2. Population Growth and Demographic Transition

2. 1 Theory

GDP function or national production function $\xrightarrow{}$ encapsulates technology used to make inputs into outputs

$$Y = F(K, L) \qquad (1)$$

Y: National output — GDP

K: Capital stock - Physical capital

L: Labor force = population size

GDP per capita:

$$\text{GDP per capita} \leftarrow \frac{Y}{L} = F\,(K, L)/L \qquad (2)$$

Q: Will increase in L raise GDP_PC?

A: It depends.

Note:

- Increase in L will increase GDP (numerator) which augments GDP_PC
- Increase in L increases labor force (denominator) which lowers GDP_PC

Whether rise in L helps or hurts a country depends on the relative size of the two effects.

To get a definite answer we need to differentiate (2) with respect to L:

$$d\,(Y/L)/dL = \frac{F_L \cdot L - F(K,L)\cdot 1}{L^2} = \frac{F_L \cdot \cancel{L}}{L^{\cancel{2}}} - \frac{F(K,L)}{L^2} = \frac{\underbrace{F_L}_{\to MPL}}{L} - \frac{\underbrace{F(K,L)/L}_{\to APL}}{L}$$

=

$$= \frac{1}{L}(MPL - APL)$$

Hence, if MPL > APL, a rise in L increases GPD_PC.

Note: Whether the inequality holds depends on specific functional form of the production function!

Example 1: Linear production function: Y = K L (Note; this is an IRS function)

MPL = K

APL = K

Therefore: MPL – APL = 0

Hence: A rise in L does not change GDP_PC

Example 2: $Y = K^{0.5} L^{0.5}$ Note: this is a Cobb-Douglas function with CRS)

MPL= $K^{0.5}\, 0.5 L^{-0.5} = 0.5 \frac{K^{0.5}}{L^{0.5}} = \frac{K^{0.5}}{2L^{0.5}}$

APL= $K^{0.5} \cdot L^{-0.5} = \frac{K^{0.5}}{L^{0.5}}$

Therefore, MPL – APL = $\frac{1}{2} K^{1/2} L^{-1/2} - K^{1/2} L^{-1/2} = (1/2 - 1)(K^{1/2} \cdot L^{-1/2}) = -1/2 (K^{1/2} \cdot L^{-1/2})$

Hence: A rise in L will lower GDP_PC

Conclude: Pop. growth may lower GDP_{PC}

2.2 Demographic Transition

First Stylized Fact:

Developed countries tend to have low population growth, while less developed countries tend to have high population growth.

Q: Why?

First, some definitions:

a. Crude Birth Rate (CBR) = $\underbrace{\frac{\text{live births}}{1{,}000}}_{\text{live births per 1,000 ppl}} = \frac{\text{live births}}{\text{pop}} \cdot 1{,}000$

$= \boxed{\sum_a ASBR_a * \frac{POP_a}{POP}}$

$ASBR_a$: Age-specific birth rate of age cohort "a"

Pop_a: population size of age cohort 'a'

b. Crude Death Rate (CDR) = $\frac{\text{deaths}}{\text{pop.}} \times 1{,}000$ → deaths per 1,000 ppl

$= \sum_a ASDR_a * \frac{POP_a}{POP}$

$ASDR_a$: Age-specific death rate of age cohort "a"

c. Total Fertility Rate (TFR) = $\frac{\sum_a ASBR_a \cdot (\text{length of time interval of cohort } a)}{1000}$

→ # of years in each cohort (ex. 5yrs for 10-15)

Note: For international comparison of fertility, use TFR nor CBR because only the TFR is independent of a country's age distribution.

d. Female Survival Rate (SR) =

SR = Percentage of newborn girls that survive until age 49. SR ∈ [0,1]

Note: 49 is considered end of child-bearing age for women

e. Replacement Level of Fertility (RLF)

RLF ~~RLR~~ = $2/SR$

Note that RLF measures the number of children per family needed to replace the previous generation: RLF $\geq$ 2

Example: if SR = 1 (all girls make it to age 49), then RLF=2 (the average family needs to have 2 children to replace the parent generation)

Example: if SR = .5 (only 50% of girls make it to age 49), the RLF = 4 (the average family needs to have 4 children to replace parent generation)

f. Crude Rate of Natural increase (CRNI) = CBR − CDR

g. Population Growth = $\frac{CRNI}{10}$ + net immigration rate

h. Net immigration rate = immigration rate − emigration rate

Note: since most LDCs have few immigrants, net immigration rate is zero or negative for emigration LDCs.

