Tutorial 1

Applied Stats/Quant Methods 1

Week 3, Fall 2021

1 Conditional Probabilities

(From DeGroot, p.145) Each student in a certain high school was classified according to her year in school (freshman, sophomore, junior, or senior) and according to the number of times that she had visited a certain museum (never, once, or more than once). The proportions of students in the various classifications are given in the following table:

	Never	Once	More than once
Freshmen	0.08	0.10	0.04
Sophomores	0.04	0.10	0.04
Juniors	0.04	0.20	0.09
Seniors	0.02	0.15	0.10

(a) If a student selected at random from the high school is a junior, what is the probability that she has never visited the museum?

(b) If a student selected at random from the high school has visited the museum three times, what is the probability that she is a senior?

(a) If a student selected at random from the high school is a junior, what is the probability that she has never visited the museum?

(b) If a student selected at random from the high school has visited the museum three times, what is the probability that she is a senior?

2 Joint and Conditional Probabilities

(Adapted from DeGroot, p.146) Suppose that in the population of US college students the joint distribution of test scores Y on mathematical and musical aptitudes is bivariate normal with the following parameters:

$$\mathbf{Y} \sim \mathcal{N}\left(\begin{bmatrix}350\\300\end{bmatrix}, \quad \begin{bmatrix}20 & 15\\15 & 25\end{bmatrix}\right)$$

Use R to draw 2,500 simulations from this joint probability density function and answer the following questions (use the `mvrnorm` function in the `MASS` library):

(a) Approximately what proportion of college students obtain a score greater than 355 on the mathematics tests? (The mean math score is 350)

(b) If a student's score on the music test is between 290 and 295, what is the approximated probability that his score on the mathematics test will be greater than 350?

(c) If a student's score on the mathematics test is lower than 340, what is her expected test score on the music test?

The following code simulates a population of 2500 colleges students and creates a dataset named `simulation`. The first column (math) of the dataset gives the math scores and the second column (music) gives the music scores of students.

```
library(MASS)
means <- c(350, 300)
covariance <- matrix(c(20,15,15,25), ncol=2)
simulation <- data.frame(mvrnorm(2500, means, covariance))
names(simulation) <- c("math", "music")
```

(a) Approximately what proportion of college students obtain a score greater than 355 on the mathematics tests? (The mean math score is 350).

(b) If a student's score on the music test is between 290 and 295, what is the approximated probability that his score on the mathematics test will be greater than 350?

(c) If a student's score on the mathematics test is lower than 340, what is her expected test score on the music test?

3 Cumulative Distribution Functions

(From Maindonald and Braun, p. 99) The function `pexp(x, rate=r)` can be used to compute the probability that an exponential variable is less than `x`. Suppose the time betwen accidents at an intersection can be modeled by an exponential distribution with a rate of 0.05 per day. Find the probability that the next accident will occur during the next three weeks.

4 Central Limit Theorem

(From Maindonald & Brown, 2010, p.99) Use R to generate a random sample of size 100 for variable Y from a normal distribution.

(a) Calculate the mean and standard deviation of Y

(b) Use a loop to repeat the above calculation 50 times. Store the 50 means in a vector named `av`. Calculate the standard deviation of the values of `av`.

(c) Create a function that performs the calculations described in (b). Run the function a few times, and plot one of the distributions of 50 means in a density plot. What form does this density have?

(d) Change the underlying distribution of Y to a χ^2 distribution (function `rchisq` in `R`) and repeat steps (a) through (c). What is the form of the density now? Explain how this exercise relates to the central limit theorem.