1. Find the likelihood ratio test for testing $H_0: \mu = \mu_0$ against $H_1: \mu = \mu_1$, based on a random sample of size $n$ from $N(\mu, \sigma^2)$. Suppose, $H_0$ will be rejected if the likelihood ratio is less than 1. Find the rejection region in terms of the sample mean.

(20)

The likelihood function is

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2}$$

The log-likelihood is

$$\ell(\mu, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

$$\frac{\partial \ell}{\partial \mu} = \frac{n}{2\sigma^2} - \frac{1}{\sigma^2} \sum (x_i - \mu) = 0$$

$$\hat{\mu} = \frac{1}{n} \sum (x_i - \mu) = -\frac{1}{2\sigma^2} \sum (x_i - \mu)^2 = -\frac{n}{2}$$

The likelihood ratio is

$$\Lambda = \frac{L(\mu_1, \sigma_1)}{L(\mu_0, \sigma_0)} = \left( \frac{\hat{\mu}_1}{\hat{\mu}_0} \right)^n e^{-\frac{n}{2}}$$

$$= \left[ \frac{\sum (x_i - \mu_1)^2}{\sum (x_i - \mu_0)^2} \right]^{\frac{1}{2}}$$

$H_0$ will be rejected if $\Lambda < 1$. Then, the rejection region is

$$\sum (x_i - \mu_1)^2 < \sum (x_i - \mu_0)^2$$

$$\sum x_i^2 + n \mu_1^2 - 2 \mu_1 \sum x_i < \sum x_i^2 + n \mu_0^2 - 2 \mu_0 \sum x_i$$

$$n(\mu_1^2 - \mu_0^2) < 2(\mu_1 - \mu_0) \sum x_i$$

$$\left( \frac{\mu_1 + \mu_0}{2} \right) < \frac{\sum x_i}{n}$$

or

$$\frac{\sum x_i}{n} > \frac{\mu_1 + \mu_0}{2}$$
A study addressed the relationship between attitudes of children toward their fathers and their birth order. Fifteen firstborn and 15 second-born males (independent samples) were given a questionnaire dealing with these attitudes. The scores in the scale of 0 to 100 are:

Firstborn: 40, 41, 44, 49, 53, 54, 56, 61, 62, 64, 65, 67, 67

Second-born: 23, 25, 38, 43, 44, 47, 49, 54, 55, 58, 58, 60, 66, 66, 72

A large score means that the child supports the role of the father in the family.

Use large-sample approximation for a nonparametric method to test the hypothesis of no difference between the attitudes of first and second-born children at 0.05 level of significance. State the null and alternative hypotheses in terms of the distribution functions. ($\Phi(1.96) = 0.975$).

\( H_0 : F(x) = G(x) \quad \text{for all } x \)

\( H_1 : F(x) < G(x) \text{ or } F(x) > G(x) \quad \text{for at least one } x \)

\( n = 15, \quad m = 15, \quad N = m + n = 30 \)

Let \( T_y \) be the rank sum of the second sample,

\[ T_y = 1 + 2 + 3 + 4 + 6 + 7.5 + 9 + 10.5 + 15 + 17 + 19.5 + 21 \]

\[ + 26.5 + 26.5 + 30 \]

\[ T_y = 214 \]

\[ U_y = T_y - \frac{m(m+1)}{2} = 214 - \frac{15(16)}{2} = 94 \]

\[ E(U_y) = \frac{mn}{2} = \frac{15(15)}{2} = 112.5 \]

\[ V(U_y) = \frac{mn(n+1)}{12} = \frac{15(15)(30+1)}{12} = 81.25 \]

\[ Z = \frac{U_y - E(U_y)}{\sqrt{V(U_y)}} = \frac{94 - 112.5}{\sqrt{81.25}} = -1.77 \]

We cannot reject \( H_0 \) at .05 level.
3. In an experiment to investigate warping of copper plates, the two factors studied were temperature and the copper content of the plates. The response variable was a measure of the amount of warping, and the data were as follows:

<table>
<thead>
<tr>
<th>Copper Content(%)</th>
<th>40</th>
<th>60</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>17</td>
<td>20</td>
<td>16</td>
<td>21</td>
</tr>
<tr>
<td>75</td>
<td>12</td>
<td>9</td>
<td>18</td>
<td>13</td>
</tr>
<tr>
<td>100</td>
<td>16</td>
<td>12</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>125</td>
<td>21</td>
<td>17</td>
<td>23</td>
<td>21</td>
</tr>
</tbody>
</table>

(a) Write down the analysis of variance model and define each of its components.

\[
Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijk} \\
\epsilon_{ijk} \sim N(0, \sigma^2)
\]

\(Y_{ijk}\) is the measurement for the \(i\)-th level of factor A, \(j\)-th level of factor B, and \(k\)-th level of error.

