

- ①:T人、binary形質、k種類の間隔尺度付ジェノタイプから得られる2xk分割表のトレンド統計量を $X^2_{trend}$ とする。  
 ②:形質、ジェノタイプから得られるピアソン相関係数をrとする。  
 ⇒この時、 $X^2_{trend}$ , r の関係は次式で表すことができる。

$$X^2_{trend} = Tr^2. \dots \textcircled{1}$$

・形質(y)をジェノタイプ(x)で線形回帰した際の回帰係数を a, b とする。

$$y = a + bx.$$

・この時、回帰係数 b の分散は次式で表される。(S<sub>ij</sub> は i, j の偏差積和)

$$Var(b) = Var\left(\frac{S_{xy}}{S_{xx}}\right) = Var\left[\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}\right] = \left[\frac{1}{\sum(x_i - \bar{x})^2}\right]^2 \sum(x_i - \bar{x})^2 Var(y_i) = \frac{1}{\sum(x_i - \bar{x})^2} \frac{\sum(y_i - \bar{y})^2}{T} = \frac{S_{yy}}{TS_{xx}}.$$

・トレンド統計量の定義より、

$$X^2_{trend} = \frac{b^2}{Var(b)} = \left[\frac{S_{xy}}{S_{xx}}\right]^2 \frac{TS_{xx}}{S_{yy}} = T \left[ \sqrt{\frac{S_{xy}^2}{S_{xx}S_{yy}}} \right]^2 = Tr^2.$$

1. Tests for Liner Trends in Proportions and Frequencies. P.Armitage. *BIOMETRICS*, Sep 1955.

分割表上の変数から、①式に基づきトレンド統計量の数式を導いてみる。

	genotype distribution (x)							total	
order of genotype	0	1	2	...	s <sub>i</sub>	...	s <sub>k-1</sub>	-	
phenotype (y)	1	n <sub>0</sub>	n <sub>1</sub>	n <sub>2</sub>	...	n <sub>i</sub>	...	n <sub>k-1</sub>	t
	0	N <sub>0</sub> -n <sub>0</sub>	N <sub>1</sub> -n <sub>1</sub>	N <sub>2</sub> -n <sub>2</sub>	...	N <sub>i</sub> -n <sub>i</sub>	...	N <sub>k-1</sub> -n <sub>k-1</sub>	T-t
total	-	N <sub>0</sub>	N <sub>1</sub>	N <sub>2</sub>	...	N <sub>i</sub>	...	N <sub>k-1</sub>	T

$$\begin{aligned} \sum x_i &= \sum s_i N_i. & S_{xx} &= \sum x_i^2 - \frac{(\sum x_i)^2}{T} = \frac{T \sum s_i^2 N_i - (\sum s_i N_i)^2}{T^2}. \\ \sum x_i^2 &= \sum s_i^2 N_i. & S_{yy} &= \sum y_i^2 - \frac{(\sum y_i)^2}{T} = \frac{t(T-t)}{T^2}. \\ \sum y_i &= \sum (0 \cdot (N_i - n_i) + 1 \cdot n_i) = t. & S_{xy} &= \sum x_i y_i - \frac{\sum x_i \sum y_i}{T} = \frac{T \sum s_i n_i - t \sum s_i N_i}{T^2}. \\ \sum y_i^2 &= \sum (0^2 \cdot (N_i - n_i) + 1^2 \cdot n_i) = t. \\ \sum x_i y_i &= \sum (0 \cdot s_i (N_i - n_i) + 1 \cdot s_i n_i) = \sum s_i n_i. \end{aligned}$$

$$X^2_{trend} = Tr^2 = T \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{T(T \sum s_i n_i - t \sum s_i N_i)^2}{t(T-t)(T \sum s_i^2 N_i - (\sum s_i N_i)^2)}. \dots \textcircled{2}$$

2x3分割表においてs<sub>i</sub>={0,1,2}と置いた場合。

$$X^2_{trend 2 \times 3} = \frac{T(T(n_1 + 2n_2) - t(N_1 + 2N_2))^2}{t(T-t)(T(N_1 + 4N_2) - (N_1 + 2N_2)^2)}.$$

## ○: 量的形質への適用

EIGENSTRATにおけるトレンド統計量算出方法

- ・質的形質(eigenstrat.c)、量的形質(eigenstratQTL.c)共に、①式に基づき統計量を算出している。

(右のサブルーチンを使用)

