Table of Contents

Mathematics....2
01. Fundamentals....2
02. Functions....2
03. Sequence....2
04. Limits....2
05. Differentials....3
06. First Order Condition....4
07. Second Order Condition....4
08. Graph Sketching....4
09. Multi-variable functions....5
10. Series....5
11. Integration....5
12. (Constrained) Optimisation....6
13. Vectors/Matrices....6
14. Hessian....7
15. Exam Application....7
16. Taylor Series....8
17. Difference Equations....8
18. Differential Equations....8
Statistics....10
01. Fundamentals....10
02. Single Probability (and Distribution)....10
03. Joint Probability....11
04. Moments....11
05. Sampling....11
06. Estimators....11
07. Standard Error & CI....12
08. Hypothesis Tests....13
09. OLS Regression: Basics....13
10. OLS Regression: Deriving α, β....13
11. OLS Regression: Proof that α, β are unbiased....14
12. OLS Regression: Extras....14
13. OLS Regression: Assumptions....15

Explain the Math

Mathematics

01. Fundamentals

Quadratic Equation: $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ | Completing the Square: $a\left(x+\frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$

Combinations: $nCk = \binom{n}{k} = \frac{n!}{(n-k)!k!}$ | Permutations: $nPk = \binom{n}{k} = \frac{n!}{(n-k)!}$

Set Notation

$\mathbb{N}$(natural) $\subset \mathbb{Z}$(integers) $\subset \mathbb{Q}$(rational) $\subset \mathbb{R}$(real)

$\mathbb{R}_+$: positive real numbers ; $\mathbb{R}_{++}$: non-negative real numbers

$x \in S$: x is an element of set S | Number of members in set S: $|S|$ or $\sum_{s \in S} 1$

2^S: All subsets of S

$X = \{s \in S \mid s \ldots\}$ [i.e. X is the set of members S which satisfy property…]

Cartesian product of two sets $(A \times B)$ is the set of all **ordered pairs** (a, b)

$\mathbb{R}^2$: set of real numbers (x, y) producing Cartesian graph

$[a, b) = \{x \mid a \leq x < b\}$

Image of a set S under f: $f(S) = \{f(x) \mid x \in S\}$

02. Functions

$\mathbb{R} \to \mathbb{R}$ $[\boldsymbol{f}(\boldsymbol{x})]$: Graph$(\boldsymbol{f}) = \{(\boldsymbol{x}, \boldsymbol{f}(\boldsymbol{x})) \mid \boldsymbol{x} \in \mathbb{R}\}$ which is $\subseteq \mathbb{R}^2$

$\mathbb{R}^n \to \mathbb{R}$ $[\boldsymbol{f}(\boldsymbol{x})]$ *e.g.* Graph$(\boldsymbol{f}) = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in \mathbb{R}^3 \mid \boldsymbol{z} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y})\}$

Homogenous of degree k if $f(\lambda x_1, \lambda x_2 \ldots) = \lambda^k f(x_1, x_2 \ldots)$ for all $\lambda > 0$ and $(x_1, x_2 \ldots) \in \mathbb{R}^n$

Intermediate Value T.: if f is continious on $[(a, b])$it'll take all values between $f(a), f(b)$ [NE

03. Sequence

Sequences with values in set A can be expressed as $f: \mathbb{N} \to A$

Convergence (to L): "there exists K such that $n > K \Rightarrow |a_n - L| < \epsilon$"

(i.e. eventually all terms of the sequence are within the ϵ-neighbourhood of L)

Hence $\lim_{n \to \infty} a_n = L$ or $a_n \to L$

Divergence (to $+\infty$): "for any bound $M > 0$ there exists $K > 0$, such that $n > K \Rightarrow a_n > M$"

04. Limits

"a is a limit point of set S if for every ϵ>0, the ϵ-neighbourhood of a includes elements of S

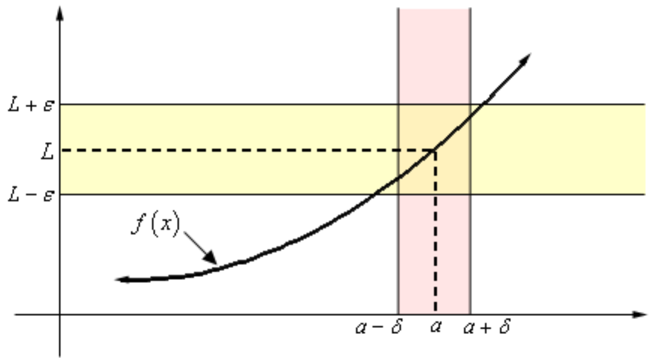
i.e. Numbers that are or are infinitely close to an element in S [NE]

- Bounded: there exists a number N such that $\boldsymbol{x} < |N|$ for all $\boldsymbol{x}$ in S
- Closed: all limit points in S are elements of S [NE]
- Compact: both closed and bounded

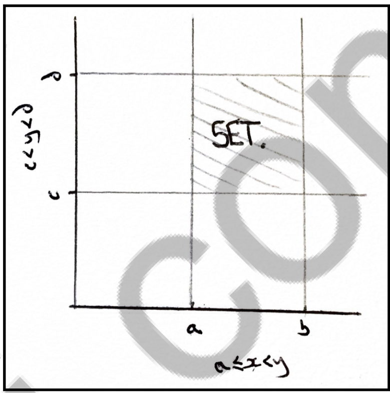
- <u>Open:</u> any point $\boldsymbol{x}$ in S there exist $\epsilon > 0$ such that $(\boldsymbol{x}+\epsilon, \boldsymbol{x}-\epsilon) \subseteq S$ (i.e. all points in S are interior)

Sets have limit points, functions have limits: $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = L$. If $= f(\boldsymbol{a})$ then continuous

L is limit of $f(\boldsymbol{x})$ if, as $\boldsymbol{x}$ approaches to $\boldsymbol{a}$, "given any $\epsilon > 0$ there exists $\delta > 0$ s.t. $\|\boldsymbol{x}-\boldsymbol{a}\| < \delta \Rightarrow |f(\boldsymbol{x}) - L| < \epsilon$" (i.e. points close to $\boldsymbol{a}$ are mapped close to L) [NE]



e.g. $[a,b)\times(c,d)$ has limit points $[a,b]\times[c,d]$



Tip: If $\lim_{x\to a}\frac{f(x)}{g(x)}$ results in $\frac{0}{0}$ use Bernoulli's rule $= \frac{\left(\frac{df(a)}{dx}\right)}{\left(\frac{dg(a)}{dx}\right)}$

05. Differentials

$\frac{\partial}{\partial x} f(x)^n \to n\, f'(x)\, f(x)^{n-1}$ | $f(x) = x^n$ *then* $f^k(x) = \begin{cases} n(n-1)\ldots(n-k+1)x^n - k & \text{if } 0<k\leq n \\ 0 & \text{if } 0<n<k \end{cases}$

