.

2007 Exam – Problem 3 (c)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(c) (2 points) Why is the allowable increase to the Aluminum supply constraint right-hand side equal to 1E+30 in the sensitivity report?

Not a binding constraint. Optimal solution is already not utilizing all available aluminum. An increase of any amount of aluminum does not report us any value or extra profits.

2007 Exam – Problem 3 (d)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(d) (3 points) Assuming all other data are unchanged, what would the new optimal profit be, if we realize we have to perform preventive maintenance to the machines, which will decrease the available machine hours per month by 200 hours?

Our profits will decrease by 0.5 * 200 = $100 / month.

2007 Exam – Problem 3 (e)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(e) (3 points) Assuming all other data are unchanged, if we could increase our supply of one of the metals by 100 tons per month, which one should we choose? What would be the resulting increase in optimized profit?

Increase supply of copper by 100 tons: highest shadow price and 100 tons is within its allowed increase.
The profits would increase by 318.6 * 100 = $31,860 / month.

2007 Exam – Problem 3 (f)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(f) (3 points) Imagine you find out that to produce alloy A you need 5 hours of machine hours per ton (instead of 4). How would the profits per month change? Would your answer be different if you found out that you needed 1 machine hour per month (instead of 4)?

Would not change the result: additional restrictive constraint on the production of alloy A, don't produce any.

If we only needed 1 machine hour, maybe worth producing A. Cannot tell.

2007 Exam – Problem 3 (g)

Adjustable Cells

Cell	Name	Final Value	Reduced Cost	Objective Coefficient	Allowable Increase	Allowable Decrease
B10	A	0	-1.1	105	1.1	1E+30
C10	B	632.4	0.0	45	12.5	2.1
D10	C	284.6	0.0	95	2.0	4.2
E10	D	664.0	0.0	140	14.2	1.8

Constraints

Cell	Name	Final Value	Shadow Price	Constraint R.H. Side	Allowable Increase	Allowable Decrease
B16	Aluminum Supply	381.03	0	600	1E+30	219.0
B17	Copper Supply	400	318.6	400	156.9	202.9
B18	Magnesium Supply	800	21.3	800	442.1	128.6
B19	Machine Time	8000	0.5	8000	1161.3	2711.9

(g) (3 points) How would the optimal profit change, if we increase the maximum Copper supply per month by 200 tons? Be as specific as possible.

Maximum allowable increase is 156.9.
Profit would be at least 156.9 * 318.6 = $49,998 higher per month.
Profits cannot be higher than 200 * 318.6 = $63,720 / month.

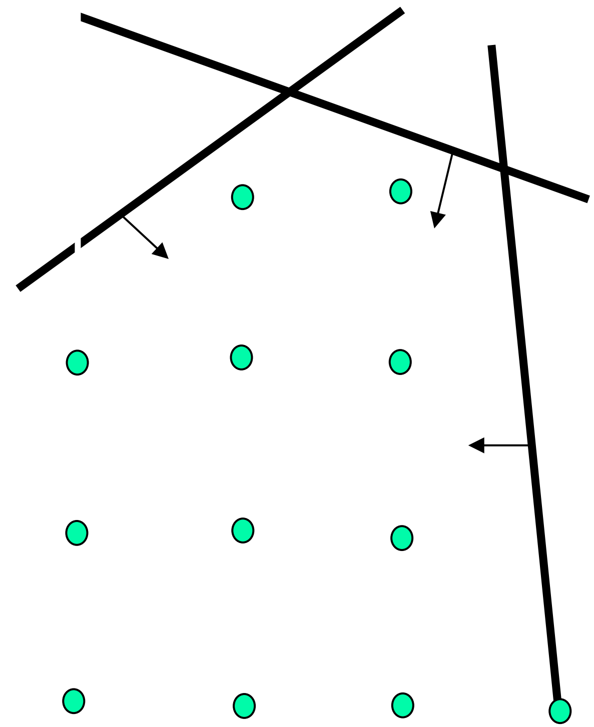
Outline

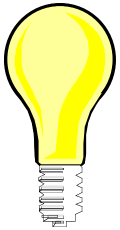
- **Topic 1 :** Probability Theory
- **Topic 2 :** Decision Analysis
- **Topic 3 :** Discrete Random Variables
- **Topic 4 :** Continuous Random Variables
- **Topic 5 :** Covariance and Correlation
- **Topic 6 :** Regression
- **Topic 7 :** Linear Optimization
- ***Topic 8 :*** *Discrete Optimization*
- **Topic 9 :** Nonlinear Optimization

