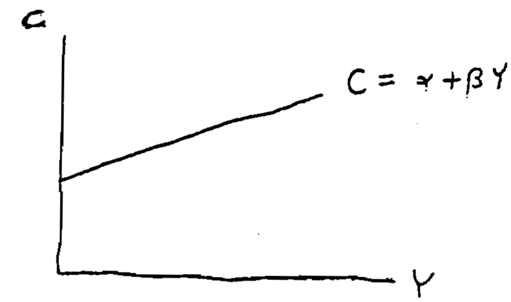
REGRESSION UNDER IDEAL COND.

SIMPLE REGRESSION

- Attempt to measure relationship between two variables

e.g. Suppose you believe in a <u>Consumption fn.</u>
$C = \alpha + \beta Y$
- <u>exact</u> relationship



In fact, observe this.
More or less a st. line
- not exactly.

And you want to "fit" a line to it.

<u>Defn:</u> A <u>simple</u> (2-variable) <u>regression model</u> is given by

$$y_i = \alpha + \beta x_i + \epsilon_i \qquad i = 1, \ldots, N$$

where i indexes observations
y = dependent variable
x = explanatory (indep) variable
ϵ = error term
α = intercept of line
β = slope

Outline for part 2.

2. Regression Under Ideal Conditions

a) Simple Regression
   i) Algebra of Least Squares
   ii) Statistical Properties of LS

b) Multiple Regression
   i) Algebra of LS
   ii) Statistics of LS

c) Multiple Regression Re-examined
   i) Algebra of LS
   ii) Statistics of LS
   iii) Asymptotic Properties
   iv) Hypothesis Testing

In a) and b) we'll derive important concepts with summation (Σ) notation. In c) we'll derive the same concepts again, and more — in a more rigorous fashion with Matrix Notation. In this section we will employ all the tools learned in parts 0. and 1. (Matrix Algebra Review and Statistical Review).

Fitting a line $\Longleftrightarrow$ estimating α & β

Define <u>fitted line</u> as $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

$\hat{\alpha}$ = intercept of fitted line
$\hat{\beta}$ = slope

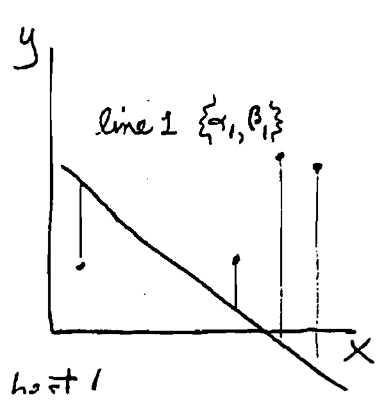
<u>Note:</u> Given N observations on X & Y;
In fitting a line, you're presuming that the relationship is in truth <u>linear</u>.
You use 2 d.f.:
in choosing a line with a particular <u>slope</u>
& " " " <u>intercept</u>

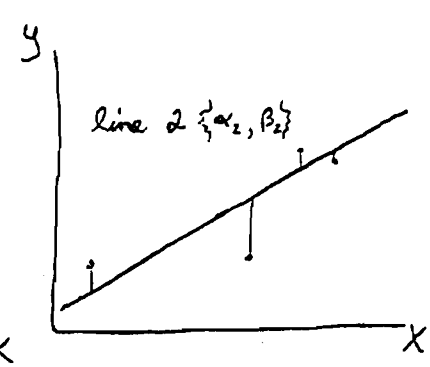
Question: Criterion for choice of line.

→ <u>Least Squares</u>: Choose $\hat{\alpha}$ & $\hat{\beta}$ by <u>minimizing</u>

$$SSE(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

<u>Picture</u>: look at hypothetical data.
Draw arbitrary line (1).
$\forall i$ calculate $(y_i - \hat{y}_i)$,
square & add.





Line 1 terrible: Line 2 better: L.S. best!

Why L. S. ?

i) e.g. why not least errors ; $\sum_i (y_i - \hat{y}_i)$

-- big positive errors could cancel out
big negative errors ! draw picture ! --- line 1

ii) e.g. why not least absolute errors ; $\sum_i |y_i - \hat{y}_i|$

Reasons :

1) Easy.
2) Intuitively, want to penalize $SSE(\hat{\alpha}, \hat{\beta})$ more severely for points farther away from line (outlyers). Hence L.S.E. rather than L.A.E.

* 3) Desireable statistical properties.
- MLE, consistent, efficient, Blue, as. eff., ...