Mathematical Intuition

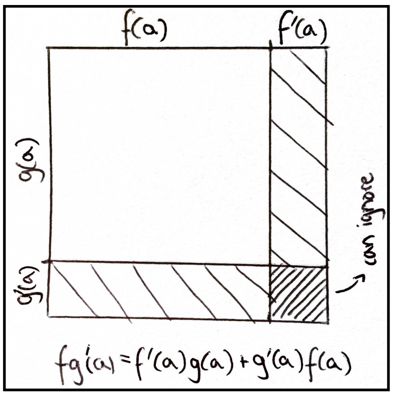
<u>What is differentiation?</u>

- Not "instant rate of change" of $f(x)$ [i.e. $\frac{rise}{run}$ between two close points], that's an oxymoron
- Instead "best constant approx. around a given point" [i.e. slope of line tangent at a single point]
- So dx is not infinitely small or 0 but instead some value h that approaches 0: $\left.\frac{df}{dx}\right|_{x=a} = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}$

Note: Ignore any terms including $dx^{n>1}$ as they are even smaller than dx. Also, must be continuous

<u>Where does the general rule come from?</u>

$\frac{\partial}{\partial x}x^n = \frac{(x+dx)^n - x^n}{dx} = \frac{x^n + nx^{n-1}dx + [\text{terms inc } dx^{n>1}] - x^n}{dx} = \frac{nx^{n-1}dx + [\text{terms inc } dx^{n>1}]}{dx} = \frac{nx^{n-1}dx}{dx} = nx^{n-1}$



Basic Algebra

Sum Rule: $(f+g)'(a) = f'(a) + g'(a)$

Product Rule: $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$

Chain Rule: $(g\circ f)(x) = g(f(x))$ and $(g\circ f)'(x) = g'(f(x)) \times f'(x)$

Inverse Product Rule: $\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a)+f(a)g\prime(a)}{[g(a)]^2}$

Constant Rule: $(cf)'(a) = cf'(a)$

06. First Order Condition

"If $f'(a) = 0$ then a is interior local maximum/minimum" [for single variable, for multi see p7 **Hessian**]

Mathematical Intuition

Extreme Value Theorem: function attains max and min i.e.

"if f is continious on $[a, b]$ there exist points c and d such that $f(c) \leq x \leq f(d)$"

If $f(c)$ is min and <u>interior</u> then $f(c+h) - f(c) \geq 0$ for $h \neq 0$

$\frac{f(c+h)-f(c)}{h} \geq 0$ for $h > 0$

$\frac{f(c+h)-f(c)}{h} \leq 0$ for $h < 0$

Similar if $f(d)$ is max

Characteristics

Monotonicity is if function's first derivative never changes sign (need not be continuous)

- <u>Increasing:</u> $a > b$ implies $f(a) > f(b)$ hence $f'(x) > 0$
- <u>Decreasing:</u> $a > b$ implies $f(a) < f(b)$ hence $f'(x) < 0$
- <u>Non-Increasing:</u> $a > b$ implies $f(a) \leq f(b)$ hence $f'(x) \leq 0$
- <u>Non-Decreasing:</u> $a > b$ implies $f(a) \geq f(b)$ hence $f'(x) \geq 0$
- <u>Constant:</u> $f(a) = f(b)$ for interval (a, b) hence $f'(x) = 0$

07. Second Order Condition

- If $f'(a) = 0$ and $f''(a) > 0$, a is a local minimum
- If $f'(a) = 0$ and $f''(a) < 0$, a is a local maximum

- <u>Convex function:</u> $f(\lambda a + (1-\lambda)b) \leq \lambda f(a) + (1-\lambda)f(b)$ hence $f''(x) \geq 0$ for all x
- <u>Concave function:</u> $f(\lambda a + (1-\lambda)b) \geq \lambda f(a) + (1-\lambda)f(b)$ hence $f''(x) \leq 0$ for all x
- <u>Strictly X function:</u> removes the $=$ sign from above. Will mean there is only one max/min
- <u>Quasi X function:</u> $f(\lambda a + (1-\lambda)b)[\leq \max f(x) \ \ or \ \ \geq \min f(x)]$
  [see p6 on **Hessian** for more]

These characteristics are conditional on…

- Continuous and twice differentiable functions
- **Convex domain** (for every $x, y \in S$ and $\lambda \in (0,1)$, $\lambda x + (1-\lambda)y$ is also in S)
- **Note:** Convex preferences are caused by quasi-concave utility functions!

08. Graph Sketching

Sketching $y = f(x)$

[0] figure out the largest possible domain if not give [1] rewrite if top heavy or complete the square if it can't be factorised [2] x, y asymptotes [3] intersection with x, y axis [4] $y, x \to \pm\infty$ [5] x near asymptotes $(+ \text{or} -)$ [6] where increasing and where decreasing [7] max, min if existent [8] where concave and where convex

<u>Algebraic Approach</u>: turn into quadratic with y being coefficient of x^n. Turning point when one solution $(b^2 - 4ac = 0)$, range when two $(b^2 - 4ac > 0)$.

Sketching $y^2 = f(x)$

[1] sketch $y = f(x)$ for reference [2] erase below x axis [3] transfer points on $y = 0,1$ $x = 0$ and value of turning points (sub in) [4] draw $\infty \frac{dy}{dx}$ at x axis intersect unless $f'(x) = 0$ [5] draw above reference for $x < 1$and below for $x > 1$ [6] reflect in x axis. Asymptotes: $y = a$ to $y = \pm\sqrt{a}$; $x = a$ stays $x = a$

09. Multi-variable functions

Typical examples

$\max \pi = pf\big(K^*(p,r,k), L^*(p,r,k)\big) - \big(rK^*(p,r,k) + wL^*(p,r,k)\big)\ s.t. K, L \geq 0$

$B = \{\boldsymbol{x} | \boldsymbol{p} \cdot \boldsymbol{x} \leq m$ and $\boldsymbol{x} \geq 0\}$ Find $(\boldsymbol{a}) \in B$ so that $u(\boldsymbol{a}) \geq u(\boldsymbol{x})$

Vectors

$\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$; Distance $\|\boldsymbol{a} - \boldsymbol{b}\| = \sqrt{(a_1 - b_1)^2 + \cdots (a_n - b_n)^2}$;

r-nneighbourhood $= \{x \in \mathbb{R}^n | \|\boldsymbol{x} - \boldsymbol{a}\| < r\} = (\boldsymbol{a} - r, \boldsymbol{a} + r)$

Mathematical Intuition

$\lim_{\boldsymbol{x} \to \boldsymbol{a}} \frac{f(\boldsymbol{x}) - f(\boldsymbol{a})\,[\mathbb{R}]}{\boldsymbol{x} - \boldsymbol{a}\,[\mathbb{R}^n]}$ Doesn't make any sense!

