Below are the three final problems in the Probability course.
All of them require analytical solution, but you can submit them in digital or in handwritten form. In any case, you should clearly explain all steps of your solution.
Unlike the previous home assignments, this one should be submitted not in pairs, but **individually**.

Prediction [35 points]
In a study of patients who had undergone surgery of breast cancer, 81 out of 306 have died within 5 years (Haberman, 1976). These 81 patients had on average 7.5 positive axillary nodes detected, while the 225 patients who survived 5 years or longer had on average 2.8 nodes detected.
Let us assume that within each group the number of positive axillary nodes follows a geometric distribution.
Estimate probability of survival within 5 years for a patient with 2 positive axillary nodes detected.

Likelihood [35 points]
Let $X_1, \dots, X_n$ be an i.i.d. sample from a distribution with density function

$$f(x,\theta) = \begin{cases} \frac{2x}{\theta} e^{-\frac{x^2}{\theta}}, if\ x > 0 \\ 0, if\ x \le 0 \end{cases}$$

where $\theta > 0$ is an unknown parameter.
1. Use the method of maximum likelihood to express a formula for an estimator of θ.
2. Apply this formula to estimate θ from the sample $(0.5, 0.5, 1)$.

Hypothesis [35 points]
Let $X_1, \dots, X_n$ and $Y_1, \dots Y_m$ be two independent samples from Bernoulli distribution (1 or 0) with probability of 1 (success) being p_X and p_Y respectively.
You have 100 observations of X with 60 successes in total and 150 observations of Y with 70 successes in total.
With 1% significance level test the null hypothesis $p_X = p_Y$ vs the alternative $p_X > p_Y$.
P.S. It is OK to use normal approximation here, since sample sizes 100 and 150 are “large enough”.