Examples

Alan T. Arnholt

11/13/2018

Contents

1 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. Consider Table 1.

2 Including Plots

You can also embed plots such as Figure 1.

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.

3 Inline `R` Code

Table 1: First six rows of 'mtcars'

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

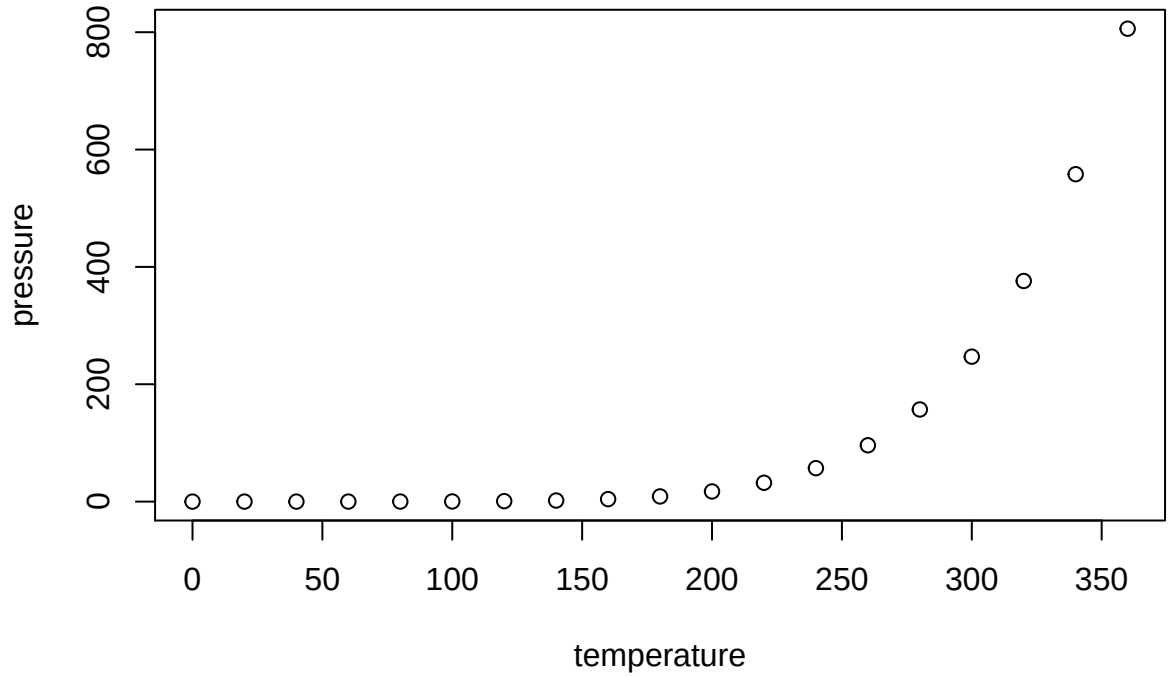


Figure 1: Scatterplot of ‘pressure‘ versus ‘temperature‘

```
mtcars$cyl <- factor(mtcars$cyl)
library(tidyverse)
SI <- mtcars %>%
  group_by(cyl) %>%
  summarize(Mean = mean(qsec), SD = sd(qsec))
SI
```

```
# A tibble: 3 x 3
  cyl    Mean    SD
  <fct> <dbl> <dbl>
1 4      19.1  1.68
2 6      18.0  1.71
3 8      16.8  1.20
```

The mean `qsec` time for six cylinder cars is 17.98 seconds.

4 Chi-Square Test?

Two different professors teach an introductory statistics course. Table 2 shows the distribution of final grades they reported. We wonder whether one of these professors is an “easier” grader.

- Will you test goodness-of-fit, **homogeneity**, or independence?
- Write the appropriate hypotheses.

H_0 : The distribution of grades is the same for the two professors. H_A : The distribution of grades is different for the two professors.

- Find the expected counts for each cell.

