Propagation of Errors (Section 3.6)

- Usually when doing experimental calculations you must combine several measurements to get the final result.
- Each individual measurement has some error associated with it.
- How do these combine?

Propagation of Errors (addition)

Let x_i and y_i be two random variables or measurements with mean values of $\bar{x}$ and $\bar{y}$ and associated variances s_x^2 and s_y^2, respectively, where

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad \text{and} \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$1 \leq i \leq n$. Also let C_1 and C_2 be two known constants.

Addition/subtraction of random variables

Consider the general linear combination $z_i = C_1 x_i + C_2 y_i$. We would like to derive the equations that relate $\bar{z}$ and s_z^2 to our statistical estimates for x and y.

We begin by determining the mean value of z:

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i.$$

Propagation of Errors (addition)

Substituting the functional dependence of z on x and y:

$$\bar{z} = \frac{C_1}{n}\sum_{i=1}^{n} x_i + \frac{C_2}{n}\sum_{i=1}^{n} y_i,$$

$$\bar{z} = C_1\bar{x} + C_2\bar{y}.$$

Next, we determine the variance associated with z:

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i^2 - \bar{z}^2)$$

(from Equation (3.30)). Squaring z_i,

$$z_i^2 = C_1^2 x_i^2 + C_2^2 y_i^2 + 2C_1C_2x_iy_i.$$

Squaring Equation (3.36),

$$\bar{z}^2 = C_1^2\bar{x}^2 + C_2^2\bar{y}^2 + 2C_1C_2\bar{x}\bar{y}.$$

Propagation of Errors (addition)

Substituting the expressions for z_i^2 and $\bar{z}^2$ into that for s_z^2, we obtain

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(C_1^2(x_i^2 - \bar{x}^2) + C_2^2(y_i^2 - \bar{y}^2) + 2C_1C_2(x_iy_i - \bar{x}\bar{y})\right)$$

$$= C_1^2 s_x^2 + C_2^2 s_y^2 + \frac{2C_1C_2}{n-1}\sum_{i=1}^{n}(x_iy_i - \bar{x}\bar{y}).$$

The last term in the above equation,

$$s_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_iy_i - \bar{x}\bar{y}),$$

… is called what?

Propagation of Errors (addition)

Substituting the expressions for z_i^2 and $\bar{z}^2$ into that for s_z^2, we obtain

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(C_1^2(x_i^2 - \bar{x}^2) + C_2^2(y_i^2 - \bar{y}^2) + 2C_1C_2(x_iy_i - \bar{x}\bar{y})\right)$$

$$= C_1^2 s_x^2 + C_2^2 s_y^2 + \frac{2C_1C_2}{n-1}\sum_{i=1}^{n}(x_iy_i - \bar{x}\bar{y}).$$

The last term in the above equation,

$$s_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\sum_{i=1}^{n}(x_iy_i - \bar{x}\bar{y}),$$

… is called what?

the <u>covariance</u> of x and y.

Propagation of Errors (add/subtract)

- If x and y are <u>independent</u>, then $E(s_{xy}^2) = 0$

When adding two random variables x and y such that $z_i = C_1 x_i + C_2 y_i$, the *associated variance* is calculated as

$$s_z^2 = C_1^2 s_x^2 + C_2^2 s_y^2 + 2C_1C_2 s_{xy}^2. \tag{3.37}$$

- If we <u>subtract</u> two variables we get a similar result:

$$\bar{z} = C_1\bar{x} - C_2\bar{y}$$

and

$$s_z^2 = C_1^2 s_x^2 + C_2^2 s_y^2 - 2C_1C_2 s_{xy}^2.$$

Propagation of Errors (subtraction)

$$\bar{z} = C_1 \bar{x} - C_2 \bar{y}$$

and

$$s_z^2 = C_1^2 s_x^2 + C_2^2 s_y^2 - 2 C_1 C_2 s_{xy}^2.$$

- A positive covariance (can be +/-) <u>reduces</u> the error in the difference between variables.

