CSE 5370: Bioinformatics
Homework 1 Answers

Kushaal Kashyap Narasimhan , Aravind Ravi Kiran Kashyap
Shreyas Jagadeep Shete , Nischitha Sadananda

January 31, 2022

1 Genetics Questions

1.1 Mendelian Genetics

	RY	Ry	rY	ry
ry	RrYy	Rryy	rrYy	rryy
ry	RrYy	Rryy	rrYy	rryy
ry	RrYy	Rryy	rrYy	rryy
ry	RrYy	Rryy	rrYy	rryy

- What fraction of peas from this cross will be round yellow (phenotype)?
  Answer: 4 / 16

- What fraction of peas from this cross will be round green (phenotype)?
  Answer: 4 / 16

- What fraction of peas from this cross will be wrinkled yellow (phenotype)?
  Answer: 4 / 16

- What fraction of peas from this cross will be wrinkled green (phenotype)?
  Answer: 4 / 16

1.2 Genome Wide Association Studies (GWAS)

- Genome-wide association studies (GWAS) look for associations between single nucleotide polymorphisms (SNPs) and phenotypes across many subjects. Suppose we test 2.6 million SNPs for association with a phenotype,

such as height. Assuming each association between a SNP and the height phenotype is an independent hypothesis, and we want our effective p-value to be .05. What is our Bonferroni-corrected p-value? [4 points]

Answer: : Bonferroni corrected p values = $\alpha \div n$

0.05 ÷ 2.6 million

1.92307692e × 10 ∧ -8

- If we were to perform a GWAS with the same number of SNPs but instead look for correlation with 1000 metabolites, what would be our new Bonferroni-corrected p-value? [4 points]
  Answer: 0.05 ÷ 1000
  0.00005

- Bonferroni correction is noted to be conservative. What does it mean in terms of type I and type II errors? [3 points]

  Answer:Bonferroni correction is used to reduce type 1 error but it will become vulnerable to type 2 error as it fails to reject the null hypothesis when it is supposed to be done , hence it is noted to be conservative.

2 Statistics Questions [30 points total]

2.1 Drug Approval

- State the null and alternative hypotheses. [8 points]
  Answer: Null Hypothesis : The drug lowers the blood pressure of the patients
  Ho = $\mu = \mu \circ = 115$

  Answer: Alternative Hypothesis : The drug does not lower the blood pressure of the patients
  Ho = $\mu \neq \mu \circ \neq 115$

- What type of test is appropriate to test statistical significance? What is the p-value resulting from this test? [8 points]
  Answer: Sample Size n = 20 which is less than 30 and standard deviation us given ,it is going to be Z test that us appropriate to test statistical significance.

$$Z = G \left(\frac{\overline{X} - \mu \circ}{\frac{\sigma}{n}} \right)$$

$\overline{X} = 115$
$\mu \circ = 120$
$n = 20$
$\sigma = 15$
substituting the values we get

Z = -1.49 , —Z— = 1.49 using the online calculator we get P value = 0.06
We can approve the drug as 0.06 is greater than 0.05

- What conclusions can be drawn from this p-value at a significance threshold of 0.05? Should this drug be put on the market? What factors about the experimental design would you consider in making this decision? (No need to do additional calculations) [9 points]
  Answer: Yes , this drug can be put on the market .

  The sample of 20 mice taken for experiment will have similar result in the population and that in human beings is a factor considered in making the decision.

2.2 Other Questions

- What is the advantage of a non-parametric test vs a parametric test? [3 points]
  Answer The advantages of non-parametric test are it can be used on small sample size , they make fewer assumptions about the data and they are useful in analyzing data that are inherently in ranks or categories.

- If a non-parametric test makes fewer assumptions, why should we ever use a parametric test? [2 points]

  Answer We should use parametric test as it highlights better variance in the distribution between two arms but non parametric test makes fewer assumptions.

3 Difficulty Adjustment

Your answers to this section will be used to adjust the difficulty of future assignments in the class.

- How long did this assignment take you to complete?
  Answer: This assignment took me more than a day to complete.For the Questions other than the programming I had to go through the lecture video again to solve and google to understand the basics which took me roughly 3 hrs . For the programming assignment it took me a days work as i have been using windows and no access to mac. I had to try different ways to install and due to limited lecture in proceeding with colab.

- If the assignment took you longer than the designed 3 hours, which parts were overly difficult?
  Answer:Setting up of the environment could have been explained in detail during face to face class and took more than day to get everything setup.