(b) Complete the following analysis of variance table:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMP</td>
<td>3</td>
<td>156.09</td>
<td>52.03</td>
<td>7.67</td>
</tr>
<tr>
<td>COPPER</td>
<td>3</td>
<td>698.35</td>
<td>232.78</td>
<td>34.32</td>
</tr>
<tr>
<td>TEMP*COPPER</td>
<td>9</td>
<td>113.78</td>
<td>12.64</td>
<td>1.86</td>
</tr>
<tr>
<td>ERROR</td>
<td>10</td>
<td>108.50</td>
<td>10.85</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>31</td>
<td>1076.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Use the Bonferroni method to construct a 95% confidence interval for the difference between the means of the 40% and 100% copper contents. (t(0.05/2|10) = 3.08).

\[
\text{The } 100(1-\alpha) \% \text{ Confidence Interval is: } \bar{Y}_{40} - \bar{Y}_{100} \pm t(\alpha/2) \cdot s_p \sqrt{\frac{2}{n}}
\]

\[m = \binom{4}{2} = 6, \text{ df } = 16, s_p = 6.78, \text{ s.e.m. } = 0.5/12 + (0.05/12) = 3.08\]

\[
(28.25 - 15.5) \pm 3.08 \sqrt{6.78 \cdot \frac{2}{4(2)}}
\]

\[12.75 \pm 4.01 \]

\[\left[8.74 , 16.76\right]\]
4. Consider a 2x2 contingency table and let \( n_{ij} \) be the cell frequency of the \( i \)-th row and \( j \)-th column. Suppose the null hypothesis \( H_0 \) is defined by the following table of joint probabilities:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( p )</td>
<td>( p(1-p) )</td>
</tr>
<tr>
<td>2</td>
<td>( p(1-p) )</td>
<td>( (1-p)^2 )</td>
</tr>
</tbody>
</table>

(a) Find the maximum likelihood estimate of \( p \) in terms of the cell frequencies.
(b) Suppose, \( n_{11} = 26, n_{12} = 15, n_{21} = 7, \) and \( n_{22} = 37 \). Find the likelihood ratio test statistic for testing \( H_0 \) against the alternative hypothesis that in this 2x2 table the probabilities of off-diagonal cells are equal.
(c) Using the cell frequencies given in (b), find the likelihood ratio test statistic for testing \( H_0 \) against the alternative hypothesis that in this 2x2 table the cell probabilities are unrestricted (except they add to 1).

\[
\Lambda = \frac{\hat{L}(\hat{p})}{\hat{L}(\hat{p}_{\text{MLE}})} = \left( \frac{37/85}{1 - 37/85} \right)^{24} \left( \frac{37/85}{1 - 37/85} \right)^{22} \cdot \left( \frac{37/85}{1 - 37/85} \right)^{22} \cdot \left( \frac{37/85}{1 - 37/85} \right)^{10} \cdot \left( \frac{37/85}{1 - 37/85} \right)^{22} = 1.1902 \cdot \left( 10^{-5} \right) 
\]

-2 \log \Lambda \approx 19.7

should be compared to \( \chi^2 \) with 1 df.

(c) With \( n_{11} = 26, n_{12} = 15, n_{21} = 7, \) and \( n_{22} = 37 \). Then

\[
\Lambda = \left( \frac{37/85}{1 - 37/85} \right)^{24} \left( \frac{19/85/15/85}{7/85} \right)^{22} \cdot \left( \frac{37/85}{1 - 37/85} \right)^{22} = 1.1902 \cdot \left( 10^{-5} \right) 
\]

-2 \log \Lambda \approx 22.68

and it should be compared to \( \chi^2 \) with 2 df.
5. Consider a regression model with independent normal errors and constant variance. Assume that the intercept is equal to zero and we are dealing with a model with no intercept.