Propagation of Errors (division)

- Okay, how about multiplication and division?
- Let $z_i = x_i / y_i$. where x_i and y_i are normally distributed.
- Note that z is <u>not</u> normally distributed!
- It only approaches this <u>if</u> $S_x/\bar{x}, S_y/\bar{y} \ll 1.$
- Let's look at this more closely.
- Let $x_i = \bar{x} + x'_i, \;\; y_i = \bar{y} + y'_i,$

  (prime = deviation)

relative error

Propagation of Errors (division)

- Suppose $S_x/\bar{x}, S_y/\bar{y} \ll 1.$
- This means that x'_i, y'_i are small.

Expanding $(1 + y'_i/\bar{y})^{-1}$ using the binomial expansion,

$$z_i = \frac{\bar{x}}{\bar{y}}\left(1 + \frac{x'_i}{\bar{x}}\right)\left(1 - \frac{y'_i}{\bar{y}} + \left(\frac{y'_i}{\bar{y}}\right)^2 - \left(\frac{y'_i}{\bar{y}}\right)^3 + \cdots\right)$$

$$= \frac{\bar{x}}{\bar{y}}\left(1 + \frac{x'_i}{\bar{x}} - \frac{y'_i}{\bar{y}} + O\left(\frac{x'_i y'_i}{\bar{x}\bar{y}}, \left(\frac{y'_i}{\bar{y}}\right)^2\right)\right),$$

and ignoring the second- and higher-order terms in the expansion, we get

$$z_i \approx \frac{\bar{x}}{\bar{y}}\left(1 + \frac{x'_i}{\bar{x}} - \frac{y'_i}{\bar{y}}\right).$$

Propagation of Errors (division)

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_i \approx \frac{1}{n}\sum_{i=1}^{n} \frac{\bar{x}}{\bar{y}}\left(1 + \frac{x_i'}{\bar{x}} - \frac{y_i'}{\bar{y}}\right).$$

The last two terms in the above equation sum to 0, therefore

$$\bar{z} \approx \frac{\bar{x}}{\bar{y}}.$$

Now, calculating the associated variance, we obtain

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i^2 - \bar{z}^2) = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\frac{\bar{x}}{\bar{y}}\left(1 + \frac{x_i'}{\bar{x}} - \frac{y_i'}{\bar{y}}\right)\right)^2 - \left(\frac{\bar{x}}{\bar{y}}\right)^2\right]$$

$$= \left(\frac{\bar{x}}{\bar{y}}\right)^2 \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\frac{x_i'}{\bar{x}}\right)^2 + \left(\frac{y_i'}{\bar{y}}\right)^2 + 2\frac{x_i'}{\bar{x}} - 2\frac{y_i'}{\bar{y}} - 2\frac{x_i'y_i'}{\bar{x}\bar{y}}\right],$$

Propagation of Errors (division)

$$s_z^2 = \left(\frac{\bar{x}}{\bar{y}}\right)^2 \left[\frac{1}{\bar{x}^2} \cdot \frac{1}{n-1}\sum_{i=1}^{n}(x_i')^2 + \frac{1}{\bar{y}^2} \cdot \frac{1}{n-1}\sum_{i=1}^{n}(y_i')^2 - \frac{2}{\bar{x}\bar{y}} \cdot \frac{1}{n-1}\sum_{i=1}^{n}x_i'y_i'\right].$$

Now,

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i')^2, \qquad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i')^2, \qquad \text{and} \qquad s_{xy}^2 = \frac{1}{n-1}\sum_{i=1}^{n}x_i'y_i'.$$

When dividing two random variables x and y such that $z = x/y$, the variance associated with z is given by

$$s_z^2 = \left(\frac{\bar{x}}{\bar{y}}\right)^2 \left[\frac{s_x^2}{\bar{x}^2} + \frac{s_y^2}{\bar{y}^2} - 2\frac{s_{xy}^2}{\bar{x}\bar{y}}\right]. \tag{3.43}$$

So for division you add the <u>fractional</u> or <u>relative</u> variance!

