Lab Assignment 4

STA 100 | A. Farris | Spring 2021

A pdf copy of your assignment is due at the end of the day (5:00pm PT) Monday, May 17. Submission of the pdf will be through gradescope. Please put in the effort to make it look reasonably professional – you're encouraged to use R Markdown. Note that specific tasks for you are highlighted. You are free to work in groups, but you must submit your own writeup, and run your own code.

Goodness-of-fit for Mendelian inheritance

It is hardly necessary here to motivate the importance of Mendelian notions of inheritance in the history of Biology.[1].

Chi-squared tests of goodness-of-fit are a natural tool for evaluating claims of the kind inherent in Mendelian inheritance. Mendel reported his results in 1865; our tests for goodness-of-fit were not developed until 1900. Thus, these tools were not available to Mendel himself. Nonetheless, they were applied for these purposes very shortly after the development of the tests[2].

It is therefore of historical interest for us to evaluate several claims of Mendel's, on the basis of his own reported experimental results.

Experiment 1

In his first experiment, Mendel differentiated his pea plants by two characteristics, 'form' and 'albumen'. Form here could be round (A) or wrinkled (a), and albumen could be yellow (B) or green (b). Thus, each plant can be categorized according to whether it displays one or the other of each of these characteristics, or according to whether it shows a mix of these characteristics as a hybrid. For example, then, 'AB' would characterize peas that are yellow and round, and 'AaB' would characterize peas with a mix of round yellow and wrinkled yellow types.

In Figure 1 we find a table of Mendel's results, counting the number of plants obtained in one generation of each type.[3]

For these categories Mendel hypothesized the following ratios:

Table 1: Relative ratios under the Null hypothesis

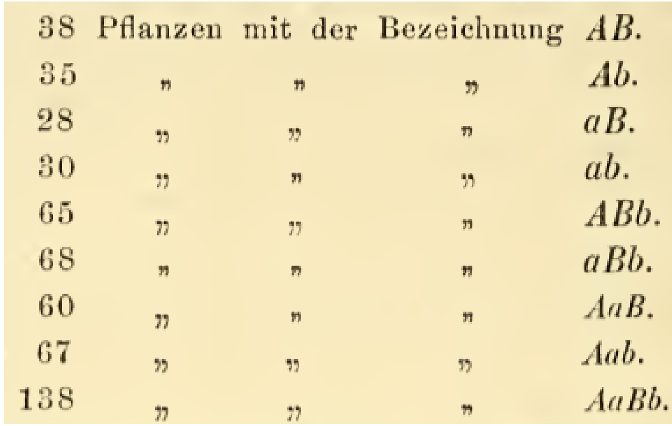
Category	AB	Ab	aB	ab	ABb	aBb	AaB	Aab	AbBb
NullRatio	1	1	1	1	2	2	2	2	4

Thus he proposed that the population proportion for AaBb should be four times that of Ab, the proportion of AB should be half that of ABb, etc. Our Null hypothesis is that this claim is correct.

[1] Mendel, Gregor. 1866. Versuche über Plflanzen-hybriden. Verhandlungen des naturforschenden Ver-eines in Brünn, Bd. IV für das Jahr 1865, Abhand-lungen, 3–47. http://www.esp.org/foundations/genetics/classical/gm-65.pdf

[2] Harris, J. Arthur. 1912. A Simple Test of the Goodness of Fit of Mendelian Ratios. The American Naturalist, Vol. 46, No. 552 (Dec., 1912), pp. 741-745 https://www.jstor.org/stable/2456130

[3] https://library.si.edu/digital-library/book/versucheberpflan00mend



38	Pflanzen	mit	der Bezeichnung	*AB.*
35	„	„	„	*Ab.*
28	„	„	„	*aB.*
30	„	„	„	*ab.*
65	„	„	„	*ABb.*
68	„	„	„	*aBb.*
60	„	„	„	*AaB.*
67	„	„	„	*Aab.*
138	„	„	„	*AaBb.*

Figure 1: Mendel's observed counts for Experiment 1

Table 2: Observed vs. Expected values for Experiment 1 (continued below)

	AB	Ab	aB	ab	ABb	aBb
NullProportion	0.0625	0.0625	0.0625	0.0625	0.125	0.125
Expected	33.1	33.1	33.1	33.1	66.1	66.1
Observed	38	35	28	30	65	68
Discrepancy	0.737	0.114	0.775	0.284	0.0191	0.0532

	AaB	Aab	AbBb
NullProportion	0.125	0.125	0.25
Expected	66.1	66.1	132
Observed	60	67	138
Discrepancy	0.567	0.0116	0.25

From Table 2 above, we find that the peas in categories AB (round and yellow), aB (wrinkled and yellow), and AaB (yellow and both round and wrinkled) stand out somewhat in terms of their discrepancies between values observed in the sample and those expected under the Null hypothesis.

