1. (a)

$$\begin{aligned} P(X < x) &= P(F^{-1}(U) < x) \\ &= P(U < F(x)) \\ &= F(x) \end{aligned}$$

So $X \sim F$.

(b) For $Z = X - Y$:

$$\begin{aligned} P(Z = z) &= P(X - Y = z) \\ &= \int_0^1 P(X = z + y \,|\, Y = y) P(Y = y) dy \\ &= \int_0^1 I(0 < z + y < 1) dy \\ &= \int_0^1 I(-z < y < 1 - z) dy \\ &= \int_{\max\{0, -z\}}^{\min\{1, 1-z\}} dy \\ &= \max\{0, |1 - z|\} \end{aligned}$$

For $Z = \min\{X, Y\}$:

$$\begin{aligned} P(Z < z) &= P(Z < z \,|\, Y \geq z) P(Y \geq z) + P(Z < z \,|\, Y < z) P(Y < z) \\ &= P(X < z \,|\, Y \geq z) P(Y \geq z) + P(Y < z) \\ &= I(0 < z < 1)(2z - z^2) \end{aligned}$$

(c)

$$\begin{aligned} \mathrm{E}(Y) &= \mathrm{E}(e^X) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{x - \frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-1)^2 + \frac{1}{2}} dx \\ &= e^{\frac{1}{2}} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-1)^2} dx \\ &= e^{\frac{1}{2}} \end{aligned}$$

$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{V}(e^X) \\ &= \mathrm{E}((e^X)^2) - \mathrm{E}(e^X)^2 \\ &= \mathrm{E}(e^{2X}) - e \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{2x - \frac{x^2}{2}} dx - e \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-2)^2 + 2} dx - e \\ &= e^2 \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{1}{2}(x-2)^2} dx - e \\ &= e^2 - e \end{aligned}$$

2. (a) We can calculate the sum of squares for estimate $\hat{\beta}$ as

$$\|Y - X\hat{\beta}\|^2 = Y'Y - 2Y'X\hat{\beta} - \hat{\beta}'X'X\hat{\beta}$$

Since this is a convex function, we will find the minimum when the gradient with respect to $\hat{\beta}$ is 0:

$$\begin{aligned} \nabla_{\hat{\beta}}(Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta}) &= -2X'Y + 2X'X\hat{\beta} \\ 0 &= -2X'Y + 2X'X\hat{\beta} \\ X'Y &= X'X\hat{\beta} \\ (X'X)^{-1}X'Y &= \hat{\beta} \end{aligned}$$

So $\hat{Y} = X\hat{\beta} = X(X'X)^{-1}X'Y = HY$ is the least squares estimate.

(b)

$$HX = X(X'X)^{-1}X'X = X$$

(c)

$$\begin{aligned} H' &= (X(X'X)^{-1}X')' \\ &= X(X(X'X)^{-1})' \\ &= X((X'X)^{-1})'X' \\ &= X((X'X)')^{-1}X' \\ &= X(X'X)^{-1}X' \\ &= H \end{aligned}$$

(d)

$$H^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H$$

(e) Since HY is a linear combination of the columns of X, it must be in the column space of X. Further, since

$$X'HY = X'X(X'X)^{-1}X'Y = X'Y$$

We can conclude it is actually the projection of Y into this space.

(f) Using the SVD of $X = U\Sigma V$:

$$\begin{aligned} rank(X) &= rank(U) \\ &= rank(UU') \\ &= rank(H) \\ &= tr(H) \\ &= d \end{aligned}$$

The last step is since the rank of an idempotent matrix (which H is) is its trace.

3.
   1. There are 30 rolls in the data set
   2. The data is stored in unicode format, so the '+' operator does a string join instead of addition. The end result is a string with all the numbers in order.
   3. The sum this time is correct and gives us a total of 92
   4. Running this on my home computer, the first time took .00472 seconds, and the second .00352
   5. I didn't do the hello world part because I can't get on AWS for the moment, but the reduce is the same as count()