Propagation of Errors (multiplication)

- For multiplication you get the same:

If $z_i = x_i y_i$ and $S_x/\bar{x}, S_y/\bar{y} \ll 1$, then

$$\bar{z} \approx \bar{x}\bar{y}$$

and

$$\frac{s_z^2}{\bar{z}^2} = \left[\frac{s_x^2}{\bar{x}^2} + \frac{s_y^2}{\bar{y}^2} + 2\frac{s_{xy}^2}{\bar{x}\bar{y}}\right].$$

Note the (+) sign

Propagation of Errors

- What about a more general functional relationship?
- Suppose $z = f(x, y)$, where f is some messy function.
- Let $x_i = \bar{x} + x_i', \; y_i = \bar{y} + y_i',$
- We shall <u>expand</u> f in a 2-D Taylor series: **(can do n-dimensions)**

$$f(x_i, y_i) = f(\bar{x}, \bar{y}) + x_i' \left.\frac{df}{dx}\right|_{\bar{x},\bar{y}} + y_i' \left.\frac{df}{dy}\right|_{\bar{x},\bar{y}} + O\left(x_i'^2, y_i'^2, x_i' y_i'\right).$$

(Ignore 2nd order terms!)

Propagation of Errors

- The deviations sum to zero:

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} f(x_i, y_i) \approx \frac{1}{n}\sum_{i=1}^{n} f(\bar{x}, \bar{y}) + \left.\frac{df}{dx}\right|_{\bar{x},\bar{y}} \frac{1}{n}\sum_{i=1}^{n} x_i' + \left.\frac{df}{dy}\right|_{\bar{x},\bar{y}} \sum_{i=1}^{n} y_i' \approx f(\bar{x}, \bar{y}).$$

Propagation of Errors

The variance associated with $f(x, y)$ is given by

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i^2 - \bar{z}^2)$$

$$\approx \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(f(\bar{x},\bar{y}) + x_i' \left.\frac{df}{dx}\right|_{\bar{x},\bar{y}} + y_i' \left.\frac{df}{dy}\right|_{\bar{x},\bar{y}}\right)^2 - (f(\bar{x},\bar{y}))^2\right]$$

$$\approx \frac{1}{n-1}\sum_{i=1}^{n}\left[x_i'^2\left(\left.\frac{df}{dx}\right|_{\bar{x},\bar{y}}\right)^2 + y_i'^2\left(\left.\frac{df}{dy}\right|_{\bar{x},\bar{y}}\right)^2 + 2x_i'y_i'\left(\left.\frac{df}{dx}\right|_{\bar{x},\bar{y}}\right)\left(\left.\frac{df}{dy}\right|_{\bar{x},\bar{y}}\right)\right.$$

$$\left. + 2f(\bar{x},\bar{y})\left(x_i' \left.\frac{df}{dx}\right|_{\bar{x},\bar{y}} + y_i' \left.\frac{df}{dy}\right|_{\bar{x},\bar{y}}\right)\right].$$

The last term sums to zero and we are left with…

Propagation of Errors

$$s_z^2 \approx s_x^2 \left(\left. \frac{df}{dx} \right|_{\bar{x},\bar{y}} \right)^2 + s_y^2 \left(\left. \frac{df}{dy} \right|_{\bar{x},\bar{y}} \right)^2 + 2 s_{xy}^2 \left(\left. \frac{df}{dx} \right|_{\bar{x},\bar{y}} \right) \left(\left. \frac{df}{dy} \right|_{\bar{x},\bar{y}} \right).$$

- Remember this, it is very powerful when you need to estimate the error from combining multiple random variables!
- Reduces to the earlier expressions for addition, subtraction, multiplication and division. (Exact for +/- since 2^{nd} deriv.=0)

Q1: If x is a Gaussian distributed variable, then is the variable $y = 2/x$ Gaussian too?