Instead take **partial derivative w.r.t variable $\boldsymbol{x_i}$** i.e. $f_i = \frac{\partial f(\boldsymbol{x})}{\partial x_i} = \lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{f(x_1, \dots x_i + h, \dots x_n) - f(\boldsymbol{x})}{h}$ [= 0 for FOC]

Chain Rule: If $t \to f(x(t), y(t))$ [i.e. can be expressed as single variable function], then $\frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$

10. Series

$\sum_{i=0}^{n} a_i = a_0 + a_1 \dots + a_n$; Partial Sum $= S_n$;

Geometric Series: $a_n = a_0 c^n$

$\sum_{i=0}^{n} a_0 c^i = \frac{a(1 - c^{n+1})}{1 - c}$ [if $c \neq 1$]

- $a_n \to 0$ if $|c| < 1$
- $a_n \to a$ if $c = 1$
- a_n diverges [fluctates between a and $-a$] if $c = -1$
- a_n diverges to ∞ if $c > 1$
- a_n diverges with growing magnitude and flipping sign if $c < -1$

11. Integration

$\int f(x)^n \to \frac{1}{n\, f\prime(x)}\, f(x)^{n-1}$

Mathematical Intuition [NE]

Let $f: [a, b] \to \mathbb{R}$ be continuous and $F: [a, b] \to \mathbb{R}$ be defined as $F(x) = \int_a^x f(t)dt$

Since continuous, $\int_a^b f(x)dx$ has max and min: $m(b - a) \leq \int_a^b f(x)dx \leq M(b - a)$

$m \leq \frac{1}{b-a}\int_a^b f(x)dx \leq M$; any $m \leq x \leq M$ attainable i.e. $f(c) = \frac{1}{(b-a)}\int_a^b f(x)dx$ [Mean Value T. [NE]]

$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} \frac{\int_x^{x+h} f(t)dt}{h} = \lim_{h \to 0} f(c)$[for some $c \in [x, x+h]$] $= f(x)$

Basic Properties

$\int_a^b f(x)dx = -\int_b^a f(x)dx$

$\int_a^b f(x)dx + \int_b^c f(x)dx = \int_a^c f(x)dx$

$\int kf(x)dx = k\int f(x)dx$

$\int udv = uv - \int vdu$ [always use 'simpler' function for u, unless $\ln x$, then use it as u regardless]

12. (Constrained) Optimisation

$\max f(\boldsymbol{x})\ s.t. \text{"}constraint\text{"}$. By def. $f_{x_i}(\boldsymbol{x}^*) = 0$
Necessary but <u>not sufficient</u> condition for <u>interior</u> solutions [see p6 for more conditions i.e. **Hessian**]

Lagrange: Ensures a constraint binds
$\mathcal{L}(\boldsymbol{x}, \lambda, m) = f(\boldsymbol{x}) - \lambda(\text{"}constraint\text{"})$ [λ-term is punishment for exceeding or reward for 'under'-ceding]
How do we determine λ? By def. $\mathcal{L}_i(\boldsymbol{x}^*, \lambda^*) = 0$ and $\mathcal{L}_\lambda(\boldsymbol{x}^*, \lambda^*) = 0$. Solve equations, compare solutions
λ^* can be thought of as 'willingness to pay to violate constraint by one unit' (e.g. MU of income)
For multiple constraints [i.e. $g^i(\boldsymbol{x})$]: $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \sum_{i=0}^m \lambda_i g^i(\boldsymbol{x})$

Envelope Theorem
Let x^* depend on a, which is not a choice variable [i.e. exogenous]
$V(a) = f(x^*(a), a) = \max_x f(x, a)$ s.t. $g(x, a) = 0$
i.e. $V(a) = \mathcal{L}(x^*(a), \lambda^*(a), a) = f(x^*(a), a) - \lambda^* g(x^*(a), a)$
$\frac{\partial V(a)}{\partial a} = \frac{\partial \mathcal{L}^*}{\partial a} = \frac{\partial \mathcal{L}}{\partial x}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}} \times \frac{\partial x^*}{\partial a} + \frac{\partial \mathcal{L}}{\partial \lambda}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}} \times \frac{\partial \lambda^*}{\partial a} + \frac{\partial \mathcal{L}}{\partial a}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}} \times \frac{\partial a}{\partial a}$
Note: By FOC def. $\frac{\partial \mathcal{L}}{\partial x}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0$; $\frac{\partial \mathcal{L}}{\partial \lambda}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0$; also $\frac{\partial a}{\partial a}=1$
Hence simplifies to $\frac{\partial V(a)}{\partial a} = \frac{\partial \mathcal{L}}{\partial a}\Big|_{\substack{x=x^*\\ \lambda=\lambda^*}}$ where $\mathcal{L}(x, \lambda, a) = f(x, a) - \lambda\, g(x, a)$
"Rate of change of the optimal value (V) with respect to the parameter (a) = rate of change of $\mathcal{L}$ with respect to parameter evaluated at the optimal solution $\left(\frac{\partial \mathcal{L}}{\partial a}\Big|_*\right)$. No indirect effect!

13. Vectors/Matrices

$\boldsymbol{x} = (x_1, x_2) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ | $\boldsymbol{e^1} = (1,0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ | $\boldsymbol{e^2} = (0,1) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
Addition: $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$ | Scalar: $c\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}$ | Product: $x \cdot y = x_1 y_1 + x_2 y_2$

Linear Transformations
$T(0) = 0$
$T(\boldsymbol{x} + \boldsymbol{y}) = T(\boldsymbol{x}) + T(\boldsymbol{y})$
$T(c\boldsymbol{x}) = cT(\boldsymbol{x})$
$(T \circ U)(\boldsymbol{x}) = T(U(\boldsymbol{x}))$
Note: $\boldsymbol{x} = x_1 \boldsymbol{e^1} + x_2 \boldsymbol{e^2}$ so $T(\boldsymbol{x}) = x_1 T(\boldsymbol{e^1}) + x_2 T(\boldsymbol{e^2})$
i.e to describe T it is sufficient to describe what T maps each e^i to
e.g. Rotation 90° clockwise: $x_1 T(\boldsymbol{e^1}) + x_2 T(\boldsymbol{e^2}) = x_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Matrix Rules
$A(BC) = (AB)C$ | $AB \neq BA$ | $(AB)^{-1} = B^{-1}A^{-1}$ | $(AB)_{ij} = A_{i1}B_{1j} + \cdots + A_{in}B_{nj}$
Identity Matrix: $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ [where $AI = A$] | Transpose Matrix: $\begin{bmatrix} x \\ y \end{bmatrix}^T = \begin{bmatrix} x & y \end{bmatrix}$

If $ax_1 + bx_2 = y_1$ and $cx_1 + dx_2 = y_2$ then $A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
Hence, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ where $A^{-1} = \frac{1}{ad-bc} \times \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ [if det $(ad - bc = 0)$ then A is singular]

3x3 Matrix

- Minor: 2x2 determinant when eliminating an element's row & column e.g. $\begin{pmatrix} a & b & c \\ d & \boldsymbol{e} & f \\ g & h & i \end{pmatrix}$
- Cofactor: Minor multiplied by the element's $(-1)^{i+j}$: $\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$
- Determinant: Sum of a row/column when Cofactor is multiplied by its element
- Inverse: Find Cofactors of elements (do not $\times$ elements); Transpose $\begin{pmatrix} a & d & g \\ b & e & h \\ c & f & i \end{pmatrix}$, divide by det

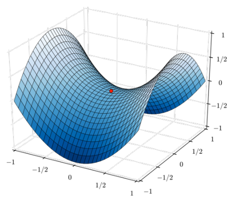
Gaussian Elimination: Apply row operations to $\left[\begin{matrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{matrix} \middle| \begin{matrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{matrix}\right]$ until $LS = I$ then $RS = A^{-1}$

