Erick Castillo
Horsep Akopyan
Austin Venezia
Ryan Agustin.

Obs Val.

	white	not white	
part of jury	11	1	12
not part of jury	5	6	11
	16	7	23

Lab: Contingency Tables

1. Twenty-three people were called for jury duty, of which sixteen are white. From this panel of 23 individuals, a jury of twelve people was selected. The jury contains only one non-white person. Does this give statistically significant evidence of a bias against non-whites in the selection process?

(a) Use Pearson's chi-square test to address this question.

Exp. Val.

	white	not white	
part of jury	8.347	3.65	12
not part of jury	7.652	3.347	11
	16	7	23

$$\frac{(11-8.348)^2}{8.348} + \frac{(5-7.652)^2}{7.652} + \frac{(1-3.65)^2}{3.65} + \frac{(6-3.347)^2}{3.347}$$

↳ $0.8425 + 0.9191 + 1.923 + 2.103$

≈ 5.788

p-value: 0.016.

There is reasonably strong evidence to suggest that there was bias against selecting non-whites to be in the jury.

(b) Give two important drawbacks to using Pearson's test here.

- The sample size is too small in this case.
- This type of test relies on the X^2 distribution, which cannot consider a one-tail alternative.
- Some values in the expected value table are < 5.

(c) Use Fisher's exact test, based on the hypergeometric distribution.

H_0: No bias in selection process. Any differences are caused by randomization.

Calculate p-value where $P(W \geq Sel \geq 11)$

$$P(W \geq Sel \geq 11) * 2 = 2 * \left[\frac{\binom{16}{11}\binom{7}{1}}{\binom{23}{12}} + \frac{\binom{16}{12}\binom{7}{0}}{\binom{23}{12}} \right]$$

$= 0.04792$

This p-value indicates there is moderate evidence to suggest that there was bias in the creation of this jury.

2. From one of Mendel's many experiments in genetics involving pea plants, the following results were reported for 529 offspring of cross-bred genetically hybrid parents:

Observed values:

		Form (smooth or wrinkled)			
		AA	Aa	aa	
Color (yellow or green)	BB	38	60	28	126
	Bb	65	138	68	271
	bb	35	67	30	132
		138	265	126	529

(a) Test the hypothesis that form and color are independent characteristics. (Find the *p*-value and make a conclusion.)

Expected Vals:

	AA	Aa	aa
BB	$\frac{138 \cdot 126}{529} = 32.87$	63.12	30.01
Bb	70.70	135.76	64.55
bb	34.43	66.12	31.44

$$T.S. = \frac{(38-32.87)^2}{32.87} + \dots + \frac{(30-31.44)^2}{31.44} = 1.857$$

$$pchisq(T.S., 4, lower.tail = FALSE) \approx 0.762$$

There is strong evidence suggesting that the form and color of the leaves are independent.

(b) Test the hypothesis that form and color are independent characteristics and that the three genotypes of each factor occur in the ratios 1:2:1. (Find the *p*-value and make a conclusion.)

Recall $H_0: \pi_{ij} = \pi_{i\cdot}\pi_{\cdot j} \quad i \in \{1, 2, \dots, I\}, \; j \in \{1, 2, \dots, J\}$.

Given the above ratio, we can construct the expected probability contingency table

	AA	Aa	aa
BB	$\frac{1}{16}$ 33.0625	$\frac{1}{8}$ 66.125	$\frac{1}{16}$ 33.0625
Bb	$\frac{1}{8}$ 66.125	$\frac{1}{4}$ 132.25	$\frac{1}{8}$ 66.125
bb	$\frac{1}{16}$ 33.0625	$\frac{1}{8}$ 66.125	$\frac{1}{16}$ 33.0625

as seen on the left. The left values are the probability of landing in that cell. The values on the right are the expected values, obtained after multiplying the probabilities by 529.

Calculating the test statistic follows from before...

$X^2 = 2.8109$ w/ p-value: 0.58995.

So with this p-value, I can conclude that there is sufficient evidence to suggest that the characteristics are independent and preserve the 1:2:1 ratio.

(c) Mendel's pea experiments were conducted in the mid-1800's, before the development of inferential statistical methods such as those used in parts (a) and (b). Ronal Fisher, in a comprehensive 20th century review of all of Mendel's experiments, found a combined overall chi-square value of 42, on 84 degrees of freedom. Use a normal approximation to estimate the significant probability (p-value). Then use software to determine the p-value based on the exact chi-square distribution. Are they close? (Hint: Be careful here—it is common to make a mistake in determining these p-values correctly!)

$X^2 = 42$, $df = 84$

w/ software : 0.9999646

w/ Normal Approx.

recall the mean of the X^2 dist. is the df while the variance is 2 * df.

$$z = \frac{42 - 84}{\sqrt{2 \cdot 84}} = -3.24037. \Rightarrow 1 - \text{pnorm}(-3.24037) = 0.999564.$$ The values are close.

I would reach the same conclusion given either.

(d) Draw an appropriate conclusion about Mendel's experiments based on the answers you found in part (c).

Utilizing either p-value, I would conclude that there is very strong evidence to suggest that both phenotypes are independent from one another.