Note ; (i) $\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \bar{y} \sum x_i - \bar{x} \sum y_i + \sum \bar{x}\bar{y}$

$$= \sum x_i y_i - N\bar{x}\bar{y} - N\bar{x}\bar{y} + N\bar{x}\bar{y}$$

$$= \sum x_i y_i - N\bar{x}\bar{y}$$

(ii) $\sum_{i=1}^{N} (x_i - \bar{x})^2 = \sum x_i^2 - N\bar{x}^2$

similar algebra

$$y_i = \alpha + \beta x_i + \epsilon_i \qquad i = 1, \ldots, N$$

Algebra of LS: will derive algebra 1st; then to the statistics

THEOREM: The LS estimators are

$$\hat\beta = \frac{\sum_{i=1}^N (x_i - \bar x)(y_i - \bar y)}{\sum_{i=1}^N (x_i - \bar x)^2} \; /\!/ \; \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^N x_i y_i - N \bar x \bar y}{\sum_{i=1}^N x_i^2 - N \bar x^2}$$

$$\hat\alpha = \bar y - \hat\beta \bar x$$

Proof #1 (Calculus):

Minimize $SSE(\hat\alpha, \hat\beta) = \sum_{i=1}^N (y_i - \hat\alpha - \hat\beta x_i)^2$ wrt $\hat\alpha, \hat\beta$

$$\frac{\partial SSE}{\partial \hat\alpha} = \sum_{i=1}^N 2(y_i - \hat\alpha - \hat\beta x_i)(-1) = -2 \sum_{i=1}^N (y_i - \hat\alpha - \hat\beta x_i) = 0$$

$$\frac{\partial SSE}{\partial \hat\beta} = \sum_{i=1}^N 2(y_i - \hat\alpha - \hat\beta x_i)(-x_i) = -2 \sum_{i=1}^N x_i (y_i - \hat\alpha - \hat\beta x_i) = 0$$

NORMAL EQUATIONS:

$$\sum_{i=1}^N (y_i - \hat\alpha - \hat\beta x_i) = 0 \quad \left[= \sum_i (y_i - \hat y_i) \right]$$

$$\sum_{i=1}^N x_i (y_i - \hat\alpha - \hat\beta x_i) = 0 \quad \left[= \sum_i x_i (y_i - \hat y_i) \right]$$

1st normal equation says

$$\sum_{i=1}^N y_i - N\hat\alpha - \hat\beta \sum_{i=1}^N x_i = 0$$

$$N\hat\alpha = \sum_i y_i - \hat\beta \sum_i x_i$$

$$\hat\alpha = \bar y - \hat\beta \bar x$$

2nd says $\sum_{i=1}^{N} x_i y_i - \hat{\alpha} \sum_i x_i - \hat{\beta} \sum_i x_i^2 = 0$

$$\sum_i x_i y_i - (\bar{y} - \hat{\beta}\bar{x}) \sum_i x_i - \hat{\beta} \sum_i x_i^2 = 0$$

$$\sum_i x_i y_i - \bar{y} \sum_i x_i = \hat{\beta}\left[\sum_i x_i^2 - \bar{x} \sum_i x_i \right]$$

$$\sum_i x_i y_i - N\bar{x}\bar{y} = \hat{\beta}\left[\sum_i x_i^2 - N\bar{x}^2 \right]$$

$$\hat{\beta} = \left(\sum_i x_i y_i - N\bar{x}\bar{y}\right) \Big/ \left(\sum_i x_i^2 - N\bar{x}^2\right)$$

Proof #2 (non calculus):

Let $\hat{\alpha}, \hat{\beta}$ = values given by THEOREM (sol'n to normal equations)

Let $\tilde{\alpha}, \tilde{\beta}$ = any other values whatever

$$SSE(\tilde{\alpha}, \tilde{\beta}) = \sum_i (y_i - \tilde{\alpha} - \tilde{\beta} x_i)^2$$

$$= \sum_{i=1}^{N} \left\{ (y_i - \hat{\alpha} - \hat{\beta} x_i) + \left[(\hat{\alpha} - \tilde{\alpha}) + (\hat{\beta} - \tilde{\beta}) x_i \right] \right\}^2$$

$$= \sum_i (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 + \sum_i \left[(\hat{\alpha} - \tilde{\alpha}) + (\hat{\beta} - \tilde{\beta}) x_i \right]^2$$

$$+ 2 \sum_i (y_i - \hat{\alpha} - \hat{\beta} x_i)\left[(\hat{\alpha} - \tilde{\alpha}) + (\hat{\beta} - \tilde{\beta}) x_i \right]$$

$$= SSE(\hat{\alpha}, \hat{\beta}) + \sum_i \left[(\hat{\alpha} - \tilde{\alpha}) + (\hat{\beta} - \tilde{\beta}) x_i \right]^2$$

$$\left.\begin{array}{l} + 2(\hat{\alpha} - \tilde{\alpha}) \sum_i (y_i - \hat{\alpha} - \hat{\beta} x_i) \\ + 2(\hat{\beta} - \tilde{\beta}) \sum_i x_i (y_i - \hat{\alpha} - \hat{\beta} x_i) \end{array}\right\} \text{zero}$$