Let us apply above concepts to 2014 numbers for HIC, MIC and LIC assuming net immigration rate = 0 due to lack of data:

→ assuming 0 net immigration

	CBR	CDR	CBR-CDR = CRNI	Pop. Growth	SR	RLF	TFR	TFR $\geq$ RLF?
LIC	36	9	27	2.7%	70%	2.86	4.8	Yes
MIC	19	7	12	1.2%	85%	2.35	2.4	Yes
HIC	11	8	3	0.3%	95%	2.11	1.7	No

Conclusion: Differences in CBRs, not CDRs drive the observed differences in population growth between DCs and LDCs

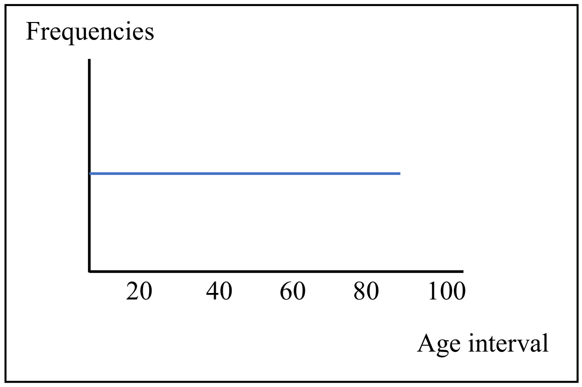
Rule: if TFR < RLF, then population growth of a country will be negative *in the long run*.

Q: We observe that TFR < RLF in HIC group, yet population growth is still positive for HIC. Why?

In the short run, even if TFR < RLF, population growth can be positive due to population momentum

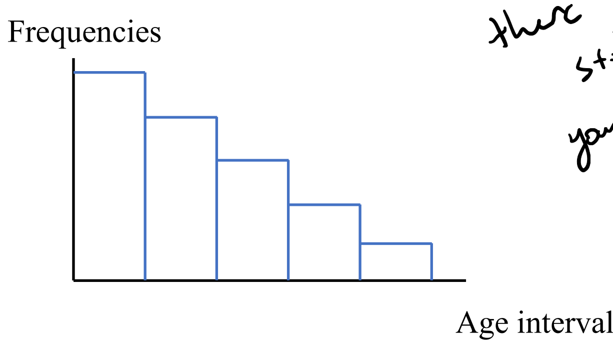
Definition: Population momentum means an age distribution that is skewed toward young.

Example: Age distribution without population momentum



constant

Example: Age distribution with population momentum



there are still many young ppl in DCs

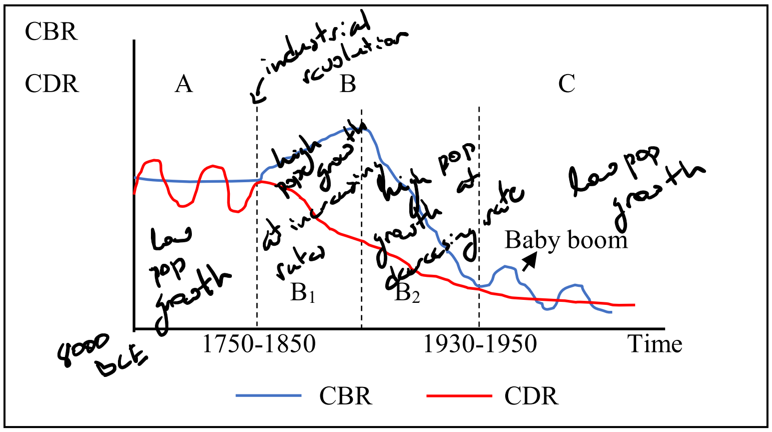
Second Stylized Fact:

HICs LICs

DCs	LDCs
Low CBR	High CBR
Low CDR	High CDR

To understand this difference, we need to look at the demographic history of today's DCs. When we do this, we observe a phenomenon known as "**demographic Transition**"

for today's HICs



There are three stages A, B, and C

A: Traditional stage:

Both CBR and CDR are high, with CDR more volatile than CBR.

Population Growth was cyclical and very low! Between 8000 BC-1650, world population increased from 5 million to 500 million, a growth rate of 0.05% per year.

B: Transitional stage:

Transitional state has 2 phases: phase 1 and phase 2.

In phase 1, CDR is falling while CBR is rising!

Population growth is positive at increasing rate (B_1): accelerating growth

In phase 2, both CDR and CBR are falling but CBR is falling faster.

