Polygenic threshold theory: [Human Molecular Genetics, 4rd Ed, Chapter 3]

- History: debate betweeen Mendelians and biometricians (Francis Galton). A paper by R.A. Fisher in 1918 settled the debate: characters governed by a large number ofindependent Mendelian factors (polygenic characters) would display precisely the continuous nature, quantitative variation, and family correlations described by the biometricians. DS Falconer extended this model to cover dichotomous characters.

- Polyngenic theory: if a trait is determined by many loci (the sum of effects of all loci), then the trait distribution is approximately normal. The goal of this theory is to understand the influence of genetic and environmental factors, and the pattern of inheritance. However direct regression analysis is difficult, because neither genotype nor environmental variables are directly measured.

- Regression to the mean: one main observation in quantitative trait studies is: the offsprings of the individuals with extreme phenotypes, tend to have traits in the intermediate between the parent and the population mean. This is called "regression to the mean". This can be explained by that the parent may have many alleles that have extreme values (comparing with population mean), but because of random mating, the other parent has a smaller number of such alleles.

- Dicontinuous characters: assume there is an underlying continuous variable: susceptibility, and that there exists a threshold - when the suspectibility of an individual is higher than the threahold, the disease may develop. Susceptibility is a random variable for individuals, similar to a quantitative character, while the liability threshold is fixed for a group. To model the risks: let t be the threshold relative to the population mean in the unit of standard deviation (i.e. the Z score), then the risk is:

$$P(X \geq t) = 1 - \Phi(t) \tag{1}$$

  where $\Phi(\cdot)$ is the CDF of normal distribution. Suppose we have a mutation in a gene, the individuals carrying this mutation have a different distribution of susceptibility - suppose its mean is μ and its standard deviation is the same (the mean is 0 for people without the mutation). Let Z_μ be the Z-score of the mean (i.e. measure of μ with the unit of standard deviation), then the risk of the mutant group is:

$$P(X \geq t) = P(X' \geq t - Z_\mu) = 1 - \Phi(t - Z_\mu) \tag{2}$$

  The relative risk is thus:

$$\text{RR} = \frac{1 - \Phi(t - Z_\mu)}{1 - \Phi(t)} \tag{3}$$

  For example, at $t = 2$, the risk is about 2.3%, and suppose $Z_\mu = 0.5$, the risk is 6.7%, and the RR about 2.9.

- Model different risks in different groups: e.g. some diseases have higher risk in one gender than the other. This can be modeled by allowing different thresholds in different genders.

Reference: `https://www.sugarsync.com/pf/D7756315_94196427_6708686`