

BIOL 220 Problem Set 09: Hypothesis tests with a chi-squared null distribution

Answer Key

Due Thursday, March 14, 2024 before lecture

Submit your answers via Google Classroom

Working with the χ^2 goodness of fit test

Suppose you have data on the frequencies of three different groups (you can imagine your favorite kinds of things—three fish species, three types of candy, three health behaviors). You would like to know if the groups are equally represented in your data or not. Imagine this is what your data look like:

group A	group B	group C
12	11	13

💡 Use the above table to answer questions 1–3.

1. Under the null hypothesis of equal frequencies across groups, what are the expected frequencies for groups A, B, and C? [1 point]

i Answer

12, 12, 12

2. Calculate the χ^2 statistic for these data [1 point]

i Answer

```
obs <- c(12, 11, 13)
expected <- c(12, 12, 12)

chi2 <- sum(((obs - expected)^2) / expected)
chi2

[1] 0.1666667
```

3. What are the degrees of freedom for these data? [1 point]

i Answer

2

4. With the χ^2 statistic and degrees of freedom you just figured out, use `pchisq` to calculate the P -value of the null hypothesis [1 point]

i Answer

```
pchisq(chi2, 2, lower.tail = FALSE)

[1] 0.9200444
```

5. Do we reject the null hypothesis? [1 point]

i Answer

We fail to reject the null at $\alpha = 0.05$ because the P -value is greater than 0.05

Diabetes prevalence in Hawai i

[Grandinetti et al. \(2007\)](#) studied prevalence of diabetes in North Kōhala on the Island of Hawai i. They found significant disparities in diabetes prevalence across ethnic groups, with Native Hawaiian and Asian individuals at higher risk of diabetes. This increased risk could

not be explained by shared lifestyle risk factors, indicating that other, unmeasured, drivers were likely at play (side note from Andy: perhaps colonialism).

The table below summarizes the results of Grandinetti et al. (2007) with respect to diabetes prevalence by ethnic group.

	Caucasian	Filipino	Hawaiian	Japanese	Other/mixed non-Hawaiian
Yes diabetes	13	36	100	40	39
No diabetes	282	150	426	150	216

💡 Use the above table to answer questions 6–9.

6. To address the question of “are there differences in diabetes prevalence across ethnic groups?” what should our null hypothesis be? [1 point]

i Answer

H_0 : The per capita proportion of diabetes is the same across all ethnic groups

7. If you were to use a χ^2 distribution to represent this null hypothesis, what are the correct degrees of freedom? [1 point]

- 1
- 2
- 4
- 5
- 10

i Answer

$(\text{Number of rows} - 1) \times (\text{Number of columns} - 1) = 1 \times 4 = 4$

8. Use R code to make a matrix of the diabetes data. To answer this question, paste your R code into the google form [1 point]

i Answer

```
# as long as your R code makes the correct matrix, that's ok
# correct code will probably look something like

diabetes <- matrix(c(13, 282, 36, 150, 100, 426, 40, 150, 39, 216),
                  nrow = 2)

diabetes

      [,1] [,2] [,3] [,4] [,5]
[1,]   13   36  100   40   39
[2,]  282  150  426  150  216
```

9. Use `chisq.test` to conduct a χ^2 null hypothesis test. Report the test statistic and *P*-value in the google form [1 point]

i Answer

```
chisq.test(diabetes)

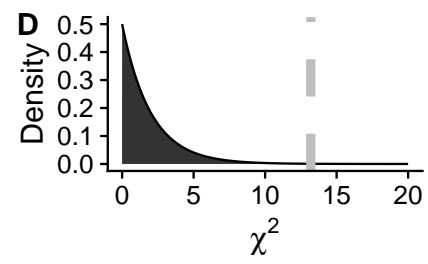
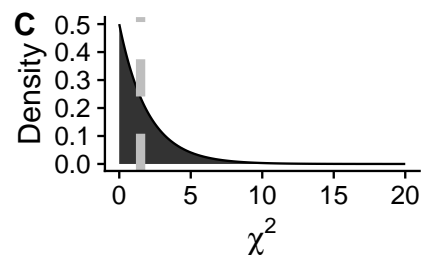
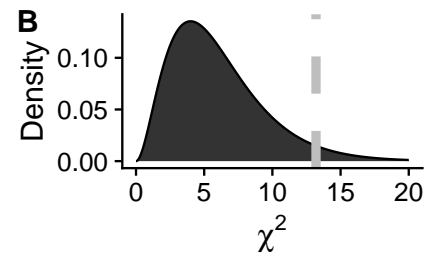
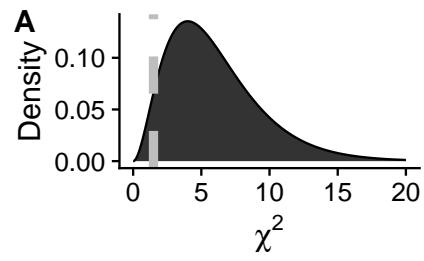
Pearson's Chi-squared test

data:  diabetes
X-squared = 38.802, df = 4, p-value = 7.654e-08
```

χ^2 distribution

💡 χ^2 distribution

10. Which of the following figures correctly shows a χ^2 sampling distribution with 6 degrees of freedom and a test statistic (vertical line) consistent with a *P*-value of 0.04. [1 point]



i Answer
B