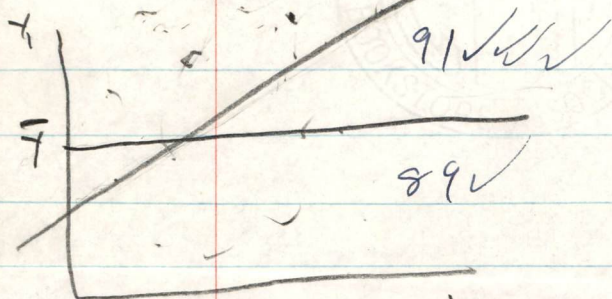


1- Plans & reasons
Kaweah analysis

10 10 10 10 10 10 10 10

MADE IN U.S.A.

If $a = \frac{b}{c}$
 does $a \pm \frac{b \pm c}{c}$; i.e., does
 $a \cdot a \pm \frac{b \cdot b}{c \cdot c}$ yes.



$y' = a + bX$
 86
 85 ✓
 84 ✓

82 ✓

80

27 ✓

75
 74 ✓

73

72

71 ✓✓
 70 ✓

68 ✓

66

$$\sigma_y \neq \sigma_{y'} + s_y$$

$$56 \sigma_y^2 = \sigma_{y'}^2 + s_y^2$$

55 ✓

54

$$r^2 = \frac{\sigma_{y'}^2}{\sigma_y^2} = .81$$

$$49 \text{ --- } 9$$

0

$$\frac{5-8}{2} = 29$$

45

44

42

39 ✓

37

34

30

$$\begin{array}{r} 60 \\ .13 \\ \hline 180 \\ 60 \\ \hline 78 \end{array}$$

8

And $r^2 = 71$ $\sigma_y^2 = \sigma_{y'}^2 + s_y^2$

$$\frac{\sigma_{y'}^2}{\sigma_y^2} = \frac{\sigma_{y'}^2}{\sigma_y^2} + \frac{s_y^2}{\sigma_y^2}$$

$$1 = r^2 + \frac{s_y^2}{\sigma_y^2}$$

$$1 - \frac{s_y^2}{\sigma_y^2} = r^2$$

61

60

58

57 ✓

78

43
 52
 39
 30
 164
 41
 205

~~350~~
 350
 338

test total
 Sec 2, 14, 25

399✓
 398✓

532✓
 330
 325

396
 395

324

460

277

317✓

390

14 = 58

452

314✓

272

388

312

271 = few
 54

386 ~~250~~

308
 306
 305
 303
 302

434

261
 260

D

423

379

376

236

298

233 = Elliot

367

227

AT 418✓

366

295

294

412

363

216

412
 408
 404
 401✓

292

305 on bands
 164 mixed
 skin 4

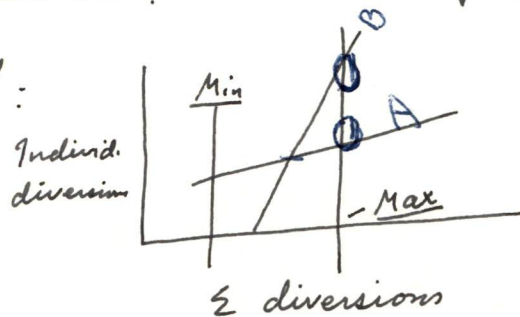
359

355

Questions for Dan, Fri. Oct. 9

1. Is basic = n ok? $\sigma_1 + \sigma_2 \stackrel{?}{=} \sqrt{\sigma_1^2 + 2\sigma_1\sigma_2(1-\rho)}$

In view of:



2. If it is ok, (or whatever is correct), how derive, & how expand to 20 sigmas?
3. If $\rho=1$, how minimize $\Sigma \sigma$: radius vector or straight lines?
4. Beyond minimizing aggreg. variability, we want to pool risks so as to equate the % variability of all. Is this a separate problem or does solving one solve the other?
5. As partial criterion, wd it make sense to see how well system approached its linear regressions?
6. Our $\Sigma \sigma$ depends on size distribn of σ s. But how about relations of σ s to $\bar{x}$ s?

$$\sigma = \sqrt{\frac{\sum d^2}{n}}$$

$$x_1 = 100$$

$$x_2 = 0$$

$$\bar{x} = 50$$

$$n = 2$$

$$\sigma = \sqrt{\frac{5000}{2}} = \sqrt{2500} = 50$$

So how can σ ever $> \bar{x}$? Easy

$$x_1 = 100$$

$$x_2 = 0$$

$$x_3 = 0$$

$$x_4 = 0$$

$$\bar{x} = 25$$

$$n = 4$$

$$\sigma = \sqrt{\frac{75^2 + 3 \times 25^2}{4}} = \sqrt{\frac{5625 + 1875}{4}} = \frac{10}{2} \sqrt{75}$$

$$= 5 \times 8.65 = 43 \frac{1}{4}$$

A numerical example showing that ~~the coeff. of variation of A-B does not nec =~~
~~coeff. of varn of A, & B, tho they are =.~~ the coeff. of variation of A-B does not nec =
coeff. of varn of A, & B, tho they are =.

A 12 8 14 6 $\bar{X} = 10$

d_A 2 2 4 4

d_A^2 4 4 16 16 $\sum d_A^2 = 40$

B 4 6 7 3 $\bar{X} = 5$

d_B 1 1 2 2

d_B^2 1 1 4 4 $\sum d_B^2 = 10$

A-B = C 8 2 7 3 $\bar{X} = 5$

d_C 3 3 2 2

d_C^2 9 9 4 4 $\sum d_C^2 = 24$

$$\sigma_A = \sqrt{\frac{40}{4}} = \sqrt{10}; \quad CV_A = \frac{\sqrt{10}}{10} = \frac{1}{\sqrt{10}}$$

$$\sigma_B = \sqrt{\frac{10}{4}} = \sqrt{\frac{5}{2}}; \quad CV_B = \frac{\sqrt{\frac{5}{2}}}{5} = \frac{1}{\sqrt{2}\sqrt{5}} = \frac{1}{\sqrt{10}}$$

$$\sigma_C = \sqrt{\frac{24}{4}} = \sqrt{6}; \quad CV_C = \frac{\sqrt{6}}{5} \gg \frac{1}{\sqrt{10}}$$

Stat. operations, files.

File key: a1D. But just put subsequent keys on tabs, to save clutter & tab space.

1. Kaweah studies, & distributaries

- Plans of work, & rationalizations

a. Range

b. Skew

c. σ

d. LCR, ~~time units & water~~ time units & ~~water~~

e. Length of season

f. Seasonal patterns

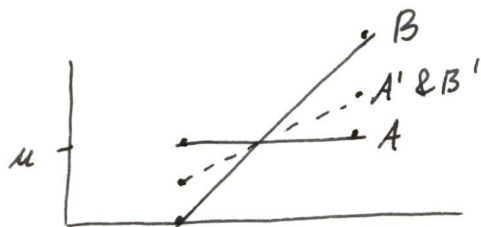
g. Clustering of deficits

h. LCR, acres & water

i. Variability & capacity.
~~System capacities & entitlements~~

j. ~~Usage of capacity~~

Confirming Brunk's statement that rotating curves won't change Σ



$$A: \bar{x} = \mu$$

$$d_1 = 0; \quad d_1^2 = 0$$

$$d_2 = 0; \quad d_2^2 = 0$$

$$\sigma_A = \sqrt{\frac{\sum d_i^2}{n}} = \sqrt{\frac{0+0}{2}} = 0$$

$$B: \bar{x} = \mu$$

$$d_1 = -\mu \quad d_1^2 = \mu^2$$

$$d_2 = \mu \quad d_2^2 = \mu^2$$

$$\sigma_B = \sqrt{\frac{2\mu^2}{2}} = \mu$$

$$\Sigma = \mu$$

Now shift values so

$$A_1 = \frac{\mu}{2}$$

$$A_2 = \frac{3\mu}{2}$$

$$B_1 = \frac{\mu}{2}$$

$$B_2 = \frac{3\mu}{2}$$

$$A': \bar{x} = \mu$$

$$d_1 = \frac{\mu}{2} \quad d_1^2 = \frac{\mu^2}{4}$$

$$d_2 = \frac{\mu}{2} \quad d_2^2 = \frac{\mu^2}{4}$$

$$\sigma_{A'} = \sqrt{\frac{\frac{1}{2}\mu^2}{2}} = \frac{\mu}{2}$$

$$\Sigma = \mu$$

$$B': \bar{x} = \mu$$

$$d_1 = \frac{\mu}{2} \quad d_1^2 = \frac{\mu^2}{4}$$

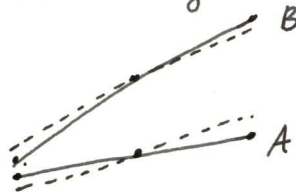
$$d_2 = \frac{\mu}{2} \quad d_2^2 = \frac{\mu^2}{4}$$

$$\sigma_{B'} = \sqrt{\frac{\frac{1}{2}\mu^2}{2}} = \frac{\mu}{2}$$

Confirming, 2

Trouble w. confirmation is, wd also work with mean deviation. & σ should be different but squared.

Let's take longer curves; & rotate just a little.



A: 7, 9, 11	$\bar{x} = 9$	$d_1 = 2, d_2 = 0, d_3 = 2$	$\sum d^2 = 8$	$\sigma = \sqrt{\frac{8}{3}}$
		$d_1^2 = 4, d_2^2 = 0, d_3^2 = 4$	$n = 3$	
B: 8, 12, 16	$\bar{x} = 12$	$d_1 = 4, d_2 = 0, d_3 = 4$	$\sum d^2 = 32$	$\sigma = \sqrt{\frac{32}{3}}$
		$d_1^2 = 16, d_2^2 = 0, d_3^2 = 16$	$n = 3$	
Σ : 15, 21, 27	$\bar{x} = 21$	$d_1 = 6, d_2 = 0, d_3 = 6$	$\sum d^2 = 72$	$\sigma = \sqrt{\frac{72}{3}} = 6\sqrt{\frac{2}{3}}$
		36, 0, 36	$n = 3$	

$$\sigma_A + \sigma_B = \sqrt{\frac{8}{3}} + \sqrt{\frac{32}{3}} = \frac{\sqrt{8} + \sqrt{32}}{\sqrt{3}} = \frac{(2 + 4)\sqrt{2}}{\sqrt{3}} = 6\sqrt{\frac{2}{3}}$$

Now rotate

A': 6, 9, 12	$\bar{x} = 9$	$d_1 = 3, d_2 = 0, d_3 = 3$	$\sum d^2 = 18$	$\sigma = \sqrt{\frac{18}{3}}$
		$d_1^2 = 9, d_2^2 = 0, d_3^2 = 9$	$n = 3$	
B': 9, 12, 15	$\bar{x} = 12$	$d_1 = 3, d_2 = 0, d_3 = 3$	$\sum d^2 = 18$	$\sigma = \sqrt{\frac{18}{3}}$
		$d_1^2 = 9, d_2^2 = 0, d_3^2 = 9$	$n = 3$	
Σ : 15, 21, 27	$\bar{x} = 21$			

$$\sigma_{A'} + \sigma_{B'} = \sqrt{\frac{18}{3}} + \sqrt{\frac{18}{3}} = 2 \times 3\sqrt{\frac{2}{3}} = 6\sqrt{\frac{2}{3}}$$

Q. E. D.

$$X = \frac{1}{b} f - \frac{1}{b}$$

$$f = a + X(n+1)$$

1

$$f = a + X$$

0 = 1

$$f = a + X$$

$$\begin{cases} X + Y = F = \Sigma \text{ diversos} \\ X = pF + b \\ \cancel{Y = gF + b} \end{cases}$$

$$Y = F - X = F - pF - b \\ = gF - b \quad \text{where } g = 1 - p$$

$$\rho \equiv \frac{\text{cov}(X, Y)}{\sigma_1 \sigma_2}$$

$$V(X + Y) = V(F)$$

$$V(X) + V(Y) + 2\text{cov}(X, Y)$$

$$\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho = \sigma_\Sigma^2$$

$$\sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2 = \sigma_\Sigma^2 + 2\sigma_1\sigma_2(1-\rho)$$

$$\sigma_1 + \sigma_2 = \sqrt{\sigma_\Sigma^2 + 2\sigma_1\sigma_2(1-\rho)}$$

$$\begin{aligned} V(X+Y) &= E([X-\mu_1 + Y-\mu_2]^2) \\ &= E[(X-\mu_1)^2 + 2(X-\mu_1)(Y-\mu_2) + (Y-\mu_2)^2] \\ &= V(X) + 2\underbrace{E[(X-\mu_1)(Y-\mu_2)]}_{\text{cov}(X, Y)} + V(Y) \end{aligned}$$

E = Expected value = mean

μ = mean

$$\sum_{i=1}^n X_i = F$$

$$E \left[\sum_i (X_i - \mu_i) \right]^2 = V(F)$$

$$E \left\{ \left[\sum_i (X_i - \mu_i) \right] \left[\sum_j (X_j - \mu_j) \right] \right\}$$

$$E \left[\sum_i (X_i - \mu_i)^2 + \sum_i \sum_{j \neq i} (X_i - \mu_i)(X_j - \mu_j) \right]$$

$$\sum_{i=1}^n V(X_i) + \sum_i \sum_{j \neq i} \text{cov}(X_i, X_j)$$

$$\sum_{i=1}^n \sigma_i^2 + \sum_i \sum_{j \neq i} \rho_{ij} \sigma_i \sigma_j = \sigma_\Sigma^2$$

$$\sum_{i=1}^n \sigma_i^2 + \sum_i \sum_{j \neq i} \sigma_i \sigma_j = \sigma_\Sigma^2 + \sum_i \sum_{j \neq i} (1 - \rho_{ij}) \sigma_i \sigma_j$$

$$\sum_{i=1}^n \sigma_i = \sqrt{\sigma_\Sigma^2 + \sum_i \sum_{j \neq i} (1 - \rho_{ij}) \sigma_i \sigma_j}$$

$$n=2$$

$$\sum_{i=1}^2 \left(\sum_{j \neq i} (1 - \rho_{ij}) \sigma_i \sigma_j \right)$$

$$= (1 - \rho_{12}) \sigma_1 \sigma_2 + (1 - \rho_{21}) \sigma_2 \sigma_1$$

$$n=3$$

$$\sum_{i=1}^3 \sum_{j \neq i} (1 - \rho_{ij}) \sigma_i \sigma_j$$

$$= (1 - \rho_{12}) \sigma_1 \sigma_2 + (1 - \rho_{13}) \sigma_1 \sigma_3$$

$$+ (1 - \rho_{21}) \sigma_2 \sigma_1 + (1 - \rho_{23}) \sigma_2 \sigma_3$$

$$+ (1 - \rho_{31}) \sigma_3 \sigma_1 + (1 - \rho_{32}) \sigma_3 \sigma_2$$

Brunk 3-7

Now find min. possible value for $\sum \frac{\sigma_i^2}{\bar{x}_i}$, in terms of $\frac{\sigma_z^2}{\bar{x}_z}$

i.e.
$$\frac{\sum \left(\frac{\sigma_i^2}{\bar{x}_i} \right)}{\frac{\sigma_z^2}{\bar{x}_z}}$$

With σ_z & $\bar{x}_z$ given, this means min. $\sum \frac{\sigma_i^2}{\bar{x}_i}$. And while at it, find max. possible.

Simple beginning, let $\rho = 1$

$$\sigma_1 + \sigma_2 = \sigma_z$$

$$\sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2 = \sigma_z^2$$

$$\bar{x}_1 + \bar{x}_2 = \bar{x}_z ; \bar{x}_1 = k_1 \bar{x}_z ; \bar{x}_2 = k_2 \bar{x}_z ; k_1 + k_2 = 1$$

$$\frac{\sigma_1^2}{\bar{x}_1} = \frac{\sigma_1^2}{k_1 \bar{x}_z}$$

This will call for some good algebra.

But we found $\sum \sigma_i^2$ a min. when σ all =. So it is very likely we wd find $\sum \left(\frac{\sigma_i^2}{\bar{x}_i} \right)$ a min. when $\frac{\sigma_i}{\bar{x}_i}$ all =. --- which is

also when $x_1 = k x_2$

$$\frac{\sigma_1^2}{\bar{x}_1} + \frac{\sigma_2^2}{\bar{x}_2}$$

Now let $\sigma_2 = a \sigma_1$
 $x_2 = b x_1$

Now $\sum = \frac{\sigma_1^2}{\bar{x}_1} + \frac{a^2 \sigma_1^2}{b \bar{x}_1}$ Now $\frac{d\sum}{da} = 0 + \frac{2a\sigma_1^2}{b \bar{x}_1} = 0$ when $a = 0$

Brunik, p. 7

It would tell us that with 20 sigmas of the size we ~~achieved~~
and here it is their relative size that is important
have, the ^{min.} ~~max.~~ possible value, if the flows were absolutely unrelated,

~~would be~~ of ~~$\frac{\sigma_y}{\sum \sigma_{x_i}}$~~

of $\frac{\sigma_y}{\sum \sigma_{x_i}}$ would be such. And we wd. cf. this with what we

actually have.