$\geq SSE(\hat{\alpha}, \hat{\beta})$ since 2nd term ≥ 0

normal equ.

Notes :

1. Define residuals : $e_i = y_i - \hat{y}_i$

LS min's $\sum_i e_i^2$

and satisfies normal equations $\begin{cases} \sum_i e_i = 0 \\ \sum_i e_i x_i = 0 \end{cases}$

2. $\sum_i y_i = \sum_i \hat{y}_i$ (GOOD CHECK)

3. $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \Rightarrow \bar{y} = \hat{\alpha} + \hat{\beta}\bar{x}$ (GOOD CHECK)

i.e. means of data are on the line

4. Divide top & bottom of $\hat{\beta}$ formula by $(N-1) \Rightarrow$

$$\hat{\beta} = \frac{\widehat{\mathrm{cov}}(x,y)}{\widehat{\mathrm{var}}(x)}$$

$\Sigma\, y_i^2 - 2y_i\bar{y} + \bar{y}^2 = \Sigma y_i^2 - n\bar{y}^2$

Decomposition of variance of y:

$\Sigma Y_i^2 - n\bar{Y}^2 \leftarrow$ $SST = \sum_i (y_i - \bar{y})^2$ (var of y, $\times N$)

$\Sigma(\hat{Y}_i^2 - 2\hat{Y}_i\bar{Y} + \bar{Y}^2) \leftarrow$ $SSR = \sum_i (\hat{y}_i - \bar{y})^2$ (var of $\hat{y}$, $\times N$, since $\bar{y} = \bar{\hat{y}}$) by 2. above

$\Sigma(Y_i^2 - 2Y_i\hat{Y}_i + \hat{Y}_i^2) \leftarrow$ $SSE = \sum_i (y_i - \hat{y}_i)^2$ (var of e; times N) so later!

Theorem: $SST = SSR + SSE$

$R+SSE = \Sigma\, \alpha^2 + 2\alpha\beta Y_i + \beta^2 Y_i^2 - 2\alpha\bar{Y} - 2\beta Y_i\bar{Y} + \bar{Y}^2 + \Sigma Y_i^2 - 2\alpha Y_i - 2\beta Y_i^2 + \alpha^2 + 2\alpha\beta Y_i + \beta^2 Y_i^2$

$= n\alpha^2 + 2\alpha\beta n\bar{Y} + \cancel{\beta\Sigma Y_i^2} - 2n\alpha\bar{Y} - 2\beta n\bar{Y}^2 + n\bar{Y}^2 + \Sigma Y_i^2 - 2n\alpha\bar{Y} - \cancel{2\beta\Sigma Y_i^2} + n\alpha^2 + 2\alpha\beta n\bar{Y} + \cancel{\beta^2\Sigma Y_i}$

$= 2n\alpha^2 + 4\alpha\beta n\bar{Y} - 4n\alpha\bar{Y} + \Sigma Y_i^2(\beta^2 - \beta + 1) + \bar{Y}^2 n(1 - 2\beta)$

$= 2n\alpha[\alpha + 2\beta\bar{Y} - 2\bar{Y}] + \ldots$

Note: $e_i \neq \epsilon_i$ [redacted]

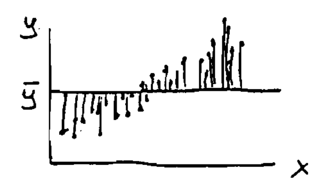
$\epsilon_i = y_i - \alpha - \beta x_i$ - truth

$e_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ - estimated

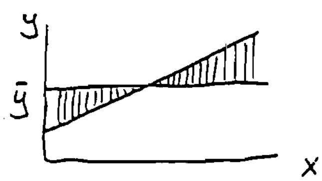
Note 5. to convert $\hat{\beta} = \dfrac{\widehat{Cov}(x,y)}{\hat{\sigma}_x^2}$ into unit-less Beta Coefficient (or correlation coefficient), simply multiply $\hat{\beta}$ by $\left(\dfrac{\hat{\sigma}_x}{\hat{\sigma}_y}\right)$!