14. Hessian

Mathematical Intuition

Multi-variable functions $f'(x^*, y^*) = 0$ insufficient condition for local min/max due to **saddle points** [See 16. Taylor Series]

1. $F(x_1 + h_1, x_2 + h_2) \approx F(x_1, x_2) + F_1(x_1, x_2)h_1 + F_2(x_1, x_2)h_2 + \frac{1}{2}[F_{11}(x_1, x_2){h_1}^2 + 2F_{12}(x_1, x_2)h_1h_2 + F_{22}(x_1, x_2){h_2}^2]$
2. $F(x_1 + h_1, x_2 + h_2) \approx F + \frac{1}{2}[F_{11}{h_1}^2 + 2F_{12}h_1h_2 + F_{22}{h_2}^2]$
3. $F(x_1 + h_1, x_2 + h_2) - F \approx \frac{1}{2}[F_{11}{h_1}^2 + 2F_{12}h_1h_2 + F_{22}{h_2}^2]$
4. $F(x_1 + h_1, x_2 + h_2) - F \approx H_{11}{h_1}^2 + 2H_{12}h_1h_2 + H_{22}{h_2}^2$
5. $F(x_1 + h_1, x_2 + h_2) - F \approx [h_1 \quad h_2]\begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix}\begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$



To be definite, graph never crosses axis i.e. quadratic has no solutions

1. $H_{11}{h_1}^2 + 2H_{12}h_1h_2 + H_{22}{h_2}^2$
2. $H_{11}(h_1/h_2)^2 + 2H_{12}(h_1/h_2) + H_{22}$
3. $b^2 - 4ac < 0 \Longrightarrow 4H_{12} - 4H_{11}H_{22} < 0 \Longrightarrow H_{11}H_{22} - (H_{12})^2 > 0 \Longrightarrow \det(H) > 0$

Conditions

Iff **ve+ Definite** i.e. $[h \quad k]H\begin{bmatrix} h \\ k \end{bmatrix} > 0$ for all $(h, k) \neq (0,0)$ (i.e. $det(H) > 0$ and $H_{11}, H_{22} > 0$) then min

Iff **ve- Definite** i.e. $[h \quad k]H\begin{bmatrix} h \\ k \end{bmatrix} < 0$ for all $(h, k) \neq (0,0)$ (i.e. $det(H) > 0$ and $H_{11}, H_{22} < 0$) then max

Iff min then $[h \quad k]H\begin{bmatrix} h \\ k \end{bmatrix} \geq 0$ for all (i.e. $det(H) \geq 0$ and $H_{11}, H_{22} > 0$) then **ve+ Semi-Definite**

Iff min then $[h \quad k]H\begin{bmatrix} h \\ k \end{bmatrix} \geq 0$ for all (i.e. $det(H) \geq 0$ and $H_{11}, H_{22} < 0$) then **ve- Semi-Definite**

If $det(H) = 0$ then indeterminate; If $det(H) < 0$ (i.e. H_{11}, H_{22} opposite signs) then saddle point

f is convex iff H is **ve+ Semi-Definite** for all (x^*, y^*) and strictly convex iff H is **ve+ Semi-Definite**
f is concave iff H is **ve- Semi-Definite** for all $(x^*, y^*)$and strictly concave iff H is **ve- Semi-Definite**

15. Exam Application

Rates of Change

0. $Q(L) = f(a, b, L)$
1. $\frac{\partial Q}{\partial L} = 0$ to get L^*. Don't know if max or min. Check with boundary solutions (or Hessian). Sub for Q^*.
2. $\frac{\partial Q^*}{\partial a}$ [rate of change] $\times$ 0.x[magnitude of change]. Note this is approximate as IRL L^* adjusts.

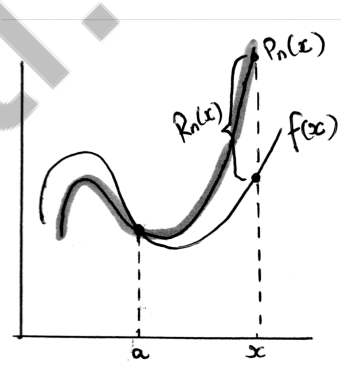
How do inputs (r, w) change factors of production (K^*, L^*)?

0. $\pi = \max pf(K, L) - (wL + rK)$
1. FOC: $\pi_K = pf_K(K^*, L^*) - r = 0 \mid \pi_L = pf_L(K^*, L^*) - w = 0$
2. SOC: $\frac{\partial \pi_K}{\partial w} = pf_{KK}\frac{\partial K^*}{\partial w} + pf_{KL}\frac{\partial L^*}{\partial w} = 0 \mid \frac{\partial \pi_L}{\partial w} = pf_{LL}\frac{\partial L^*}{\partial w} + pf_{KL}\frac{\partial K^*}{\partial w} - 1 = 0$
3. Matrix: $\begin{bmatrix} f_{KK} & f_{KL} \\ f_{KL} & f_{LL} \end{bmatrix}\begin{bmatrix} \frac{\partial K^*}{\partial w} \\ \frac{\partial L^*}{\partial w} \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{p} \end{bmatrix} \to \begin{bmatrix} \frac{\partial K^*}{\partial w} \\ \frac{\partial L^*}{\partial w} \end{bmatrix} = \frac{1}{f_{KK}f_{LL} - f^2{}_{KL}}\begin{bmatrix} f_{LL} & -f_{KL} \\ -f_{KL} & f_{KK} \end{bmatrix}\begin{bmatrix} 0 \\ \frac{1}{p} \end{bmatrix}$
4. Evaluate: $-f_{KL}$ shows if substitute or complement (i.e. if income or sub effect is stronger)
   [Note only has a solution for DRTS and this only unique if concave]

From cost function to production function

0. $C(r, w, y) \Longrightarrow \mathcal{L} = Kr + Lw - \lambda[f(K, L) - y]$
1. Via envelope theorem: $\frac{\partial C}{\partial r} = \frac{\partial \mathcal{L}}{\partial r}\Big|_{K^*, L^*, \lambda^*} = K^*$ and $\frac{\partial C}{\partial w} = \frac{\partial \mathcal{L}}{\partial r}\Big|_{K^*, L^*, \lambda^*} = L^*$
2. Combine $C_w = L^*$ and $C_r = K^*$ cancelling r, w terms

16. Taylor Series



$f(x)$ [centred @ $x = a$] $\approx P_n(x) = f(a) + f'(a)(x - a) + \frac{f''(a)}{2!}(x - a)^2$

$R_n(x) = f(x) - P_n(x) \leq \frac{f^{(n+1)}(c)}{(n+1)!}(x - a)^{n+1}$ for some c $\in [a, x]$

[Note $f(a) = P_n(a)$; choose c to maximise)

Can check if over or underestimate by looking at sign of next Taylor terms

17. Difference Equations

<u>Linear, First Order, Autonomous</u>: $x_t = bx_{t-1} + a$

$$x_t = b^t x_0 + \frac{1 - b^t}{1 - b}a = b^t\left(x_0 - \frac{a}{1 - b}\right)[\text{deviation}] + \frac{a}{1 - b}[\text{steady state}]$$

Steady state: $x^*{}_t = x_m = x_{m-1}$. Not necessary that a steady state will be reached or converged to

- If $|b| < 1$, converges to $\frac{a}{1-b}$ [Note doesn't depend on x_0!]
- If $b = 1$, constant divergence [Note $x_t = x_0 + ta$]
- If $|b| > 1$ no convergence; If $b = -1$ oscilates.