Population growth is positive at a declining rate (B_2): decelerating growth

C: Modern stage:

Both CDR and CBR are low and falling modestly, with some exceptions (high CBR during baby boom).

Population growth is low, often close to zero or sometimes even negative (e.g. Japan since 2015).

Economic interpretation of stages

A: Traditional/Preindustrial Stage:

CBR: high because:

- lack of birth control
- Child labor is needed to increase family income
- response to high CDR
- kids are needed for old age support
- Religious/moral standards

CDR: high and volatile because:

- poor medical standards
- Wars
- famine
- high infant + maternal mortality

Note: infant and maternal mortality rates are particularly high

Example: JSB had 20 children with 2 wives; he had 7 children with wife 1 of which 4 died and 3 survived to adulthood; with wife 2, he had 13 children of which 6 died and 6 survived to adulthood; of a total of 20 children, 11 died and 9 made it to adulthood;

Question: How many of Bach's children did become famous/well-known musicians?

B: Transition Stage:

Phase 1:

CBR: went up because:

- income effect: as income ↑, parents want more (normal good)
- fecundity: if health of parents improve, prob. of conception increases

CDR: declined because:

- family income ↑ => more money for healthcare services
- as gov't revenues ↑ => more spending on fresh water + sanitation
- medical innovations
- healthier populations => more resistant to disease

Phase 2:

CBR: decreased because:

- Child labor laws
- Mandatory public education
- Social security
- education of women
- Secular societies
- Change in preference: less children became more popular

CDR: decreased because:

- Same reasons as in phase 1

C: Modern Stage:

CBR: decreasing except for baby boom, reasons are similar to phase 2 of transition stage

CDR: decreasing for reasons similar to transition stage; CDRs decline at decelerating rates due to diminishing returns in health care production. + diseases

Note: Since all of today's developed countries went through demographic transition, we expect that all of today's less developed countries will have to go through the demographic transition stage as well!

2.3 Economic Theories of Population Growth and Development

2.3.1 Malthus Model

Malthus (1766-1834)

Links population growth with output (food) growth.

Basic Malthus Model:

- Food supply (X) grows linearly over time (t): X= 2 t
- Population (N) grows at a geometric rate over time: $N = 2^t$

t	1	2	3	4
X	2	4	6	8
N	2	4	8	16
"GDP_PC" = X/N	1	1	3/4	1/2

$\lim_{t \to \infty} X/N = 0$

Note: X/N is food production per capita (similar to today's GDP per capita).

As t grows large, (X/N) falls to "**the subsistence level of income**". The prediction that mankind is doomed to live at the subsistence level is the reason why economics is called "**the dismal science**".

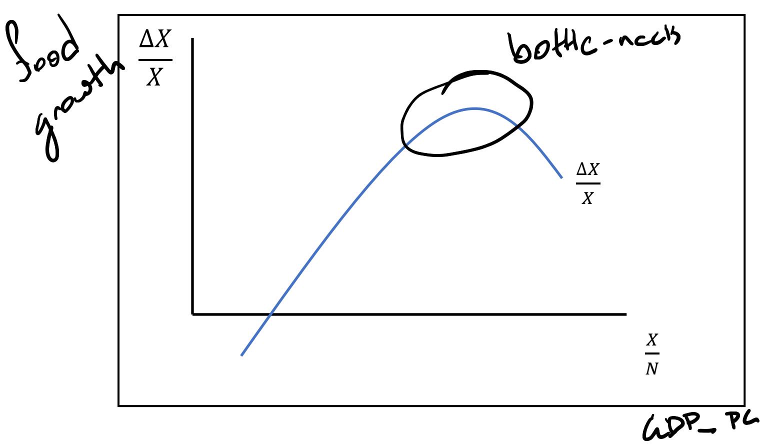
→ calory intake = calory used

Sophisticated Malthus Model:

Growth rate $= \frac{\text{change in x}}{\text{x}} = \frac{\Delta x}{x}$

$$\frac{x_t - x_{t-1}}{x_{t-1}} = \frac{\Delta x_{t-1}}{x_{t-1}} = \frac{\Delta x}{x}$$