Table 2: Grades for Alpha and Beta

	Alpha	Beta
A	3	10
B	11	11
C	14	8
D	10	1
F	4	0

```
chisq.test(tt)$exp
```

```
     Professor
Grade     Alpha     Beta
    A  7.583333 5.416667
    B 12.833333 9.166667
    C 12.833333 9.166667
    D  6.416667 4.583333
    F  2.333333 1.666667
```

- Explain why the chi-square procedures are not appropriate for this table. **Since some of the cells have expected counts less than 5, the chi-square procedures are not appropriate.**
- Solution—Permutation

4.1 Make `tt` tidy

```
ttt <- tt %>%
  broom::tidy() %>%
  uncount(n)
ttt
```

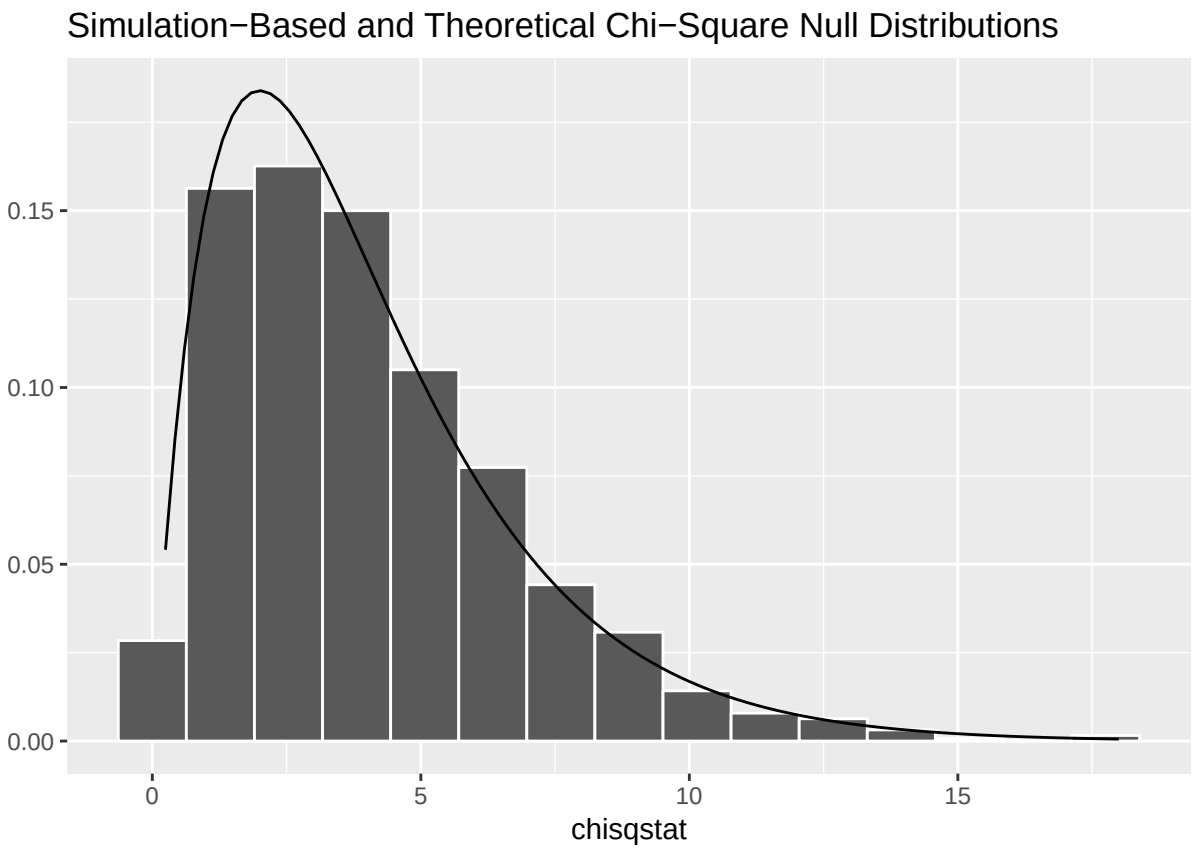
```
# A tibble: 72 x 2
   Grade Professor
   <chr> <chr>
 1 A     Alpha
 2 A     Alpha
 3 A     Alpha
 4 B     Alpha
 5 B     Alpha
 6 B     Alpha
 7 B     Alpha
 8 B     Alpha
 9 B     Alpha
10 B     Alpha
# ... with 62 more rows
```

4.2 Using `infer`

```
T1 <- xtabs(~Grade + Professor, data = ttt)
obs_stat <- chisq.test(T1)$stat
obs_stat
```

```
X-squared
 15.19121
```

```
library(infer)
null <- ttt %>%
  specify(Grade ~ Professor) %>%
  hypothesize(null = "independence") %>%
  generate(reps = 999, type = "permute") %>%
  calculate(stat = "Chisq")
visualize(null, method = "both")
```




```
get_pvalue(null, obs_stat, direction = "right")
```

```
# A tibble: 1 x 1
  p_value
    <dbl>
1 0.00200
```

```
(pvalue <- (sum(null$stat >= obs_stat) + 1)/(999 + 1))
```

```
[1] 0.003
```

4.3 Using a `for()` loop

Figure 2 shows the theoretical χ^2_4 distribution in blue, while the permutation distribution from using computation is shown as the pink density.