There would remain a possibility the sigmas might are smaller?

Brunk, p. 6

$$\frac{a^2 + b^2}{a^2 + b^2 + 2ab} = \frac{1}{1 + \frac{2ab}{a^2 + b^2}} \leq \infty \quad \text{But not less than 1}$$

To evaluate $\frac{ab}{a^2 + b^2} = \frac{1}{\frac{a}{b} + \frac{b}{a}} \leq \frac{1}{2} \quad \text{But not less than 0}$

To evaluate $\frac{a}{b} + \frac{b}{a}$. This has a minimum value of 2 ^{when $a=b$} and grows larger, up to any value, as a & b diverge

$$\frac{a}{b} + \frac{b}{a} \geq 2$$

So $\frac{a^2 + b^2}{a^2 + b^2 + 2ab}$ can vary between limits of 1 to ~~1~~ $\frac{1}{2}$

~~Not a very narrow limitation!~~

If $a=b$, it is ~~1~~ $\frac{1}{2}$

If a or b is ∞ or 0, it = 1

Now if it = 1, then $\sigma_y = \sigma_x + \sigma_z$, still, and our upper limit is very close to our lower limit.

So it is likely that when we figured the upper limit for ~~the~~ 20 items, it would not be much above the lower limit. And we could figure it, from specific values of each σ — only then what would it mean, wouldn't it be same as & identical with the result actually achieved? No.

~~7~~ "When to irrt, & how much".

'55 Wbk p. 372 fl.

Re Kammhaster.

To irrt effectively, need to
apply wt in certain ways vis-a-vis
field conditions etc. — & not at
arbitrary times as per Kuhl.
schedule.

Even more striking wd. be to figure the lowest possible variability they might have achieved for the amount of water they actually diverted. Take the mean 2 diversions, only in a few years wd. run fall below it. They cd. have provided a near-perfect steady supply!!

Of course this wd. have to be modified for seasonal inequalities.

Now to define this upper limit. Brunk says it is when $r=0$ among all components (and with Y , also, or is that impossible?)

~~at which~~ Under this condition, $\sigma_y^2 = \sigma_x^2 + \sigma_z^2$

$$\text{and } \sigma_y = \sqrt{\sigma_x^2 + \sigma_z^2} < \sigma_x + \sigma_z = \sqrt{\sigma_x^2 + 2\sigma_x\sigma_z + \sigma_z^2}$$

$$\left[\begin{array}{l} \text{i.e., if } a^2 = b^2 + c^2 \\ \text{Then } a < b + c \quad \text{That makes sense. Cf. Pythagorean.} \end{array} \right]$$

$$\text{at lower lim., } \frac{\sigma_y}{\sigma_x + \sigma_z} = 1.$$

$$\text{at upper lim., } \frac{\sigma_y}{\sigma_x + \sigma_z} = \frac{\sqrt{\sigma_x^2 + \sigma_z^2}}{\sqrt{\sigma_x^2 + 2\sigma_x\sigma_z + \sigma_z^2}} = \frac{\sigma_x^2 + \sigma_z^2}{\sigma_x^2 + \sigma_z^2 + 2\sigma_x\sigma_z}$$

This resolves itself into putting a numerical value on $\frac{a^2 + b^2}{a^2 + b^2 + 2ab}$ in general, for any number — or in our case 20 — of items; and any ratio of a to b , etc. — if this makes a difference, it will complicate things.

So: If $y = x + z$

and $x_n = k_x y_n$; $z_n = k_z y_n$

and $k_x + k_z = 1$

Then: $\sigma_y = \sigma_x + \sigma_z$ and so on, the proof is freely expandible.

Now Brunk says if we split up y any way but x & z , then we get a Σ of σ 's which exceeds σ_y .

If so, this argues that in general, a fractional system of water rights does as much as is statistically possible in minimizing aggregate variability.

This conclusion is worth giving, & emphasizing for its own sake. But it no way lessens the interest in knowing how much Kwh. distribn. increases aggregate variability in system.

To give meaning to the "how much", tho. it wd. be well to find an upper limit to possible increases in ^{aggregate} variability. This seems to be when the components are completely unrelated — when $r=0$ among them all.

The fact that there is this upper limit makes a given Δ of aggreg. vblty. much more striking. 15% above the lowest possible doesn't seem very bad; but it may be halfway between the lowest & the highest.

Alternative ways of organizing Kuch evidence:

1. Areas getting better & worse $\frac{\sigma}{\bar{x}}$ than Kwh.
2. a-f served under " " " " " "
3. a-f " " " " " ", weighted by the amount that they are worse.
4. Group water users into a few classes of $\approx$ size - by wte cons.n — take simple mean of c.v.s.

These are all inelegant & clumsy. - altho ~~possibly~~ useful as expository tools for popular consumption, hence probly worth working up.

76 $y = x + z$, each year. And $x = k_x y$ and $z = k_z y$, (then $k_x + k_z = 1$)

Then $\sigma_y \stackrel{?}{=} \sigma_x + \sigma_z$

$$\sigma_y = \sqrt{\frac{\sum d_y^2}{n}} = \sqrt{\frac{\sum d_x^2}{n}} + \sqrt{\frac{\sum d_z^2}{n}}$$

The ns cancel

$$\sqrt{\sum (y_n - \bar{y})^2} \stackrel{?}{=} \sqrt{\sum (x_n - \bar{x})^2} + \sqrt{\sum (z_n - \bar{z})^2}$$

But $(y_n - \bar{y}) = (x_n - \bar{x}) + (z_n - \bar{z})$ each year

$$\text{Or } d_y = d_x + d_z$$

$$\sqrt{\sum d_y^2} = \sqrt{\sum d_x^2} + \sqrt{\sum d_z^2}$$

$$d_x = k_x d_y$$

$$d_z = k_z d_z$$

$$\sqrt{\sum d_i^2} \stackrel{?}{=} \sqrt{\sum k_x^2 d_{iy}^2} + \sqrt{\sum k_y^2 d_{ix}^2} = \sqrt{k_x^2 \sum d_{iy}^2} + \sqrt{k_y^2 \sum d_{ix}^2} = k_x \sqrt{\sum d_{iy}^2} + k_y \sqrt{\sum d_{ix}^2}$$

$$\rightarrow = (k_x + k_z) \sqrt{\sum d_y^2} = \sqrt{\sum d_y^2}$$

Brink, p. 2

if $y = x_1 + x_2$

a) if all indep., then $\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$

b) if all propul., then $\sigma_y = \sigma_{x_1} + \sigma_{x_2}$

c) Most cases would fall in between. What is in between?

Beginning with (b), perfect proportionality, that is lower limit of $\sum \sigma_{x_n} = \sigma_y$.

As lose proportionality, $\sum \sigma_{x_n}$ grows, until with perfect independence, $\sum \sigma_{x_n} > \sigma_y$. Because then $\sigma_y^2 = \sigma_1^2 + \sigma_2^2$

$$\text{But } \sigma_1 + \sigma_2 = \sqrt{\sigma_1^2 + 2\sigma_1\sigma_2 + \sigma_2^2} \\ > \sigma_y$$

9-29-59. Summary of discussion with Brunk (& Benham)

$$\text{If } Y = X_1 + X_2 + \dots + X_n$$

Then

$\sigma_y = \sigma_{x_1} + \sigma_{x_2} + \dots + \sigma_{x_n}$ if $\sum$ the x 's are all perfectly correlated with each other

But if they are not so correlated, then $\sigma_y < \sum \sigma_{x_n}$

[Because if not perfectly correl-d, some go up when others go down, & they must be larger, in general, to = Y on all occasions - so?]

Brunk tends toward use of variance, σ^2 , or $\frac{\sum d^2}{n}$ (over)

With variance have useful relationship, if $Y = X_1 + X_2$

then $\sigma_y^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2$ — but only if variables are indep., not correl-d.

Which with σ wd. be simpler:

$$\sigma_y = \sqrt{\sigma_{x_1}^2 + \sigma_{x_2}^2} = \sqrt{\frac{\sum d_1^2}{n} + \frac{\sum d_2^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum d_1^2 + \sum d_2^2}{n}} \text{ etc.}$$

$$\text{So } \sum d_y^2 = \sum d_{x_1}^2 + \sum d_{x_2}^2$$

$$\sigma \equiv \sqrt{\frac{\sum d^2}{n}} = \frac{\sqrt{\sum d^2}}{\sqrt{n}}$$

$$\sigma^2 = \frac{\sum d^2}{n}$$

$$V = kX$$

~~004~~

$$X + Y = Z$$

$$(X+Y)^2$$

$$V(X) + V(Y)$$

$$V(X) + 2\text{cov}(X, Y) + V(Y) = V(Z)$$

$$\sigma_Z^2 = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\text{cov}(X, Y)}$$

$$2\rho\sigma_X\sigma_Y$$

$$-\sigma_X\sigma_Y \leq \text{cov}(X, Y) \leq \sigma_X\sigma_Y$$

$$\sigma_X - \sigma_Y \leq \sigma_Z \leq \sigma_X + \sigma_Y$$

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_X\sigma_Y}$$

$$\sigma_X^2 + 2\sigma_X\sigma_Y + \sigma_Y^2 = \sigma_Z^2 + 2\sigma_X\sigma_Y(1-\rho)$$

$$\sigma_X + \sigma_Y = \sqrt{\sigma_Z^2 + 2\sigma_X\sigma_Y(1-\rho)}$$

$$X = -kY$$

$$V(X) = k^2 V(Y)$$

$$[(1-k)Y]$$

$$(1-k)^2 \sigma_Y^2$$

$$(k-1)^2 \sigma$$

Idea - take data as % of Σ , thus artificially get const. Σ , make all correlations negative, learn something

Brunk 2-7

Ask Dan: when x_i number 20, & $p=0$, how express $\Sigma \sigma$?

B.

Brunk, 2, - 6

Max possible $\leq \sigma$.

First, let $\rho = -1$

Now if $\sum x_i = 2$ (instead of 20)

$$\leq \sigma = \sqrt{\sigma_z^2 + 4\sigma_x\sigma_y}$$

Knowing σ_z^2 , σ_x , & σ_y , we can give this a numerical value.

However, $\rho = -1$ is implausible (unless we go on % basis, & this may be compelling reason to do so).

It wd. be entirely reasonable to suppose that negative ρ is ~~not~~ beyond the limit we wish to consider. ~~That one~~ That one ^{seems only} should regularly get less water in yrs of greater abundance ~~if it is~~ possible if one deliberately set out to maximise aggregate variability. We shall content ourselves with seeing how badly the system could possibly inflate aggregate variability without going so far as to do anything so flamboyantly designed to Δ agg. variability as to regularly reduce one when others were Δ g. ~~This $\rho = -1$ as min might seem as tho we were just getting a measure of how well~~

Then, with $\rho = 0$,

$$\leq \sigma = \sqrt{\sigma_z^2 + 2\sigma_x\sigma_y}$$

Now it just remains to work out this formula with No. of $x_i = 20$.

It may seem that our measure is "just" a measure of how well correlated the flows are. But it is not immediately evident that correlation reduces agg. variability, nor by how much. It is worth emphasizing for its own sake that fractional division of flow minimizes agg. variability.

Brunk 2-5

So we find $\frac{\sigma_z}{\sum \sigma}$ depends on 3 factors:

1. No. of σ_i
2. Relative sizes of σ_i (and this can be further divided into relative sizes of $\bar{x}_i$, and c.v._i)
3. ρ
4. And perhaps a 4th, the σ_z which is a given for the problem.

{ Kaweah division falls short on all counts.

1. There are too many individual distrib.

2. They are too =

3. Their flows are too little correlated, too independent.

One way to judge system wd. be to take linear regressions as norm, see how jumpy they were in terms of regressions.

Now we can take these one at a time, with others given, evaluate system's performance step by step.

A. Given 20 dists.;

" existing relative sizes of σ_i (and absolute sizes)

" σ_z

Then how well might a system do? i.e., what is

min. possible $\frac{\sum \sigma}{\sigma_z}$? Easy — equals 1 is min., when $\rho = 1$

Then how poorly might a system do? Max. possible $\frac{\sum \sigma}{\sigma_z}$.

(Since σ_z is given, this can be treated as simply max. possible $\sum \sigma$, if more convenient)

Brunk 2-4

At this point $\frac{\sigma_z}{\pm\sigma}$ can $= \frac{1}{2}$ if $\sigma_x = \sigma_y$.

Now we get into negative correlations, and $\frac{\sigma_z}{\pm\sigma}$ can go all the way to zero, if $\sigma_z = 0$. Probly however it could not reach this min. if $\sigma_z > 0$, and the min. in fact depends on value of σ_z . To work this out, evaluate $\frac{\sigma_z}{\pm\sigma}$ when

$$\sigma_x^2 + \sigma_y^2 = \sigma_z^2 + 2\sigma_x\sigma_y$$

$$\begin{aligned} \text{Now: } \frac{\sigma_z}{\pm\sigma} &= \frac{\sigma_z}{\sqrt{\sigma_z^2 + 4\sigma_x\sigma_y}} = \frac{\sigma_z^2}{\sigma_z^2 + 4\sigma_x\sigma_y} \\ &= \frac{1}{1 + 4\frac{\sigma_x\sigma_y}{\sigma_z^2}} \end{aligned}$$

$$\text{But } \frac{\sigma_x\sigma_y}{\sigma_z^2} = \frac{\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y} = \frac{1}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x} - 2}$$

$$\begin{aligned} \text{So } \frac{\sigma_z}{\pm\sigma} &= \frac{1}{1 + \frac{4}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x} - 2}} = 0 \text{ when } \sigma_x = \sigma_y \\ &> 0, \text{ but still small, when } \sigma_x \approx \sigma_y \\ &= 1 \text{ when } \sigma_x \gg \sigma_y \end{aligned}$$

If in fact $\rho = -1$, then $\sigma_z = 0$ necessarily, when $\sigma_x = \sigma_y$, which wd.

mean, with $\rho = -1$, $x = -y$. If $x = -y$, then z would vary, & $\sigma_z \neq 0$.

Note, even with $\rho = -1$, σ_z can $\rightarrow \pm\sigma$ when $\sigma_x \gg \sigma_y$.

Bunk 2-3

$$\sigma_x + \sigma_y = \sqrt{\sigma_z^2 + 2\sigma_x\sigma_y(1-\rho)}, \text{ where } \rho = \text{coeff. of correln. of } x \& y$$

$$\text{If } \rho = 1, \quad \sigma_x + \sigma_y = \sigma_z$$

$$\text{If } \rho = 0, \quad \sigma_x^2 + \sigma_y^2 = \sigma_z^2$$

$$\text{If } \rho = -1, \quad \sigma_x^2 + \sigma_y^2 - 2\sigma_x\sigma_y = \sigma_z^2 + 2\sigma_x\sigma_y - 2\sigma_x\sigma_y$$

Showing $\frac{\sigma_z}{\sigma_x + \sigma_y}$ is highest of all when $\rho < 0$. This is not likely in

$$\rho = \frac{\text{cov } x y}{\sigma_x \sigma_y}$$

my problem, unless & until I express diversions as σ 's, in which case all correlations become negative.

There is complete continuity here. As ρ gets smaller, from its max. value of 1, $\frac{\sigma_z}{\sigma_x + \sigma_y}$ or $\frac{\sigma_z}{\Sigma \sigma}$ gets smaller, and this continues right thru zero to $\rho = -1$, its min. value.

Intuitively, it seems to me when $\rho = -1$, σ_z cd = 0 and $\Sigma \sigma$ cd be > 0 , easily. So that $\frac{\sigma_z}{\Sigma \sigma} = 0$ is entirely possible, when $\rho = -1$, and $\sigma_x = \sigma_y$.

This then gives us our complete range of possibilities, for

$\frac{\sigma_z}{\Sigma \sigma}$. Max. possible is $= 1$, when $\rho = 1$. Gets smaller as ρ does, & as $\sigma_x \cong \sigma_y$.

When $\rho = 0$, reach a bench mark where

$$\frac{\sigma_z}{\Sigma \sigma} = \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y} \equiv \theta = \frac{1}{1 + \frac{2}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x}}} \quad 1 \geq \theta \geq \frac{1}{2}$$

But there are 2 aspects to ~~the~~ $\frac{\sigma_x}{\sigma_y}$. It is an expression of both relative variability and relative size.

So our first test only shows how poorly mgmt is doing, given $\frac{\sigma_x}{\sigma_y}$. But perhaps $\bar{x} \neq \bar{y}$. Say $\bar{x} > \bar{y}$, river is not divided evenly, altho $\sigma_x = \sigma_y$.

Thus far, all we have done is show how the limit to how far $\Sigma \sigma$ can exceed σ_z is set by relative sizes of σ_s , $\frac{\sigma_x}{\sigma_y}$.

There is also a gn. of how the σ_s are apportioned among the flows. Intuitively, if c.v.s are all =, this minimizes some element of variability.

We cd. next take ratio of means, $\frac{\bar{x}}{\bar{y}}$, and tell how it wd. affect aggreg. variability, or $\Sigma \sigma$, if $\frac{\sigma_x}{\sigma_y} = \frac{\bar{x}}{\bar{y}}$.