Growth Case

- <u>Discrete Growth</u>: $A = P(1 + r_d)^t \Longrightarrow \ln(P) + t\ln(1 + r_d) \approx \ln(P) + tr_d$ for small values
- <u>Continuous Growth</u>: $A = Pe^{tr_c} \Longrightarrow \ln(P) + tr_c$ [Discrete Growth approximates Continuous Growth!]

18. Differential Equations

Differential Equation: Relates function to its derivatives

<u>Separable</u>: $\frac{\partial y}{\partial t} = y'(t) = F'(y)G'(t) \Longrightarrow \int F'(y)\partial y = \int G'(t)\partial t \Longrightarrow F(y) = G(t) + C$

<u>Linear, First Oder, Autonomous.</u>: $\frac{\partial y}{\partial t} = y'(t) = by(t) + a \Longrightarrow y(t) = Ae^{bt} - \frac{a}{b}$ [solve for $t = 0$ to find A]

- General Solution: $y(t) = Ae^{bt} - \frac{a}{b}$
- Let $z(t) = y_1(t) - y_2(t)$. Note Complementary solution $z'(t) = bz(t)$
- Stationary Solution: y^P s.t. $y^{P\prime}(t) = 0$ for all t. Hence in form $by^P + a = 0 \Rightarrow y^P = -\frac{a}{b}$
- Stable Solution: $y^P = -\frac{a}{b}$ if unique and always converged to as $t \to \infty$. Requires $b < 0$

Worked Example

0. $q^D = a_0 - a_1 p(t)$; $q^S = b_0 - b_1 p(t)$; $\frac{\partial p}{\partial t} = \lambda(q^D - q^S)$
1. Equilibrium when $q^D = q^S \Rightarrow a_0 - a_1 p^* = b_0 - b_1 p^* \Rightarrow p^* = \frac{a_0 - b_0}{a_1 + b_1}$ [Particular solution]
2. $\frac{\partial p}{\partial t} = \lambda\{[a_0 - b_0] - [a_1 + b_1]p(t)\} = \lambda[a_1 + b_1][p^* - p(t)] = c[p^* - p(t)] \Rightarrow \frac{\partial p}{\partial t} + cp(t) = cp^*$
3. $p = p^* + Ae^{-ct}$ [General solution]

Statistics

01. Fundamentals

Sample statistics are not population statistics | Correlation does not imply causation

- **Frequentists**: observed f represent p of future outcomes: $\hat{p} = \frac{\#correct}{\#total}$
- **Bayes-ists**: $prior + new\ data \Longrightarrow improved\ belief$: $\hat{p} = P(A|prior) = \frac{P(prior|A)P(A)}{P(prior)}$

$P(X$[outcome of random variable] $= x$[a number in sample space]$)$

Measures of Central Tendency: Arithmetic mean $\left(\frac{\sum x_i}{n}\right)$; Geometric mean $\left(\sqrt[n]{x_1 \dots x_n}\right)$; median; mode

Measures of Dispersion: range $(L - S)$; IQR $(Q_3 - Q_1)$; var (see 04. Moments); skewness= $\frac{\frac{1}{n}\sum(x_i - \bar{x})^3}{\sigma^3}$ [how grouped]; kurtosis= $\frac{\frac{1}{n}\sum(x_i - \bar{x})^4}{\sigma^4}$ [how peaked]

Measures of Relationships: co-var $Cov = E([X - E(X)][Y - E(Y)]) = E(XY) - E(X)E(Y) = \sigma_{XY}$, for sample $= \frac{1}{n[-1]}\sum([x_i - \mu_x][y_i - \mu_y])$or $\iint (x_i - \mu_x)(y_i - \mu_y)\, f(x,y)dxdy$ [Note: = 0 if independent]; correlation $r = \frac{Cov(X,Y)}{\sqrt{Var(x)Var(y)}}$ for sample $= \frac{S_{XY}}{S_X S_Y}$ where $S_{XY} = \frac{1}{n-1}\sum([X - \bar{X}][Y - \bar{Y}])$

02. Single Probability (and Distribution)

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ [if mutually exclusive $\cap = 0$]
- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$ [if independent also $= P(A)P(B)$]
- Hence $P(B|A)$[i. e. $P(B)$ given A] $= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$

  Use last term for P(T1-error) questions e.g. $P(ve^+\ test\ |\ disease) = 0.99$, $P(disease) = 0.01$

  $P(not\ disease|ve^+\ test) = \frac{P(ve^+\ test\ |\ disease)P(disease)}{P(ve^+ test)} = \frac{0.99 \times 0.11}{0.99 \times 0.11 + 0.99 \times 0.11} = 0.5$

- **Permutations** (i.e. order relevant) $= nPr\ \frac{n!}{(n-r)!} = \frac{\#objects!}{\#repeats!}$
- **Combinations** (i.e. order irrelevant) $= nCr\ \frac{n!}{(n-r)!r!}$

- **Probability Density Function**: gives p for a <u>continuous</u> random variable: $P(a \le X \le b) = \int_a^b f(u)du$
- **Probability Mass Function**: gives p for a <u>discrete</u> random variable: $P(X = x) = f(x)$
- **Cumulative Distribution Function:** $F(x) = P(X \le x) = \int_{-\infty}^{x} f(u)du$ [$p \to 1$ as $x \to \infty$]

Note: $F(x) = \int_{-\infty}^{x} f(x)dx$ is incoherent! Must be $F(x) = \int_{-\infty}^{x} f(u)du$

Types of PMFs

- **Uniform:** $P(X = x|N) = \frac{1}{N}$
- **Bernoulli:** $P(X = x|p) = p^x(1-p)^{1-x}$ $[x = 0,1]$
- **Poisson:** $P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ $[x = 0, \dots, n]$
- **Geometric:** $P(X = x|n) = p^x(1-p)^{n-x}$ [" "]
- **Bino:** $P(X = x|n,p) = \binom{n}{x} p^x(1-p)^{1-n}$ [" "]

Types of PDFs

- **Uniform:** $f(x|a,b) = \frac{1}{b-a}$ $[a \le x \le b]$
- **Exponential:** $f(x|\beta) = \frac{1}{\beta}e^{-\frac{x}{\beta}}$ $[x \ge 0, \beta > 0]$
- **Normal:** $f(x|\mu,\sigma^2) =$ see tables

03. Joint Probability

$P(x,y) = P(X=x, Y=y)$ [if independent $= P(X=x)P(Y=y)$]
Can be PMassF (discrete) or PDensityF (continuous)

		Y_k			$\sum_{j=1}^{J} p_{jk}$
		1	2	3	
X_j	1	0.2	0.1	0.05	0.35
	2	0.05	0.2	0.05	0.3
	3	0.05	0.1	0.2	0.35
$\sum_{k=1}^{K} p_{jk}$		0.3	0.4	0.3	1.0

