

HW 8 – Math 58B

due Thursday, March 26, 2026

your name here

```
library(tidyverse)
library(mosaic)
library(praise)
library(infer)
```

Assignment Summary (Goals)

- chi-squared tests of independence

Q1. True or false, Part I

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.¹

- The chi-squared distribution, just like the normal distribution, has two parameters, mean and standard deviation.
- The chi-squared distribution is always right skewed, regardless of the value of the degrees of freedom parameter.
- The chi-squared statistic is always positive.
- As the degrees of freedom increases, the shape of the chi-squared distribution becomes more skewed.

Q2. True or false, Part II

Determine if the statements below are true or false. For each false statement, suggest an alternative wording to make it a true statement.²

- As the degrees of freedom increases, the mean of the chi-squared distribution increases.

¹From OpenIntro Statistics, exercise 3.37

²From OpenIntro Statistics, exercise 3.38

- (b) If you found $X^2 = 10$ with $df = 5$ you would fail to reject H_0 at the 5% discernability level. [Use `xpchisq()`.]
- (c) When finding the p-value of a chi-squared test, we always shade the tail areas in both the upper and the lower tails.

Q3. US Volunteerism I

The 2003 study on volunteerism conducted by the Bureau of Labor Statistics reported the sample percentages who performed volunteer work, broken down by many other variables. For example, respondents were categorized by age. The following reports the percentage of sample respondents in each age group who had performed volunteer work in the previous year:³

Age group	16–24 years	25–34 years	35–44 years	45–54 years	55–64 years	65 or more
% volunteer	21.9%	24.8%	34.1%	31.3%	27.5%	22.7%

- (a) Is the information (percentages) sufficient to construct a segmented bar graph for comparing the proportions of volunteers across the various age categories?
- (b) Explain why the information (the group percentages only) is **not** sufficient to conduct a chi-squared test of whether these sample proportions differ discernably across the age categories.

The sample sizes in each age group are not given in the report, but based on other information we can estimate them to be as follows:

Age group	16–24 years	25–34 years	35–44 years	45–54 years	55–64 years	65 or more
Sample size	9719	10613	12070	10959	7329	9310

- (c) Use the additional information to produce a table of counts with age groups in rows and volunteer status (yes or no) in columns. Make sure that you know how to create the 6x2 table of counts by hand. You can use the `table()` function in R.

³From ISCAM, HW 5.6

```

vol_data <- matrix(c(2128, 2632, 4116, 3430, 2015, 2113, 7591, 7981, 7954, 7529, 5314, 7197)
                  ncol=6, byrow=TRUE)

vol <- rep(rep(c("Vol", "noVol"), 6),
           times = vol_data)
year <- rep(rep(c("y1624", "y2534", "y3544", "y4554", "y5564", "y65plus"), each = 2),
           times = vol_data)

volunteer <- data.frame(vol, year)

```

- (d) Create a segmented barplot. The x-axis should be the age (look at the data frame to see what the variables are called!), the fill should be the information about whether or not each person volunteers