If $\bar{x} > \bar{y}$, but $\sigma_x = \sigma_y$, then the possible $\Sigma \sigma$, (and probly actual $\Sigma \sigma$ as well) are higher than if $\frac{\bar{x}}{\bar{y}} = \frac{\sigma_x}{\sigma_y}$.

If $\bar{x} > \bar{y}$, but $\sigma_x \gg \sigma_y$ then possible (& probable?) $\Sigma \sigma$ is reduced by ~~the~~ inequality of c.v.s.

Explore more, & summarize.

Brunk 2-1

at ^{upper} ~~lower~~ limit, $\frac{\sigma_z}{\sigma_x + \sigma_y} = 1$, when components are all ~~are~~ vary aly.

at ^{lower} ~~upper~~ limit, $\frac{\sigma_z}{\sigma_x + \sigma_y} = \left[\frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y} \right] \equiv \theta \leq 1$ when components are unrelated, $r=0$

$1 \geq \theta \geq \frac{1}{2}$ [But this can't be right, for σ_z might be = 0? But then $r_{xy} < 0, \neq 0$, contrary to postulate.]
[And as no. of items gets larger, lower limit to θ gets lower, thinks Dan]

$$\theta = \frac{1}{1 + \frac{2}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x}}}$$

When $\sigma_x = \sigma_y$, $\theta = \frac{1}{2}$

When $\sigma_x \gg \sigma_y$, $\theta = 1$

And cd. define a fn. in between, $\theta = \frac{1}{1 + \frac{2}{\frac{\sigma_x}{\sigma_y} + \frac{\sigma_y}{\sigma_x}}}$

(Naturally $\theta \rightarrow 1$ when σ_x & σ_y become very diff., for then one of them dominates & situation approaches aty.)

This tells us that if river is divided evenly among users, ^{more exactly, if the σ are =.} then $\Sigma \sigma$ can be double σ_z . ^{if $r=0$ among components} (for 2 users - & higher if more than 2).

So take the σ_x & σ_y we actually have, we can from their ratio, $\frac{\sigma_x}{\sigma_y}$, tell how far their sum can exceed σ_z . ^{if $r=0$} Then cf. this, (as a ratio), ~~with~~ with how far their sum actually does exceed σ_z .

And you have an expression for how far the actual distribn has exploited the potentialities of mischief which it has. Which mischief in this framework springs entirely from non-correlation, and probably cd. be measured that way, too, if more convenient. I might simplify my analysis by simply taking r of each flow with Σ divisions, taking weighted average thereof.

Don't dismiss -p as implausible.

There is some, recall, some srs. take more
water in dry years.

$$\frac{\sigma^2}{\bar{x}} = \frac{\sum d^2}{n \times \bar{x}} = \frac{\sum d^2}{\text{Total}}$$

$$\text{C.V.} = \frac{\sigma}{\bar{x}} = \frac{\sqrt{\sum d^2}}{\sqrt{n} \bar{x}} \leftarrow \text{This may}$$

be a distorted index, as between ~~the~~ dispersion of low and hi ~~in~~ $\sum d^2$. ~~A given~~

~~cv might come from many n & large d's, a few n & large d's small d's~~

The square root over the $\sum d^2$ is the villain. It damps the index of variability. It more than compensates for squaring the d 's, since they are smaller. Then ~~to~~ ^{by} off. it to the mean, we understate relative importance of dispersions that are wider.

So variance is a better measure - & one that math. statisticians seem to prefer anyway, and perhaps one that has its rationale all worked out.

I also might return to simple $\frac{\sum d}{\bar{x} n}$, see how that works out.

Stat. From Cecil Gregory, Rural Soc., U. Mo., 10-58

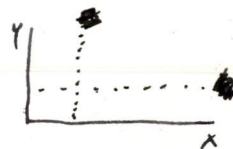
1. Coeff. of correlation, r . $\nearrow$ x & y here mean deviations -- from mean, a regression?

$$r_{12}^2 = \frac{\sum (xy)}{n \sum (x)^2 \sum (y)^2} = 1 - \frac{S_{xy}^2}{\sigma_y^2}$$

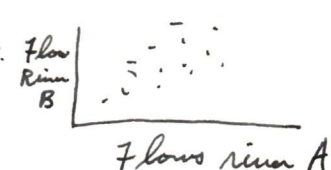
... this must mean est. of y from x .

This coeff. of correlation is independent of whether you take x on y or v.v. ?

It's meaning is seen in ~~the~~ right-most expression. S is standard error of est., i.e. deviations of points from your regression line. σ is deviations from mean. Ratio is index to your success in finding a relation. If all points fall on your line, e.g., $S = 0$, $r^2 = 1$.

? But here is bias. Cf. 2 series: 

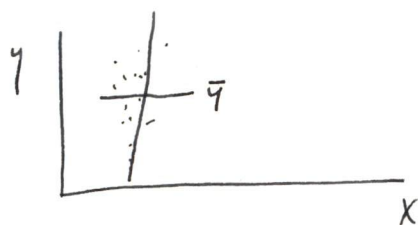
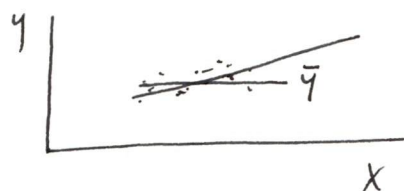
For vertical series σ is hi, for horiz. is low. ($\sigma \equiv \sqrt{\frac{\sum (d^2)}{n}}$)

~~But~~ is just one series involved here: 

σ_y is for flows of B, alone.

But in case of vertical line, interpretation tough. Are they unrelated, or extremely correlated? If x is a constant river, & y changeable, then σ_y is high, which makes correlations look good.

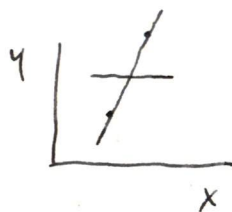
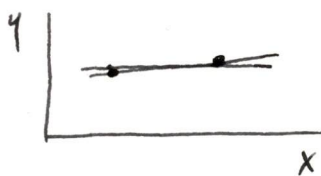
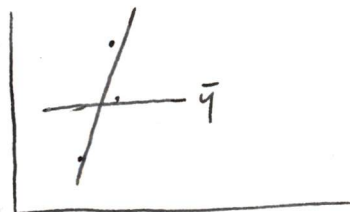
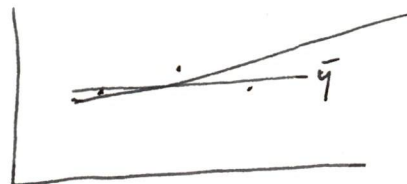
?



On right ~~side~~ ^{graph}, σ_y is smaller. But so is S_{xy} ... error of est. of y from x . For same reason - y doesn't deviate much from one value.

Might seem more improvement possible on left graph.

Take simple case, 2 points:



Probably chances for 90 improvement are =. Probably stat-ns have worked it out.

2. Formula for combining σ s:

$$\sigma_z^2 = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2 r_{12}$$

This brings in correlation coeff., r .

[Note, if $r=0$, $\sigma_z^2 = \sigma_1^2 + \sigma_2^2$, and $\therefore \sigma_z < \sigma_1 + \sigma_2$,

(by Pythagoras) so $\frac{\sigma_z}{\bar{z}} < \frac{\sigma_1}{\bar{1}}$, and likewise $\frac{\sigma_2}{\bar{2}}$, likely]

~~as before~~ [If $r=1$, we could expect $\frac{\sigma_z}{\bar{z}}$ to lie between
coeff variation of 2 originals]

Get formula for relations of coeffs of variation. It
should make clear that these tend to be reduced by
summing.

$$\sigma = \sqrt{\frac{\sum(d^2)}{n}}$$

d	d^2
0	0
4	16
2	4
3	9

$$\sum(d^2) = 29$$

$$n = 4$$

$$\frac{\sum d^2}{n} = \frac{29}{4} = 7.25 = 2.69^2$$

Now add $d_5 = 0$

$$\frac{29}{5} = 5.8 = 2.408^2$$

Now suppose $d_5 = 5$, $d_5^2 = 25$, $\sum(d^2) = 54$

$$\frac{54}{5} = 10.8 = 3.29^2$$

So one high deviation has more effect on the measure of deviation than one zero deviation. Is this a good measure then?

Job #1.

Remaining work, 6-1-59.

1. Check Persian Ditch adjustments, annual flows.
Use 40 sf as max, instead of 80 sf?
Recompute σ therefrom.

Consider omitting winter flows from annual "diversions"

2. All diversions, check reasonableness of figures. Maybe winter diversions not real? Maybe some is diverted just to spill, maintain right.

Figure needed adjustments for this.

3. Spot-check all Harold's work. Try to devise quick cross-checks.
Check his handling of Deep Ck. division between Fm's D. & TID, et hoc genus omne.

4. In winter, heavy Kwh flows must largely escape. Trace channels, find out where.

5. Inspect tables visually, write down impressions — may get as many good comments on system that way as by detailed analysis.

1. 6. Tabulate all results, inspect, & use to formulate future plans.

7. Check all Harold's Fm's Ditch data — he forgot to deduct Deep Ck. Cut, once at least.

8. Find how much wtr you can apply to an acre in one day.

9. For Jennings, choose year other than 1946, when couldn't divert full entitlement

10. Measure area within outports of WW Co. — & cf. area served.

Measure length of canals,

Remaining work, Job #1, p. 2

11. Readjust proportion of LSI D flows - have them decline & Watch Birch as whole. To get WWG. diversions.
12. Fill in "O"s, where Harold left blanks.
13. Finish major projects:
 - A. Skew
 - B. LCR
 - C. Length of season
 - D. Stability or correlation of seasonal patterns.
 - E. Clustering of deficits.
 - ✓ F. Variability & capacity.
- ✓ 14. Adjust all data for seasonality. As they stand, I'm treating winter flows equivalent to summer, etc.
Weight each month, or something of the sort.
- ✓ 15. Better than range is to compute min. as % of mean.
This only involves one extreme, & the more signif. one.
So compute $\frac{\text{min.}}{\text{mean}}$, add to range data.

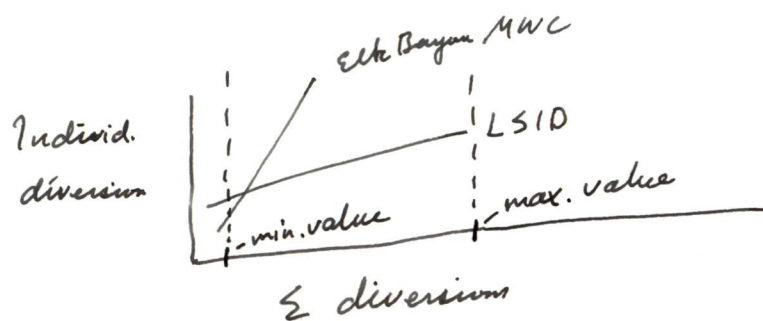
Analysis of Variance:

For a simple one see

Paschal & French, 1956

USDA Tech Bul 1141.

Correlation & regression, I



Correlation coeffs. are measures of how well actual relationships conform to a linear regression.

Above, LS & EB are each perfectly correlated with Σ , and with each other, if all points lie on lines.

So $\rho = 1$ does not mean well, I wonder if Ben's formulae take account that $x_i = k_i \Sigma$ doesn't cover situations; but $x_i = a_i + k_i \Sigma$.

~~Now EB~~ ρ is a mismeasure of how jumpy the curves are.

Even with $\rho = 1$, still have great inequality in distribn. of vblty if EB & LS regn coeffs $\neq$.

[This may be another way of observing that $cv_1 \neq cv_2$?]

With $\rho = 1$, $\frac{\sigma}{\bar{x}} \propto$ regression coeff or λ ; ~~to~~

But $\rho = 1$ is not same as $y = k \Sigma$.

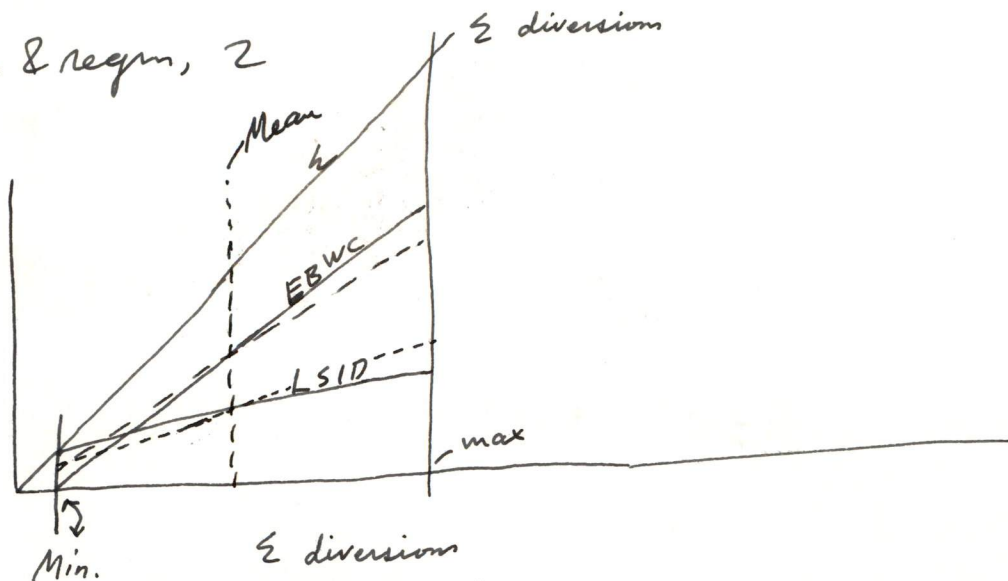
Having minimized aggreg. vblty, is still gn. of distribut'g it = ly, or pooling risks.

Circle & region, 3

Root gn: $\rho=1$, how min. $\angle \sigma$? By radius vectors or || lines?

Circle & regn, 2

Individ
divisions

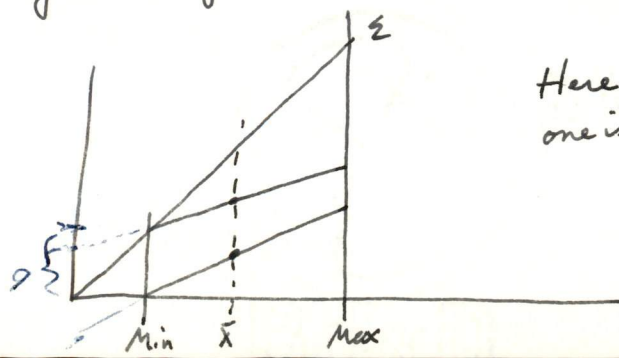


$$y_i \neq k \Sigma, \text{ but } = a + b \Sigma$$

Dani's = us told us that $\Sigma \sigma = \sigma_e$ when $p = 1$. And somehow we equated this with fractional division; but now they seem very different.

Suppose the solid lines above are original position, & (---) rep. a change toward proportionality. This rep a shift of values from larger deviations to smaller ones, and since they are all squared, should reduce aggreg. vblty, or $\Sigma \sigma$.

After we've achieved σ_y , do further shifts no longer $\Delta \Sigma \sigma$? It would seem so, until they ||. You can't get them || w/o changing Σ , but you can get them nearer to || than radius vectors:



Here they can go past ||, so lower one is steeper

Job # 1. Details remaining, July 24.

Check
Greg-
doing
this?
Yes-
will
call.

- ✓ 1. Annual data. Get 2 diversions; and Kaweah R. less 2 diversions. ~~Gregory is doing this~~
~~Gregory can do this.~~ Be sure he does not include the Kaweah R. itself as a diversion!!
2. Instruct Gregory to change weights of $2\frac{1}{2}$ & $1\frac{1}{2}$ to 3 and 1.
3. Follow up on corrections Greg. now running on Evans, Watson, and especially Tulare I.D. Retabulate corrections when they arrive.
- ✓ 4. Monthly & daily. Get Longs Canal. Adjust 2 diversions for whatever changes you make - subtract previous assumed constant flow (if any), and add actual flows. Probably Greg. can do this if you give him data, but be sure to check them - they make errors there, too.
In years of missing data, fill in with data from comparable years.
5. Check computer results:
 - a. Are initial data properly arrayed in order of size?
 - b. Are solutions mutually consistent? (a good check seems to be that coeff. of variation = 2x Lorenz Conc. Ratio]
6. Tabulate Computer results when available.
7. Ask Gregory to print out $\bar{X}$ for each series.

Kaweah analysis, summary. (annual data)

Idea was to see what happens when you take a run apart. — is there a loss of aggregate stability, due to un-pooling of risk.

To answer, take a weighted average of means x coeffs of variation.

And note diverters have only taken a little over $\frac{3}{4}$ the water,

~~presumably~~ the steadiest part, so should have gained on run.

~~Just~~ Just sum the σ s, see if they don't exceed σ for Σ diversion

	σ	
Kaweah	177,000	
Σ diversion	138,000	} — 189,000
Kwh less Σ	51,000	
Sum of individual σ s	150,000	} — 201,000

$\frac{\sigma}{\bar{x}}$ for sum of individual diversion σ is $\frac{150,000}{333,000} = .45$

Thus they have not improved on Kwh, which is .43; and have left over Kwh less Σ of .66.