More in Inserts p.7 - (come back to...)

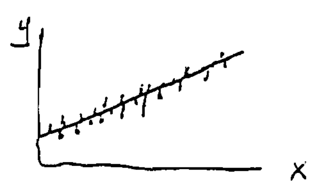
SST = total variation - about the mean



SSR = "explained" variation



SSE = "unexplained" variation



Do pictures opp. p.6

$SST = SSR + SSE$

Proof: $SST = \sum_i (y_i - \bar{y})^2$

$$= \sum_i \left[(\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \right]^2$$

$$= \sum_i (\hat{y}_i - \bar{y})^2 + \sum_i (y_i - \hat{y}_i)^2 + 2 \sum_i (\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$$

$$= SSR + SSE + 2 \sum_i (\hat{y}_i - \bar{y})(y_i - \hat{y}_i)$$

show last term equal zero:

$$\sum_i (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = \sum_i (\hat{\alpha} + \hat{\beta} x_i - \bar{y}) e_i$$

$$= (\hat{\alpha} - \bar{y}) \sum_i e_i + \hat{\beta} \sum_i x_i e_i = 0 \quad \text{by normal eq}^{n}\text{s}$$

Interpretations

1. Total variation = Explained variation + unexplained variation

2. $\mathrm{var}(y) = \mathrm{var}(\hat{y}) + \mathrm{var}(e)$

Define : $R^2 = \dfrac{SSR}{SST} = 1 - \dfrac{SSE}{SST}$ "% explained"

(i) satisfies $0 \le R^2 \le 1$

(ii) measures goodness of fit : $1 \Rightarrow$ perfect, $e_i = 0 \ \forall i$
$0 \Rightarrow$ none, $\hat{\beta} = 0$.

THEOREM: $R^2 = [\hat{\rho}(x,y)]^2$

$$R^2 = \frac{SSR}{SST}$$

Note:
$$SSR = \sum_i (\hat{y}_i - \bar{y})^2$$
$$= \sum_i (\hat{\alpha} + \hat{\beta} x_i - \bar{y})^2$$

$\hat{\alpha} = \bar{y} - \beta \bar{x}$

$$= \sum_i (\bar{y} - \hat{\beta}\bar{x} + \hat{\beta} x_i - \bar{y})^2$$
$$= \sum_i [\hat{\beta}(x_i - \bar{x})]^2$$
$$= [\hat{\beta}^2] \sum_i (x_i - \bar{x})^2$$
$$= \frac{[\sum_i (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum (x_i - \bar{x})^2}$$

since $\hat{\beta} = \dfrac{\widehat{Cov}(x,y)}{\widehat{Var}(x)}$

whereas
$$SST = \sum_i (y_i - \bar{y})^2$$

$$R^2 = \frac{SSR}{SST}$$
$$= \frac{\left[\left(\frac{1}{N-1}\right)\sum_i (x_i - \bar{x})(y_i - \bar{y})\right]^2}{\left[\left(\frac{1}{N-1}\right)\sum_i (x_i - \bar{x})^2\right]\left[\left(\frac{1}{N-1}\right)\sum_i (y_i - \bar{y})^2\right]}$$
$$= \left[\frac{\widehat{cov}(x,y)^2}{\widehat{var}(x)\ \widehat{var}(y)}\right] = \hat{\rho}^2 \checkmark$$

Note: This $\Rightarrow$ R^2 is the same whether y on x OR x on y !!!

Statistical Properties of LS

up to now, have done algebra.
Now, do statistics.
For this, we need to make assumptions.

The coefficients in a Regression are <u>not</u> generally <u>unit-free</u>.

$Y = \alpha + \beta X$; Size of coefficients $\hat{a}$ & $\hat{\beta}$ depends on <u>units</u> of measurement for X & Y.

— Can always rescale the variables X & Y w/o affecting the results!
The important thing is the size of the coeff relative to its standard error (<u>Var($\hat{b}$) later</u>).

e.g.

$$\overset{\text{tons}}{Y} = \underset{(3)}{10} + \underset{(1)}{2}\,\overset{\#}{X} \qquad R^2 = .5 \qquad F = \sim$$

Want results in terms of <u>lbs.</u>?
- Don't need to rerun regression; simply convert the size of the coeff. (& std. errors) according to how you want to measure X & Y!