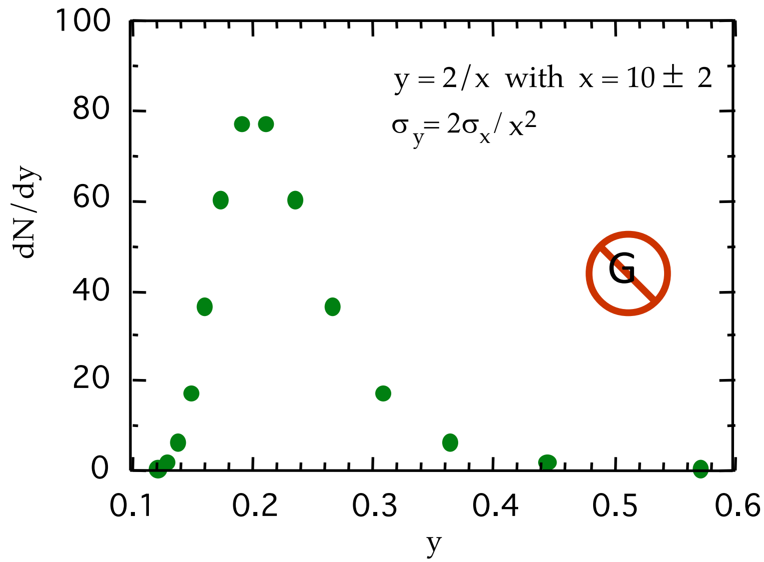
A. Yes

B. No

Q1: If x is a Gaussian distributed variable, then is the variable y = 2/x Gaussian too?

A. Yes

B. NO



Start with a gaussian with μ=10, σ=2.
DO NOT get another gaussian !
Get a *pdf* with μ = 0.2, σ = 0.04.
This new *pdf* has longer tails than a gaussian *pdf*.
Prob($y>\mu_y+5\sigma_y$) =5×10^{-3}, for gaussian $\approx 3\times10^{-7}$

Q2: Go with your gut… does systolic blood pressure share covariance with body weight?

	A	B
1	Systolic (y)	Weight (x)
2	145	210
3	155	245
4	160	260
5	155	230

A. **Positive Covariance**
B. **Negative Covariance**
C. **Zero Covariance**

Answer: Positive Covariance!

	A	B	C	D	E
1	Systolic (y)	Weight (x)		*Systolic (y)*	*Weight (x)*
2	145	210	Systolic (y)	140.25	
3	155	245	Weight (x)	266.9	784.84
4	160	260			

$$\sigma_{xy}^2$$

$$\sigma_x \approx 28$$

$$\sigma_y \approx 12$$

Q3: So, from Q1 we know that a function of a Gaussian variable can have a non-Gaussian distribution.
On the other hand, can one or more NON-Gaussian variables combine into a function which has a Gaussian distribution?

A. Yes
B. No

blood 1992 79: 1351-1358

Rheologic properties of senescent erythrocytes: loss of surface area and volume with red blood cell age

RE Waugh, M Narla, CW Jackson, TJ Mueller, T Suzuki and GL Dale

Rheologic Properties of Senescent Erythrocytes: Loss of Surface Area and Volume With Red Blood Cell Age

By Richard E. Waugh, Mohandas Narla, Carl W. Jackson, Thomas J. Mueller, Takashige Suzuki, and George L. Dale