On monthly & daily data, must abstract from seasonal movements before σ tells much — or use some measure of variation other than σ

6. On Kwh we find, as expected, annual diversions vary less than river; some srs. vary not, or even perversely; some jr. vary more than river — & even more than river less Σ diversions. This last is gross defect of system — we should at least be able to serve the last jr. as well as river less Σ .

7. There are some who need ^{cyclical} storage: Elk Bayou, TID, Parkwood, (Persian?), and ~~more~~ need seasonal storage. But those who do not need such storage, due to their wtr, occupy much of the area. Thus no compact wtr user group cd. be formed. This wd Δ costs. And of course its recharge benefits wd. go to nonmembers in part. Whereas now by deep pumping they get wtr from neighbors.

5. My thought had been that residual wth claimants would need to bring in seniors to get steadier annual flows, to ΔMP of storage. But this presumes that seniors have something that needs storing. There is no gain in bringing in seniors whose demand = supply, even if perfectly steady. But if their demand $>$ supply there is gain. But not on the steady part. Just they use storage, like any other jr. (they do need seasonal storage)

No - my thought was that use should vary with scarcity. And steady ~~flows~~ diversions do not allow of this. Thus it is that seniors are skinning cream.

Cal. ~~the~~ Dir. W & Res rpt. is deficient & misleading in this, bases on assumption of what steady annual supply can provide. Thus introduces a rigid fixed element, largely for analytical simplicity, that should be let vary. Analysts not doing their job. Can get by with less storage if let price vary with abundance.

Now if seniors could require a constant annual supply, wd. be bad enough. But on Kwh, worse, they can require constant rate of flow.

What obstructs Kwh. people getting together to store their individ
excess winter waters for summer use?

1. Might deter U.S. aid. They need to show falling wtr table.

NB anomaly: LA has excess wtr, so can bargain stronger for more.

Kwh. avoids getting wtr, to get U.S. aid.

Not entirely anomalous, tho. Idea both cases is to have large claims
to local wtr, but before developing fully, try to import.

Question: who backed CVP? How did they ever override
opposition? Probly some big specs working behind scenes - anomaly
that they should be expansive element in society, & small firms of
Modests etc. be restrictive.

2. Wd. be ~~very~~ very hard to determine relative rights, based on
natural flow, after that altered. (altho they seem to have on Kings)

3. Pumping may be cheaper than storing for surface diversion

4. Those who most need the water, & might be willing to pay,
can't get any. Must wait until need is felt by those with
good water already. This is pretty conclusion. And those
with poor wtr probably can't condemn better wtr, as there is some 'use'.

Water-rich will have higher wtr tables usually, prefer to pump

5. Could wtr-rich sell their winter wtr, keep summer? Law doesn't
envisage - another defect.

No one shd be let claim beneficial use for winter diversion. Yet no Kwh rts are liend to summer. This is what obstructs storage, seasonal. Too many interests. You can drag together landholders via. I.D.s. But how coordte holders of indept. wth rts?

They cd. join in big I.D., & perhaps wd. if under severe pressure. But U.S. comes in, rescues, is easier. Local way wd require pooling rights by ID buying them. But cd. ID condemn rts? CCID did, I think. Check this point - F. Adams may have e.g.s.

All excesses of winter claims over winter diversions are waters legally tied up, but unused. A storing agency probly cd not file on these as unused wths, w/o great litigation.

So sum up legal diversion max. Bull 49 B makes clear they are not all "legal," but based on local compromise, often. They $\Sigma = 2888$ sf plus seepage + riparian. Say ripn might claim 200 sf. Seepage allowance should be 500 sf or so probly, (altho it is only c. 50 in July). Say 300 seepage. So 3400 sf is Σ .

River rarely rises so high. Any storage by newcomer to salvage occasional floods wd. be strictly long term storage. In 1952, could have claimed c. 32,000 sf flows > 3400 sf.

Re $\leq$ diversion etc., p. 2

What is available for storage is all winter runoff, even if diverted; and a diminishing % of spring runoff.

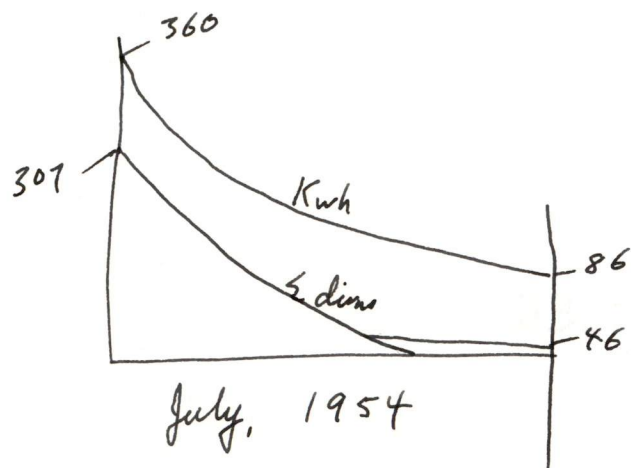
You can est. seepage rate as June or July. Kwh less $\leq$ diam, add that as const. diversion in rest of year - it wd be greater with more wt in river, but saturation of ground might offset this.

Clearly needed is small storage to catch winter flows, then spend all year recharging, empty resr by Nov.

To minimize needed resr., want to max. ^{use of} recharge areas ^{in h¹ months.} This is impeded probly by present rigid requirements. Recharge rate is the bottleneck - - - All rts by agreement are based on stream flow reaching point of diversion. You can't $\therefore$ take out any upstream w/o necessitating new downstream agreements, among 10-20 contentious parties. (But U.S. is doing it - how?)

Some of winter wt is sewed up by those with summer rts - Foothill, WW Co, ripns, LC, MR, Ham. They wd not pay for storage of their winter wt. Might sell it - bilateral mpy + ho.

Re Σ diversions etc.



In dry mo, as Kwh gets very low, higher % sinkes. Hence Σ diversions goes down faster than, or as fast as, river. Hence wd. expect in July, & for Σ divs wd. be higher than for Kwh less Σ .

Other factor is of course that I originally relied on — as Kwh gets lower, diversions take higher % of what comes along.

But by July they are taking all of what comes along anyway. So diversions drop faster than the river, ~~or~~ or as fast as it, or nearly as fast, ~~or~~ be same amt or more river wd sinkes.

Check this thru with original data. It is right.

	July 1	July 31	$\frac{\Sigma}{K}$
Kwh & River	360	86	.42
Σ	307	46	.51
Kwh - Σ	53	40	.23

Note that this Kwh less Σ is not the residual available for storage upstream. I have wasted time computing it. It has to flow, to get wd. down to ditch-heads; & it is not wasted be goes underground.

$$A = B + C$$

$$Kwh = \Sigma + Resid.$$

$$\frac{\sigma}{\bar{x}} \stackrel{?}{=} \frac{\sigma}{\bar{x}} + \frac{\sigma}{\bar{x}}$$

A	5	1	6	4	10	14	12	8
B					75	59	6	4
C					3	9	6	4

Correlate Kwh flow with Σ diversions.

Constant waste & variable use suggests reparations or someone has a fixed claim that must be let pass by.

$$\begin{array}{r} 55 \\ 55 \\ \hline 275 \\ 275 \\ \hline 30.25 \end{array}$$

Last thoughts, p. 4

Sampling may cut down work on daily flows 2 ways:

1. Take a few days
2. Take them for just a few yrs. This makes more sense, as the diff. yrs are more random, while days in year are not.

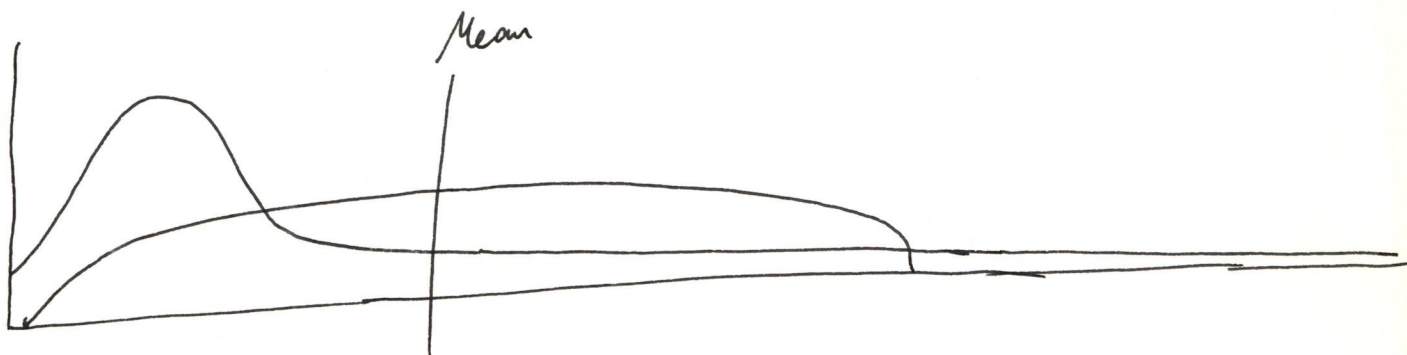
Might get by using e.g. $\frac{1}{2}$ the yrs. As to days, stratify them according to interest: 3 within the 10-day bloc, & others 2 per month.

Select years for representativeness - thus use smaller sample.

Take longer samples for more variable distributions, & v.v. - many for Elk B., few for Foothill; many for July, few for April; ~~etc.~~ higher % from high runoff yrs.

Make finite popn correction - p. 498

p. 61 etc. Techniques for establishing diffs. - use re ~~diffs~~ $\frac{\sigma}{x}$, LCR, etc., by sampling.



Flat top freq. distn. is better, if you are building storage, can capture whole flow with less capac.

~~But if no storage, as on a Kuck ditch, skewed is a better way to get a~~
~~flow. 1 is more~~

Also if no storage, flat top is better. Keeps ~~med~~ median put., whereas skew moves it down, & you can't rely on getting mean flow but less than half time.

Large part of total flow is made useless, ~~because~~ comes so bunched, in a few big years.

So:

LCR is superior to σ as index to variability.

Having data for several ^{random} days, set up hypothesis that

Use χ^2 to test for signif. of hypoth.

$\frac{\text{LCR of diversions}}{\text{LCR of Kuck}}$ As, as
more from
monthly to
daily figs.

Take these data:

Months: March - Aug

10-day interval: July 1 - 10

Days: July 3, 6, 9.

Last thoughts, p. 2

Do not do each day of a 10-day block individually. One day is enough - of its $\frac{\sigma}{x}$ with 10-day block. Since one day is likely to be like another, this is as impressive as taking them all & of σ with σ . Different is case of taking ea. company separately, they are not alike.

But check Poisson distribn, law of small nos, etc., signif. of small sample.

Spread days over season, get seasonal variations.

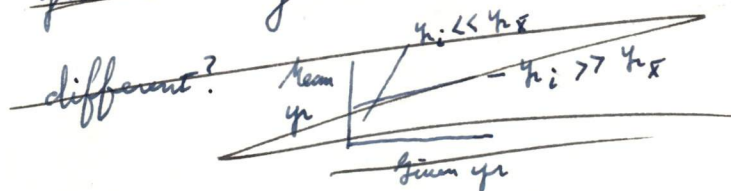
Shall I take σ of each ditch for ea. mo. (over many yrs)^{ea. monthly column}; or sum ~~over~~ ^{over the yrs & correlate} each yr's seasonal pattern w. mean[?] or Σ pattern?

If latter, will have set of Σ patterns, can do something with? I don't know what.

If take σ s, I have something specific to ea. mo., gives new info.

Correltn gives something specific for ea. yr. ~~but of ambiguity?~~ ~~Res if~~

~~yr's Σ nearly = ~~mean~~ Σ , then is harder to get his Σ than if very~~



So ask him to get both - each gives some useful info.

Monthly & daily flows - last thoughts before summary.

LCR may obviate σ , but get σ anyway bcs more familiar to most; & yields some diff. kinds of info.

~~Make freq. distn of Kameah - the bimodal I have in Am. R.~~

Kameah, 1890-1955, not notably bimodal. But very skewed.

✓ Over 550,000 afy, have $\frac{15}{66}$ years, with c. ~~19m~~ $\frac{11m}{28m}$ af. Take freq. distr. for 1920-55, which is what have for Am. R., see if similarly bimodal. If not, Kwh must be uncorrelated with σ etc.

σ alone leaves much untold. E.g. in very dry years, seniors tend to take more than usual. E.g. LSID in 1924, lowest year on record for Kwh., 101,612. No doubt their local source dried up, it was worth their while to spend more on pumping wtr.

Run correlations of σ s vs Kwh., to illustrate this.

Why did LSID increase its diversion so in 1944 et seq.? Write Don Burr, & read ~~the~~ TID vs LSID. Bulls 21 & 49 differ on this.

Dry cycle, 1924-34. See if correlations work out better then. Especially in 1929-34, after 5 years of drought.

Take a ^{weighted} σ of the σ s, to measure how unequally risk is distributed; and how inequality increases as take smaller time units. Probly it $\downarrow$ s with smaller time units, just as total σ will.

Snedecor, George W. (& Cochran, Wm G.) Stat. Methods Amer: 9a State Col. Pa., ~~1937~~ 1937.
5th ed, 1956, reprinting of 1957.

"Enumeration data" are like yes or no — e.g. so many beams are black, so many not. Hence "binomial" distribn: yes or no.

"Measurement data" are measures.

"Stratified random sampling" — break up popn into strata or parts, take from each part a random sample based on or $\propto$ its size.
— thus assure regional spreading. (Geostat?)

Agenda: make freq. distribns of river flows.

also LCR ... lowest 10% yrs have 2% flow, etc.

If this index proves more sensitive etc. than σ & skew,
try it on integrating some rivers.



Probably I shdn't use σ at all for skewed distribns —
check on that.

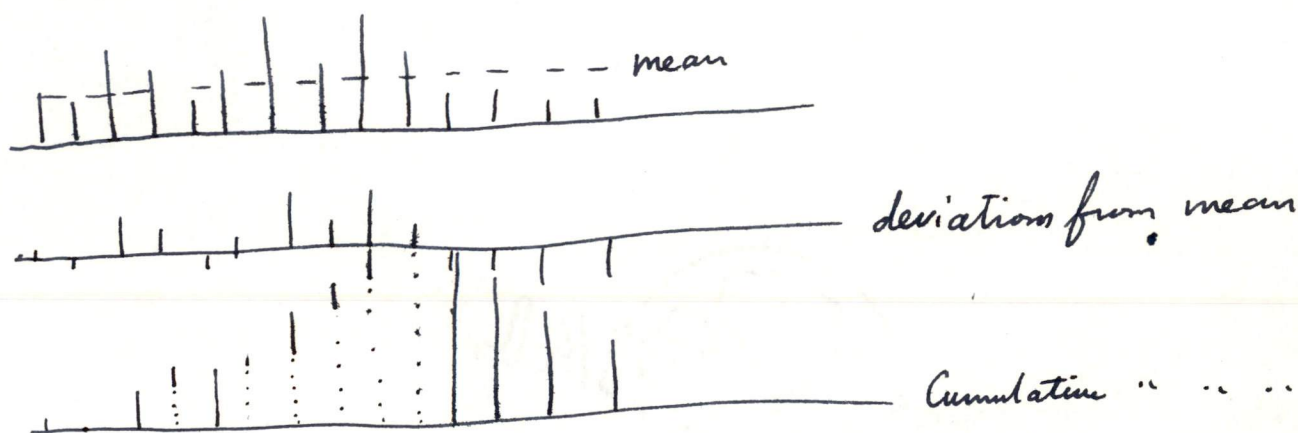
LCR is going to be better than σ , so on these
monthly & daily data, get LCR right from start.

Use max. no. of trapezoids in computing that computer
can handle, up to ~~max~~ no. of observations.

Kirk R freq. distribn. yly flows is skewed to rt, & bimodal, as I recall. Not "normal." What does this do to t significance?

μ (μ) customarily denotes mean of popn, while $\bar{x}$ is mean of sample. I have evidently been misusing $\bar{x}$.

Take cumulative deviations from mean, & freq. distribn of them, or σ or a LCR — thus index to clustering



Clustering deficits is in point when we ask re use of storage. And groundwater. But for regular surface inn from Kirk, dry y^r is no worse because it follows another dry year.

Clustering is important as to daily deficits, tho.

Snedecor, 3.

p. 31. σ of a popn is estd. from range of sample.

Thus if I take 1919-55, or whatever, as sample of flows over years, I can est. from it popn. ~~range~~ σ !!! - for all time well, recent time, anyway. And can take future flows as part of popn.

✓ Use sampling to select days, for analysis of daily variations. Thus cut work, but have valid index to whole.

— If I take 6 mos. for monthly data, why take annual data for 12 mos? Better rpt. by Σ 6 mo. data? *

— Can blame wtr its for variations only when that is limiting factor. March, April, May, irr. demands are cause of variations, supply for most users is excess (inspect monthly Σ diversions vs. monthly flows Kwh) (to check that).

Watch if you break something into 12 parts, you $\Delta \sigma$. So what is to be seen is whether $\frac{\sigma \text{ of diversions}}{\sigma \text{ of Kamaah}}$ doesn't increase.