Under the Null hypothesis, the sum of these discrepancies has an approximate χ^2 distribution with 8 degrees of freedom. From this, we calculate a p-value of 0.946, suggesting a lack of evidence against the Null hypothesis, that the true population ratios are as Mendel posited; so, these discrepancies are together not sufficiently large to rule out the possibility that they are merely due to chance.

Experiment 2

In his second experiment, Mendel differentiated his pea plants by three characteristics, 'form', 'albumen', and 'seed-coat'. Form here could be round (A) or wrinkled (a), albumen yellow (B) or green (b), and seed-coat grey-brown (C) or white (c). Thus, each plant can be categorized according to whether it displays one or the other of each of these characteristics, or according to whether it shows a mix of these characteristics as a hybrid. For example, then, 'ABC' would characterize peas that are yellow and round with grey-brown seed coat, and 'AaBc' would characterize peas having a mix of round yellow with white seed-coat and wrinkled yellow with white seed-coat types.

For this experiment, Mendel proposed the relative ratios of plants in the Table below.

Table 4: Relative ratios under the Mendel's proposal (continued below)

Category	ABC	ABc	AbC	Abc	aBC	aBc	abC	abc	ABCc	AbCc
NullRatio	1	1	1	1	1	1	1	1	2	2

Category	aBCc	abCc	ABbC	ABbc	aBbC	aBbc	AaBC	AaBc	AabC
NullRatio	2	2	2	2	2	2	2	2	2

Category	Aabc	ABbCc	aBbCc	AaBCc	AabCc	AaBbC	AaBbc	AbBbCc
NullRatio	2	4	4	4	4	4	4	8

Figure 2 below displays Mendel's observed counts for Experiment 2.

8	Pflanzen	*ABC*	22	Pflanzen	*ABCc*	45	Pflanzen	*ABbCc*
14	"	*ABc*	17	"	*AbCc*	36	"	*aBbCc*
9	"	*AbC*	25	"	*aBCc*	38	"	*AaBCc*
11	"	*Abc*	20	"	*abCc*	40	"	*AabCc*
8	"	*aBC*	15	"	*ABbC*	49	"	*AaBbC*
10	"	*aBc*	18	"	*ABbc*	48	"	*AaBbc*
10	"	*abC*	19	"	*aBbC*			
7	"	*abc*	24	"	*aBbc*			
			14	"	*AaBC*	78	"	*AaBbCc*
			18	"	*AaBc*			
			20	"	*AabC*			
			16	"	*Aabc*			

Figure 2: Mendel's observed counts for Experiment 2

Evaluate Mendel's claim on the basis of the observed counts in Figure 2, using significance level 0.1 . To do so, first report a table of observed values, expected values, and discrepancies between them. Which categories appear most discrepant? What is the Null distribution for the test statistic here (including its degrees of freedom)? Would you reject the null hypothesis here, or fail to do so?

Appendix: R Script

```r
library(pander) # may need to install package 'pander' to make the table below
Category <- c("AB","Ab","aB","ab",
              "ABb","aBb","AaB","Aab",
              "AbBb")
NullRatio <- c(1,1,1,1,
               2,2,2,2,
               4)
ATable <- rbind(Category,NullRatio)
pander(ATable, caption="Relative ratios under the Null hypothesis")
NullProportion <- NullRatio/sum(NullRatio) # converting ratios to proportions
Observed <- c(38,35,28,30,65,68,60,67,138) # from Mendel's table
Expected <-sum(Observed)*NullProportion # converting proportions to Expected counts
Discrepancy <- (Expected-Observed)^2/Expected
TestStat <- sum(Discrepancy) # chi-squared test statistic
p_value <- 1-pchisq(TestStat,9-1) # don't forget the degrees of freedom!
ATable <- rbind(NullProportion,Expected,Observed,Discrepancy)
pander(ATable,
      col.names=Category,
      digits=3,
      caption="Observed vs. Expected values for Experiment 1")
Category <- c("ABC","ABc","AbC","Abc",
              "aBC","aBc","abC","abc",
              "ABCc","AbCc","aBCc","abCc",
              "ABbC","ABbc","aBbC","aBbc",
              "AaBC","AaBc","AabC","Aabc",
              "ABbCc","aBbCc",
              "AaBCc","AabCc",
              "AaBbC","AaBbc",
              "AbBbCc")
NullRatio <- c(1, 1, 1, 1,
               1, 1, 1, 1,
               2, 2, 2, 2,
               2, 2, 2, 2,
               2, 2, 2, 2,
               4, 4,
               4, 4,
               4, 4,
               8)
ATable <- rbind(Category,NullRatio)
panderOptions('table.continues', '')
pander(ATable, caption="Relative ratios under the Mendel's proposal")
```