补充题

在一次对大学生看武侠小说的随机调查中，男生中有 a 名爱看，b 名不爱看；女生中有 c 名爱看，d 名不爱看（a、b、c、d 都很大），并记总人数 $a+b+c+d=n$。检验男女同学在爱看武侠小说方面有无显著差异，请给出检验统计量 U 的具体结果，证明

$$U^2=\frac{n(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

并指出在 $H_0: p_1=p_2$ 成立条件下 U^2 的分布，如果以 U^2 作为检验统计量，给出在显著水平 $\alpha=0.05$ 下的拒绝域。

ps：高中学理科的孩子们，对于 U^2 的形式是不是很熟悉？（至少在高考时应该很熟悉）

解：假设 $H_0: p_1=p_2$ vs $H_0: p_1\neq p_2$，检验统计量为

$$U=\frac{\bar{X}-\bar{Y}}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1),$$

其中

$$\bar{X}=\frac{a}{n_1}=\frac{a}{a+b},\quad \bar{Y}=\frac{c}{n_2}=\frac{c}{c+d},\quad \hat{p}=\frac{a+c}{n_1+n_2}=\frac{a+c}{n},$$

则

$$U=\frac{\frac{a}{a+b}-\frac{c}{c+d}}{\sqrt{\frac{a+c}{n}\left(1-\frac{a+c}{n}\right)}\sqrt{\frac{1}{a+b}+\frac{1}{c+d}}}=\frac{\frac{ad-bc}{(a+b)(c+d)}}{\sqrt{\frac{(a+c)(b+d)}{n^2}}\sqrt{\frac{n}{(a+b)(c+d)}}}$$

$$=\frac{\sqrt{n}(ad-bc)}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}。$$

可得

$$U^2=\frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)},$$

在 $H_0: p_1=p_2$ 成立条件下，U^2 为标准正态变量的平方，$U^2\sim\chi^2(1)$，拒绝域为

$$W=\{U^2\geq\chi^2_{1-\alpha}(1)\}=\{U^2\geq 3.8415\}。$$