- Convert Y to lbs. (× 2000):

$$\overset{\text{lbs.}}{Y} = \underset{(6000)}{10\,(2000)} + \underset{(2000)}{2\,(2000)}\,X \qquad R^2 = .5 \qquad F = \sim$$

Note that <u>R^2 & F</u> are not changed by scale!!
(& s.e.'s of $\hat{\beta}$ & $\hat{a}$)

→ * <u>Results</u> not complete until scale of X & Y are stated!!

★

$\exists$ a Regression Format in which Coefficients are unitless. (See Note 5. page 5)

— Beta Coefficients -- [express X & Y as standardized var.

We have the True model;

$$Y = \alpha + \beta X + \epsilon$$

and $\bar{Y} = \alpha + \beta \bar{X}$.

Subtract: $$Y - \bar{Y} = \beta(X - \bar{X}) + \epsilon$$

$\div \sigma_Y$ $$\Longrightarrow \left(\frac{Y-\bar{Y}}{\sigma_Y}\right) = \frac{\beta}{\sigma_Y}(X - \bar{X}) + \frac{\epsilon}{\sigma_Y}$$

$\cdot \frac{\sigma_X}{\sigma_X}$ $$\Longrightarrow \left(\frac{Y-\bar{Y}}{\sigma_Y}\right) = \underbrace{\left(\beta \frac{\sigma_X}{\sigma_Y}\right)}_{\boxed{\text{Beta Coeff.}}}\left(\frac{X-\bar{X}}{\sigma_X}\right) + \frac{\epsilon}{\sigma_Y}$$

<u>Note ①:</u> Beta Coefficient is ~~correlation~~ regression coeff. between <u>standardized</u> X & Y; ∴ <u>unitless</u>.

<u>Note ②:</u> Beta Coefficient can be obtained from regr. coeff., $\hat{\beta}$, by simply multiplying * $\frac{\sigma_X}{\sigma_Y}$.

from p.5 #5. Once again: $\hat{\beta} = \frac{Cov(X,Y)}{\sigma^2}$; Beta $= \hat{\beta}\frac{\sigma_X}{\sigma}$ $\left[= \hat{\rho}(X,Y)\right]$

Insert 3

Note 3: Beta Coeff. related to R^2!

i) In simple regression,

$$\text{Beta} = \hat{\rho}(X,Y) = \sqrt{R^2}$$ (see p. 7)

ii) In multiple regression,
Beta coeff.'s positively related to Partial R^2's.
(!stn)

~~iii) Beta coeff.'s are related to the sample used.~~
~~e.g.~~
~~$Y = a + b X$~~ ~~$\frac{Y - \bar{Y}}{\sigma_Y} = (\text{Beta}) \frac{X - \bar{X}}{\sigma_X}$~~

~~tural parameter interpretation~~ → ~~— β suggests what will happen to Y if X Δs.~~ ~~(prediction purpo...)~~
~~whereas ⟹ Beta shows what actually happened to Y~~
~~given the historical variations in X (...)~~

iii) These transformations of the Regression model (scale, Beta, ...)
will not affect t-ratios,
F statistics,
R^2

— ∄ no way you can transform the model & suddenly come up with significant results where you didn't before!

Statistical Properties of LS

— up to now have done <u>ALGEBRA</u>; Now do <u>STATISTICS</u>.
— For this we must make assumptions.

Definition: The regression model $y_i = \alpha + \beta x_i + \epsilon_i$, $i = 1, 2, \cdots, N$,

satisfies the <u>ideal conditions</u> if

(1) The ϵ_i are <u>iid</u> as $N(0, \sigma^2)$

(2) The x_i are fixed (<u>non-random</u>)

} This is TRUTH

Note: (1) contains <u>many</u> parts

** (i) Zero mean : $E(\epsilon_i) = 0 \quad \forall i$

so that $E(y_i) = \alpha + \beta x_i$

(ii) Independence : $\text{cov}(\epsilon_i, \epsilon_j) = 0 \quad (i \neq j)$

(iii) Constant variance : $\text{var}(\epsilon_i) = \sigma^2 \quad \forall i$

(iv) normality

(2) is a throwback to ex<u>periment</u>al settings

(x = something like amount of fertilizer, location of field, etc.)

[Can be replaced by (2') x_i and ϵ_j are independent $\forall i, j$]

Is an assumption of exo<u>geneity</u> of x (def'd elsewhere)

and represents the view:

X causes Y <u>but</u> Y does not cause X

(big ϵ, which is random part of y, does not affect x, e.g.)