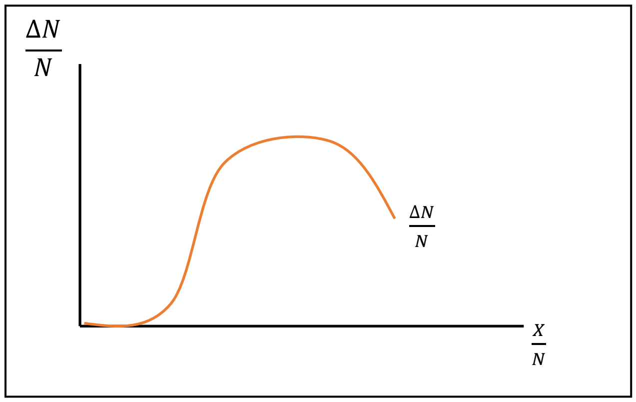
Link between output (food) growth $\frac{\Delta X}{X}$ and output per capita $\frac{X}{N}$



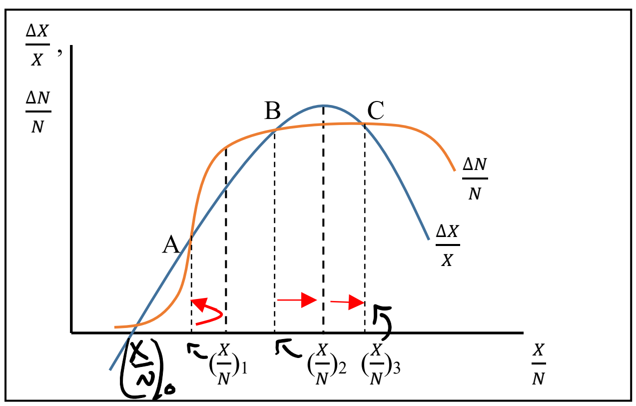
The reason for the limit to growth of X is "Bottle necks" to production:

- Scarcity of land
- Scarcity of skilled workers
- Lack of technology
- Lack of capital

Relation between per capita income $\frac{X}{N}$ and growth rate of N $\frac{\Delta N}{N}$



Graph: Sophisticated Malthus Model



Note: we discard $\left(\frac{X}{N}\right)_0$ b/c there is 0 growth in X & N

$$\text{Equilibrium} = \frac{\Delta N}{N} = \frac{\Delta X}{N}$$

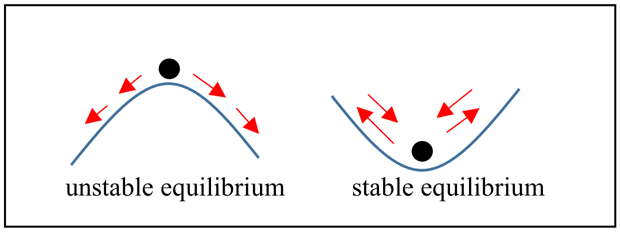
At (A): Low-income equilibrium (stable equilibrium)

At (B): Middle-income equilibrium (unstable equilibrium)

At (C): High-income equilibrium (stable equilibrium)

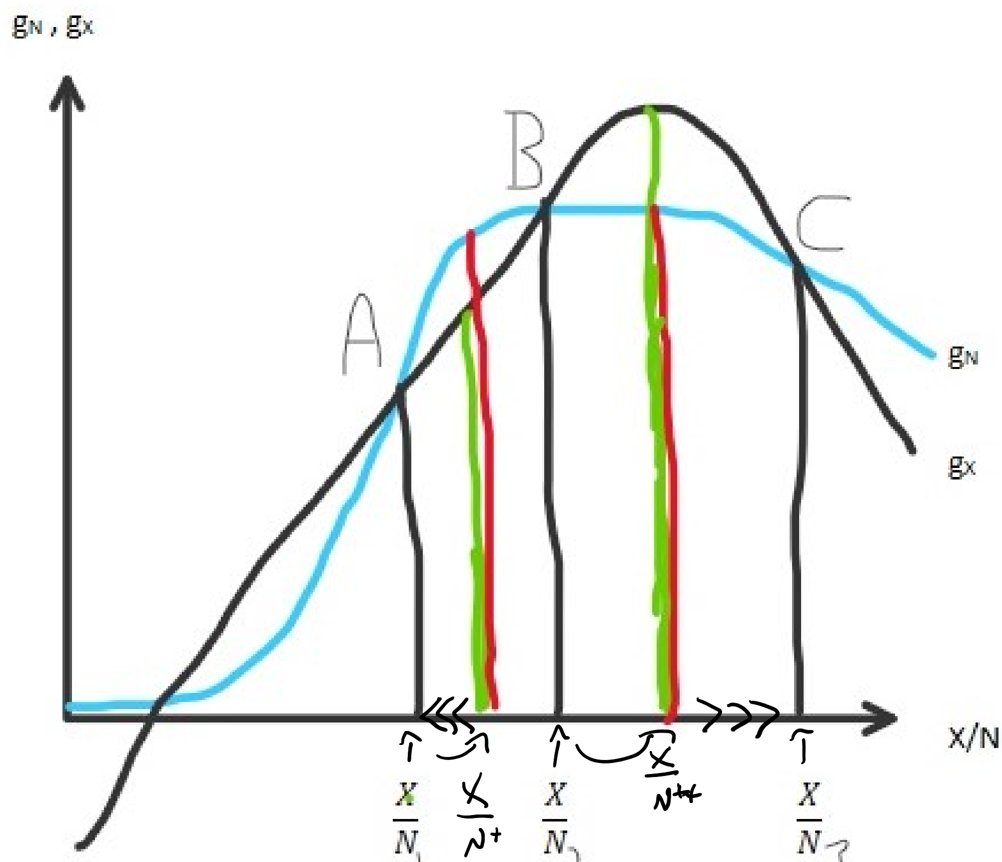
To reduce the number of equilibriums in the model, we can eliminate the unstable equilibrium. Concept of stable and unstable equilibria in physics:

Notion of stability in Physics



Notion of stability in Economics

If the economy is hit by a positive or negative shock, we can observe dynamics of the system. If the system return to initial equilibrium, then it is stable, otherwise it is unstable.



@ $\frac{X}{N^+}$: $g_x < g_n \Rightarrow \frac{X}{N}\downarrow \rightarrow$ goes back to $\frac{X}{N_1}$ $\therefore \frac{X}{N_1}$ is stable

@ $\frac{X}{N^{++}}$: $g_x > g_n \Rightarrow \frac{X}{N}\uparrow \rightarrow$ goes to $\frac{X}{N_3}$ $\therefore \frac{X}{N_2}$ is unstable

$\therefore \frac{X}{N_3}$ is stable

At $(\frac{X}{N})_1$ where $\frac{\Delta N}{N} > \frac{\Delta X}{N} \rightarrow \frac{X}{N} \downarrow \rightarrow$ go back to its original value at (A), so it is considered as a stable equilibrium.

At $(\frac{X}{N})_2$ where $\frac{\Delta N}{N} < \frac{\Delta X}{N} \rightarrow \frac{X}{N} \uparrow \rightarrow$ go further to new equilibrium value at (C), so it is considered as unstable equilibrium. Note that this result also implies that (C) is a stable equilibrium.

Conclusion:

The sophisticated Malthus model has two viable (stable) outcomes:

1) Low-income equilibrium (similar to Malthusian Trap).
2) High-income equilibrium.

Q: How to move economy from low to high equilibrium?

A: Malthus suggests to lower N by:

1) Virtue: moral restraint, which comes through education => fewer children
2) Vice: such as wars, diseases, and famines. => high CDR

Clearly, today we would focus not only on lowering N, but mostly on improving X to get to $(\frac{X}{N})_3$ at (C)

As $N\downarrow \Rightarrow \frac{X}{N}\uparrow$

As $X\uparrow \Rightarrow \frac{X}{N}\uparrow$

instead of going from $\frac{X}{N_1} \rightarrow \frac{X}{N_3}$, you can do $\frac{X}{N_1} \rightarrow \frac{X}{N_2} + \varepsilon$

↳ this will push you to the stable equilibrium at $\frac{X}{N_3}$

2.3.2 Caldwell's Theory of Intergenerational Wealth Transfers

John Caldwell (1928-2016), one of the top demographer of 20th century

In developed countries, wealth flows from old (parents) to young (children) because:

- cost of education + healthcare
- other children costs (food, housing, etc.)
- social security
- child labor laws

Therefore, small families are optimal.

In less developed countries, wealth flows from young (children) to old (parents) because:

- child labor
- old age support
- mostly elementary education
- often no schooling requirements
- non enforcement of child labor laws

Therefore, large families are optimal.

Conclusion: LDCs must reverse the wealth flow if they want to lower family size!

ask abt. colonialism

Caldwell's prediction:

In LDCs, the elite is first to adopt western life style and norms (through foreign education and access to foreign media). Since Western family size is small, the elites in LDCs will choose small family size as well.

The middle class in less developed countries imitates the elite in LDCs and thus chooses smaller family size too.

The working poor in LDCs imitate the middle class in LDCs thereby lowering family size as well ("**cultural trickledown theory**").

