Lecture 12

Minimum Mean Square Error Linear Prediction

Predicting a r.v. X by a constant a: best a?

$$\begin{aligned} \text{mmse value of } a \quad &: \quad \min_a \; E(X-a)^2 \\ &= \quad \min_a \; \{E(X^2) - 2aEX + a^2\} \end{aligned}$$

$$\text{occurs at } a = EX, \text{ minimum value } E(X^2) - (EX)^2 = \text{Var}(X)$$

Predicting the deviation $Y - EY$ of a r.v. Y from its mean as a multiple b times $X - EX$, where X is a secondly jointly distributed r.v: best b?

$$\begin{aligned} \text{mmse value of } b \quad &: \quad \min_b \; E((Y-EY) - b(X-EX))^2 \\ &= \quad \min_b \; \{E(Y-EY)^2 - 2bE(Y-EY)(X-EX) + b^2 E(X-EX)\} \\ &= \quad \min_b \; \{\text{Var}(Y) - 2b\text{Cov}(X,Y) + b^2\text{Var}(X)\} \end{aligned}$$

$$\text{occurs at } \; b = \frac{\text{Cov}(X,Y)}{\text{Var}(X)}, \; \text{minimum value } \text{Var}(Y)[1-\rho^2]$$

$$\text{where } \; \rho = \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)}$$

Regression and the Inheritance of Quantitative Traits

- Mendel gave us the story we now accept about the inheritance of **qualitative** traits. We'll revisit it. 1865.
- Galton discovered the phenomenon of **regression**. We'll revisit it. 1877.
- Pearson attempted to **reconcile** the two and failed, concluding that **quantitative** traits did not get inherited according to Mendel's laws. 1900.
- Yule **pointed out** his error. 1906.
- Fisher followed up on Yule and developed a full genetical theory for quantitative traits consistent with Mendel. 1912, 1918.

We begin with an important idea from **population genetics** due independently to Hardy and Weinberg, 1906. Suppose that A and a are two allelic forms of a gene inherited according to Mendel's laws, and suppose that individuals mate "at random" with respect to any phenotypic effects of this gene.

Suppose that the frequency of AA, Aa and aa individuals in the population is P, $2Q$ and R ($P + 2Q + R = 1$) in the parental generation. We seek the **joint distribution** of the genotypes between parent and offspring. That is, I want the full table of probabilities of the form:

$$\text{pr}(\text{father is } Aa, \text{ son is } aa)$$

Here there is no sex effect but I'll take parent = father, offspring = son for definiteness, and to relate to lab 3.

$$\text{pr}(\text{father is } Aa, \text{ son is } aa) = \text{pr}(\text{father is } Aa)\text{pr}(\text{son is } aa|\text{father is } Aa)$$

Now, pr(father is Aa) $=2Q$, but for pr(son is aa|father is Aa) we need to consider the mother.

$$\begin{aligned}
\text{pr}(\text{son is } aa|\text{father is } Aa) &= \text{pr}(\text{son is } aa, \text{ mother is } AA|\text{father is } Aa) \\
&\quad + \text{pr}(\text{son is } aa, \text{ mother is } Aa|\text{father is } Aa) \\
&\quad + \text{pr}(\text{son is } aa, \text{ mother is } aa|\text{father is } Aa) \\
&= \text{pr}(\text{m is } AA|\text{f is } Aa)\text{pr}(\text{s is } aa|\text{ m is } AA, \text{ f is } Aa) \\
&\quad + \text{pr}(\text{m is } Aa|\text{f is } Aa)\text{pr}(\text{s is } aa|\text{ m is } Aa, \text{ f is } Aa) \\
&\quad + \text{pr}(\text{m is } aa|\text{f is } Aa)\text{pr}(\text{s is } aa|\text{ m is } aa, \text{ f is } Aa) \\
&= P \times 0 + 2Q \times \tfrac{1}{4} + R \times \tfrac{1}{2} \text{ by random mating and Mendel}
\end{aligned}$$