- Joint: $P\left(X=X_j, Y=Y_k\right) = p_{jk}$
- Marginal: $P\left(X=X_j\right) = \sum_{k=1}^{K} p_{jk}$ or $\int_{\text{lower } y}^{\text{upper } y} f_{XY}(x,y)dy$
- Conditional: $P\left(Y|X=X_j\right) = \frac{p_{jk}}{\sum_{k=1}^{K} p_{jk}}$ or $\frac{f_{XY}(x,y)}{f_Y(y)}$

Note: Converting back depends. If IID observations, then j-PDF= $f(x,y) = f(x)f(y)$ and j-PMF= $P(x,y) = P(x)P(y)$. If not, much more complicated

04. Moments

	population	sample	discrete	continuous
$E(x)$	$\mu = \frac{1}{N}\sum(X)$	$\bar{x} = \frac{1}{n}\sum X$	$\frac{1}{N}\sum xP(X=x)$	$\int xf(x)\,dx$
$Var(x)$	$\sigma^2 = \frac{1}{N}\sum(X-\mu)^2$ $= \frac{1}{N}\sum(X) - \mu^2$	$\bar{\sigma}^2 = \frac{1}{n-1}\sum(X-\bar{x})^2$ $= \frac{1}{n-1}\sum(X) - \bar{x}^2$	$\frac{1}{N}\sum x^2 P(X=x)$	$\int x^2 f(x) - \mu^2\,dx$

Notes: $E(x)$ always a number; $\not\equiv \bar{x}$ [$E(x)$ value constant, sample mean changes if repeated]

Basic Rules

$E(a) = a$	$Var(a) = 0$
$E(bX) = bE(X)$	$Var(bX) = b^2 Var(X)$
$E(X+Y) = E(X) + E(Y)$	$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X,Y)$

[Note: $Cov = 0$ if independent

05. Sampling

Observation: known value [i.e. real number] of variable $(X_1 = x_1)$
Sample: collection of variables from pop. $(X_1, \ldots, X_n)$, if IID then random. Hence, it has a distribution…
$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ [latter term requires full IID to be gained from def. of $Var(\bar{X})$]

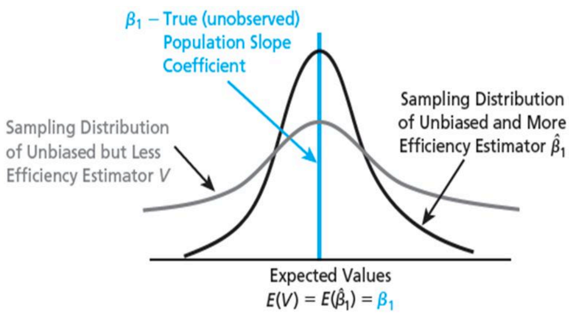
Note: μ is the population mean $\left(= \frac{1}{N}\sum X_N\right)$; $\bar{X}$ is the sample mean $\left(= \frac{1}{n}\sum X_n\right)$
$\bar{X}$[variable] $\neq \mu$[number] but $E(\bar{X})$[number] $= \mu$[number] if IID;

IRL issues: confidentiality, lying, non-response, attrition, survivorship bias (i.e. sampling on outcome)

06. Estimators

$\hat{\theta}$[estimator] is good if <u>unbiased</u>, <u>efficient</u>, and <u>random</u>

- $bias(\hat{\theta}) = E(\hat{\theta}) - \theta$. The lower, the less biased.
  If bias is consistent term, you can multiply it away
  Exam Q: Does $E(\hat{\theta}) = \int_{-\infty}^{\infty} xf(x)\,dx$
  Asymptotic unbias: as $n \to \infty, E(\hat{\theta}) \to \theta$
- Mean Squared Error= $E\left[\left(\hat{\theta} - \theta\right)^2\right] = Var(\hat{\theta}) + bias\left(\hat{\theta}\right)^2$. The lower, the more efficient.



Common Estimators

$\hat{\mu} = \bar{X} = \frac{1}{n}\sum X_i$ [sample mean; unbiased]

$\hat{\sigma}^2 = s^2$ [sample variance; biased] so adjust to…

$\hat{\sigma}^2 = S_{n-1}^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ [unbiased]

A note on notation

σ^2 = pop Var	μ = pop E
$\hat{\sigma}^2$ = pop Var est.	$\hat{\mu}$ = pop E est.
$\bar{\sigma}^2$ = sample dist. Var	$\bar{\mu}$ = sample dist. E
S^2 = sample Var [bias]	$\bar{X}$ = sample E
$\bar{\sigma}^2 = \frac{\sigma^2}{\sqrt{n}} \approx \frac{S_{n-1}^2}{\sqrt{n}}$	

07. Standard Error & CI

Confidence Intervals: range of values so $P(LC < X < UC) = p$. Use standardisation to easier calculate. Two tailed: need to divide significance level accordingly [use common sense]

Mean

$P(\bar{x} - SE < \mu < \bar{x} + SE) = p$ [where SE is Standard Error at some Confidence Level]

Note that $P(X < x) = P(Z < z) = \phi(z)$ where $X \sim N(\mu, \sigma^2)$; $Z \sim N(0,1)$; $z = \frac{x-\mu}{\sigma/\sqrt{n}}$ [$n = 1$ for single obser!]

- If $X \sim N(\mu, \sigma^2)$ then $\bar{X} \sim N\left(\bar{\mu} = \mu, \bar{\sigma}^2 = \left(\frac{\sigma}{n}\right)^2\right)$. Use $z = \frac{x-\mu}{\sigma/\sqrt{n}}$ with relevant substitutions to get $\bar{x} \pm SE$
  - can sub in $\hat{\mu}$ (i.e. $\bar{x}$) as it just 'shifts' everything
  - <u>can't</u> sub in $\hat{\sigma}$ (i.e. S_{n-1}^2) as it affects interval width (unless n is sufficiently large)
    - Use t-distribution instead of Normal to account for this
    - df = #"IID observations" − #"est. parameters for $\hat{\sigma}$" $= n - 1$ [usually have to est. $\hat{\mu}$]
- If $X \sim ?()$ or testing sample dist. apply CLT if n is sufficiently large so that $\bar{X} \sim N\left(\bar{\mu} = \mu, \bar{\sigma}^2 = \left(\frac{\sigma}{n}\right)^2\right)$
  - can sub in $\hat{\mu}$ as it just 'shifts' everything
  - <u>can</u> sub in $\hat{\sigma}$ (i.e. S_{n-1}^2) as n has to be sufficiently large to have applied CLT!
- If comparing two samples use $z = \frac{\bar{x}_a - \bar{x}_b}{\sqrt{\frac{s_a{}^2}{n_a} - \frac{s_b{}^2}{n_b}}} \sim (0,1)$