*** Note: <u>NON-REVERSIBILITY</u>

If $y_i = \alpha + \beta x_i + \epsilon_i$, $x_i = -\frac{\alpha}{\beta} + \frac{1}{\beta} y_i - \frac{1}{\beta}\epsilon_i$

But if y on x is nice, x on y is <u>not</u> (y corr. with ϵ)

NOTE: Where do errors come from?

1) Measurement errors

OK if on dependent variable

$$y_i^* = \alpha + \beta x_i \;,\quad y_i = y_i^* + \epsilon_i$$

$$\Rightarrow y_i = \alpha + \beta x_i + \epsilon_i$$

y_i^* = true value
y_i = observed

Not so good if on explanatory variable

$$y_i = \alpha + \beta x_i^*$$

$$x_i = x_i^* + v_i \;\Rightarrow\; y_i = \alpha + \beta x_i \underbrace{- \beta v_i}_{\epsilon_i \text{, corr. with } x_i}$$

2) Left out variables   eg C fn.

OK if they are insignificant (individually)
and not correlated with x ! — (more later)

3) Just plain randomness.

Properties of LS Estimators

Lemma: $$\hat{\beta} = \beta + \frac{\sum_i (x_i - \bar{x})\, \epsilon_i}{\sum_i (x_i - \bar{x})^2}$$

Proof: $$\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

but $$y_i = \alpha + \beta x_i + \epsilon_i$$
$$\bar{y} = \alpha + \beta \bar{x}$$
$$y_i - \bar{y} = \beta (x_i - \bar{x}) + \epsilon_i$$

$$= \frac{\sum_i (x_i - \bar{x})[\beta(x_i - \bar{x}) + \epsilon_i]}{\sum_i (x_i - \bar{x})^2}$$

$$= \beta \frac{\sum_i (x_i - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + \frac{\sum_i (x_i - \bar{x})\, \epsilon_i}{\sum_i (x_i - \bar{x})^2}$$

$$= \beta + \frac{\sum_i (x_i - \bar{x})\, \epsilon_i}{\sum_i (x_i - \bar{x})^2}$$

Thm: $\hat{\alpha}$ & $\hat{\beta}$ are unbiased, efficient, consistent, and asymptotically efficient.

Proof of unbiasedness:

$$E(\hat{\beta}) = \beta + E\left\{ \frac{\sum_i (x_i - \bar{x})\, \epsilon_i}{\sum_i (x_i - \bar{x})^2} \right\}$$

$$= \beta + \frac{\sum_i (x_i - \bar{x})\, E(\epsilon_i)}{\sum_i (x_i - \bar{x})^2} = \beta$$

$$E(\hat{\alpha}) = E(\bar{y} - \hat{\beta}\bar{x}) = E\left[\alpha + \beta\bar{x} + \bar{\epsilon} - \hat{\beta}\bar{x}\right]$$

$$= \alpha + \beta\bar{x} + 0 - \beta\bar{x} = \alpha \checkmark$$

Efficiency: Just claimed. But note: same problem really as estimating μ in $N(\mu, \sigma^2)$, except here

$\mu_i = \alpha + \beta x_i = E(y_i)$

proven later

Consistency: $\hat{\beta} = \beta + \dfrac{\sum_i (x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_i (x_i - \bar{x})^2}$

$= \beta + \dfrac{\widehat{\text{cov}}(x,\epsilon)}{\widehat{\text{var}}(x)}$ and $\text{cov}(x,\epsilon) = 0$!! from ideal cond.

Asymptotic efficiency: Claim

NOTE: Unbiasedness + consistency do not hinge on normality. Efficiency + as. effic. do.

Distribution of $\hat{\alpha}, \hat{\beta}$:

THEOREM: $\hat{\beta} \sim N\left(\beta, \dfrac{\sigma^2}{\sum_i (x_i - \bar{x})^2}\right)$

$\hat{\alpha} \sim N\left(\alpha, \dfrac{\sigma^2}{N} + \dfrac{\sigma^2 \bar{x}^2}{\sum_i (x_i - \bar{x})^2}\right)$

Proof: 1. Normality is obvious since l.c. of ϵ's, which are normal.