Problem with Caldwell's prediction:

The middle class and working poor in LDCs lack the economic means (income) to afford to have fewer children. Therefore, policy initiatives are needed to reverse intergenerational wealth flow:

1) Free public education for both boys and girls
2) Require school attendance
3) Enact child labor laws and enforce them
4) Public assistance for poor and middle class (temporary policy)
5) Introduce social security system

2.4 Population Policies

Common population policies in LDCs:

1) Laissez ~~faire~~/hands-off approach:
   - only parents determine family size
2) Information approach (about contraception):
   - information about contraception
   - subsidy for contraceptives
   - gov't supplies contraceptives
3) Economic incentives approach:
   - tax/fee on large families
   - subsidies for small families
   - provide low-cost housing for small families
   - increase legal age of marriage
4) Authoritarian approach:
   - gov't limits number of children
     Ex: China's 1 child policy
   - Enforcement through:
     - economic incentives (fees/taxes)
     - forced abortion (forced sterilization)
       ↳ India

– unintentional consequences:
  – selected abortions for girls ⇒ skewed sex ratio
  – adoption of girls abroad

- Note: China never had a uniform one-child policy: TFR in cities = 1.2 and in rural = 2.5
- Unintentional consequences of one-child policy: skewed sex ratio: 106.3 males:100 females. For young (<15), it is even worse: 117 boys to 100 girls. Why?
    - see above
    -

Handout: Was one-child policy successful? TFR Comparison: India and China

2.5 Empirical Analysis: Estimation and Hypothesis Testing

See *Table: Ranking of Countries by TFR*

2.5.1 Estimation

Linear, bivariate estimation model

Estimation method: OLS

Under classic assumption: OLS estimates are BLUE

Y: dependent variable: GDP_PC in 2013/14 (in PPP$)

X: independent variable: TFR (average 1960-2013)

Log transform

- to minimize effect of outliers
- to interpret estimated coefficients as (scale free) elasticities

$$\log Y = \alpha + B \log X + \varepsilon$$

2.5.2 Testing

Every test begins with a (null) hypothesis. Typically, that's a statement that we want to reject. In our case, the null hypothesis is straightforward:

H_0 (Null Hypothesis): There is no statistically significant correlation between X (here TFR) and Y (here GDP_PC).

We also need an alternative hypothesis in case we reject the null hypothesis:

H_a (Alternative Hypothesis): The correlation between X and Y is statistically significant.

Note: if we can reject H_0, we accept H_a.

Now, we can address the following question:

Q: Is the estimated relationship between X and Y statistically and economically significant (meaningful)?

2.5.2.1 Statistical Significance

To answer this question, we use a simple Goodness of Fit test (Wald test).

The Wald test statistic is given by

$$\text{Wald} - \text{test statistic} = \frac{R^2}{1-R^2} \frac{T-K}{K-1}$$

Where

- R^2 is the coefficient of determination (it measures the fraction of the variation in Y explained by the variation in X, hence: $R^2 \in [0,1]$
- T: sample size
- k: number of regressors (= number of estimated coefficients)

Since $R^2 = .55$, $T = 187$ and $k = 2$

$$\text{empirical F} - \text{test statistic} = \frac{.55}{.45} \frac{187-2}{1} = \frac{185(.55)}{.45} = 226.1$$

empirical Wald-test statistic

Note: To determine statistical significance, we must compare the *empirical* Wald-test statistic to the corresponding *theoretical* test statistic!

First, note that our empirical Wald-test statistic is a random variable.

Q: What is random variable?

A random variable is a mapping (function) between random outcomes and the corresponding probabilities.