Esp. in June - August. Watch bias due to seepage.

Now there is a hypothesis, to test by sampling & χ^2 test.

Snedden, 4

1ⁿ July's of hi Kwh flows, is more seepage, but less % seepage.

1ⁿ low July's, $\therefore$, Σ divms will decline less than Kwh. That's ok.

2, or 3 problems:

a, How select days - a block, a sample, monthly-stratified sample?

b, Avoid spring mos.?

c,

b, Spring mo., with unused flow, shd be no constraint on any diverter. Hence expect Σ diversions to be quite steady.

My points will be that $\Sigma \sigma$ of diversions $> \sigma \Sigma$ diversions;
and $\sigma \Sigma \text{ divms} + \sigma [Kwh - \Sigma] > \sigma Kwh$.

Summer mos., greater $\frac{\sigma}{x}$ for Kwh.

Proably also greater Δ stability by joining many rivers of CV -
try that again.

a, Select days in summer - say June 1, 15, July 1, 15.

A day may be too short a unit. Except as regards predictability,
is no great advantage to steady flow, when one below ditch capax.
For small cos, unsteady flows even better, but not worth handling

very small flows.

At any rate there is a unit of interest, between month & day, which is period you can go between irms, c. 10 days. Just one 10-day pd. is all the work it is worth, [& maybe more]. Say June 10-19. Then as a day, take June 15.

I think for a given day, Kwh wd. be almost as reliable as for month. But a ditch much less so. Bes many are subject to cutoffs. Now this daily σ is signif. in that it brings out the day-to-day uncertainty under which many cos. operate, & spares me from a separate measure of how many days were perilously near cutoff line, etc.

So take these data:

Months: March - Aug.

10-day interval: June 20 - 29

Days: June 22, 25, 28

p 37-8. You can est. σ from a sample, quite efficiently. —?

Therefore I should cut my workload way down on Kaueah, & increase scope, by sampling.

Esp. on the 10-day interval, much work, if compute for every year. As to take sampling of years, specify confidence limits in footnote. But you must have a normal popn (p. 31) I don't. Maybe he tells how to adjust for this, later. On p. 36 he mentions that applies pretty well to abnormal ones, too.

p 38, conversion factor from range of sample to σ of popn depends only on sample size, regardless of popn size! How come?

p. 43 Sample σ differs from σ Thess: divide by $n-1$, instead of n .
 $n-1$ is "degrees of freedom".

Anders, 7

If there are 2 sources of variability in a series, they ~~can~~ may tend to offset as well as augment each other, leaving no net increase.

Example. E.g. July, 1942

E.g. a series of July flows is affected by both annual & seasonal variation.

If these offset as often as they augment, result will be higher σ , due to squaring; but mean deviation unaffected. So?

Cfg σ of parts vs σ of Σ . wdn't $\sigma \Sigma$ be larger because squaring larger figures? No

x_i	x_j	Σx
5	3	8
4	4	8
3	5	8

Perfect offsetting

$x_i \quad d_i \quad d_i^2 \quad x_j \quad d_j \quad d_j^2 \quad \Sigma x \quad d_{\Sigma x} \quad d_{\Sigma x}^2$

3 -1 1 3 -1 1 6 -2 4

4 0 0 4 0 0 8 0 0

5 1 1 5 1 1 10 2 4

Σ 12 12 2 12 12 2 24 14 8

μ 4 4 8

$$\sigma_i = \sqrt{\frac{2}{3}}$$

$$\sigma_j = \sqrt{\frac{2}{3}}$$

$$\sigma_{\Sigma x} = \sqrt{\frac{8}{3}} = 2 \sqrt{\frac{2}{3}} = \sigma_i + \sigma_j$$

Suppose we augment one, reduce another:

$$\begin{array}{r|l|l} 4 & + & 2 \\ 4 & + & 0 \\ 4 & - & 2 \end{array} = \begin{array}{l} 6 \\ 4 \\ 2 \end{array}$$

If we begin with no variability, we get some.

$$\sum(d^2) \ll (\sum|d|)^2$$

If take σ for each diversion (simple annual figures only say) and cf. to σ of $\sum$ diversions, wdn't latter likely be much longer, due to effect of squaring? No, bcs we take $\sqrt{\quad}$ also.

What makes the diff. is if we somehow reduce n , & allocate same deviations among fewer n . But I don't know when that wd even happen.

$$\begin{array}{r|l|l} d & & d \\ 3 & 1 & +1 \\ 4 & 0 & +0 \\ 5 & 1 & +1 \end{array}$$

12

2

$$\begin{array}{r|l} d \\ 4 & \frac{2}{3} \\ 4 & \frac{2}{3} \\ 6 & \frac{4}{3} \end{array}$$

14

Here we $\Delta \sigma$ (by the coeff $\frac{2}{\sqrt{3}}$) by adding a variable which both offsets & augments. E.g. is complicated by a change of mean. Do another, keep same mean.

$$\sigma \quad \sqrt{\frac{2}{3}} \quad \sqrt{\frac{2}{3 \times 3}} = \frac{1}{3}\sqrt{2} \quad \sqrt{\frac{16 \times 4}{9 \times 3}} = \frac{1}{3}\sqrt{8} = \frac{2}{3}\sqrt{2} = 2\sqrt{\frac{1}{3} \frac{2}{3}} = \frac{2}{\sqrt{3}} \sqrt{\frac{2}{3}}$$

Sned, 9

A.	d_A	B	A+B	d_{A+B}
2	2	-2	0	4
2	2	+2	4	0
6	2	+0	6	2
6	2	+0	6	2
<hr/>			$M = 4$	4

$$\sigma_A = \sqrt{\frac{16}{4}} = 2 \quad \sigma_{A+B} = \sqrt{\frac{24}{4}} > 2$$

~~To be continued when kids ~~quitter~~.~~

Sampling problem: out of flows from 1919-55, or 37 years, how many must I take to have 95% confidence in results?
(Adjust for abnormally distrib. popn)

— Cd take as sample all yrs when annual flow within 10% of mean. But that is biased — omits deviant yrs.

Cd. abstract from annual variations thus: mult. each month by ~~mean~~ $\frac{\text{mean annual flow}}{\text{current " "}}$. Thus abstract from annual variation

leave just seasonal. Then take σ , of Kwh. vs. distrib.

But I don't know that this gains anything, for my needs, over simpler plan.

Snedecor, 10

Upper case letters rep. data; h_c rep. deviation. p 41

s is the sample s.d.;

σ " " popn " ".

Students' t is a tabulated constant telling us how confident we can be that our sample mean is within some % of popn mean.

$$t \equiv \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Tabulated p. 46 Useful mainly for samples under 30.

$\bar{x} - \mu$ is deviation of est. mean from popn

$s/\sqrt{n}$ is sample standard error — an approximation to $\frac{\sigma}{\sqrt{n}}$, standard error.

$$\left[\frac{\sigma}{\sqrt{n}} \text{ w.d.} = \sqrt{\frac{\sum d^2}{n^2}} = \frac{1}{n} \sqrt{\sum d^2} \right]$$

$$s/\sqrt{n} \equiv s_{\bar{x}}$$

But what I want is to est. σ from s , & draw large enough sample to know confidence is 95% within some range.

From p. 46, I'd guess a sample of 7 or 8 wd. be ~~plenty~~ enough. I still can't believe, tho., that size of popn can be ignored.

Sned, 11

a random sample wd. be foolish, when we have info. to get a better-than-random sample (p. 492)

E.g. select 5 yrs whose annual flows $\div 5 = \mu$; & which represent full array. 2 above mean, 2 below, one on it. ~~Asymmetrical~~
Let their s also = σ of popn. ^{Get some symmetry, but remember distn} is bimodal.

This is zillions better than any random sample. It is a stratified sample, without any randomness, but very representative.
Esp. as popn is abnormal.

Remember to allow for sample- s tending below σ . ~~to~~
But so long as we are just cffg sample vs. sample, is of no concern.

Any relation of bimodal annual freq. distn, & bimodal seasonal pattern? Very probably. Bimodality over yrs argues that winter ~~main~~ runoff & ^{spring} snow melt are independently determined.

This may be useful in figuring optimal storage, since snowmelt is nicely timed to demand; it is winter runoff that needs storage.

p. 508.

One needs larger samples where σ is larger, for same reliability.

So for Elk Bayou I might need twice as many as for Foothill D. He gives a formula that allows for both variability & cost of sampling in determining opt. sample sizes. To be sure, this formula is for a stratified sample's parts, but I shd be able to apply same principle to diff. ditches.

Stratified sampling is esp. recommended for "highly skewed" distns, e.g. income concn. Mine is only slightly skewed, but maybe stratifying wd. help here, too.

p 509, e.g., shows you take more of your samples from larger strata

p. 512 You can sample in 2 stages. E.g. flow of ~~diff rivers~~ several rivers, river is ~~some~~ nat. unit, in which you subsample.

Generally it is assumed popn is ∞ or very large. If sample exceeds 10%, however, must make "finite popn correction." - p. 498

p. 61 etc. A big problem in using techniques: they all use some est. of σ as means of sizing sample for given confidence in ~~some~~ value of mean, or $\bar{x}$, or etc. But what if σ itself is what you want to estimate?

Ask Dan

Maybe some hope in technique for est'g differences — differences of σ are what I am after. p 49 et seq. Thus I might use pairing: take 6 years, say, & pair σ of Kwh vs ditches for ~~each of the~~ the 6-yps of data.

Some users evidently enjoy 100% regularity over several weeks, ff. or preceded by extreme fluctuation. They need the 100% regularity less than to rid selves of the extreme flucsts.

They get the 100% reg. by imposing hi irreg. on jrs; & get the extreme flucsts from srs. Thus all parties lose more than they gain from system. But they cling so strongly to the injustice they can impose on others, they are oblivious to the greater benefits they might gain.

~~of~~ Sum To of flow supplied under conditions more & less favorable than entire river.

Give each a moment-arm; its m.d - Kuk md e.g.; we then take ratio of better to worse + than - Kuk R supplies.

Before plunging into new monthly-daily sigma study, check:

- Lib. 1. How many years needed to make sigma significant? ^{check stat book Jack cited or ask Cecil} ~~What are the~~
~~t-test & Chi² tests?~~ Can LCR be more accurate with ~~given~~ ^{given} sample than σ ? In either case stratify, by size. Due to abnormal distribn, might be better off with LCR, if sample?
- X 2. On the daily, how many & which days to choose?
One a month?
- X 3. Tabulate results of .53-.54 correlation, may get ideas.
- X 4. Consider summing each month's flows over considerable time, then correlating each yearly pattern with pattern of sum, & summing correlation coeffs. Check with Gregory.
Might use one method to check the other.
Either method involves us only in laying out monthly flows in one format. Just needs a different computer program. To get both.
- X 5. Lay out Ditch Co. tables for Lake in simplest form.

Analyze one ditch in detail, days' flow received under conditions of great uncertainty, defined as 50% or more change of cutoff

also note days' flow ~~to~~ so low as to waste as much irn labor as wtr is worth.

& No. days with 50% or more risk of flows below ec. level.

& No. days " " " " change [to great as require change of men]
of over 25%

Figure entitlement, some big year, see if lacked capacity to take it. Or just look for possible flow situation when wd lack capab. to handle.

? Figure max possible entitlement, at highest recorded river flow

Monthly flows.

What more of variability will ~~allow~~ test effects of splitting?

What is good? (More flow in summer - another question)

.. .. in terms of stability?



One season should most resemble others.

Measure: take $\frac{\sigma}{\bar{x}}$ of each month, over several yrs, for Kwh & distrib. If get same Δ instability as for annual, means is no added criticism. But probly months will work out more unstable.

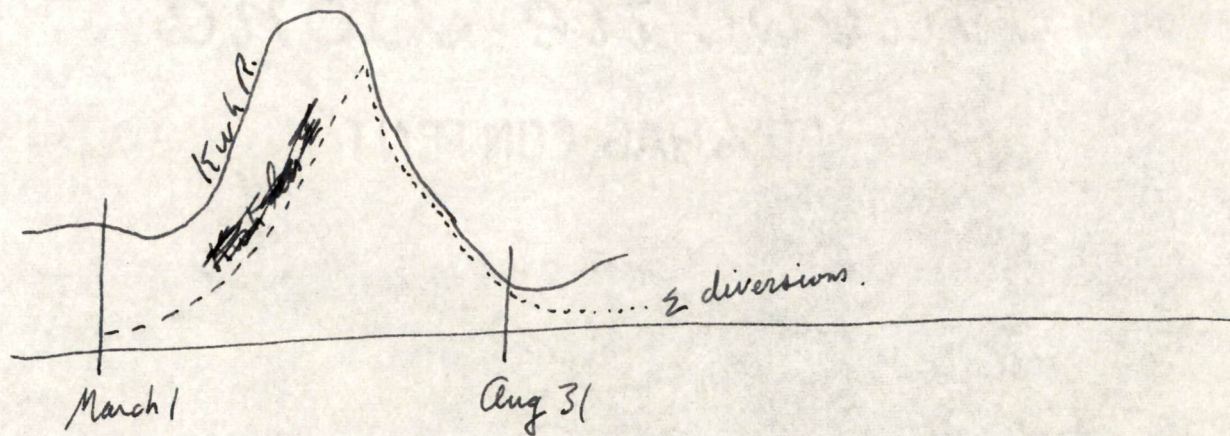
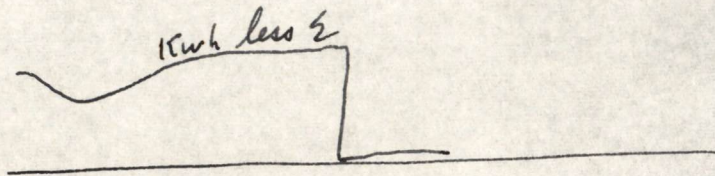
On daily flows, use same technique.

In test, point out stability is not always desirable, should adjust to Δ demand. But this system not sensitive to Δ demand, its only virtue is certainty of supply.

Actually, stability is not so important in times of surplus flows but in summer.

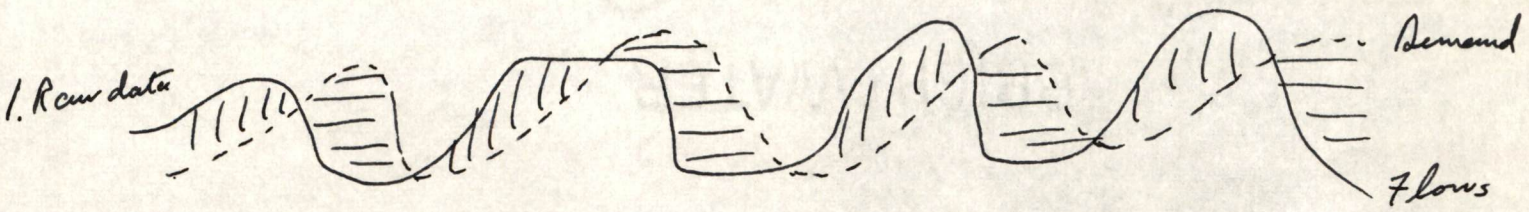
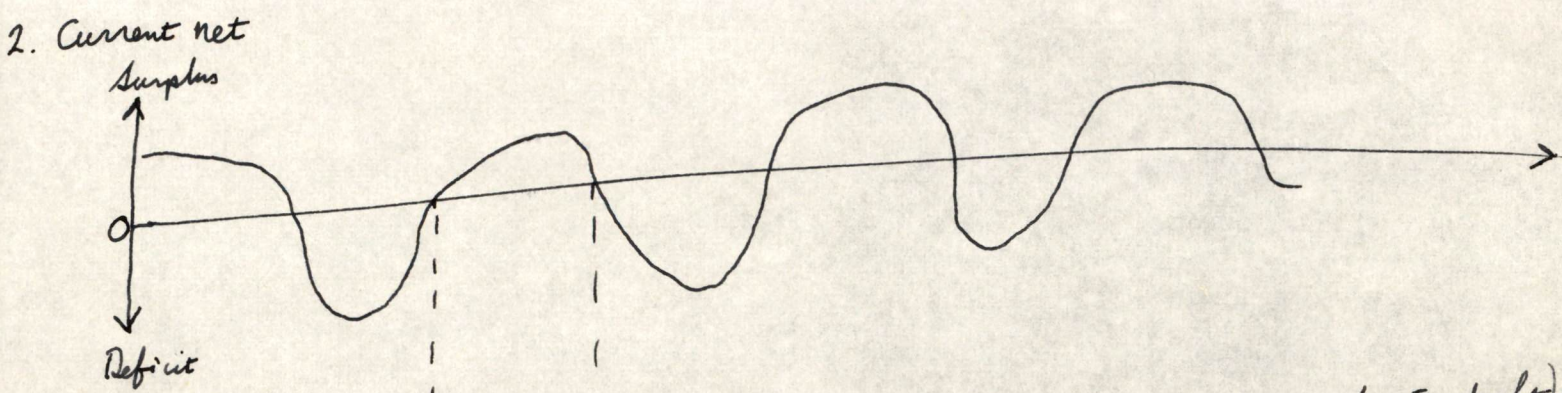
↓ This should obviate "stability of seasonal pattern" study

Monthly flows.

