1 Linear and non-linear regression

Problem 1. *You are a mechanical locksmith and you are trying to find out how the shaft machining error is related to the machine tool parameter setting. You have compiled a multivariate linear model. The model expresses the relationship between the production error (the difference between the ideal shaft diameter and the actual shaft diameter, ProdError) and the setting of ten different continuous machine parameters (P1-P10). Below is the output you received:*

```
summary(lm(ProdError ~ P1+P2+P3+P4+P5+P6+P7+P8+P9+P10),data=d)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.05270    0.09576  -0.550   0.5835
X1           0.01298    0.08924   0.145   0.8847
X2           0.01596    0.10939   0.146   0.8843
X3          -0.02865    0.09079  -0.316   0.7531
X4           0.04611    0.09548   0.483   0.6303
X5           0.14151    0.09343   1.515   0.1334
X6          -0.02375    0.10277  -0.231   0.8178
X7           0.25522    0.10516   2.427   0.0172 *
X8           0.06672    0.08972   0.744   0.4590
X9           0.09949    0.10171   0.978   0.3306
X10         -0.04003    0.09317  -0.430   0.6685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9039 on 89 degrees of freedom
Multiple R-squared:  0.1145,        Adjusted R-squared:  0.01502
F-statistic: 1.151 on 10 and 89 DF,  p-value: 0.3346
```

(a) Decide whether at least one of the machine parameters (independent variables) is useful for estimating a manufacturing error (ProdError). In other words, formally decide whether you can decline $H_0 : \beta_1 = \beta_2 = \cdots = \beta_{10} = 0$. Justify correctly.

(b) Let us compare the full model constructed above with the intercept model and with the model that employs only the variable $P7$ identified as the most relevant. Let us compare them with F-test through an ANOVA run. Interpret the ANOVA table below.

```
lm.const<-lm(ProdError ~ 1,data=d) # the intercept model
lm.sel<-lm(ProdError ~ P7,data=d) # the P7 model
anova(lm.const,lm.sel,lm.all)
Analysis of Variance Table

  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     99 82.114
2     98 76.076  1    6.0384 7.3911 0.007879 **
3     89 72.711  9    3.3647 0.4576 0.899016
```