The rheologic properties of senescent erythrocytes have been examined using two models of red blood cell (RBC) aging. In the rabbit, aged erythrocytes were isolated after biotinylation, in vivo aging, and subsequent recovery on an avidin support. Aged RBCs from the mouse were obtained using the Ganzoni hypertransfusion model that suppresses erythropoiesis for prolonged periods of time allowing preexisting cells to age in vivo. In both cases, the aged erythrocytes were found by ektacytometry to have decreased deformability due to diminished surface area and cellular dehydration. The aged rabbit erythrocytes were further characterized by micropipette methods that documented an average surface area decrease of 10.5% and a volume decrease of 8.4% for the cells that were 50 days old. Because both the surface area and volume decreased with cell age, there was little change in surface-to-volume ratio (sphericity) during aging. The aged cells were found to have normal membrane elasticity. In addition, human RBCs were fractionated over Stractan density gradients and the most dense cells were found to have rheologic properties similar to those reported for the aged RBCs from rabbits and mice, although the absolute magnitude of the changes in surface area and volume were considerably greater for the human cells. Thus, stringent density fractionation protocols that result in isolation of the most dense 1% of cells can produce a population of human cells with rheologic properties similar to senescent cells obtained in other species. The data indicate that progressive loss of cell area and cell dehydration are characteristic features of cell aging.

The average lifespan of rabbit and mouse erythrocytes is 60 days.[9] Aged erythrocytes from rabbits were isolated with a recently reported method that involves the biotinylation of erythrocytes with N-hydroxysuccinimido biotin.[21] Specifically, 2.5 kg rabbits were injected on 3 consecutive days with 7.5 mg/kg phenylhydrazine to produce a reticulocytosis; control experiments have shown that 70% to 85% of all preexisting cells are destroyed by this procedure (data not shown). Ten days later, the rabbit RBCs are biotinylated in vitro by reaction with N-hydroxysuccinimido biotin as previously described.[21] These biotinylated cells have been shown to have a normal in vivo survival.[21] After reinfusion, the biotinylated cells are allowed to age in vivo for various periods of time. The animal is then bled again and the biotinylated RBCs are recovered by binding to avidin-coated, plastic Petri dishes as described.[22] The recovered cells are removed from the Petri dishes

Aged erythrocytes from the mouse were isolated with the Ganzoni hypertransfusion procedure.[17,18] With this technique, a large number of starting mice are split into two equal groups, and one group is terminally bled to allow the other group of animals to be hypertransfused. Two weeks later, one half of the surviving animals are terminally bled to hypertransfuse the remaining mice; this procedure is repeated every 2 weeks for approximately 60 days. The model is based on the observation that hypertransfused animals will not synthesize new erythrocytes, and, therefore, the RBC population of the few surviving animals will have a continuously increasing mean age over the 60-day experiment.

