Due April 19, 2017

Problem Set 3

1) Prof. X sends two Lazy students to score a bunch of haploid yeast colonies for two properties. The Lab expects these two properties to be independent. They also expect each property, controlled by alleles A/a and B/b, to have the two variants present in equal proportion. In short the probability table should look like:

| Allele 1 \ Allele 2 | B | b |
|---|---|---|
| A | $p_{AB} = \frac{1}{4}$ | $p_{Ab} = \frac{1}{4}$ |
| a | $p_{aB} = \frac{1}{4}$ | $p_{ab} = \frac{1}{4}$ |

a) Student I decides to make up data. He presents the following table for numbers of colonies (his imagined total number is 1000).

| 1 \ 2 | B | b |
|---|---|---|
| A | 251 | 252 |
| a | 244 | 253 |

He essentially took 250 for each entry and added some made up error.

Prof. X looks at the data and thinks it fits too well. How can he use $\chi^2$ statistic [redacted] for goodness of fit to 'verify' that something is fishy? He believes the [redacted] probabilities are a quarter each and wants to do a test at 5% significance level

b) Student II is not to be outdone by her competitor. She is going to do some experiments and leave the plates all over for everyone to see. However she only scores 40 colonies and gets genuine data

| 1\2 | B | b |
|---|---|---|
| A | 7 | 11 |
| a | 14 | 8 |

Now she multiplies it by 25.

| 1\4 | B | b |
|---|---|---|
| A | 175 | 275 |
| a | 350 | 200 |

[redacted] Well, it looks suspicious: multiples of 25. The numbers are also very different from eachother.

She then adds or subtract a random integer [redacted] between [0,20] to bring the numbers close to 250. At the end She produces: [redacted]

| X \ Y | B | b |
|---|---|---|
| A | 195 | 281 |
| a | 320 | 204 |

She claims it is evidence of the probabilities be a $\frac{1}{4}$ each, with the appropriate amount of noise.

Should the professor buy her argument? What p-value would he get for her rather large $\chi^2$?

c) Why did scaling up [redacted] noisy data not work? [redacted] Typically $\chi^2$ of d.f. $\nu$ is order $\nu \pm \sqrt{2\nu}$.

If I take a 2x2 contingency table with known probabilities and scale every frequency up by a factor $\lambda$, [redacted] how does the fake $\chi^2$ statistic[redacted] change?

2) Do exercise 7.8 (Bayesian linear regression in 1d) (a)-(d), from the Murphy book.