

## Homework 5: Inference and Multivariate Statistics

For all questions below, you should show all work needed to reach your answer. You may collaborate with your classmates and consult external resources, but you should write and submit your own answer. **Any classmates with whom you collaborate should be credited at the top of your submission. Similarly, if you consult any external references, you should cite them clearly and explicitly.**

1. Suppose  $\mathbf{X}$  and  $\mathbf{Y}$  are independent  $N(0, 1)$  random variables. Derive the distribution of the random vector  $\begin{bmatrix} X + Y \\ X - Y \end{bmatrix}$ .
2. Suppose  $X_1, \dots, X_n, X_{n+1}$  are identically and independently distributed  $N(\mu, 1)$  random variables.
  - (a) Derive the conditional distribution of  $X_{n+1} \mid X_1, \dots, X_n$ .
  - (b) Suppose  $X_1, \dots, X_n$  are unobserved, but the sample mean  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  is observed. Derive the conditional distribution of  $X_{n+1} \mid \bar{X}_n$ .
  - (c) Suppose instead that for all  $i \neq j$ , the correlation between  $X_i$  and  $X_j$  is  $\rho$  for some  $0 < \rho < 1$ . Derive the conditional distribution of  $X_{n+1} \mid \bar{X}_n$ .
3. In your own words, explain why using factor analysis to draw causal conclusions may be incorrect. Give an example to illustrate the issue.
4. Complete the **Factor Analysis R Lab** and submit via Gradescope.