Discrete Optimization

- Feasible region is discrete
  - integer (# of people to hire, # of planes to build)
  - binary (sell division *i*? use factory *i* to serve customers?)

  $x_i = 1$, if decision *i* taken (0 otherwise)
- Not as "easy" to solve as Linear Optimization
- Many problems are discrete, but a linear approximation is sufficient (e.g. pottery manufacturing)
- <u>No shadow price</u> / sensitivity analysis!!
- Can be used to model complicated costs / constraints





..Example..

An electrical utility company each day is deciding which generators to start up.

It has three generators (see below).

There are two periods in a day, and the number of megawatts needed in the first period is 2900. The second period requires 3900 megawatts. Unused electricity left over from period 1 can be used in period 2.

It wants to minimize total cost.

Formulate and solve this problem!

Generator	Fixed costs per period ($)	Cost per period per megawatt used ($)	Max capacity in each period (MW)
A	3000	5	2100
B	2000	4	1800
C	1000	7	3000

Formulation

- Define Decision Variables
  - X_{A1} = machine A to be used in period 1
  - X_{A2} = machine A to be used in period 2
  - X_{B1} = machine B to be used in period 1
  - X_{B2} = machine B to be used in period 2
  - X_{C1} = machine C to be used in period 1
  - X_{C2} = machine C to be used in period 2
  - Y_{A1} = amount of electricity produced by machine A in period 1
  - Y_{A2} = amount of electricity produced by machine A in period 2
  - Y_{B1} = amount of electricity produced by machine B in period 1
  - Y_{B2} = amount of electricity produced by machine B in period 2
  - Y_{C1} = amount of electricity produced by machine C in period 1
  - Y_{C2} = amount of electricity produced by machine C in period 2

Objective Function & Constraints

- Objective Function
  minimize $3000\ (X_{A1} + X_{A2}) + 2000\ (X_{B1} + X_{B2}) + 1000\ (X_{C1} + X_{C2}) +$
  $+5\ (Y_{A1} + Y_{A2}) + 4\ (Y_{B1} + Y_{B2}) + 7\ (Y_{C1} + Y_{C2})$

- Subject to:
  - Capacity:
    - Machine A: $Y_{A1} <= 2100\ X_{A1}$; $Y_{A2} <= 2100\ X_{A2}$
    - Machine B: $Y_{B1} <= 1800\ X_{B1}$; $Y_{B2} <= 1800\ X_{B2}$
    - Machine C: $Y_{C1} <= 3000\ X_{C1}$; $Y_{C2} <= 3000\ X_{B2}$
  - Demand:
    - Period 1: $Y_{A1} + Y_{B1} + Y_{C1} >= 2900$
    - Period 2: $Y_{A2} + Y_{B2} + Y_{C2} + Y_{A1} + Y_{B1} + Y_{C1} - 2900 >= 3900$
  - Binary: $X_{A1}, X_{A2}, X_{B1}, X_{B2}, X_{C1}, X_{C2}$ = {1 if used; 0 otherwise}
  - Non-negativity:
    $Y_{A1}, Y_{A2}, Y_{B1}, Y_{B2}, Y_{C1}, Y_{C2} >= 0$

Excel Solution

Objective Function:	Fixed	+	Variable	=	Total
	10000		30400		40400

Cost:

	A	B	C
Fixed Cost per period	3000	2000	1000
Cost per megawatt	5	4	7

Decision Variables and Constraints:

		Xij (0 or 1)	Yij		Limit * Xij	Limit	Total						Demand
Period 1	A	1	2100	<=	2100	2100	3900					>=	2900
	B	1	1800	<=	1800	1800							
	C	0	0	<=	0	3000							
Period 2	A	1	1100	<=	2100	2100	2900	+	1000	=	3900	>=	3900
	B	1	1800	<=	1800	1800							
	C	0	0	<=	0	3000							

Other constraints:		
Xij	binary	
Xij, Yij	>=	0

Logical Constraints

- Recall our Decision Variables
  - X_{mi} = machine m to be used in period i, $m \in \{A, B, C\}$, $i \in \{1,2\}$
  - Y_{mi} = amount of electricity produced by machine m in period i

- For each of the following statements / conditions, write down the corresponding linear constraint(s) that should be added to the model

1. If we use machine A in period 1, then we must also use it in period 2
   - $X_{A2} >= X_{A1}$
2. If we use machine A in period 1, then we must also use it in period 2, and vice-versa
   - $X_{A2} >= X_{A1}$; $X_{A1} >= X_{A2}$
   - Or, more compactly, $X_{A2} = X_{A1}$
3. Machine B should never be used in both periods
   - $X_{B1} + X_{B2} <= 1$
   - This could be phrased in other ways (e.g. "if machine B is used in period 1, it should not be used in period 2")
4. Machine C should be used in exactly one period
   - $X_{C1} + X_{C2} = 1$

More Logical Constraints

- Recall our Decision Variables
  - X_{mi} = machine m to be used in period i, $m \in \{A, B, C\}$, $i \in \{1,2\}$
  - Y_{mi} = amount of electricity produced by machine m in period i
- For each of the following statements / conditions, write down the corresponding linear constraint(s) that should be added to the model

5. Machines A and B must always be used together
   - $X_{A1} = X_{B1}$; $X_{A2} = X_{B2}$
6. If we use either A or B in period 1, then we must use C in period 2
   - $X_{A1} <= X_{C2}$; $X_{B1} <= X_{C2}$
   - How about $X_{A1} + X_{B1} <= 2 X_{C2}$?
7. We must use <u>at least</u> 2 machines in period 1
   - $X_{A1} + X_{B1} + X_{C1} >= 2$
8. We must use <u>at most</u> two machines in period 2
   - $X_{A2} + X_{B2} + X_{C2} <= 2$
9. (Trickier) Recall that the maximum capacity for A is 2100. Using words, explain what the following constraint might be able to achieve:
   - $X_{A2} <= 1 - (Y_{A1} - 0.95 * 2100) / (0.05 * 2100)$

Outline

- **Topic 1 :** Probability Theory
- **Topic 2 :** Decision Analysis
- **Topic 3 :** Discrete Random Variables
- **Topic 4 :** Continuous Random Variables
- **Topic 5 :** Covariance and Correlation
- **Topic 6 :** Regression
- **Topic 7 :** Linear Optimization
- **Topic 8 :** Discrete Optimization
- ***Topic 9 :*** *Nonlinear Optimization*

Steps in a Formulation

0. Same steps as LP, but objective and constraints can be non-linear (eg. Minimize x^2+y^2; $x*y+y<=5$)

1. Define the decision variables.

2. Write the objective as a function of these vars. Determine whether max or min.

3. Write the constraints as functions of these vars. Either **≤ , ≥ , =**.

4. Determine the variable restrictions,

   e.g. non-negative, integer. Be careful of units!

Example of a Sensitivity Report

Microsoft Excel 10.0 Sensitivity Report
Worksheet: [Book1]Products
Report Created: 12/9/2004 10:19:06 PM

Adjustable Cells

Cell	Name	Final Value	Reduced Gradient
B8	units	58.82	0
C8	units	23.53	0
D8	units	125.00	0
B9	units	58.82	0
C9	units	23.53	0
D9	units	125.00	0

Constraints

Cell	Name	Final Value	Lagrange Multiplier
B19	machine 1 capacity LHS	578.68	0
B20	machine 2 capacity LHS	180.59	0
B21	product A limit LHS	117.65	0
B22	product B limit LHS	47.06	0
B23	product C limit LHS	250.00	0
B11	Price A units	$ 105.88	$ -
B12	Price B units	$ 35.29	$ -
B13	Price C units	$ 250.00	$ -

All Lagrange Multipliers are zero!