The R code will look something like:

```

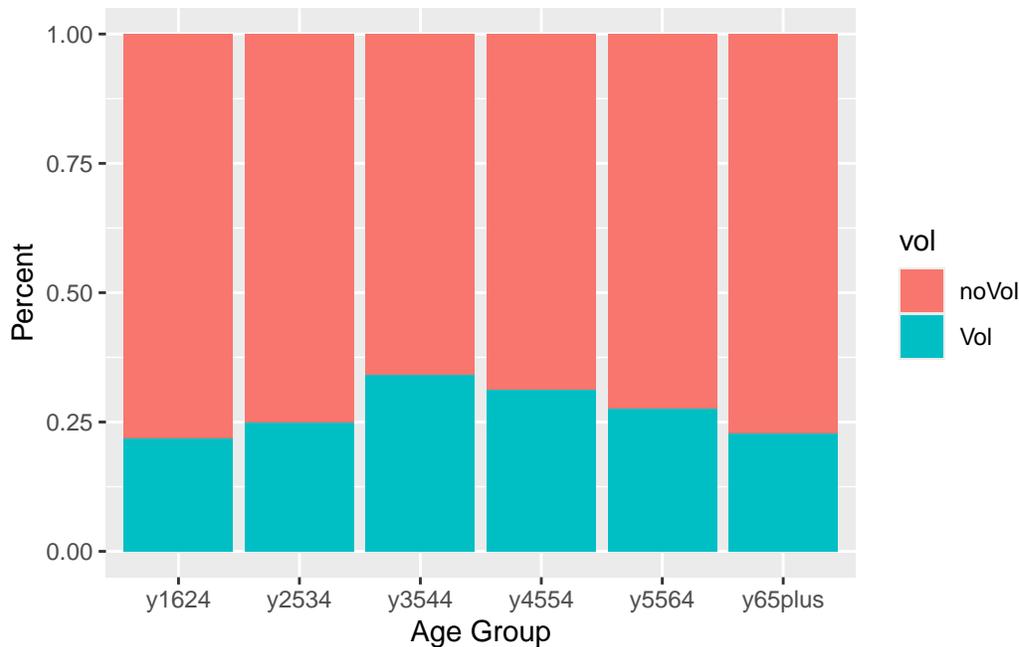
ggplot(____) +
  geom_bar(aes(x = ____, fill = _____), position = "fill") +
  labs(x = "Age Group",
       y = "Percent")

```

```

ggplot(volunteer) +
  geom_bar(aes(x = year, fill = vol), position = "fill") +
  labs(x = "Age Group",
       y = "Percent")

```



- (e) Conduct the chi-squared test. Report the hypotheses, check of technical conditions (for the distribution to work out, we need at least 5 observations in each cell of the table), sampling distribution, chi-squared test statistic, and chi-squared p-value. (Provide the details of your calculations and/or relevant computer output.) Summarize your conclusion.

The R code for running a chi-squared test will look something like:

```
your_data |>
  table() |>
  chisq.test()
```

- (f) Construct a 6×2 table with the same row and column headings as in (c), but containing only + and - signs indicating whether the observed count is larger (+) or smaller (-) than expected in that cell. Does the table reveal a pattern? Explain what that pattern suggests about the relationship between age group and volunteerism.

Note that if we keep (using <-) the output of `chisq.test`, we can pull out the observed and expected tables from the output. Try the following: `name_of_your_X2_output$observed` and `name_of_your_X2_output$expected`.

- (g) Run a randomization test which uses the same test statistic (X^2) but does not use the chi-squared probability (mathematical) distribution to find a p-value. Note that the change from “chi-squared test” to “randomization test” is **not** the statistic you calculate

(the statistic is the same as the full chi-squared test above!). The difference is how the p-value is calculated (mathematical function vs. computer simulation). Ask if that doesn't make sense! The data are permuted (shuffled) to find the sampling distribution of the X^2 statistic assuming H_0 is true.

Q4. US Volunteerism II

Reconsider the previous question about volunteerism. Suppose that the sample sizes had all been smaller by a factor of 100 (so that the entire study included only about 600 subjects) but that the conditional proportions of volunteerism within each age group had all turned out the same.⁴

- (a) How (if at all) would you expect the segmented bar graph to change? Explain.
- (b) How (if at all) would you expect the test statistic to change? Explain.
- (c) How (if at all) would you expect the p-value to change? Explain.
- (d) How (if at all) would you expect your conclusion to change? Explain.
- (e) Repeat the mathematical chi-squared analysis (no need to repeat the randomization test) with the greatly reduced sample size (round the observed counts in the new table to the nearest integer). Confirm or correct your answers to (b)–(d) in light of your analysis.

```
praise()
```

```
[1] "You are posh!"
```

⁴From ISCAM, HW 5.7