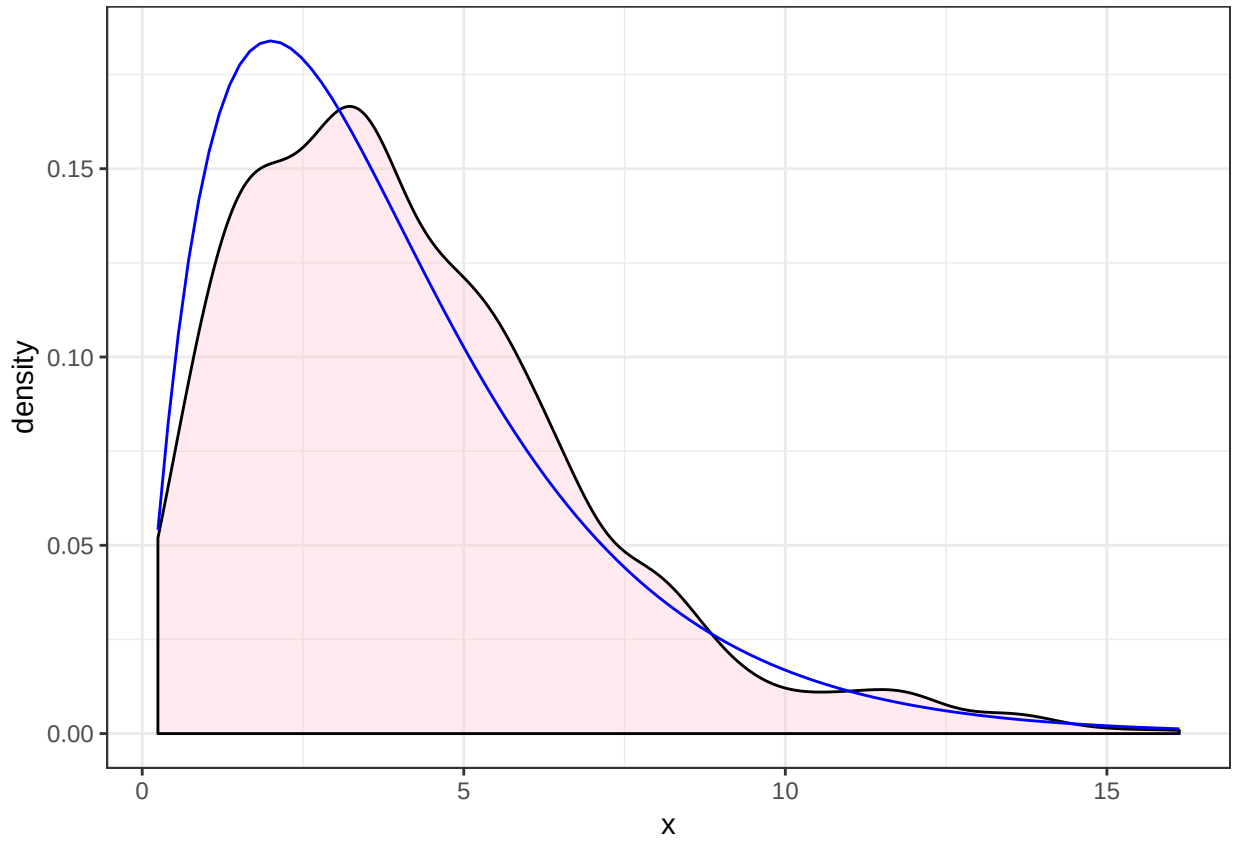


Figure 2: Theoretical and computation null distribution

```
sims <- 10^3 - 1
x2 <- numeric(sims)
for(i in 1:sims){
  TT <- xtabs(~Grade + sample(Professor), data = ttt)
  x2[i] <- chisq.test(TT)$stat
}
pvalue <- (sum(x2 >= obs_stat) + 1)/(sims + 1)
pvalue
```

```
[1] 0.002
```

```
DF <- data.frame(x = x2)
ggplot(data = DF, aes(x = x)) +
  geom_density(fill = "pink", alpha = 0.3) +
  theme_bw() +
  stat_function(fun = dchisq, args = list(df = 4), color = "blue")
```