All constraints are non-binding around the close proximity of the optimal solution.

Optimal solution occurs in the interior of the feasible region.

2007 Exam Problem 5

Problem 5. (20 points) Exeter Investments (EI) has decided to construct a new portfolio which will be comprised of IT (information technology) and energy companies' stocks. For this reason, EI collected the expected return, standard deviations and return correlations data for companies A, B, C, D, E, and F, which belong to those sectors. This data is displayed in Tables 1 and 2.

Table 1

Company	Expected Return (%)	Standard Deviation (%)	Sector
A	14	19	IT
B	17	28	IT
C	12	17	Energy
D	4	7	IT
E	19	35	Energy
F	25	45	IT

Table 2. The correlation coefficients between returns of stocks

	A	B	C	D	E	F
A	1					
B	0.7	1				
C	0.2	-0.4	1			
D	0	0	0	1		
E	0	0	0	0	1	
F	0	0	0	0	0	1

2007 Exam Problem 5

a) **(5 points)** EI would like the portfolio to have a maximum risk of 15%. Write down a nonlinear optimization problem to determine the weights that yield the maximum possible expected return satisfying this target risk requirement.

Decision Variables: w_i—what percentage of the portfolio do I invest in stock i for i=A,B…F

$$\max 14w_A + 17w_B + 12w_C + 4w_D + 19w_E + 25w_F \qquad \text{(Portfolio return)}$$

$$\sqrt{\begin{array}{l} 19^2 w_A^2 + 28^2 w_B^2 + 17^2 w_C^2 + 7^2 w_D^2 + 35^2 w_E^2 + 45^2 w_F^2 \\ + 2w_A \cdot w_B \cdot 0.7 \cdot 19 \cdot 28 - 2w_B \cdot w_C \cdot 0.4 \cdot 28 \cdot 17 + 2w_A \cdot w_C \cdot 0.2 \cdot 19 \cdot 17 \end{array}} \le 15 \qquad \text{(Risk maximum)}$$

$$w_A + w_B + w_C + w_D + w_E + w_F = 1 \qquad \text{(Invest whole portfolio)}$$

$$w_A, w_B, w_C, w_D, w_E, w_F \ge 0 \qquad \text{(Non-negative weights)}$$

b) **(3 points)** In looking at the sensitivity report with respect to the optimization problem in a), the shadow price associated with the constraint on a maximum risk of 15% was found to be [redacted]0.0143, but the sign in front of it cannot be read. That is, this shadow price is either 0.0143 or -0.0143. Which one of these possible values is correct?

+:Raising the risk maximum makes the feasible region bigger (we are less restricted). Since we are maximizing, this means we may find a new optimum with a higher return. Thus the shadow price is >=0.

2007 Exam Problem 5

(c) (**4 points**) Additionally EI wants to ensure that the risk corresponding to the part of the portfolio consisting of EI companies does not exceed 17%. Incorporate this requirement into your formulation.

A, B, D, F are IT companies. We add the constraint:

$$\sqrt{19^2 w_A^2 + 28^2 w_B^2 + 7^2 w_D^2 + 45^2 w_F^2 + 2w_A \cdot w_B \cdot 0.7 \cdot 19 \cdot 28} \leq 17$$

(d) (**5 points**) As of December 10 close, the market capitalization of energy stocks represented 12.08% of the overall S&P 500 market capitalization. EI would like to make sure that in their portfolio, the relative value of energy stocks is within +/- 2% of this S&P 500 benchmark. Show how EI can incorporate this requirement into their formulation.

C and E are Energy companies. We add the constraints:

$$0.1008 \leq w_C + w_E \leq 0.1408$$

Office Hour today at 4:15PM in the same room (E62-262)!