2. Means are obvious, from unbiasedness.

3. $\mathrm{var}(\hat{\beta}) = \mathrm{var}\left(\dfrac{\sum_i (x_i - \bar{x})\epsilon_i}{\sum_i (x_i - \bar{x})^2} \right)$

$$= \frac{1}{\left[\sum_i (x_i - \bar{x})^2\right]^2} \mathrm{var}\left[\sum_i (x_i - \bar{x})\epsilon_i\right] \quad \text{(since } X \text{ nonrandom)}$$

$$= \frac{1}{\left[\sum_i (x_i - \bar{x})^2\right]^2} \sum_i \mathrm{var}\left[(x_i - \bar{x})\epsilon_i\right] \quad \text{(since } \mathrm{cov}(\epsilon_i, \epsilon_j) = 0\text{)}$$

$$= \frac{1}{\left[\sum_i (x_i - \bar{x})^2\right]^2} \sum_i (x_i - \bar{x})^2 \mathrm{var}(\epsilon_i)$$

$$= \sigma^2 \frac{\sum_i (x_i - \bar{x})^2}{\left[\sum_i (x_i - \bar{x})^2\right]^2}$$

$$= \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} = \frac{\sigma_\epsilon^2}{\underline{\underline{N}}\, \sigma_x^2}$$

Similarly for $\hat{\alpha}$.

* NOTE: Variance depends on (i) spread of X's

If all X's are the same ($x_i = \bar{x}\ \forall i$), var $= \infty$

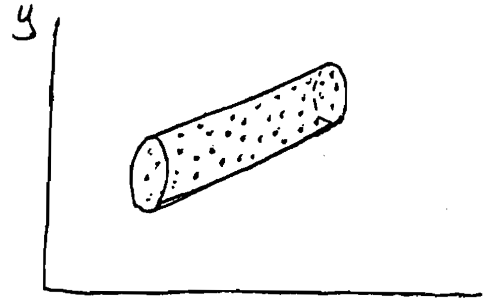
Variance also depends on (ii) σ_ϵ^2.

(WIRE THRU PIPE analogy.)

Wire thru Pipe Analogy

Think of all points being inside a pipe and fitting the regression line as fitting a wire inside this pipe.

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2} = \frac{\sigma_\epsilon^2}{N \sigma_x^2}$$



- refers to the <u>variance of the slope</u> of the wire within the pipe.

<u>Note:</u> i) If pipe is shorter, ∃ more $Var(\hat{\beta})$.

ii) If pipe has wider diameter, ∃ more $Var(\hat{\beta})$.

Observe: i) length of pipe reflects σ_x^2
-- more variability in x → more information → better expl. power re: y → <u>lower</u> $Var(\hat{\beta})$

ii) diameter of pipe reflects σ_ϵ^2
-- wider diameter → more noise → higher $Var(\hat{\beta})$.

iii) More observations, N larger;
-- more points in pipe, pipe more tightly packed: more <u>resistance</u> to variation in slope of wire

Estimation of σ^2:

Note: $\sigma^2 = var(\epsilon_i) = E(\epsilon_i^2)$ since $E(\epsilon_i) = 0$

Obvious "estimator" $\frac{1}{N}\sum_i \epsilon_i^2$ if ϵ_i were known.

Would like to replace ϵ_i by e_i = LS residual

$(\epsilon_i = y_i - \alpha - \beta x_i$,
$e_i = y_i - \hat{\alpha} - \hat{\beta} x_i)$

Estimator $\hat{\sigma}^2 = \frac{1}{N}\sum_i e_i^2$

Unforunately this is biased: $E(\hat{\sigma}^2) = \frac{N-2}{N}\sigma^2$

Unbiased estimator

$$s^2 = \frac{1}{N-2}\sum_{i=1}^{N} e_i^2 \qquad \left[= \frac{N}{N-2}\hat{\sigma}^2 \text{ so } E = \frac{N}{N-2}\,\frac{N-2}{N}\,\sigma^2 = \sigma^2 \right]$$

Intuition: 2 d.f. used up in estimating $\hat{\alpha}, \hat{\beta}$

(just as in $N(\mu, \sigma^2)$ case, $(\frac{1}{N-1})$ since 1 used up for μ)

THEOREM: s^2 is unbiased, efficient, consistent, asymp. efficient; and

$\text{MSE} = \frac{\text{SSE}}{n-2}$

$$\frac{(N-2)\,s^2}{\sigma^2} \sim \chi^2_{N-2} \qquad \text{or} \qquad s^2 \sim \frac{\sigma^2}{(N-2)} * \chi^2_{N-2}$$

pf later

Tests of Hypotheses

May wish to test $H_0: \beta = c$

vs $H_A: \beta \neq c$ (or $\beta > c$, or $\beta < c$)

Test statistic: $\dfrac{\hat{\beta} - c}{s.e.(\hat{\beta})} = \dfrac{\hat{\beta} - c}{\sqrt{\dfrac{s^2}{\sum_i (x_i - \bar{x})^2}}} \sim t_{N-2}$

Reasonable since $N(0,1)$ if σ^2 known; t_{N-2} if not.

