1. Consider a complete factorial experiment involving two factors, the first of which is at three levels and the second of which is at two levels. The experimental data are as follows:

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
<th>Time parameter (n)</th>
<th>Number of events (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>0</td>
<td>-1</td>
<td>50</td>
<td>75</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>25</td>
<td>40</td>
</tr>
</tbody>
</table>

Note that the number of events, Y, is a Poisson response variable; it has the Poisson distribution with mean $n\lambda(x)$. Now we consider the corresponding Poisson linear experimental model (Poisson regression model), in which $G = G_1 + G_2$, where $G_1$ is the space of quadratic polynomials in $x_1$ and $G_2$ is the space of linear functions of $x_2$.

(a) Determine the dimensions of $G_1$, $G_2$ and $G$.
(b) Determine which of these spaces are identifiable.
(c) Determine which of these spaces are saturated.
(d) Consider the basis $1, x_1, x_1^2 - 2/3$ of $G_1$. Determine the corresponding maximum likelihood estimate in $G_1$ of the regression coefficients. Hint: Observe that if the column vectors of a square matrix $X$ are suitably orthogonal and nonzero then $X^T X$ is an invertible square matrix $D$ and hence $X^{-1} = D^{-1} X^T$.
(e) Consider the basis $1, x_1, x_1^2 - 2/3, x_2$ of $G$. Determine the corresponding maximum likelihood equations for the regression coefficients.
(f) The actual numerical values of the maximum likelihood estimates of these coefficients are given by $\hat{\beta}_0 = 0.211$, $\hat{\beta}_1 = 0.121$, $\hat{\beta}_2 = -0.007$, and $\hat{\beta}_3 = -0.121$. Determine the corresponding variance and use it to test the goodness-of-fit of the model.
(g) Under the assumption that the Poisson regression function is in $G$, carry out the likelihood ratio test of the hypothesis that this function is in $G_1$.

2. Let $Y_1, Y_2$, and $Y_3$ be independent random variables such that $Y_1$ has the binomial distribution with parameters $n_1=400$ and $\pi_1$, $Y_2$ has the binomial distribution with parameters $n_2=600$ and $\pi_2$, and $Y_3$ has the binomial distribution with parameters $n_3=800$.
and \( \pi_3 \). Consider the logistic regression model
\[
\logit(\pi_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2
\]
for \( k = 1, 2, 3 \), where \( x_1 = -1, x_2 = 0 \), and \( x_3 = 1 \). Suppose that the observed values of \( Y_1, Y_2, \) and \( Y_3 \) are 60, 150, and 160 respectively.

a) Determine the design matrix and its inverse.

b) Determine the maximum likelihood estimates of \( \beta_0, \beta_1, \) and \( \beta_2 \).

c) Determine a 95% confidence interval for \( \beta_1 \).

d) Carry out the Wald test of the hypothesis that \( \beta_1 = 0 \).

Consider the submodel \( \beta_1 = 0 \) or equivalently, that \( \logit(\pi_k) = \beta_0 + \beta_2 x_2^2 \) for \( k = 1, 2, 3 \).

e) Determine the maximum likelihood estimates of \( \beta_0 \) and \( \beta_2 \) under the submodel.

Hints: explain why the data for the first and third design points can be merged.

f) Carry out the likelihood ratio test of the submodel.

3. a) Use least-squares approximation, as treated in the textbook, to determine a decent quadratic approximation to the function \( \sqrt{x} \) in the interval \( 1 \leq x \leq 2 \).

b) Determine the maximum magnitude of the error of approximation (or, at what point between 1 and 2, you observe the maximum error of approximation)

4. Consider the two-sample model, in which \( Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2} \) are independent random variables, \( Y_{11}, \ldots, Y_{1n_1} \) are identically distributed with mean \( \mu_1 \) and finite, positive standard deviation \( \sigma_1 \), and \( Y_{21}, \ldots, Y_{2n_2} \) are identically distributed with mean \( \mu_2 \) and finite, positive standard deviation \( \sigma_2 \). Here \( n_1 \) and \( n_2 \) are known positive integers, and \( \mu_1, \sigma_1, \mu_2 \) and \( \sigma_2 \) are unknown parameters. DO NOT assume that the indicated random variable are normally distributed or that \( \sigma_1 = \sigma_2 \). Let \( \overline{Y}_1, S_1, \overline{Y}_2 \) and \( S_2 \) denote the corresponding sample means and sample standard deviations as usually defined.

a) Derive a reasonable formula for confidence intervals for the parameter \( \mu_1 - \mu_2 \) that
is approximately valid when \( n_1 \) and \( n_2 \) are large.

b) Derive a reasonable formula for the \( P \)-value for a test of the hypothesis that the parameter in a) equals zero.

c) Calculate the numerical value of the 95\% confidence interval in a) when 
\[ n_1 = 50, \overline{Y}_1 = 18, S_1 = 20, n_2 = 100, \overline{Y}_2 = 12, S_2 = 10. \]
d) Determine the \( P \)-value for the test in b) when \( n_1 \) and so forth are as in c).

e) Derive a reasonable formula for confidence intervals for the parameter \( \frac{\mu_1}{\mu_2} \) that is 
approximately valid when \( n_1 \) and \( n_2 \) are large and \( \mu_2 \neq 0 \).

f) Explain why \( \frac{\sigma_2^2}{\sqrt{n_2} \mu_2} \approx 0 \) is required for the validity of the confidence intervals in e).

g) Derive a reasonable formula for the \( P \)-value for a test of the hypothesis that the parameter in e) equals 1.

h) Calculate the numerical value of the 95\% confidence interval in e).

i) Determine the \( P \)-value for the test in g).

j) Discuss the relationships of your answers to c), d), h) and i).