This brief review contains notes on three main technical topics that we'll need in this course. For more extensive discussion of these topics, please consult any good textbook.

1. Optimization

The Derivative

A function *f: D⇒ R* is a rule that assigns to each number in the domain *D* a single number in the range *R*. For example, a function f(x)=2x assigns to each real number x the value of 2x.

Two useful properties of functions are:

(1) Its first derivative, denoted by f′(x)
- Represents how the function changes at point *x* if we increase *x* by a very small amount (i.e. it is the slope of the function at point *x*).
- Tells us whether the function is increasing (f′(x)>0) or decreasing (f′(x)<0) at point *x*

(2) Its second derivative, denoted by f″(x)
- Represents how the slope of the function changes at point *x* if we increase *x* by a very small amount (i.e. it is the curvature of the function at point *x*).
- Tells us whether the slope is increasing (f″(x)>0) or decreasing (f″(x)<0) at point *x*

For example, for the function f(x)=2x, the first derivative is 2 and the second derivative is 0. Therefore, the function is increasing for each x in its domain and it has a constant slope 2 that does not depend on the value of x.

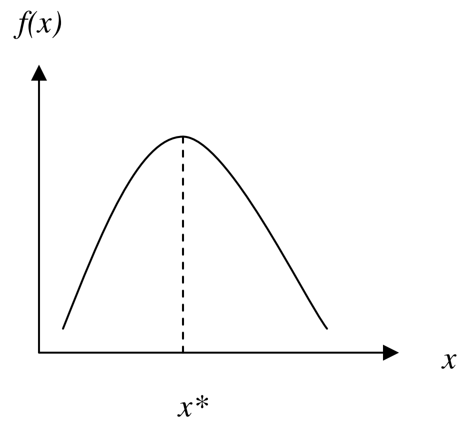
Some common formulas for derivatives are presented in the following table:

Function	Example	Derivative Rule	Derivative for Example
c (constant)	5	0	0
x^n	x^3	nx^{n-1}	$3x^{3-1}$
ln(f(x))	ln(2x)	f′(x)/f(x)	2/2x
f(x)+g(y)	$x+x^2$	f′(x)+g′(x)	1+2x
f(x)g(y)	xy^2	f′(x)g(x)+ f(x)g′(x)	y^2+2xy
f(x)/g(y)	x/y	$[f'(x)g(x)- f(x)g'(x)]/g(x)^2$	$[y-x]/y^2$
f(g(x))	$(x^2)^2$	f′(g(x))g′(x)	$2(x^2)^{2-1}2x$

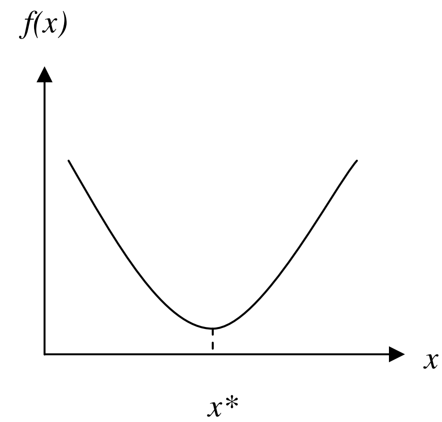
Single Variable Optimization

Consider a function *f* that depends on a single variable *x*. We say that *f* achieves a unique <u>maximum</u> at *x** if f(x*)>f(x) for all *x*.

Similarly, we say that *f* achieves a unique <u>minimum</u> at *x** if f(x*)<f(x) for all *x*.



Maximum



Minimum

When *f* is differentiable, the maximum and minimum can be defined using the first-order and second-order conditions:

	Maximum at x^*	Minimum at x^*
First-order condition:	$f'(x^*)=0$	$f'(x^*)=0$
Second-order condition:	$f''(x^*)<0$	$f''(x^*)>0$

In words, the slope of the function at both the minimum and the maximum point is zero (i.e. a tangent line to the function is flat). For the maximum, as we increase *x* the slope becomes negative; for the minimum, as we increase *x* the slope becomes positive.

For example, the function $f(x)=-x^2$ achieves its maximum at x=0 and the function $f(x)=x^2$ achieves its minimum at x=0.