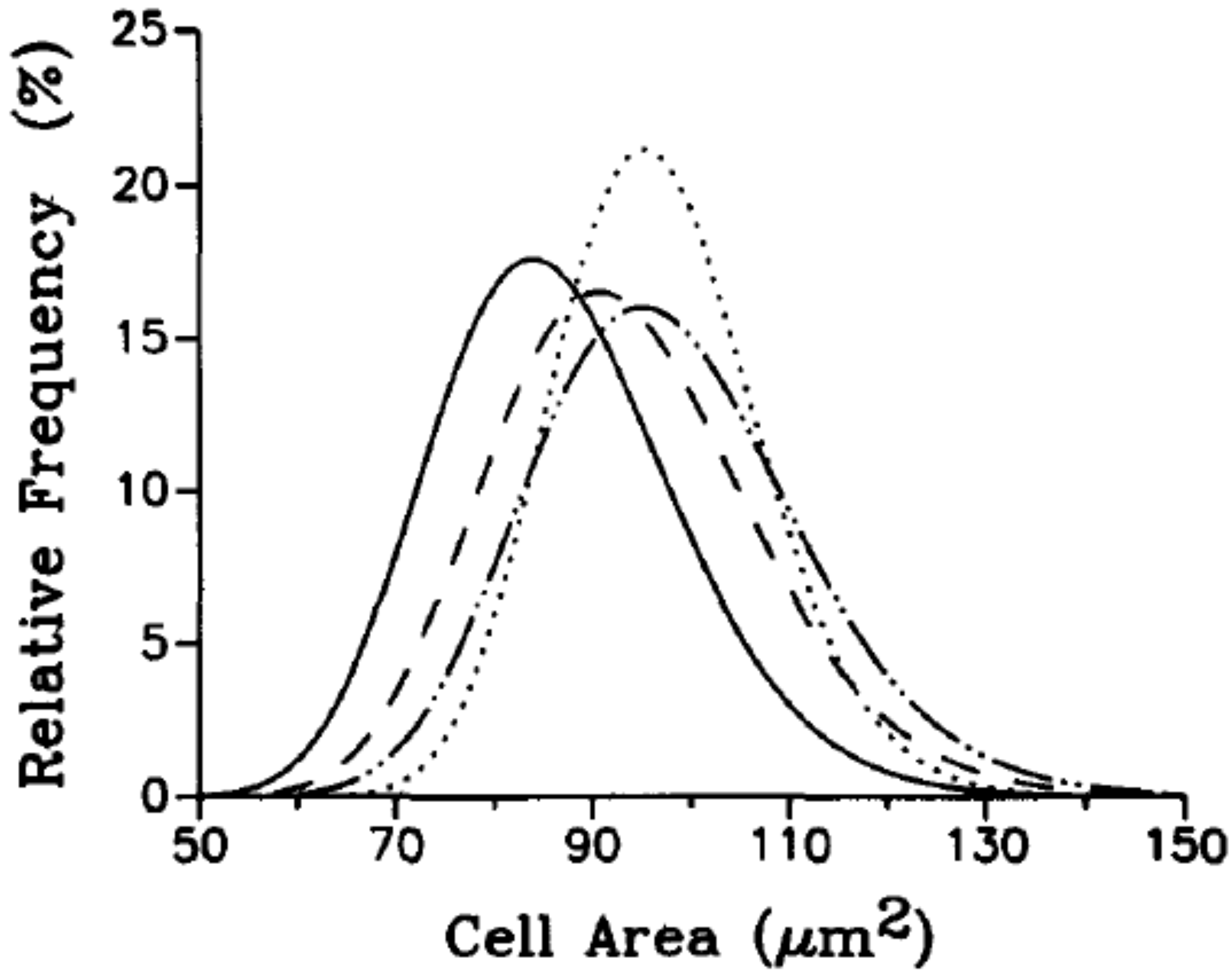


Fig 5. Analysis of cellular surface area by micropipette method. The control sample (· · · ·) is a whole blood control drawn from the rabbit 60 days after the end of phenylhydrazine treatment. The remaining curves illustrate the distributions of area for the aged cohorts of biotinylated, rabbit RBCs drawn at Day_{50} (———), Day_{24} (---), and Day_8 (— · · —). Data were obtained on cells from 11 different rabbits, three sampled at Day_8, four at Day_{24}, and four at Day_{50}. Clearly, the cells lose surface area as they age.