		Offspring genotype			
		AA	Aa	aa	Marginal
Parental	AA	$P(P+Q)$	$P(Q+R)$	0	P
genotype	Aa	$2Q \times \frac{1}{2}(P+Q)$	$2Q \times \frac{1}{2}$	$2Q \times \frac{1}{2}(Q+R)$	$2Q$
	aa	0	$R(P+Q)$	$R(Q+R)$	R
Marginal		$(P+Q)^2$	$2(P+Q)(Q+R)$	$(Q+R)^2$	

Joint distribution table

The other entries are built up similarly. Check the row marginal probabilities: as assumed. Check the column marginal probabilities: o.k. for AA, aa; for Aa: either use summing to 1 or:

$$\begin{aligned}
2(P+Q)(Q+R) &= 2PQ + 2Q^2 + 2PR + 2QR \\
&= PQ + Q(P + 2Q + R) + 2PR + QR \\
&= P(Q+R) + Q + R(P+Q)
\end{aligned}$$

If we let $P + Q = p$ — the A-allele **gene frequency** we have derived the Hardy-Weinberg equilibrium **genotype** frequencies in the offspring generation:

$$\begin{array}{ll} AA & p^2 \\ Aa & 2p(1-p) = 2pq \text{ where } q = 1-p \\ aa & q^2 \end{array}$$

If we now suppose that the population was **already** in Hardy Weinberg equilibrium via random mating, the joint distribution of parental-offspring genotypes is:

		Offspring genotype			
		AA	Aa	aa	
Parental	AA	p^3	p^2q	0	p^2
genotype	Aa	p^2q	pq	pq^2	$2pq$
	aa	0	pq^2	q^3	q^2
		p^2	$2pq$	q^2	

Following Yule we'll put $p = q = 1/2$.

		Offspring genotype		
		AA	Aa	aa
Parental	AA	1/8	1/8	0
genotype	Aa	1/8	1/4	1/8
	aa	0	1/8	1/8

Now we consider a **quantitative trait** governed solely by this gene (a bit unrealistic, but we'll proceed, with Yule). The population mean value of this trait will be something, but we are interested in fluctuations about this mean generated by Mendelian inheritance.

Suppose that individuals with genotypes AA, Aa and aa are α, β and γ units from the mean. Then $\alpha+2\beta+\gamma = 0$ (since genotype frequencies are 1:2:1). The population **variance** of the trait for both parent and offspring is:

$$\text{Variance} = \frac{1}{4}\alpha^2 + \frac{1}{2}\beta^2 + \frac{1}{4}\gamma^2$$

and the parent-offspring **covariance** is:

$$\begin{aligned} \text{Cov(Par,Off)} &= \frac{1}{8}\alpha^2 + \frac{1}{8}\alpha\beta + \frac{1}{8}\beta\alpha + \frac{1}{4}\beta^2 + \frac{1}{8}\beta\gamma + \frac{1}{8}\gamma\beta + \frac{1}{8}\gamma^2 \\ &= \frac{1}{8}(\alpha^2 + 2\alpha\beta + 2\beta^2 + 2\beta\gamma + \gamma^2) \end{aligned}$$

Thus the population linear **regression coefficient** b of offspring on parent trait is

$$b = \frac{\text{Cov(Par,Off)}}{\text{SD(Par)SD(Off)}} \quad = \quad \frac{\frac{1}{8}(\alpha^2 + 2\alpha\beta + 2\beta^2 + 2\beta\gamma + \gamma^2)}{\frac{1}{4}(\alpha^2 + 2\beta^2 + \gamma^2)}$$

$$
\begin{aligned}
\text{i.e. } b &= \frac{1}{2} + \frac{\beta(\alpha+\gamma)}{\alpha^2+2\beta^2+\gamma^2} \\
&= \frac{1}{2} \qquad \text{if } \beta = 0
\end{aligned}
$$

If there was dominance, say $\alpha = \beta$, we'd find $\gamma = -3\alpha$ and

$$\frac{\text{Cov(Par,Off)}}{\text{SD(Par)SD(Off)}} = \frac{1}{3}$$

Interestingly, Pearson thought (following Mendel) there should be dominance, got this 1/3, whereas 1/2 is observed, and thought Mendel must fail for QTs. Yule pointed out that additivity restores Mendel.