2. Uncertainty

Review of Statistical Concepts

The set of possible outcomes in a statistical problem is represented by a sample space *S*. A random variable *x* is a function that associates with each element *s* in *S* a real number *x*(*s*). For example, the elements of *S* may be the books in a library and *x*(*s*) may specify the number of pages in the book *s*. With the sample space comes a probability function *p* that assigns a probability *p*(*s*) to each element *s* in *S*. In our earlier example with the library books, *Prob*(*x*=100) is the probability that the book *s* has exactly 100 pages.

Each random variable has a mean denoted by $\bar{x}$, also called its expectation and denoted by E[*x*]. The formula for calculating expectations is:

$$E[x] = \sum p(s)x(s) = \bar{x}$$

Each random variable also has a variance given by

$$Var(x) = E[(x - \bar{x})^2] = \sum p(s)(x - \bar{x})^2$$

Variance is one measure of the degree of randomness of *x.*

To illustrate, consider the following example:

p(*s*)	*x*(*s*)	$p(s)x(s)$	$(x(s)-\bar{x})^2$	$p(s)(x(s)-\bar{x})^2$
1/2	0	0	4,000,000	2,000,000
1/3	3,000	1,000	1,000,000	333,333
1/6	6,000	1,000	16,000,000	2,666,667
		$\sum p(s)x(s) = \bar{x}$ =2,000		$Var(x) = \sum p(s)(x-\bar{x})^2$ =5,000,000

Given two random variables *x* and *y*, the covariance of *x* and *y* is:

$$Cov(x, y) = E[(x - \bar{x})(y - \bar{y})]$$

Covariance measures the extent to which *x* and *y* are correlated (move together).

Given two random variables *x* and *y* and two real numbers α and β, we can form a new random variable $\alpha x + \beta y$. Its expectation is given by:

$$E[\alpha x + \beta y] = \alpha E[x] + \beta E[y]$$

and its variance is equal to:

$$Var[\alpha x+\beta y]=\alpha^2 Var[x]+\beta^2 Var[y]+2\alpha\beta Cov(x,y)$$

The two random variables *x* and *y* are statistically independent if for all numbers α and β, $\Pr ob(x=\alpha$ and $y=\beta)=\Pr ob(x=\alpha)\Pr ob(y=\beta)$. Statistical independence represents the idea that knowing the value of one of the variables provides no information about the value of the others. If *x* and *y* are statistically independent, then *Cov*(*x,y*)=0 and *Var*(*x+y*)=*Var*(*x*)+ *Var*(*y*).

Risk Premium and Certainty Equivalent

A risk-averse individual would prefer receiving a certain income of $\bar{I}$ to receiving a random income with expected value of $\bar{I}$. To illustrate, suppose that the individual preferences are given by the utility function *u*(*I*), where *I* is a random variable with expected value of $\bar{I}$. Then, for the risk-averse individual we have that:

$$u(\bar{I})>\sum p(I)u(I)=E[u(I)]$$

The amount that the risk-averse person is willing to pay to make the switch is called the risk premium and is denoted by *RP*:

$$u(\bar{I}-RP)=E[u(I)]$$

One of the central results of decision theory under uncertainty is that the risk premium can be estimated by a simple formula:

$$RP=0.5rVar(I)$$

This result is proved below. r is a parameter of the decision maker's personal preference called the coefficient of absolute risk aversion[1]. For risk-averse individuals, *r* >0 and it indicates that these individuals are willing to pay to avoid risk. For risk-neutral individuals, *r*=0. Therefore, their risk premium is also 0, which implies that they are indifferent between receiving a certain income of $\bar{I}$ and receiving a random income with expected value of $\bar{I}$.

The amount $\bar{I}-RP$ that is left after the risk premium is paid is called the certainty equivalent and is denoted by *CE.* It is the amount of income, payable for certain, that the person regards as equivalent in value to the original, random income:

$$u(CE)=E[u(I)]$$

[1] In general, *r* may depend on $\bar{I}$, but we will assume that it does not. Intuitively, we are assuming that the degree of risk aversion does not depend on the person's income.