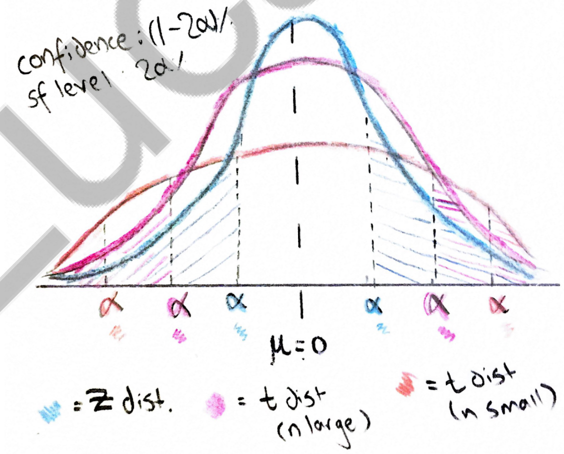
Variance

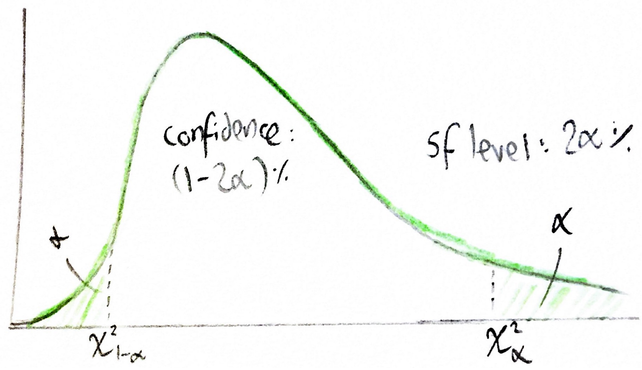
If $X \sim N(\mu, \sigma^2)$then $\frac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$ [i.e. chi-squared (asymmetric!) with $n-1$ d.o.f.]

Hence $P\left(\frac{(n-1)S_{n-1}^2}{\sigma^2} < \chi_{\alpha,n-1}^2\right) = a$ which rearranges to $P\left(\frac{(n-1)S_{n-1}^2}{\chi_{\alpha,n-1}^2} < \sigma^2\right) = a$

Proportion

- If $X \sim B(n, p)$ and $np, n(1-p)$ are large (i.e. > 10) can approx. $X \sim N(np, npq)$; divide by n [so we get a proportion not number]
- Hence, $z = \frac{x - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ [sub in $\hat{p}$ i.e. $\bar{p}$ for x]





If H_0 is in Critical region (i.e. $|H_0| > t_{crit}$), as shown by α it's rejected

08. Hypothesis Tests

(1) Define H_0 and H_1; (2) Assume H_0 is true; (3) Define rejection rule; (4) Test and draw conclusion

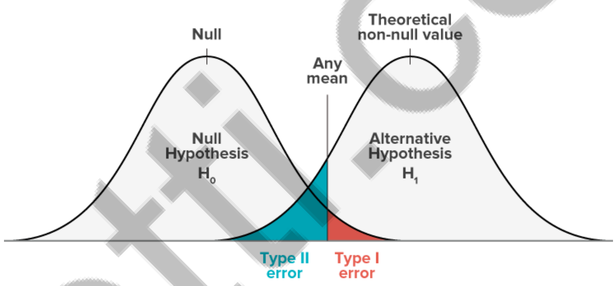
Need to know distribution | Acceptance region + Critical region $= 1$ | two-tailed tests has sig level $\div 2$ |
$\Rightarrow$ Use standardization, with $\hat{?}$ replacing x [i.e. sample data] and $?_0$ replacing $?$; see diagram above

<u>About the mean</u>: Use $\hat{\mu}$ (and $\hat{\sigma}^2$) to see if $\mu_0 = \mu$: Reject if $\left|\frac{\hat{\mu}-\mu_0}{\sqrt{\frac{\hat{\sigma}^2}{n}}}\right| > t_{crit}$ [normal or t tables]

"At the $x\%$ sig. level, the X-tailed hypothesis level is [$<$ *or* $> t_{crit}$], hence [do *or* don't]reject H_0"
<u>About the variance:</u> Use $\hat{\sigma}^2$ to see if $\sigma_0 = \sigma$: Reject if $\left|\frac{(n-1)S_{n-1}^2}{\sigma^2}\right| > t_{crit}$ [chi-squared tables]
<u>About proportions:</u> Use $\hat{p}$ to see if $p_0 = p$. If assume normal [continuity correction].
If not... Reject if $\left|\frac{\hat{p}-p_0}{\left(\frac{p_0(1-p_0)}{N}\right)}\right| > t_{crit}$ [normal tables]

Errors

	H_0 true	H_0 false
Don't reject H_0	No error	Type II error
Reject H_0	Type I error	No error



H_0: specific claim about population | H_1: alternative
$P(TI) = P(\text{reject } H_0 | H_0 \text{ false}) = \alpha$ [i.e. sig. level]
power$= 1 - P(TII) = 1 -$ ['Acceptance Range' of H_0 in H_1 distribution] $= P(\text{reject } H_0 | H_1 \text{ true})$

09. OLS Regression: Basics

Correlation: X causes Y; units matter | Regression: X related to Y; units free
$Y_i = \alpha + \beta X_i + \mu_i$ can be estimated with $\widehat{Y_i} = \hat{\alpha} + \hat{\beta} X_i$
Residual: $e_i = Y_i - \widehat{Y_i}$ with standard deviation $\sqrt{\frac{\sum(e_i)^2}{n-1}}$
Standardize: $X_i \rightarrow \frac{X_i - \bar{X}}{S_X} = \tilde{X}$ [Note: now $\tilde{\beta} = r$]

10. OLS Regression: Deriving $\hat{\alpha}, \hat{\beta}$

Want to minimise residuals i.e. (i) $\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i)$; (ii) $\sum|Y_i - \hat{\alpha} - \hat{\beta}X_i|$; (iii) $\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i)^2$
(i) Offsetting effect, infinite number of solutions
(ii) Not differentiable, harder to observe stat properties
(iii) Weight extremes more but... unique solution, will also equilibrate others. Use this

$$\min_{\hat{\alpha},\hat{\beta}} SSR = \sum(e_i)^2 = \sum\left(Y_i - \widehat{Y_i}\right)^2 = \sum\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2$$

- $\frac{\partial SSR}{\partial \alpha} = \frac{\partial SSR}{\partial e_i} \times \frac{\partial e_i}{\partial \alpha} = 2\sum(e_i) \times -1 = 0$
- $\sum(e_i) = 0$
- $\sum(Y_i - \hat{\alpha} - \hat{\beta}X_i) = 0$
- $n\bar{Y} - n\hat{\alpha} - \hat{\beta}n\bar{X} = 0$
- $\bar{Y} - \hat{\alpha} - \hat{\beta}\bar{X} = 0$
- $\bar{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$

- $\frac{\partial SSR}{\partial \beta} = \frac{\partial SSR}{\partial e_i} \times \frac{\partial e_i}{\partial \beta} = 2\sum(e_i) \times -\sum(X_i) = 0$
- $\sum(e_i X_i) = 0$
- $\sum(X_i Y_i - \hat{\alpha}X_i - \hat{\beta}X_i^2) = 0$
- $\sum(X_i Y_i) = \hat{\alpha}\sum(X_i) + \hat{\beta}\sum(X_i^2)$
- $\sum(X_i Y_i) = (\bar{Y} - \hat{\beta}\bar{X})\sum(X_i) + \hat{\beta}\sum(X_i^2)$
- $\sum(X_i Y_i) = \bar{Y}\sum(X_i) - \hat{\beta}\bar{X}\sum(X_i) + \hat{\beta}\sum(X_i^2)$
- $\sum(X_i Y_i) - \bar{Y}\sum(X_i) = \hat{\beta}\left[\sum(X_i^2) - \sum(X_i)\right]$
- $\sum(X_i Y_i) - n\bar{X}\bar{Y} = \hat{\beta}\left[\sum(X_i^2) - n\bar{X}^2\right]$
- $\hat{\beta} = \frac{S_{XY}}{S_X{}^2}$ [Intuition: slope $= r\frac{S_y}{S_x} = \frac{S_{xy}}{S_x S_y} \times \frac{S_y}{S_x} = \frac{S_{xy}}{S_x^2}$