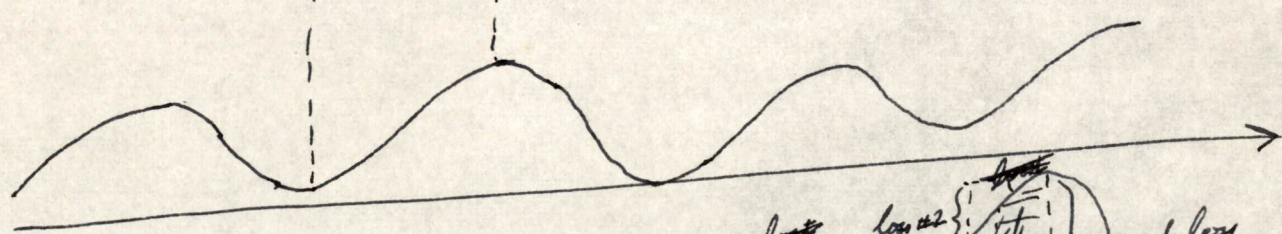


As with daily flows, no reason to expect Kwh less Σ to be very variable.

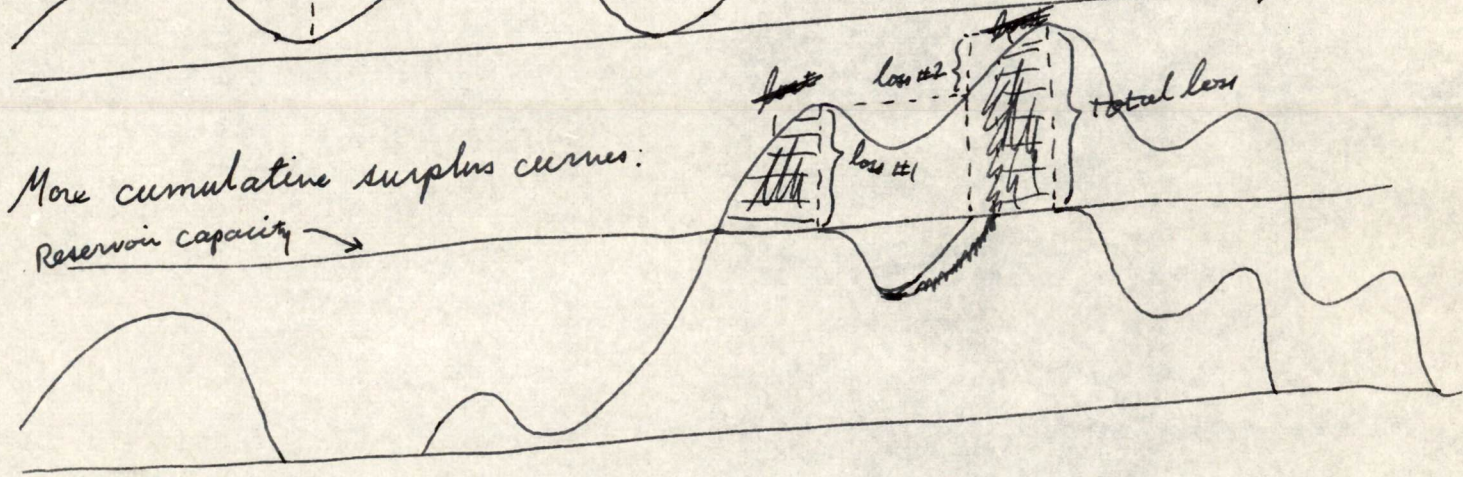
Only annually is demand steady over time, so there is where we find Σ diversions steady, & Kwh less Σ very variable.

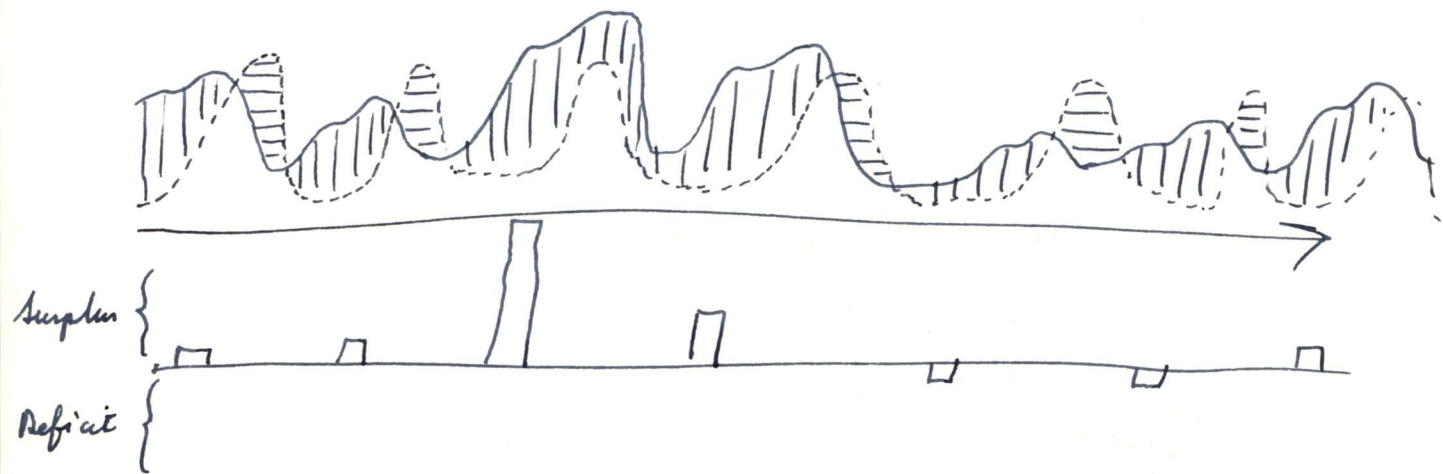


3. Cumulative surplus (there is no cumulative deficit) (unless we count ground water draft)

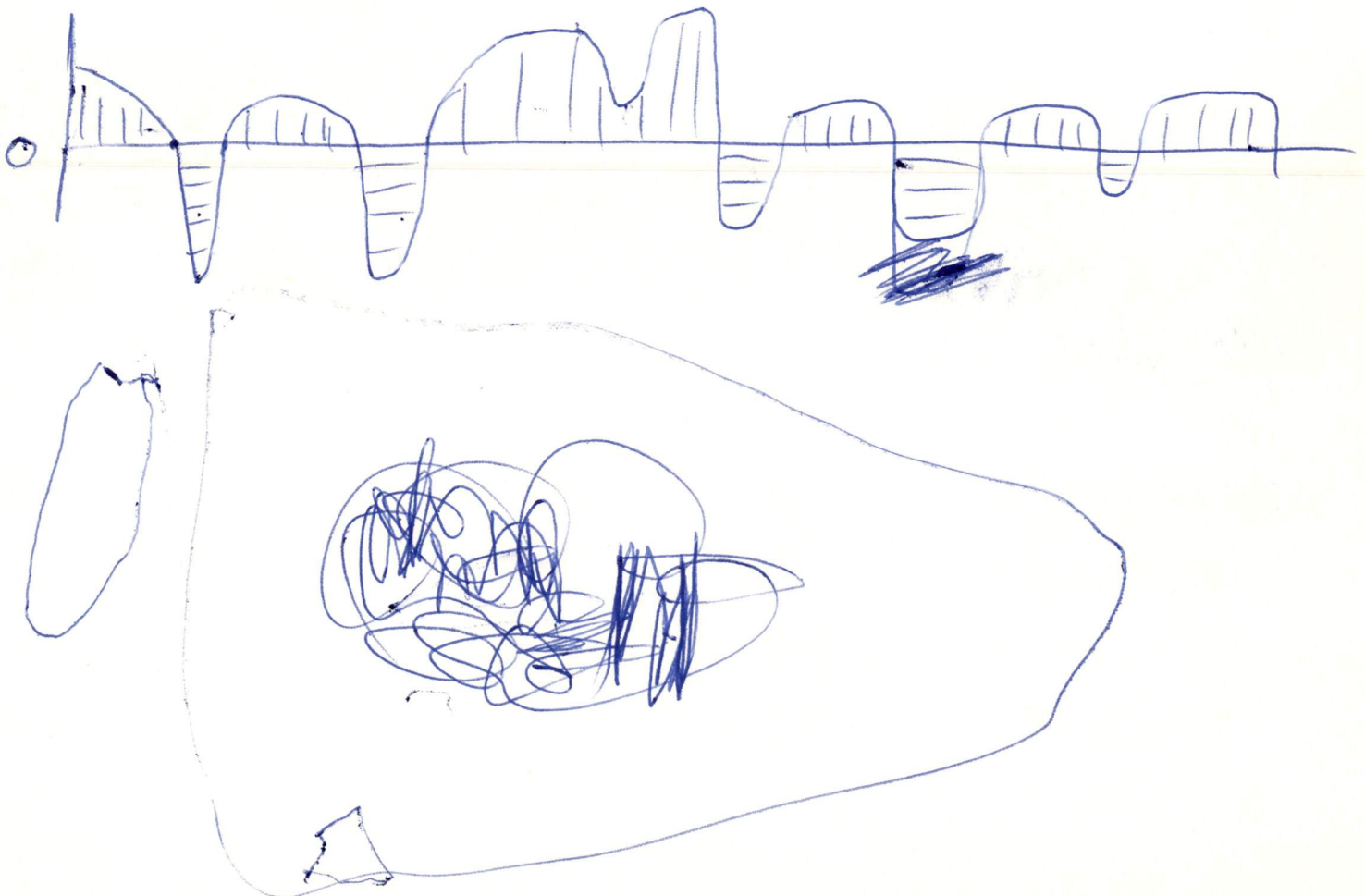


3 A. More cumulative surplus curves:
Reservoir capacity →





(If res is so small is always full, is no good at all. Now as demands increase, if ~~there are small res~~ if system storage is of that kind, its productivity increases)



A. Variability of Kaweah R. & diversions. (Absinse)

The following indices of variability are to be applied to ~~the following~~
~~time periods & units~~: daily, monthly, & yearly flows, as ff.:

- (3) Daily: ^{21 + days} March 1 - Sept. 30, for 1952 & 1955, of results of the 2 years;
 and
 for April, 1954, and July, 1954, separately. of results of the 2 mo.
 (Second-feet) (March 1 - Aug. 31) 1949-50 to 1954-55
- (2) Monthly: For each individual water year (~~Oct 1 - Sep. 30~~, 1949-55.
 Record & tabulate for easy reference.
 Then average the results. (not the original data)
 (Second-foot-days) (acre-feet)
- (1) Yearly: 1939-⁵⁵~~54~~. & Acre-feet

1. Range
 - a. Max rate of flow during period of analysis. with year, of occurrence ^{month, or day}
 - b. Min
 - c. Range (= Max - Min)
 - d. ~~Range~~ Mean
 - e. Range ÷ Mean
 - f. (e) ÷ Kaweah R. (e) --- not all Kwh R., just the sum of diversions,
 the steady part, the cream that has
 been skimmed.

2. Skew
 - a. Median
 - b. Skew (= $\frac{3(\bar{X} - Md)}{s}$) ~~but not~~ ~~the~~ ~~mean~~
 - c. Skew ÷ Kaweah skew (± diversions)

3. ~~Mean~~ Standard deviation
 - a. $\sum (\text{deviations squared})$
 - b. $\sqrt{\sum d^2 / n} = \text{std. deviation.}$
 - c. $\sqrt{\sum d^2 / n} \div \bar{X} = \text{form. coeff. of variation}$
 - d. $\frac{\text{coeff. var.}}{\text{form. d.}} \div \text{Kwh R., } (\pm \text{ diversions})$
- Monthly, for 1951-52
 and lowest year 1948-49

Variability of Kameah R & diversions, 2

4. LCR (if machine can do.) Time units & ^{flow} ~~flow~~; e.g. highest 10% of days have 72% of flow. May have advantages over $\frac{\sigma}{\bar{x}}$.

5. Length of season, 1950-55.

a. No. of days w/o flow, March 1 - Sep 30, each yr, 1950-55. ✓

b. Mean ✓

c. ~~Mean dev~~ $\frac{\sigma}{\bar{x}}$

d. ~~Mean % dev~~ $\frac{\sigma}{\bar{x}}$ coeff. of variation

b. Seasonal patterns.

a. $\bar{x}_{1953, 1954}$ [March 1 - Sep 30] Monthly data only.

b. Above, $\div$ Kwh R. ($\leq$ diversions)

? This is not a measure of variability

7. "Adequacy of aggregate supply" (for cfsn & basic reference purposes)

a. ata/acre, 1939-55. This calls for extra division of consuming units, because there are areas of overlap. See separate instruction sheet, "7 - AFA measurements".

b. LCR, acres & wtk

8. Index of clustering of deficits. - a big job.

~~over~~ a. Take $\frac{1}{2}$ + the mean of March 1 - Sep. 30 as base

b. Take ~~deficits below~~ 10-day moving average of net

deficits below + the half-mean, as 90% of $\frac{1}{2}$ mean.

~~ignore excess~~ Daily excesses above half-mean are relevant

when they ~~can~~ reduce net 10-day deficits, but not otherwise.

1 ~~can~~ Count no daily surplus above ~~daily~~ capacity.

~~to no can~~

c. Cube the 10-day moving averages, ~~from the cubes~~

d. Take daily average of cubes. This is index of clustering of

deficits.
e. $\frac{\sigma}{\bar{x}}$ Kwh R. ($\leq$ diversions)

~~A better base is one by which 2 diversions has no deficit. Then~~

B. Variability & capacity.

1. System capacities & entitlements

a. Max recorded rate of flow, 1949-55 (Take sum, record at bottom)

b. Date of (a).

c. Capacity. ^(Record sum) ~~from any misc. info. (Take sum, record)~~

d. Max possible entitlement, ~~too~~ river flow = ∞

e. Max. plausible

f. " " " river at max. At any time. ∞
per acre, any time (showing jrs can't absorb
floods, yet srs can't share)

2. Usage cfd to capacity

a. Mean rate of flow, ^{March 1 - July 31} ~~year-round~~, 19 ~~49-54~~ ⁵⁰⁻⁵⁵

b. (a) as % capacity.

c. (b) as % Kaweah

3. Potential benefits from storage

a. Savings lost w/ - cumulative excess of entitlements over capacity. Take a big year (1952), & contrast 2 cos - But I find none w/o a max. Who is entitled to floods? No one? Developer?

b. % share of Kaweah R. flow, highest yr (1952) & lowest.

Then within one year (1954,) highest vs lowest month.

Ignore other times, (3b only)

The following data must be prepared
for machine processing

~~extra sheets for~~

daily
data

April 1946 (some of the extra)
July 1946 (sheets, we have Watchman.)
April 1954 extra sheets
July 1954 " " " " *completed*

Persian Ditch (all periods)

~~Jennings Ditch (all periods)~~

Σ of diversions (all periods)

Kaweah R. - Σ of D. (all periods)

There remains a very slight inelegance, that they shd have used $n-1$, rather than n , to compute σ . But then, perhaps they did. Should they then have used $n-1$ to compute $\bar{x}_i$? Probly not, but ask someone.

1. But it will affect c.v. What wd. be right?

2. It also makes a false $\bar{X}_{zi}$, if that matters. Bes. all yrs of n.d. were effectively treated as 0.

1. Effect on c.v. wd. be negligible, for c.v. is indep. of ~~size of~~ ~~frequency~~.
no. of items, or frequency.

Except insofar as years of zero flow were excluded from n , (if that was done.) Check Elk Bayou, #12. The years of zero flow are entered as 0. — 4 of them, 37 entries in all, which checks with orig. data. Check Elk Bayou mean, see what was used as divisor.

Σ of 37 yrs = 157,667 (from Jake's sheets).

$$\text{Reported } \bar{X}_i = \frac{5,359}{1.258} = 4,259$$

$$\frac{157,667}{4,259} = 37.$$

Showing: a. When entry was zero, Ronnie correctly included it in n , to get $\bar{X}$

b. Ronnie xcribed Jake's figures correctly.

But there remains the problem that some years of n.d. were actually of zero flow. However this is not really a problem, ~~the~~ unless I failed to make the best estimate I could. It is worth mentioning somewhere in note that likely a large share of n.d. years were extreme ones, so data understate σ s of X_i .

~~Probably~~ some $\bar{x}_i$ are ^{of d. to $\bar{x}_{zi}$} relatively overstated, ~~bes~~ when were ~~the~~ ~~n.d.~~, they divided by a lower n

E.g. on Long's canal there are only 27 entries on computer sheet. Probly they set up program to use n as no. of entries. We submitted 27 years of data - but no instructions what do w. yrs. of n.d.

Taking mean of Σ diversions, of course, there wd. be no yrs of flow = 0, nor n.d.

check Long's Canal. Did computer use $n = 27$ or $n = 37$?

Σ of 27 entries for Long's = 119,906

$$\frac{119,906}{\textcircled{27}} = 4,441 \leftarrow \text{This is the one computer used. - i.e. } n = 27$$

$$\frac{119,906}{37} = 3,241$$

This is then a reason, and I hope the full reason why

$$\Sigma \bar{x}_i > \bar{x}_{\Sigma i}$$

What I must do?