(a) Obtain the least-squares estimate of the regression slope, $\hat{\beta}_1$.
(b) Find the distribution of the estimated slope.
(c) Show the following:

$$\sum(y_i - \hat{\beta}_1 x_i)^2 = \sum(y_i - \hat{\beta}_1 x_i)^2 + (\hat{\beta}_1 - \beta_1)^2 \sum x_i^2$$

(d) Find the likelihood ratio test statistic for testing $H_0: \beta_1 = \beta_{10}$ against $H_0: \beta_1 = \beta_{10}$.

Using part (c) express the likelihood ratio test statistic in terms of $\left(\beta_1 - \beta_{10}\right)^2$.

\[
\begin{align*}
\text{(a)} & \quad \sum e^2 = \sum (y_i - \hat{\beta}_1 x_i)^2 \\
\frac{\partial \sum e^2}{\partial \hat{\beta}_1} &= -2 \sum \Sigma x_i y_i + 2 \hat{\beta}_1 \sum x_i^2 \Rightarrow \\
\text{Note that} \quad \frac{\partial^2 \sum e^2}{\partial \hat{\beta}_1^2} &= 2 \sum x_i^2 > 0, \quad \text{we have minimum.} \\
\text{(b)} & \quad \text{Since } y_i \sim N(\hat{\beta}_1 x_i, \sigma^2), \quad \hat{\beta}_1 \text{ is normal with} \\
& \quad \text{E}(\hat{\beta}_1) = \sum x_i, \quad \text{Var}(\hat{\beta}_1) = \sum x_i^2 / \Sigma x_i^2 = \hat{\beta}_1, \quad \text{Var}(\hat{\beta}_1) / \text{Var}(x) = \sigma^2 / \sum x_i^2 \\
\text{(c)} & \quad \sum (y_i - \hat{\beta}_1 x_i)^2 = \sum (y_i - \hat{\beta}_1 x_i) + (\hat{\beta}_1 - \beta_1) x_i \\
\text{But} \quad \sum (y_i - \hat{\beta}_1 x_i)(\hat{\beta}_1 - \beta_1) x_i = (\hat{\beta}_1 - \beta_1) \left[ \sum y_i x_i - \hat{\beta}_1 \sum x_i^2 \right] \\
& \quad = (\hat{\beta}_1 - \beta_1) \left[ \sum y_i x_i - \hat{\beta}_1 x_i \right] = 0 \\
\text{Then we have} \quad \sum (y_i - \hat{\beta}_1 x_i)^2 = \sum (y_i - \hat{\beta}_1 x_i)^2 + (\hat{\beta}_1 - \beta_1)^2 \sum x_i^2 \\
\text{(d)} & \quad \text{The likelihood function is} \\
\ell_k (\beta_i, \sigma^2) &= (2\pi)^{n/2} (\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum (y_i - \hat{\beta}_1 x_i)^2} \\
\text{Since} \quad \sigma^2 = \frac{1}{2} \sum (y_i - \hat{\beta}_1 x_i)^2, \text{then} \quad \ell_k (\hat{\beta}_1, \sigma^2) = (\sigma^2)^{-n/2} e^{-\frac{1}{2}} \\
\text{The likelihood ratio for testing} \quad H_0: \beta_1 = \beta_{10} \text{ against} \beta_1 \neq \beta_{10} \text{ is} \\
\Lambda = \frac{\ell_k (\beta_{10}, \hat{\sigma}_0^2)}{\ell_k (\hat{\beta}_1, \hat{\sigma}_1^2)} = \left[ \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} \right]^{-n/2} = \left[ \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} \right]^{-n/2} \\
\text{Using (c)} \quad \frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2} = \frac{\sum (y_i - \hat{\beta}_1 x_i)^2}{\sum (y_i - \beta_{10} x_i)^2} = \frac{\sum (y_i - \beta_{10} x_i)^2}{\sum (y_i - \hat{\beta}_1 x_i)^2} = 1 + \left( \hat{\beta}_1 - \beta_{10} \right)^2 / \hat{\sigma}_1^2 \\
-2 \log \Lambda = n \log \left( 1 + \left( \hat{\beta}_1 - \beta_{10} \right)^2 / \hat{\sigma}_1^2 \right).