11. OLS Regression: Proof that $\hat{\alpha}, \hat{\beta}$ are unbiased

- Let deviation of mean be $X_i^* = X_i - \bar{X}$ where $\sum(X_i^*) = 0$
- Note that $\sum(X_i^* X) = \sum(X_i^*[X_i^* + \bar{X}]) = \sum(X_i^*)^2 + \bar{X}\sum(X_i^*) = \sum(X_i^*)^2$
- Hence $\hat{\beta} = \frac{S_{xy}}{{S_x}^2} = \frac{\sum([X_i-\bar{X}][Y_i-\bar{Y}])}{\sum(X_i-\bar{X})^2} \times \frac{n-1}{n-1} = \frac{\sum(X_i^* Y_i) - \bar{Y}\sum(X_i^*)}{\sum(X_i^*)^2} = \frac{\sum(X_i^* Y_i)}{\sum(X_i^*)^2} = \frac{\sum(X_i^*[\alpha+\beta X_i+\mu_i])}{\sum(X_i^*)^2} = \beta + \frac{\sum(X_i^*\mu_i)}{\sum(X_i^*)^2}$
- $E(\hat{\beta}) = E(\beta) + E\left(\frac{\sum(X_i^*\mu_i)}{\sum(X_i^*)^2}\right) = E(\beta) + \frac{E(\mu_i)\sum(X_i^*)}{\sum(X_i^*)^2} = E(\beta)$ [Unbiased!]
- $E(\hat{\alpha}) = E(\bar{Y} - \hat{\beta}\bar{X}) = \bar{Y} - \bar{X}E(\hat{\beta}) = \alpha + \beta\bar{X} - \beta\bar{X} = \alpha$ [Unbiased!]

Distribution of $\hat{\beta} \sim N[E(\beta), V(\beta)]$ can be used to set up Hypothesis test: $zort = \frac{(b-\beta_{H_0})}{V(\hat{\beta})}$

$$V(\hat{\beta}) = V\left(\beta + \frac{\sum(X_i^*\mu_i)}{\sum(X_i^*)^2}\right) = V\left(\frac{\sum(X_i^*\mu_i)}{\sum(X_i^*)^2}\right) = \frac{1}{\left[\sum(x_i^*)^2\right]^2} V\left(\sum(X_i^*\mu_i)\right) = \frac{1}{\sum(x_i^*)^2} V(\mu_i) = \frac{\sigma^2}{\sum(X_i-\bar{X})^2} = \frac{\left[\frac{\sum(e_i^2)}{n-2}\right]}{\sum(X_i-\bar{X})^2}$$

12. OLS Regression: Extras

Transformation/Function-Forms

Non-Linear:

- Linear: $E(Y|X) = \alpha + \beta X$ ($\beta = \frac{\partial y}{\partial x}$ Elastic: constant unit between Y and X)
- Non-1 $E(\ln Y\,|X) = \alpha + \beta X$ ($\beta = \frac{\partial y}{\partial x}\frac{1}{y}$ is Semi-elastic: constant unit between $\ln Y$ and X)
- Non-2 $E(\ln Y\,|\ln X) = \alpha + \beta \ln X$ ($\beta = \frac{\partial y}{\partial x}\frac{x}{y}$ is Semi-elastic: constant unit between $\ln Y$ and $\ln X$)

Conditional Expectation Function: $E(Y|X_1, X_2)$

- Normally $E(\hat{Y}_i) = a = \bar{Y}$. Now $E(\hat{Y}_i) = a_K = \bar{Y}|(X_1, X_2)$ where K is distinct combination of X
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$
- Using $\frac{\partial SSR}{\partial \hat{\beta}_i} = 0$ get $\hat{\beta}_0 = \hat{Y}_i - \hat{\beta}_1\bar{X}_1 + \hat{\beta}_2\bar{X}_2$

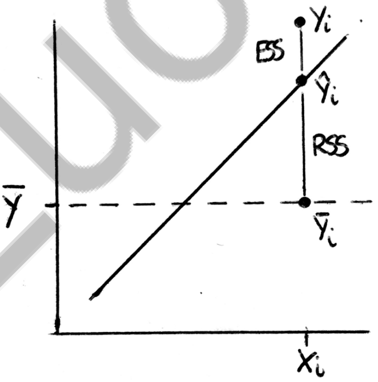
Dummy Variable: $X_i = 0,1$

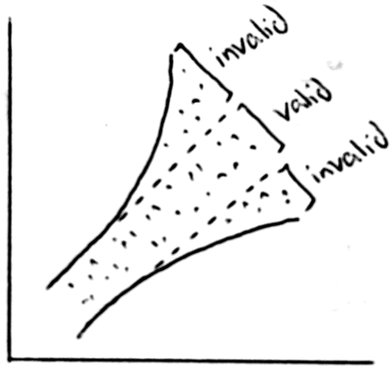
- E.g. $E(wage|[edu, sex]) = \hat{\beta}_0 + \hat{\beta}_1 edu + \hat{\beta}_2 sex$
- Where $\hat{\beta}_2 = E(wage|[edu, fem]) - E(wage|[edu, mal])$ [Note: same slope, different intercept]

R-Squared

$R^2 = \frac{RSS}{TSS}$ = % of variation explained by regression model

- *TSS*: Total Sum of Squares: $\sum(Y_i - \bar{Y})^2 = \sum(Y_i^*)^2$. Can be broken down to…
- *RSS*: Regression Sum of Squares: $\sum(\hat{Y}_i - \bar{Y})^2$ | *ESS*: Error Sum of Squares: $\sum(Y_i - \hat{Y}_i)^2 = \sum {e_i}^2$





13. OLS Regression: Assumptions

- If upheld least square estimate is BLUE (Best Linear Unbiased Estimator i.e. least sample variance)

A1. Linear Functional Form: $Y_i = \alpha + \beta X_i + \mu_i$ NOT $Y_i = \alpha + \beta X_i^2 + \mu_i$ or $Y_i = \alpha + \alpha\beta X_i + \mu_i$

A2. X_is are Non-stochastic with variation in X: [i] Sample variation in X_i (i.e. not same value) and [ii] $X_i s$ are fixed in repeated sampling and Y_is randomly drawn (i.e. treat X_is as constant)

A3. Zero conditional mean of error: $E(\mu_i|X_i) = E(\mu_i)X_i = 0$

A4. Homoskedastic errors: constant conditional variance: Errors (μ_i) don't increase as X_i increases: $Var(\mu_i|X_i) = \sigma^2 = E(\mu^2|X) - [E(\mu|X)]^2 = E(\mu^2|X)$

A5. Independent Ys: No correlation in errors (e.g. square foot and square meters)