The advantage of focusing on the certainty equivalent is this. In general, maximizing the expected utility function $E[u(I)]$ is difficult because it depends on the whole probability distribution of income *p*(.). However, since $u(CE)=E[u(I)]$, maximizing *u*(*CE*) is equivalent to maximizing $E[u(I)]$. Moreover, if *u* is well-defined (i.e. the inverse of *u* is a positive transformation of *u*), then maximizing *CE* is the same thing as maximizing *u*(*CE*). The reason for this is that utility functions only rank outcomes and any positive transformation will not affect this ranking. Therefore, the problem of maximizing $E[u(I)]$ can be reduced to the problem of maximizing *CE*. This is obviously a much simpler problem since *CE* can be computed easily using the mean and variance of *I* as follows:

$$CE = \bar{I} - 0.5rVar(I)$$

Proof that $CE = \bar{I} - 0.5rVar(I)$

The second-order Taylor approximation to a function *u* for any point *z* is:

$$u(z) = u(\bar{x}) + (z-\bar{x})u'(\bar{x}) + 0.5(z-\bar{x})^2 u''(\bar{x})$$

where $\bar{x} = E[x]$.

Substituting *x* for *z* in this equation and computing the expectation, we have:

(1)
$$\begin{aligned} &E[u(x)] = u(\bar{x}) + E[x-\bar{x}]u'(\bar{x}) + 0.5E[(x-\bar{x})^2]u''(\bar{x}) \\ &= u(\bar{x}) + 0.5Var(x)u''(\bar{x}) \end{aligned}$$

where the second line follows since $E[x-\bar{x}]$=0 and $E[(x-\bar{x})^2] = Var(x)$.

Next, for *CE* we can use the first-order Taylor approximation (since we expect that *CE* will be close to $\bar{x}$) as follows:

(2)
$$u(CE) = u(\bar{x}) + (CE-\bar{x})u'(\bar{x})$$

Lastly, we can combine equations (1) and (2) as follows:

$$\begin{aligned} &u(CE) = E[u(x)] \\ &\Rightarrow u(\bar{x}) + (CE-\bar{x})u'(\bar{x}) = u(\bar{x}) + 0.5Var(x)u''(\bar{x}) \\ &\Rightarrow CE = \bar{x} - 0.5\frac{u''(\bar{x})}{u''(\bar{x})}Var(x) = \bar{x} - 0.5rVar(x) \end{aligned}$$

3. Regression Analysis

Regression analysis can be used to describe a relationship between a dependent or outcome variable and one or more independent variables. For example, our goal can be to describe how workers' productivity depends on the type of compensation and workers' ability. As a first step, we may specify the model as follows:

$$Y_i=a+bD_i+cA_i+u_i$$

where i indexes workers, Y_i represents worker i's productivity, D_i is an indicator variable equal to 1 if worker i is a piece-rate worker and 0 if he is a salary worker, and A_i is worker i's ability. u_i represents all other factors that explain variation in productivity across workers besides how workers are paid and their ability.

The coefficients a, b, and c can be interpreted as follows:

(1) $a=E[Y_i|D_i=0,A_i=0]$, so a represents the average productivity of a salary worker ($D_i=0$) of ability 0 ($A_i=0$).
(2) $b=\partial E[Y_i]/\partial D_i$, where the symbol ∂ indicates a partial derivative and indicates that all other factors are held constant. Since D takes the value of 1 or 0, b represents the average difference in productivity between piece-rate and salary workers, holding all other variables constant. In our model, b tells us the average difference between productivity of piece-rate and salary workers of same ability.
(3) $c=\partial E[Y_i]/\partial A_i$, so b tells us the average difference in productivity between two workers who differ in their ability by a small amount, but who are otherwise paid in the same way.

Once we specified the model, we need to collect data on Y, D, and A. Suppose that we obtained the following data:

Worker (i)	Productivity (Y_i)	Piece-Rate (D_i)	Ability (A_i)
1	10.82	1	3
2	13.18	1	4
3	10.42	1	2
4	11.90	1	3
5	10.20	1	3
6	10.17	0	3
7	7.68	0	2
8	7.38	0	2
9	4.43	0	1
10	8.06	0	2

Our next task is to estimate coefficients a, b, and c, given the data. This can be accomplished in a number of ways, but the most commonly method used is the ordinary least squares (OLS). The OLS method is to choose parameters a, b, and c to minimize the sum of squared difference between the observed and predicted values of Y:

$$\text{Min}_{a,b,c} \sum_i (Y_i - a - bD_i - cA_i)^2$$

The first-order conditions for a, b and c will give us values that best 'fit' the data.