- ・量的形質に対する統計量もdf=1で処理している。

⇒線形回帰の傾きの検定と考えれば、量的形質への適用は妥当かもしれないが、df=1の根拠は？

- ・回帰係数bの帰無仮説(b=0)に対する検定は

$$\beta = \frac{b}{s_b} \sim t_{T-2}.$$

に基づき実施される。

( $s_b$ はbの標本標準偏差、 $t_{T-2}$ はdf=T-2のt分布)

- ・サンプル数Tが十分に大きい時、

$$Tr^2 = \frac{b^2}{Var(b)} \approx \left( \frac{b}{s_b} \right)^2 \sim \chi^2_{(1)}.$$

```
double compute_chisq(double *source, double *target)
{
    int n;
    double sum1, sumx, sumxx, sumy, sumyy, sumxy, numer, denom1, denom2;
    double corr;

    sum1 = 0.0; sumx = 0.0; sumxx = 0.0; sumy = 0.0; sumyy = 0.0; sumxy = 0.0;
    for(n=0; n<NSAMPLES; n++)
    {
        if(outlier[n]) continue;
        if(source[n] < -99.0) continue;
        if(target[n] < -99.0) continue;
        sumx += source[n];
        sumxx += source[n]*source[n];
        sumy += target[n];
        sumyy += target[n]*target[n];
        sumxy += source[n]*target[n];
        sum1 += 1.0;
    }
    if(sumxx == 0.0) return -1.0;
    if(sumyy == 0.0) return -1.0;
    numer = sumxy/sum1 - (sumx/sum1)*(sumy/sum1);
    denom1 = (sumxx/sum1 - (sumx/sum1)*(sumx/sum1));
    denom2 = (sumyy/sum1 - (sumy/sum1)*(sumy/sum1));
    if(denom1 <= 0.0) return -1.0;
    if(denom2 <= 0.0) return -1.0;
    corr = (numer/sqrt(denom1*denom2));
    return (sum1*corr*corr);
}
```

偏差積和の計算

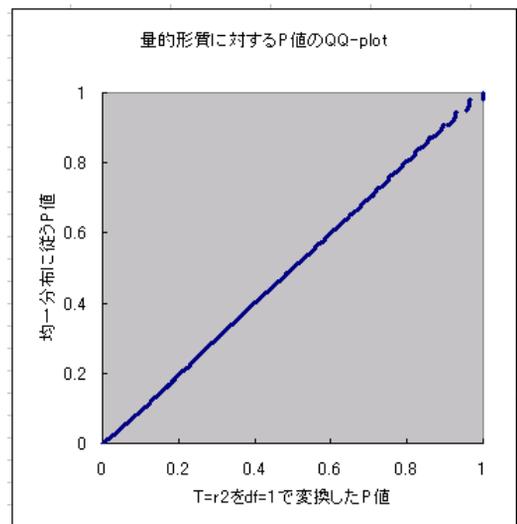
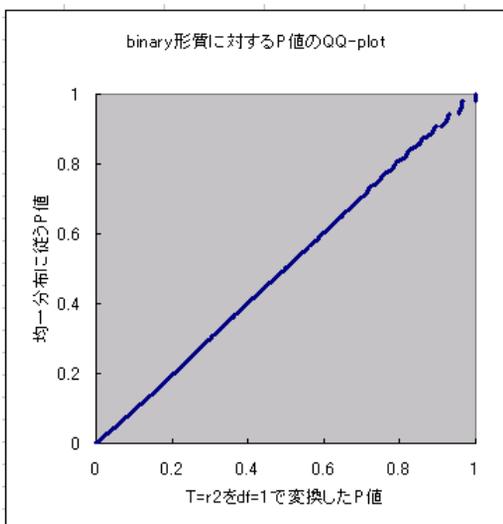
相関係数rの計算

$X^2_{trend}=Tr^2$ の計算

## ○: シミュレーションデータによる解析結果

・T=1000, 10,000SNPs, MAF=0.2~0.8 のジェノタイプデータ。Binary形質(case:control=1:1)、量的形質(正規分布より生成)の各々のフェノタイプデータを生成。

- ・両形質に対し $Tr^2$ を計算、df=1の $\chi^2$ 分布に基づきP値に変換し、QQ-plotを作成した。



⇒共にほぼ均一なP値の分布が得られた。