This won't much affect σ_i - unless years of n.d. were exceptionally far from a near mean. ~~the~~

Prime effect $\therefore$ is simply to spoil a good check. But this I could easily remedy. Divide each $\bar{x}_i$ by $\frac{37}{n_i}$, then add up these adjusted means.

C.V. vs. L.C.R.

~~1.258~~

~~- 0.143~~

~~1.115~~

- LCR vs C.V., annual Kwh figures.
cv runs a little less than double
LCR, except Kwh less Σ
also, Elk Bayou runs exactly $\frac{1}{2}$.
These are the highly skewed ones.

Suggesting c.v. is more sensitive
to skew than is LCR

Which varies more, LCR or C.V.?

1. LCR

Range is from .080, (LS10) to .626, ~~(Parkway)~~ (Elk Bayou)

$$\text{LCR range} = .546$$

2. CV

Range from 0.143 (LS10) to 1.258 (Elk Bayou)

$$\text{CV range} = 1.115$$

Not much choice between them, on a C.V. or % basis.

Nor do they seem to affect rank much.

• Relation of coeff. of variation to LCR:

On Kameah R. distr. annual diversions,
 $CV \approx 2 \times LCR$ for every one except
one: Tulare I.D.

Check that one, which has
hi CV but low LCR

This suggests T.I.D. is a flattop?
Q mode is to right of mean

7 reg. distribns, annual diversions, Evans & Watson Ditches, & Tulare I.D.

Evans

Watson

Tulare

000 at

000

2-3 11

2-3 11

0-5 111111

3-4

3 1

5-10 11

4 111

4 11

10-15 11

5 1

5 11

15-20 11111

6 1111

6 111111

20-25 111

7 1111

7 $\bar{x}$ 1111

25-30 1

8 $\bar{x}$ 11111111

8 11111111

30-35 111

9 11111

9 1111

35-40

10 111

10 111

40-45 1

11 1

11 111

45-50 1

12 1111

12 11

50-55

13 1

13

55-60 11

14

14

60-65 11

15

15

65-70

16 1

16

70-75 1

75-80 11

80-85

85-90

90-95

95-100 1

100-105 111

$\bar{x}$

Tulane: Lower $\frac{22}{35}$ yrs have $\frac{15+15+25+81\frac{1}{2}+66+28+99=335}{\text{above} + 42+47+115+125+72+155+98+309}$ of the flows.

" .63 " " $\frac{335}{1298} = .258$ of the flows.

Evans: Lower $\frac{18}{36}$ yrs have $\frac{5+13+6+26+30+\cancel{34}^{33}}{\text{above} + 27+47+32+11+51+13+17}$ of the flows

" .50 " " $\frac{113}{311} = .363$

Watson: Lower $\frac{17}{36}$ yrs have $\frac{5+3+10+11+39+30}{\text{above} + 60+38+31+35+25}$ of the flows.

" .412 " " $\frac{\cancel{289}^{98}}{287} = .341$

No ambiguity. Tulane is much more conc., should have higher LCR. Must be error.

Why is variability of flow troublesome? That is what we want to put finger on & more.

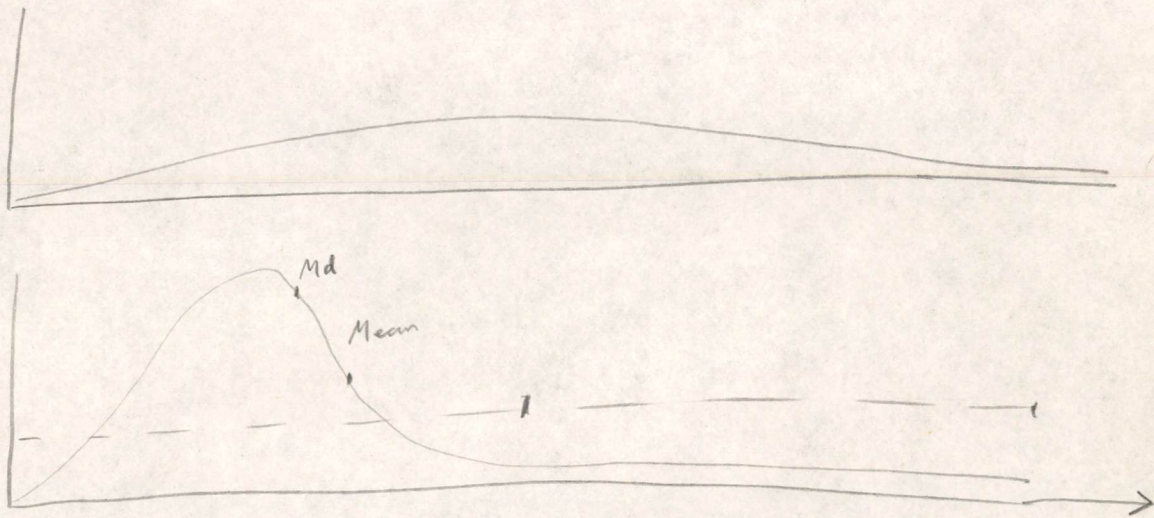
a. Unpredictability

b. A need for capacity

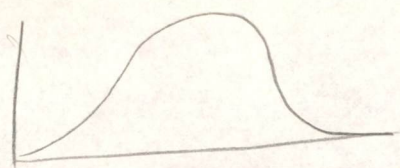
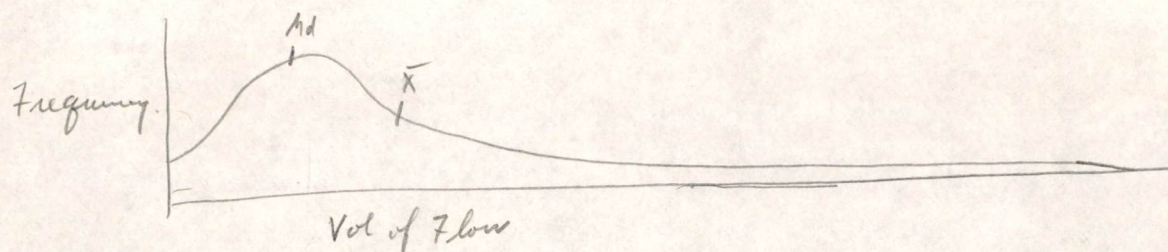
c. Unreliability

d. Hi % supply comes in ~~at~~ small % time. Skew tells us something about this, that σ doesn't. Maybe LCR is what we need?

As to (a), ^{2(c)} recipient of steady flow might be worse off than if he shared in fluctuations of river, if he comes close to being shut off completely - to where a 1% drop of river is 100% cutoff for him.



The skewed distribution has a lower σ than the even dist, bcs most values are close to mean. But it has higher concn. Skewness to left means that - to it means less concn.



$$\frac{25}{5} = 5$$

$$\frac{60}{6} = 10$$

4

7

6

9

2

11

$$39 \div 6 = 6.5$$

$$7, 9, 11 = 27$$

$$4, 6, 2 = 12$$

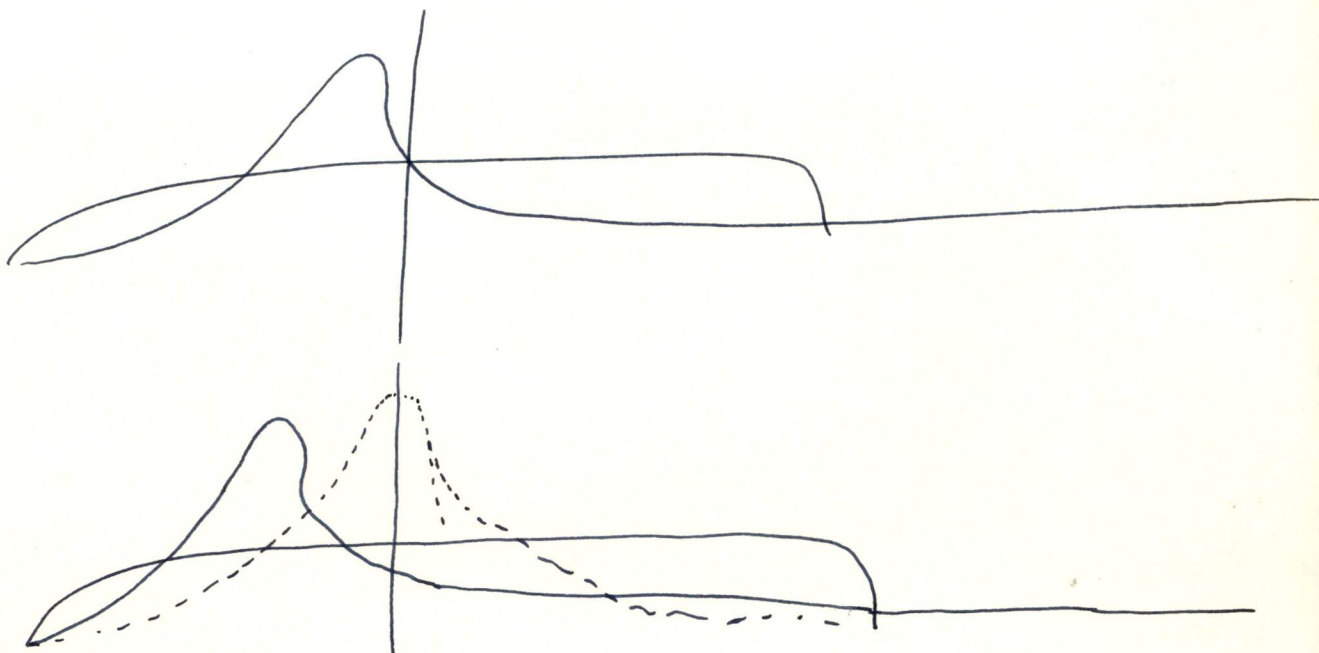
The greater the skew, the higher % of flows^{volume} comes in a few years.

Skew then is an easier-to-compute index to LCR?

How differ from deviation?



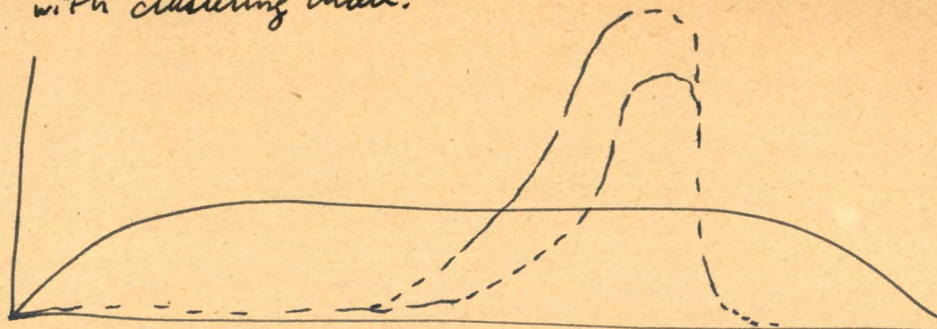
Both skew & dispersion contribute to ln LCR. Dispersion makes lowest 10% have small share; skew makes lowest 50% have small share.



Skew is worse than flat top, ^{of same mean & σ ,} because median moves left, too, & so no. of
 days you can depend on being mean is less.
 But σ gives no distinction

LCR does. — $\therefore$ better.

Lorenz curves & freq. distributions. Summarizing, try to integrate with clustering index.



With given LCR , $\propto \frac{\sigma}{\bar{x}}$, wd you rather have concn due to ^{positive} skew or dispersion? That is, a central tendency at low end of wide range, vs no central tendency & shorter range? It is between these 2 that LCR & $\frac{\sigma}{\bar{x}}$ do not distinguish.

along Kameah R., early days, wish to min. distribn costs, skew better - & let floods escape. As wth scarcer, even dispersion better, expand works to higher capac. capture all wth. Today, even dispersion wd be preferable.

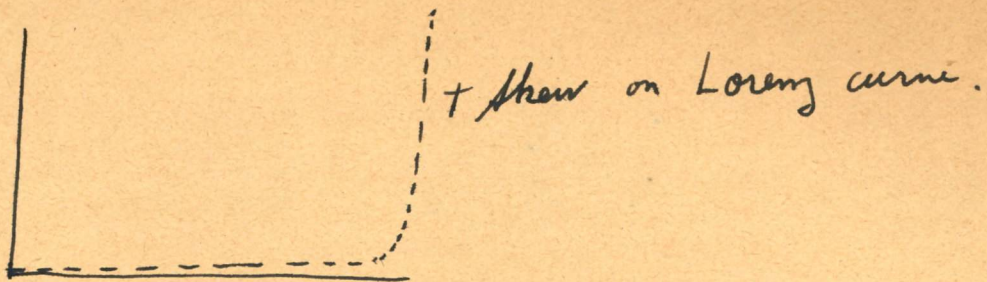
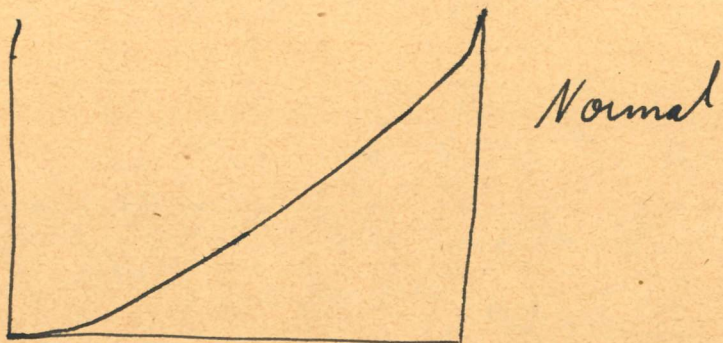
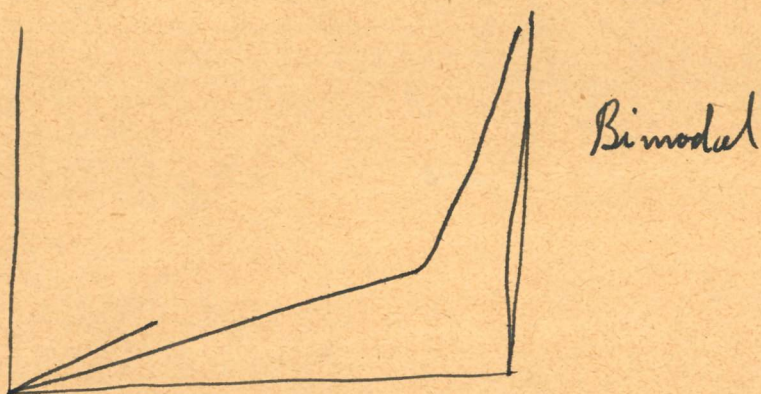
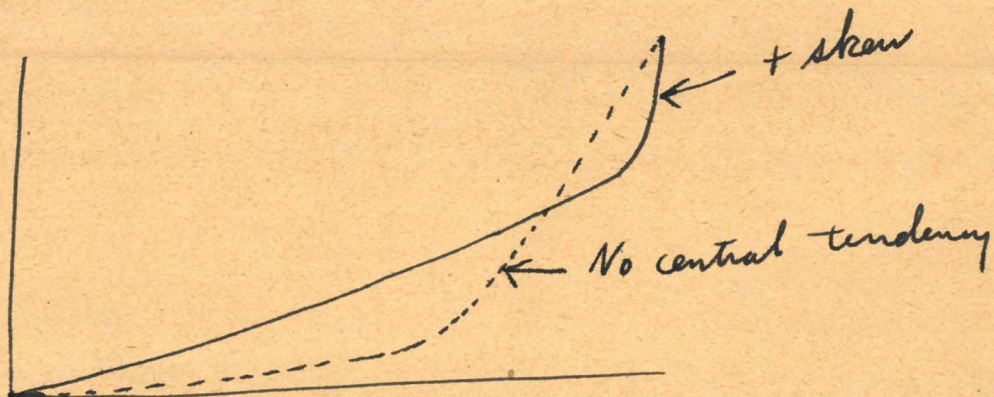
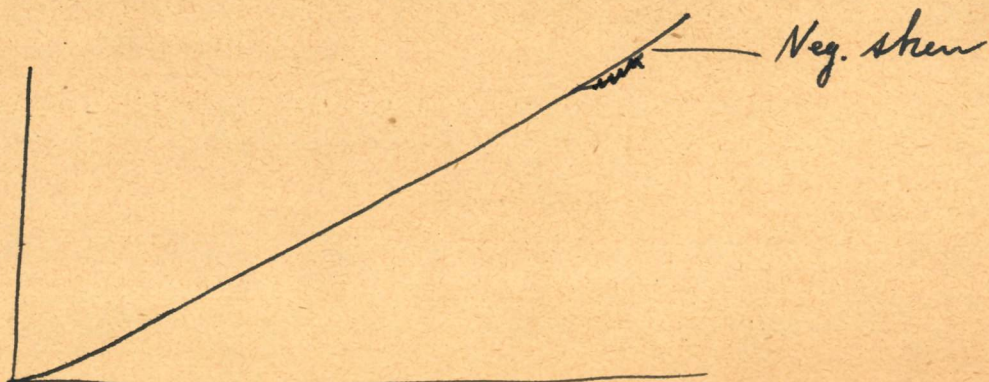
So there is a preference here. Now, can it be detected from LCR , or $\frac{\sigma}{\bar{x}}$? If not, must determine skew.