Example 1: Coin flip

Random outcomes: head or tail; corresponding probabilities: .5, .5

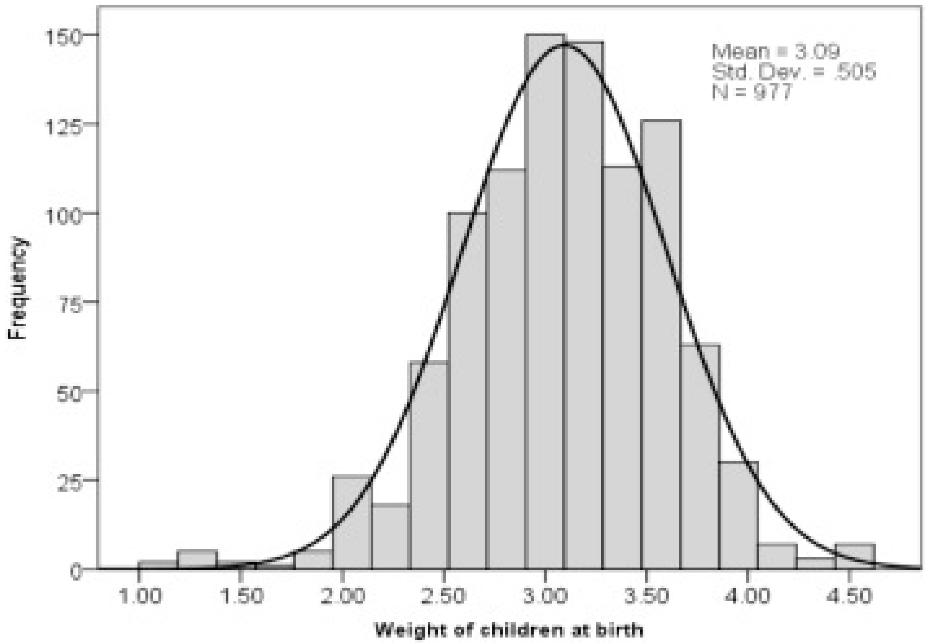
Example 2: Throwing a dice

Random outcomes: 1, 2,3,4,5,6; corresponding probabilities: 1/6 each

Example 3: Birth weight

Random outcome: weight at birth in kg; corresponding probability: approximately a Normal distribution

Figure 1: Shape of Birth Weight distribution in Oman



Note: Examples 1 and 2 are examples of uniform distributions

Q: Why is our empirical F-test statistic a random variable?

A: If we repeated the estimation procedure with different subsamples of, say, 100 countries chosen at random, we would get a different Wald-test value for each estimation. If we plotted the frequencies (probabilities) of the obtained Wald-test values, the plotted frequencies would be in the shape of a distribution.

Q: What is the *specific* shape of this distribution?

A: It is an F-distribution.

Q: Why F-distribution?

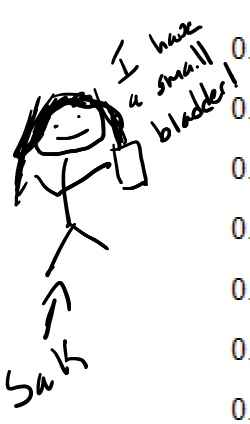
A: Recall that the most important element of the empirical Wald test is R^2, the coefficient of determination.

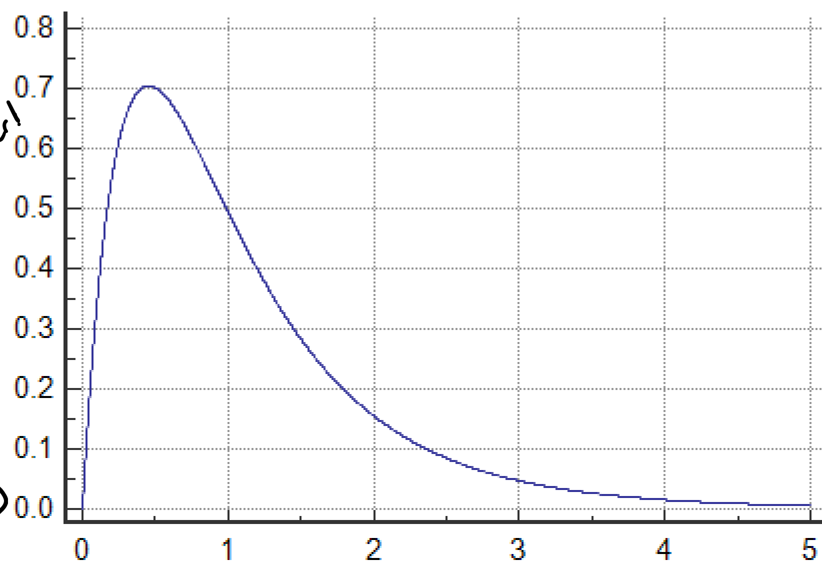
R (=r), the correlation coefficient, is normally distributed;

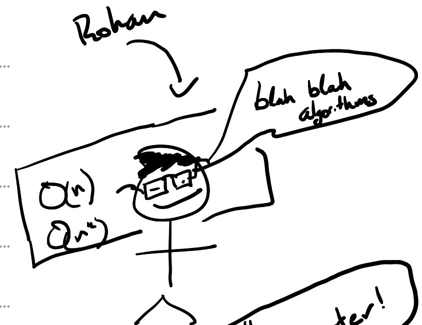
R^2 is Chi-Squared distributed (the product of two Normal distributions is a Chi-Square distribution)

