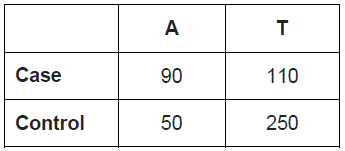
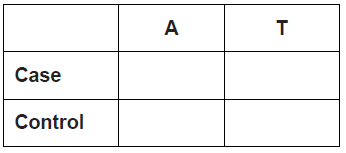
**P1. Association Studies (40 points)**

Consider the following data case-control data. We will perform a chi-square test for association with a SNP.



1. Fill in the following table with the counts you would expect if you assumed independence. Show your work.



1. Now, compute the Chi-Square statistic and state the conclusion for the p-value with cutoff of 0.05.

**P2. eQTL analysis (60 points)**

1. For Alzheimer’s disease, search for a list of associated SNPs in GWASCatalog. Submit an excel file with all significant SNPs (p\_value < 5E-8), their mapped genes, their distance to closest genes, and p-values. Note: Distance to downstream genes will be negative; Distance to upstream genes will be positive; SNPs located within genes will have distance as 0.
2. Make a bar plot showing the number of SNPs located on the upstream of genes, on the downstream of genes, and within genes respectively. (Hint: context column in the downloaded data; distance to upstream genes indicates that snp is located on the downstream of the gene; if one snp has both downstream and upstream genes, pick the one with smaller distance.)
3. Make a bar plot showing the chromosome distribution of all significant SNPs.
4. Make a histogram showing the distance from significant SNPs to closest genes.
5. Using BRAINEAC database, check whether any of the significant SNPs located on the upstream of genes are eQTLs (p\_value<0.05). If yes, submit a table with 4 columns: eQTL, associated genes, p\_value and tissues.