Usual case: $c = 0$ so test $H_0: \beta = 0$ via

$$t \equiv \hat{\beta} \Big/ \sqrt{\frac{s^2}{\sum_i (x_i - \bar{x})^2}} \quad \text{vs } t_{N-2}$$

similarly for $\hat{\alpha}$

Equivalent:

$$F = \frac{SSR/1}{SSE/(N-2)} = (N-2)\frac{R^2}{(1-R^2)} \sim F_{1,N-2} = \frac{SSR/1}{SSE/n-2} = \frac{MSR}{MSE}$$

In fact $F \equiv t^2$ which is why tests are equivalent!!

Further note: This is same as test of $\rho = 0$:

$$\sqrt{F} = t_{N-2} = \frac{\hat{\rho}}{\sqrt{\frac{1-\hat{\rho}^2}{N-2}}} \quad \text{since } R^2 = \hat{\rho}^2$$

Different assumptions, same test!!!

Presentation of Results — show GLM output → [overhead]

Usually give $\hat{\alpha}, \hat{\beta}$; either t's or s.e's ; R^2 ; N ; F

s^2

~~Example 1 : Consumption function - see handout~~

~~15 yearly observations, 1950-1964~~

~~y = personal consumption expenditures ($millions)~~

~~X = personal disposable income ($millions)~~

~~Note - excellent fit MPC = .93 (t - huge)~~

~~intercept = 0 (s.e. = 2.29)~~

~~[typical of yearly C fn.]~~

~~Yet - most economists think this is misspecified~~

~~(eg permanent y hypothesis discussion)~~

~~Also there is a question of exogeneity~~

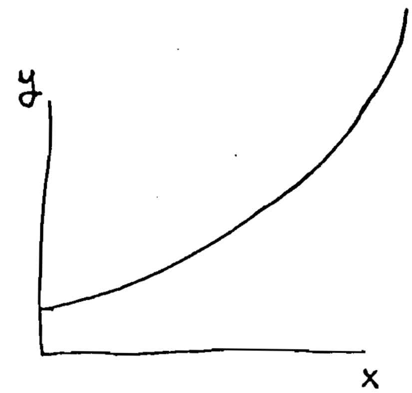
~~since C+I+G+(X-M) = y~~

Functional Form !.- Linear → constant slope (severe restriction!)

Linear means wrt parameters.

0. quadratic – problem I

1. Semilog



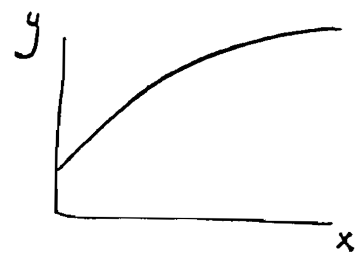
$y = e^x$

$$\ln y = \alpha + \beta x + \epsilon$$

$$\left[\text{or } \; y = e^{\alpha + \beta x + \epsilon} = (e^{\alpha} e^{\epsilon}) e^{\beta x}\right]$$

[β → curvature]

2. Semilog



$y = \ln x$

$$y = \alpha + \beta \ln x + \epsilon$$

[β → curvature]

3. Double log (log linear)

[Cobb-Douglas type fn]

$$\ln y = \alpha + \beta \ln x + \epsilon$$

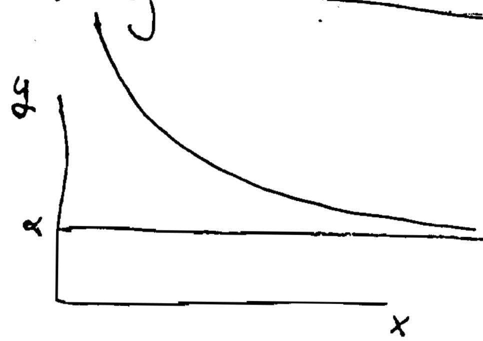
$$\text{or } \; y = e^{\alpha}(x^{\beta})e^{\epsilon}$$

Note β = elasticity

Goes thru origin
Bend depends on β

4. Reciprocal

$$y = \alpha + \beta \frac{1}{x} + \epsilon$$



** Note on local linearity + extrapolation (prediction)

↳ [estimate β with info. provided by x; ∴ fitted rel. is useful for interpret. only in relevant range! e.g. 1) is close to linear small x; fit line & predict ...]