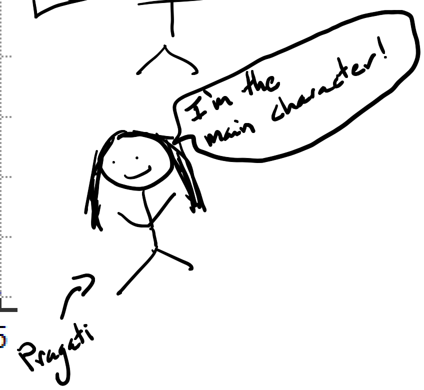
$R^2/(1-R^2)$ is F-distributed (the ratio of two Chi-Square distributions is an F-distribution)

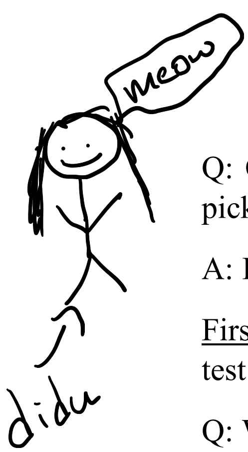
Figure 2: Shape of F-distribution (probability density function)











Q: Of the many values of F (on the horizontal axis in Figure 2), which one should we pick as the comparison value for our empirical Wald test statistic?

A: It depends.

First, we need to determine α (alpha), the critical value aka the significance level of our test. Typically, most researchers pick α=0.05 (5 percent) or α=0.01 (1 percent).

Q: Why such small values of α?

A: The critical value (α) determines the type I error of the test. Recall that the type I error is the probability of rejection the null hypothesis when it is actually true. We don't want to do this and that's why we pick a small value for α.

In this class, we pick α = 0.05 (5 percent).

<u>Second</u>, the theoretical F-distribution has 2 degrees of freedom:

- the number of regressors in the regression – 1, i.e. k-1
- the sample size, i.e. T

See handout: From the F-distribution table (for α=0.05), we can now pick the theoretical F-test value.

First we pick the appropriate column by calculating the value for k-1. Since k=2, we get 2-1=1. Hence we pick the first column in the F-distribution table.

Second, we pick the appropriate row by finding the row that is closest to our sample size. Since T=187 in our sample, we pick 120 as the closest row number.

Using the first column and row = 120, we find the theoretical F-test value 3.92 which we round to 3.9 for simplicity.

Now, we can finally complete our Wald test for statistical significance:

- if our empirical Wald test value > theoretical F-distribution value, we reject H_0 and accept H_a
- if our empirical Wald test value < theoretical F-distribution value, we cannot reject H_0

Here: Since the empirical Wald-test value = 226 > theoretical F-distribution value = 3.9

we reject the null hypothesis of no correlation between X (TFR) and Y (GDP_PC).

Hence, the correlation between TFR and GDP_PC is statistically significant.

2.5.2.2 Economic Significance

To test for economic significance, we simply compared the absolute value of the β estimate to some predetermined value.

Since the β estimate is an elasticity (due to the log-log specification of the estimation regression), we want the estimated elasticity to be sufficiently different from zero. Typically, elasticities > 0.5 will meet that threshold.

Hence, our test for economic significance is:

| Estimated β | > 0.5

Since the estimated β is -1.9 in our case, the absolute value of this estimate is 1.9 which is greater than 0.5.

Thus, the relationship between TFR and GDP_PC is economically Significant.

2.5.3 Interpretation of β Estimate

Since our β estimate of _______ measures the value of an elasticity due to the log-log specification of our estimation regression, it has the following interpretation:

A 1% increase in TFR is associated with a _____% _________ in GDP_PC.

Note that the rise in TFR corresponds to a ___________ in GDP_PC since the β estimate is _____________________.

2.5.4 Correlation and Causation

Important fact: Correlation does not imply causation!

Why not?

- Spurious regression (omitted variable bias)
  Example: Fire damage is correlated with number of fire engines fighting the fire. This does not imply that a greater number of fire engines causes more fire damage. Instead, a third (omitted) variable, the size of the fire, is causing the fire damage.

- Reversed causality
  Example: Students who are smokers have lower grades. That does not mean that smoking must cause low grades. Indeed, the opposite could be true: low grades may cause students to smoke. Also, a third (omitted) variable, extrovert personality, may be the cause of smoking as extroverts tent to have lower grades and are more likely to smoke.

- Measurement error and bias
  - Measurement error
    - Mistakes in measuring dependent and/or independent variables
      - Example: Survey questions → survey error/bias
  - Selection bias
    - The sample is not representative of the underlying population
      - Example: Assume we only observed TFR for HICs. Since all HICs have low TFRs, we will not find a significant correlation between TFR and GDP_PC for this group.