Problem 1:

Consider the following linear regression model

$$y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\boldsymbol{y} \in \mathbb{R}^n$ is an n-dimensional response vector, $\mathbf{X} \in \mathbb{R}^{n \times p}$ is an $n \times p$-dimensional matrix ($p < n$) for explanatory variables, $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ is the random vector, and $\mathbf{W}$ is a known matrix. It is known that the MLE of $\boldsymbol{\beta}$ is $\boldsymbol{\beta}\hat{} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$, the distribution of $\boldsymbol{y}$ is

$$\boldsymbol{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}),$$

and the residual vector is $\hat{\boldsymbol{\epsilon}} = [\mathbf{I} - \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}]\boldsymbol{y}$. Suppose that $\text{rank}(X) = p$ such that the inverse of $\mathbf{X}^\top\mathbf{X}$ exists. Let $\boldsymbol{z} = \mathbf{U}\boldsymbol{y}$.

1- Show that $\mathbf{z}$ and $\boldsymbol{\beta}\hat{}$ are independent if and only if $\mathbf{UX} = \mathbf{0}$.
2- Assume UX = 0. Show that if U is an n × n projection matrix then ||z||^2
follows the χ2-distribution with tr(U) degrees of freedom.

Problem 2:

Fill in the following blanks. Please find the answer of the problems and fill it in the blanks.

(a) Suppose $X_1 \sim N(3,4)$, $X_2 \sim N(2,6)$ and $X_3 \sim N(1,7)$. Let $Y = X_1 + X_2 - 3X_3$. Then, $E(Y) =$ __________, $V(Y) =$ __________, $Y \sim$ __________, and $P(Y > 5) =$ __________.

(b) Suppose we flipped a die 1000 times, and the counts of 1, 2, 3, 4, 5 and 6 are 160, 170, 155, 163, 180, 172. Assume we want to test whether the die is balanced. Then, the value of the Pearson χ^2 statistic is __________. Under H_0 it approximately follows __________ distribution. Based on this data, we __________ H_0.

(c) Assume $x_1, \cdots, x_{100}$ are observations from the Poisson distribution with mean θ. Suppose we have $\bar{x} = 3.5$. Assume we want to test $H_0 : \theta = 3$ versus $H_1 : \theta \neq 3$. Then, the value of $-2\log\Lambda$ is __________. Under H_0, the test statistic approximately follows __________ distribution. Based on this data, we __________ H_0.

(d) Suppose $X_1, \cdots, X_n$ are iid with common PDF $f_\theta(x) = \theta/(1+x)^{\theta+1}$ for $x > 0$, where $\theta > 0$ is an unknown parameter. The sufficient statistic is $T =$ ______________, the MLE is $\hat{\theta} =$ __________, and the Fisher information is $I(\theta) =$ __________. Let θ_0 be the true parameter. The approximate distribution of $\hat{\theta}$ is $\sqrt{n}(\hat{\theta} - \theta_0) \sim$ ______________.

(e) Suppose $X_1, \cdots, X_n$ are iid with common PDF $f_\theta(x)$, where $f_\theta(x) > 0$ for all $x > 0$ and $f_\theta(x) = 0$ for $x < 0$. Let $\hat{\theta}$ be the MLE and θ_0 be the true parameter. Assume $I(\theta) = \theta^2$. Then, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{D}$ __________, and $\sqrt{n}(\hat{\theta}^2 - \theta_0^2) \xrightarrow{D}$ __________. If we observed $\hat{\theta} = 4$, then the 95% confidence interval for θ is approximately __________________.