In addition to obtaining estimates that best fit the data in our sample, we also want to know how likely it is to obtain similar estimates if we were to sample another group of workers. Some variation in the estimates is expected because of idiosyncratic differences between workers, and this sampling noise can be conveniently captured by estimating variances of estimates.

Most statistical software packages, such as Stata, will routinely produce estimated coefficients of the model and their variances. For example, the Stata output for our model is as follows:

```
------------------------------------------------------------------------------
productivity |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  piecerate |      1.635   .6780478     2.41   0.047     .0316715    3.238328
    ability |      2.125   .4205074     5.05   0.001     1.130658    3.119342
      _cons |      3.294   .9212855     3.58   0.009     1.115506    5.472494
------------------------------------------------------------------------------
```

The first column (Coef.)reports the estimated coefficients b, c, and a, respectively. Therefore, the average productivity of a salary worker with no ability is about 3.3 (i.e. a=3.294). A piece rate worker of same ability as a salary worker has on average 1.6 higher productivity (b=1.635). Lastly, workers' productivity increases by about 2.1 for every additional unit of ability, independently of how workers are paid (c=2.125).

The second column (Std. Err.) indicates the standard errors of estimates b, c, and a, respectively. These standard errors, in conjunction with the estimated coefficients, can be used to test hypotheses about the value of coefficients. One commonly used hypothesis is to test whether there is any relationship between the outcome and a given independent variable (i.e. whether the coefficient is equal to zero).

For example, we may be interested to know whether how workers are paid has an impact on workers' productivity (i.e. whether b=0). Given the assumption that u has a normal distribution, we can use the t-test to test this hypothesis. The t-statistic is conveniently displayed in the third column of the table (t). To determine whether this t-statistic falls inside the region of t-values that are consistent with hypothesis b=0 or inside the region that is not consistent with the hypothesis, we need to consult the Student t table. As a rule of thumb, if the t-statistic is greater than 2 or smaller than -2, the hypothesis that the coefficient is equal to zero (i.e. no relationship exists) can be rejected at the 5 percent confidence level. When this is the case, we say that the coefficient is statistically significant. In our example, all of our three estimates coefficients are statistically significant (i.e. different from zero) because their t-values exceed 2 in the absolute value. Therefore, both how workers are paid and workers' ability significantly influence the average productivity of the workers.

The fourth column (P>|t|) gives us the exact confidence level at which we can reject the hypothesis that the coefficient is equal to zero. This value is also known as the p-value. Again, as a rule of thumb, we can conclude that the coefficient is significant if its p-value is equal to or less than 5 percent. The last two columns give us the expected interval for values of coefficients if we were to draw new samples of workers. The interval is specified at the 5 percent confidence level. For example, the piece rate workers are on average 1.6 units more productive than salary workers, controlling for the worker's ability, but this estimate can range between 0.03 and 3.3 in other samples.

The estimates presented in the table give us opportunity to say at least three things about the impact of each independent variable on the outcome variable:

1. Whether the relationship exists or not (statistical significance);
2. Whether the relationship is positive or negative; and
3. How strong is the relationship (economic significance).

In our example, the results indicate that how workers are paid influences their productivity (t-value>|2|). Moreover, the results show that piece-rate workers are more productive than salary workers (b=1.6>0). This difference seems to be economically significant: the percentage gain between piece rate and salary workers of similar ability is 1.6/3.3=0.48, or 48 percent.

While the regression analysis is useful in describing relationships between variables, it is not necessarily informative about whether the independent variables causally affect the outcome variable. In other words, the regression results tell us that the outcome and a given independent variable are correlated, but we cannot say without further assumptions that this correlation indicates a causal relationship. The main reason for this is known as the omitted variable problem. For our example, we know that how workers are paid and their productivity are positively correlated, controlling for workers' ability, but how can we be sure that there are no other variables that can explain this correlation? Put differently, are salary workers a good control group for piece rate workers, even if they had same ability?

Empirical strategies, such as randomized experiments and difference-in-differences, deal specifically with the issue of how to uncover causal relationships.