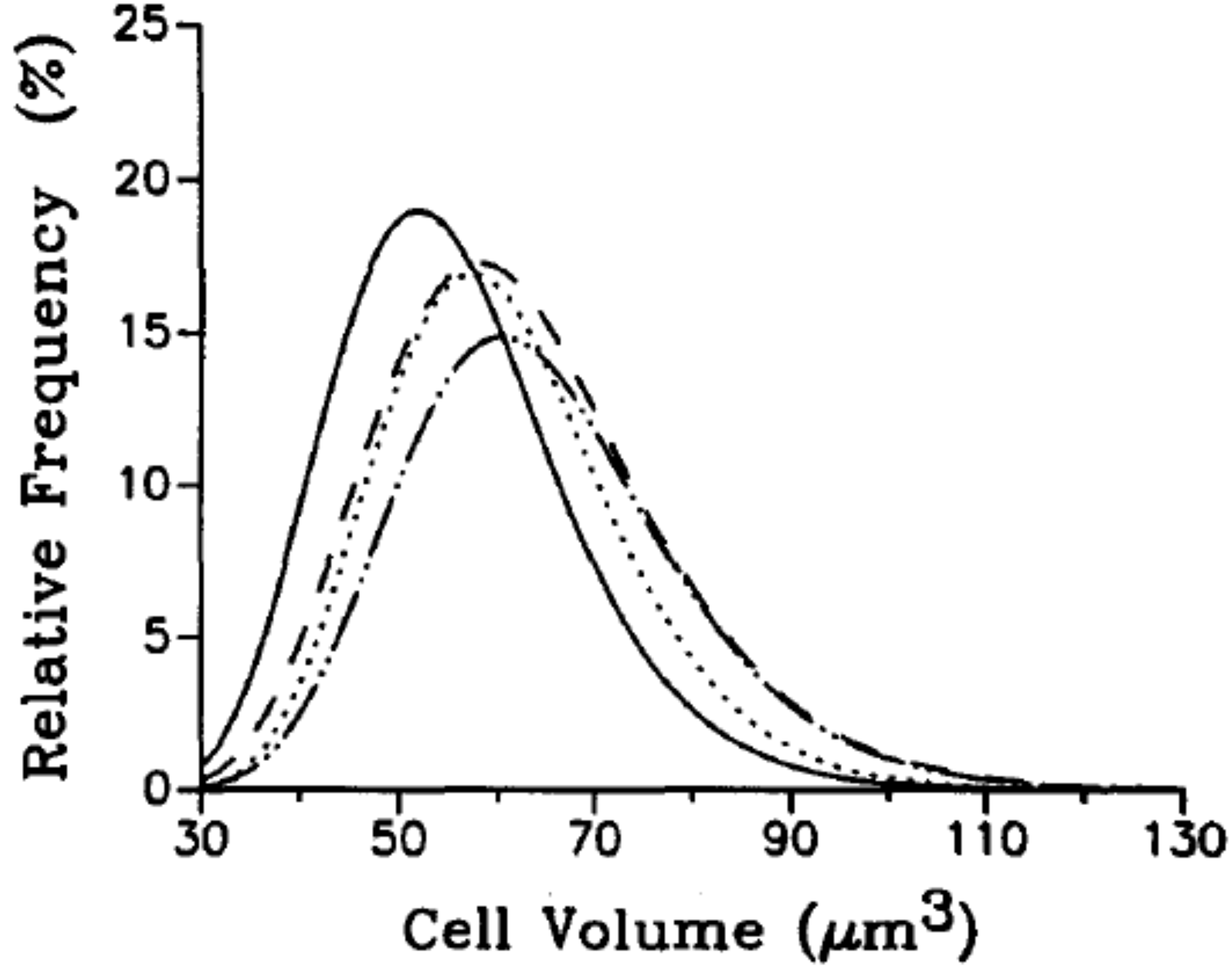


Fig 6. Analysis of cellular volume by micropipette method. Fitted distribution curves for the volume data were generated similarly to those in Fig 5. Distributions are shown for the various aged samples from a total of 11 rabbits (three at Day_8, four at Day_{24}, and four at Day_{50}) and the Day_{50} whole blood sample (results pooled from four different rabbits). A progressive loss of cellular volume is evident. See Fig 5 legend for symbols.

