1. One of the most used and useful consequences of the material you are learning is the two-sample t-test. You will now flesh out a procedure for applying this test. First, suppose you have two independent samples:

$$\begin{aligned} X_1, \ldots, X_n &\stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2) \\ Y_1, \ldots, Y_m &\stackrel{\text{iid}}{\sim} N(\mu_2, \sigma_2^2). \end{aligned}$$

   (a) If the population variances σ_1^2 and σ_2^2 are known, then derive the sampling distribution of $\bar{X} - \bar{Y}$. Propose a Z test (where the statistic Z has a standard normal distribution under the null hypothesis) to test $H_0 : \mu_1 = \mu_2$. Argue that the assumption of normal sampling distributions is not required to apply this test. What is?

   (b) The form of the t-test when the population variances are not known, depends on whether we can assume these variances are equal. Describe a statistical test (null hypothesis, statistic, and critical value) for drawing a conclusion about the variances.

   (c) Suppose in part 1b, you conclude $H_0 : \sigma_1^2 = \sigma_2^2$ is true. Derive an unbiased estimate for the common variance $\sigma^2 = \sigma_1^2 = \sigma_2^2$ and use it to formulate a t-statistic for testing $H_0 : \mu_1 = \mu_2$. State the number of degrees of freedom.

   (d) Suppose in part 1b, you conclude $H_A : \sigma_1^2 \neq \sigma_2^2$ is true. The distribution of statistic

$$t_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}},$$

   where S_X^2 and S_Y^2 are the sample variances, is well approximated by a t-distribution with degrees of freedom

$$v = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{(S_X^2/n)^2}{n-1} + \frac{(S_Y^2/m)^2}{m-1}}.$$

   (The last part is a fact; I'm not asking you to prove it.) Use the preceding procedure to test for differences in the following two samples:

```
> x
 [1] 20.624691 10.248618 20.922304 14.169417 15.008047 15.016957  6.846616
 [8] 15.362315 16.434458 16.399544 20.876594 21.719867  9.442184 12.534024
[15]  6.579588 13.197646  9.370132 18.868558 14.850241 13.073043
> y
 [1] 12.565221  1.147250  2.491917  2.487184 10.555273  9.063544 -9.520266
 [8]  7.059711 21.043265  1.842692  8.598662 18.550844 10.327906 33.767452
[15]  9.875084 -4.741950 10.345144  8.932715
```

2. Suppose $X_i \sim N(i, i^2), i = 1, 2, 3$ are independent. Use the X_is to construct statistics with the indicated distributions

   (a) chi squared with 3 degrees of freedom

   (b) t distribution with 2 degrees of freedom

   (c) F distribution with 1 and 2 degrees of freedom

3. Let $X \sim F_{p,q}$, then

   (a) Derive the pdf of X.

   (b) Compute the mean and variance of X, showing your work.

   (c) Show that $\frac{1}{X}$ has $F_{q,p}$ distribution.

4. In this problem, you will prove independence of S^2 and $\bar{X}$ when $X_j \sim N(\mu, \sigma^2)$ are iid. We will do so by working with the following transformation of X_j. Let $a_{ij}, b_{rj}, j = 1, \dots, n; i = 1, \dots, k; r = 1, \dots, m$ be constants and define

$$
\begin{aligned}
U_i &= \sum_{j=1}^{n} a_{ij} X_j, \qquad && i = 1, \dots, k, \\
V_r &= \sum_{j=1}^{n} b_{rj} X_j, \qquad && r = 1, \dots, m.
\end{aligned}
$$

   (a) Show $\text{Cov}(U_i, V_r) = \sigma^2 \sum_j a_{ij} b_{rj}$.

   (b) For the case $n = 2$, use change of variable to show that $\text{Cov}(U_i, V_r) = 0$ implies independence of U_i and V_r. (**Note:** While independence always implies 0 covariance, the converse is not always true, so the fact that the converse is true here is a special result for linear functions of Normal random variables.) **Hint:** Note that transform $Z_j = \frac{X_j - \mu}{\sigma}$ allows us to translate results obtained for Z_j to X_j, so it is sufficient to prove this result for Z_1 and Z_2. Furthermore, recall that two random variables are independent iff their joint pdf factors into a product of marginals.

   (c) The result of 4b extends to $n > 2$, so assuming U_i and V_r are independent for arbitrary integer n, show that $\bar{X}$ is independent of $X_j - \bar{X}$ for all j.

   (d) Write S^2 as a function of $X_j - \bar{X}$ and use (don't prove) the result that these random vectors $(U_1, \dots, U_k)$ and $(V_1, \dots, V_m)$ are independent iff U_i and V_r are independent for all pairs (i, r) (can be shown through a proof similar to the one of 4b), to show that $\bar{X}$ and S^2 are independent statistics.