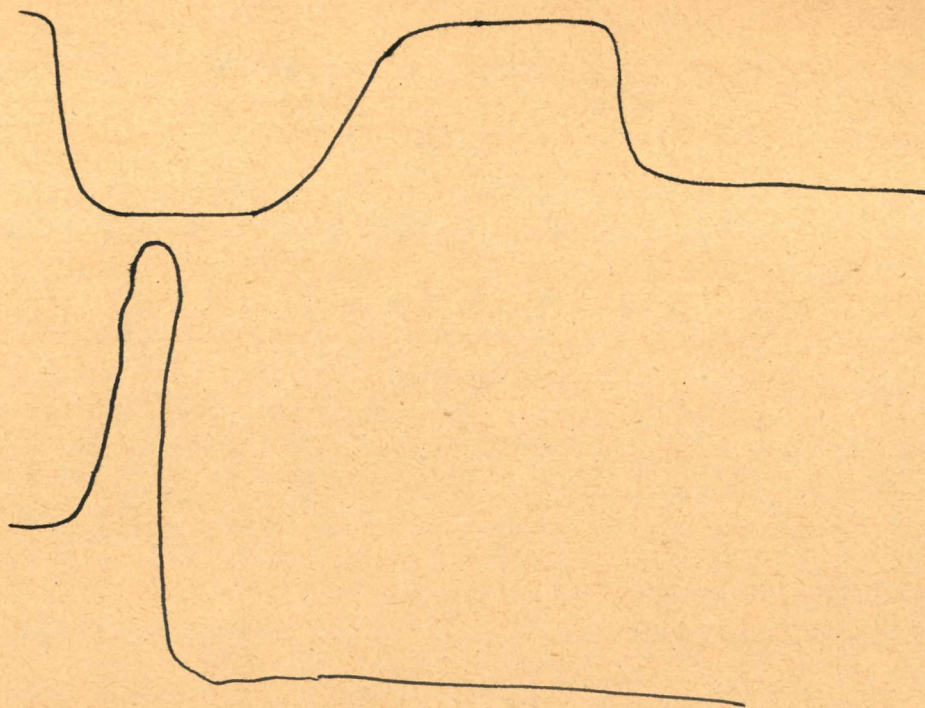
$\frac{\sigma}{\bar{x}}$ is sensitive to skew. + skew means lower mean, higher $\frac{\sigma}{\bar{x}}$, cfd to less skewed or - skewed distrs. of same σ

→ Given LCR data by deciles, can we distinguish 2 causes of hi LCR ?

If it's skew, then:

- Range hi of $\bar{x}$.
- Skewed curve is more conc. at top
Dispersed bottom, unless
skew goes back to 0, as my cases do, in which case if LCR is same, skewed, being more conc. at top & btm, must be less so in middle.



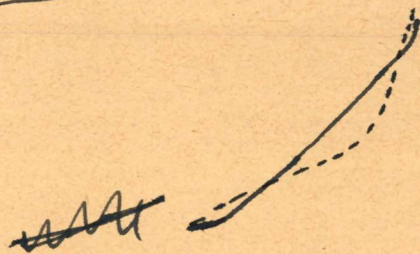


~~Q~~ To get straight Lorentz curve, need hi freq. on freq. curve.

To make Lorentz curve hook, need low freq.

So to get this shape
areas to make hooks.

calls for 2 low freq.

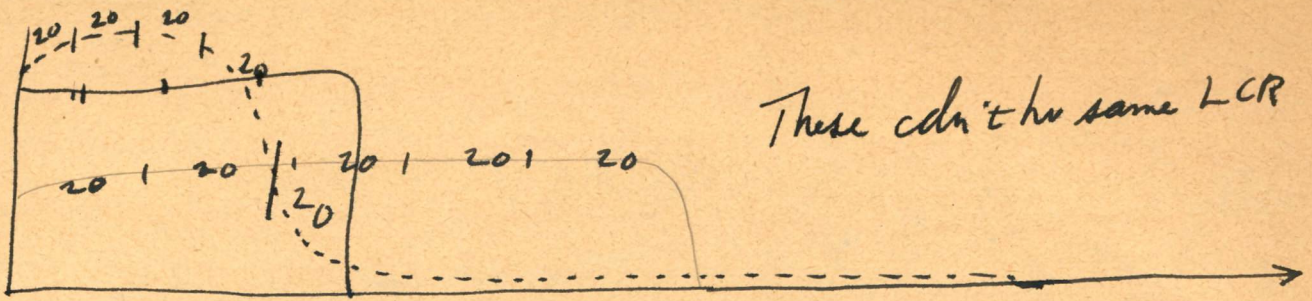


Needn't be

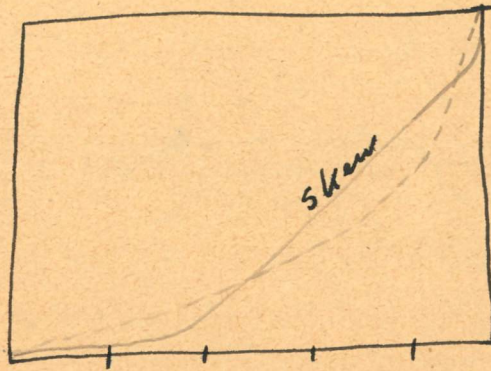
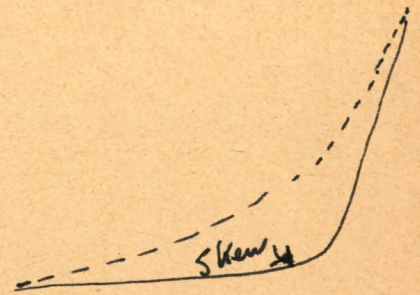
straight lines out beyond hooks, but rapidly curving, due to low freq:
a normal curve  with strong central tendency but long wings.

Nodes are where Lorentz is straight; valleys where it hooks. So 2 straight sections separated by a hook is bimodal.

Same n & Σf LCR



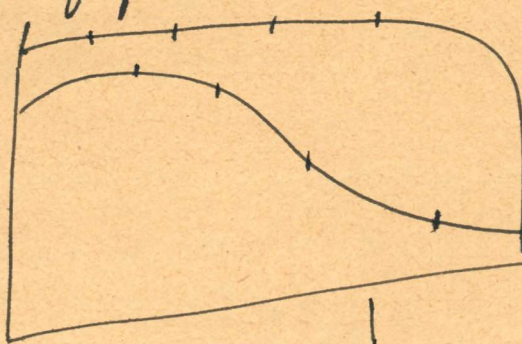
These can't have same LCR



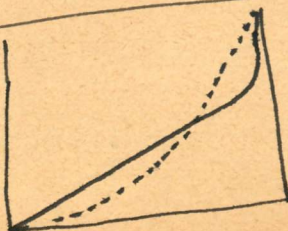
Skewed curve will have sharp hook, upper right hand corner of its LCR.
Bes 100th percentile & thereabouts add so much more than 91st. rate
of addition to Σ changes so rapidly.

Where the percentiles are wider apart. slope of ~~LCR~~ Lorenz curve
changes quicker

Impossible for skew LCR to be as low as unskewed unless unskewed
begins higher freq. - which means low range, too.



Skewed is hooked. But they
cross only once.

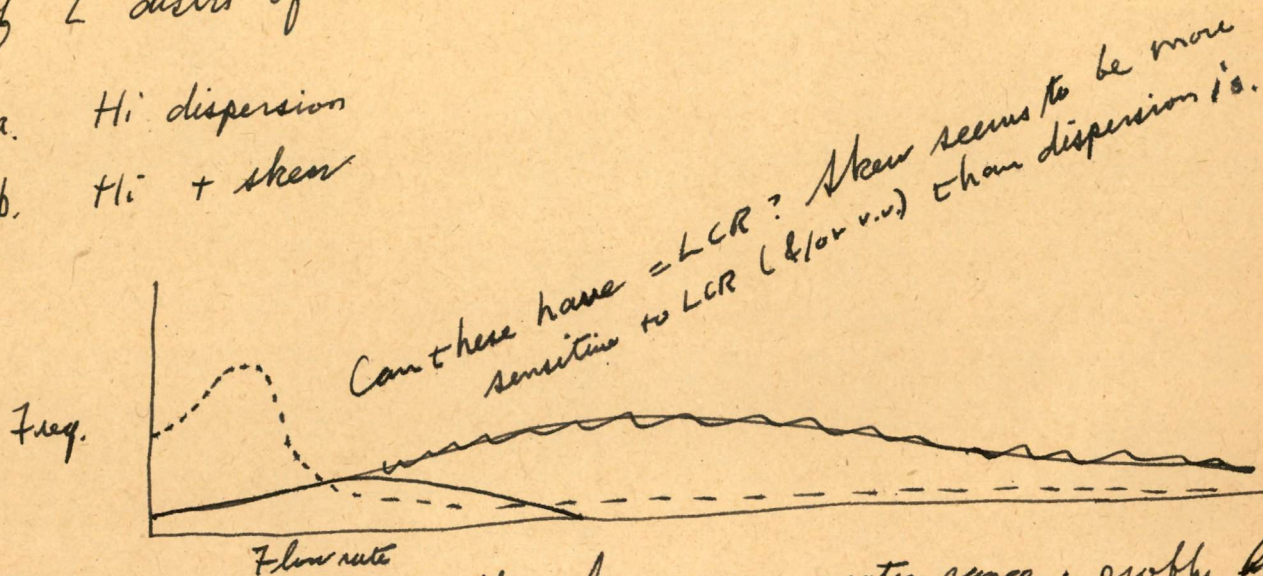


So is easy to identify skew on Lorenz curve. Could even develop a skew index
from it

Only lack of LCR is, doesn't tell you if conc. due to hi dispersion, or skew. Does it matter?

cf 2 dists of = LCR:

- a. Hi dispersion
- b. Hi + skew



The skew is more predictable; ~~lower mean~~ greater range; probly lower σ . Skew is a worse problem than lack of central tendency, but greater range. Will lose more of wth $\therefore$ Range expresses this in part. Other hand, capture higher 90% of wth with given works capacity.

In stat. practice, skew is the problem, skew is hard to measure accurately, $\frac{\sigma}{\bar{x}}$ doesn't do it adequately, so need LCR.

Likewise for all CV rivers, better analyze again with LCR (altho hi correlation suggests this not problem)

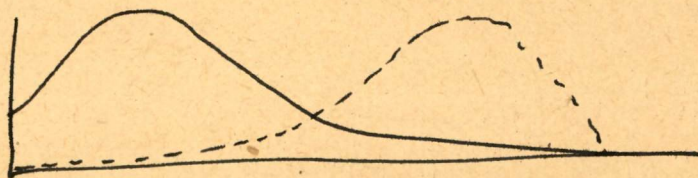
LCR also has a definite meaning you can tabulate: top 10% have

50%, etc.

Is - skew preferable to 0 skew?

cf. LCR, + & - skewed distributions.

2 distrs hv same σ , but + skew has lower $\bar{x}$, hence higher $\frac{\sigma}{\bar{x}}$ if symmetrical

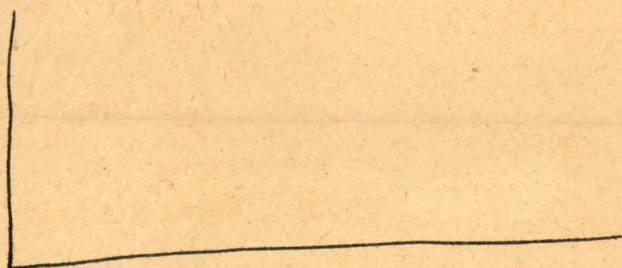


LCR

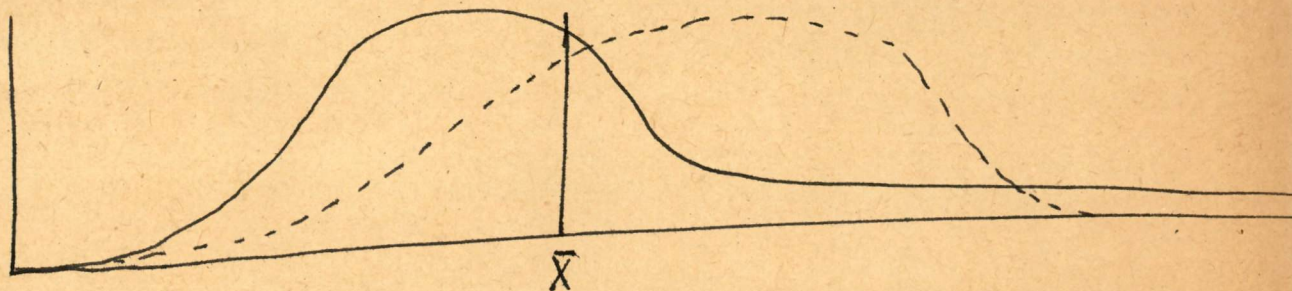
1. + skew



2. - skew



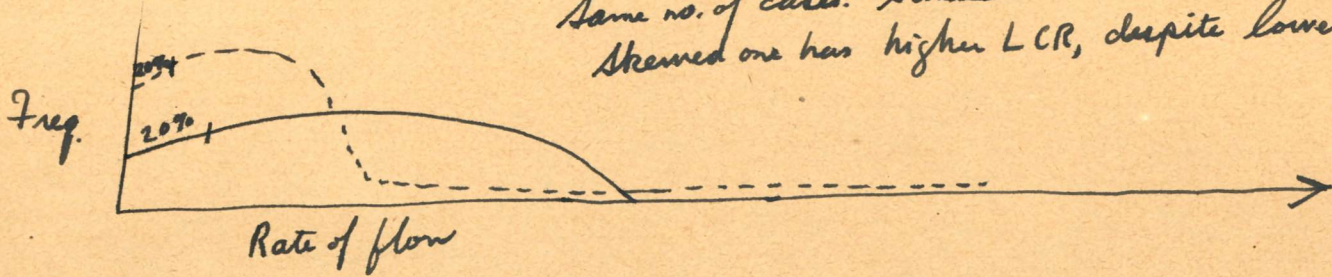
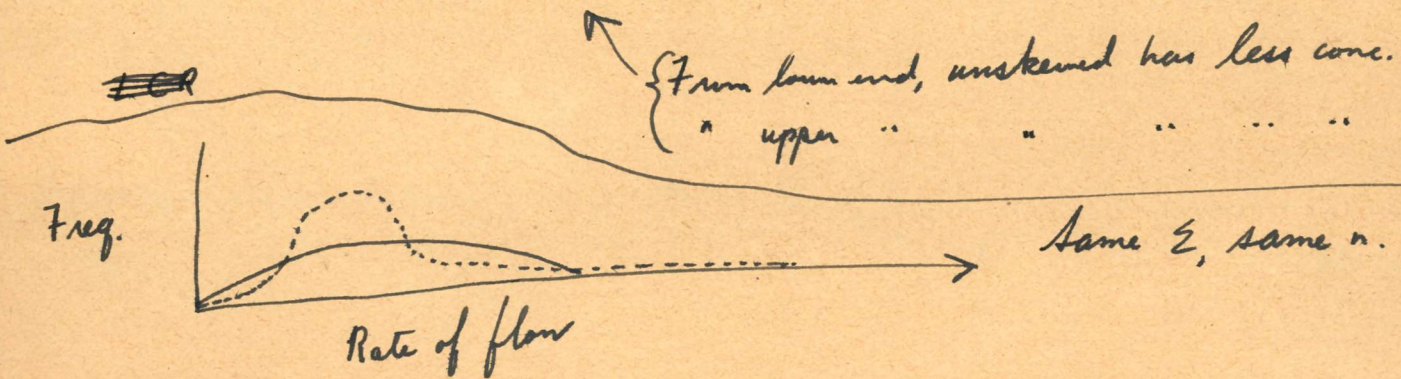
cf. + & - skew, same mean & same σ (& hence same $\frac{\sigma}{\bar{x}}$)



The - skew is more desirable, except for very small developments
The + skew has a higher LCR, & higher range
So skew definitely affects LCR



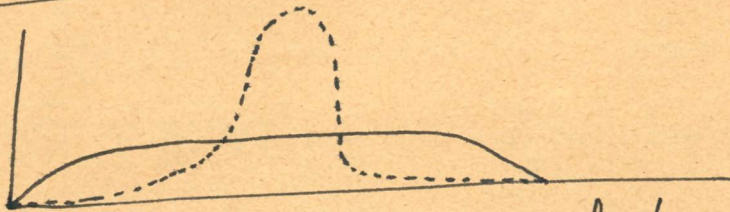
Same no. of cases. ~~Same LCR?~~ Same Σ .
 Skewed one has higher LCR, despite lower σ (probably)



From lower end, unskewed has more conc.
 " upper " " " less conc., but not for long.



As drawn, probably skewed has less concn on balance, but is not very skewed & has stronger by far mode.



Both ends, ---- has less concn. And the more skewed to right, the more so.



From left, unskewed much more concn.
 From rt