Cornell University

$$S = \frac{4\pi V^{2/3}}{(4\pi/3)^{2/3} A}$$

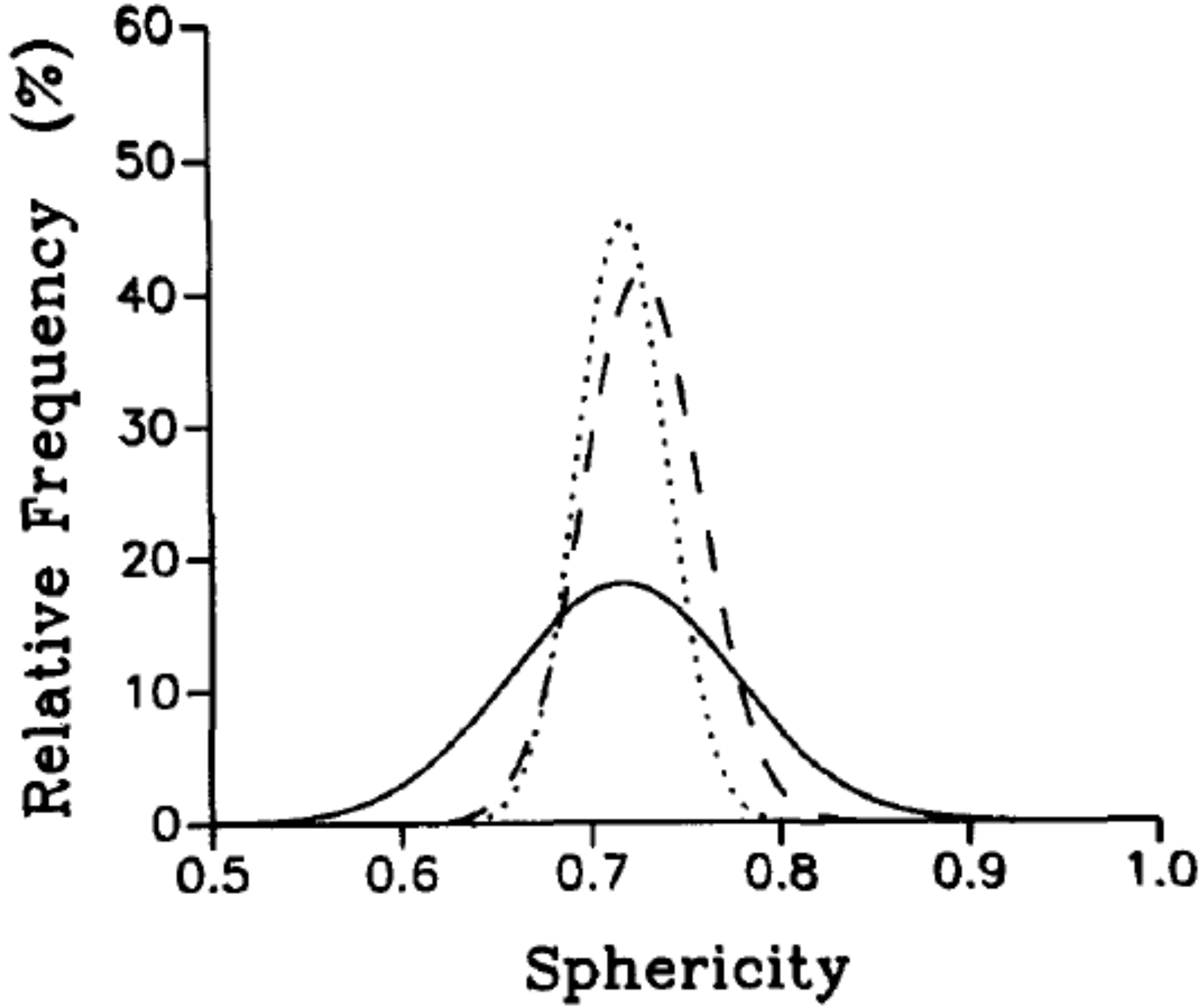


Fig 7. Distributions of sphericity for density-fractionated human cells. The mean sphericities for the different populations were not significantly different, but the variance of the oldest population was significantly increased. Each curve represents measurements on 65 to 77 cells, all obtained from a single individual. (——) MCHC > 37 g/dL; (---) MCHC = 31 g/dL; (· · · ·) whole blood.

Systematic Errors

- The result of

1. A mis-calibrated device, or
2. A measuring technique which always makes the measured value larger (or smaller) than the "true" value.

- **Example:** Using a steel ruler at liquid nitrogen temperature to measure the length of a rod.

➜ The ruler will contract at low temperatures and therefore overestimate the true length.

- Careful design of an experiment will allow us to eliminate or to correct for systematic errors.