Analyzing Male-Female Wage Differentials using Dummy and Interaction Variables

(a) Using the data set Wage1.dta,
(i) Compare the average wage of men and women in the sample. Is there a difference between male and female wages?

Mean Wage Males $7.099489
Mean Wage Females $4.587659

(ii) Show graphically the distribution of wages for males and females separately. Describe your graphs.
(b) Perform a hypothesis test on the difference between the mean wage of men and women. Be sure to formally state your null and alternative hypothesis.

\[ H_0: \bar{Wage}_{\text{Male}} - \bar{Wage}_{\text{Female}} \leq 0 \]

\[ H_1: \bar{Wage}_{\text{Male}} - \bar{Wage}_{\text{Female}} > 0 \]

(i) Calculate the critical wage differential.

\[ \bar{Wage}_{\text{Male}} - \bar{Wage}_{\text{Female}} = 2.51183 \]

From STATA
Two-sample t test of hourly wages between males and females assuming equal variances

<table>
<thead>
<tr>
<th>Group</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Err.</th>
<th>Std. Dev.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>274</td>
<td>7.09949</td>
<td>.251367</td>
<td>4.160858</td>
<td>6.604626  7.594352</td>
</tr>
<tr>
<td>1</td>
<td>252</td>
<td>4.58766</td>
<td>.159335</td>
<td>2.529363</td>
<td>4.273855  4.901462</td>
</tr>
<tr>
<td>combined</td>
<td>526</td>
<td>5.89610</td>
<td>.161026</td>
<td>3.693086</td>
<td>5.579768  6.212437</td>
</tr>
<tr>
<td>diff</td>
<td>251183</td>
<td>.3034092</td>
<td>1.915782</td>
<td>3.693086</td>
<td>5.579768  6.212437</td>
</tr>
</tbody>
</table>

\[ \text{diff} = \text{mean}(0) - \text{mean}(1) \]
\[ t = 8.2787 \]
\[ \text{degrees of freedom} = 524 \]

Ho: diff = 0
Ha: diff < 0  Ha: diff != 0  Ha: diff > 0
\[ \Pr(T < t) = 1.0000 \]  \[ \Pr(T > t) = 0.0000 \]  \[ \Pr(T > t) = 0.0000 \]

Critical wage differential = \((1.6477668)*(.3034092) = .4999476 \)

(ii) Is the observed difference between male and female wages statistically significant? Explain fully.

Sample difference of $2.51183 is greater than the critical difference of .4999476, so reject the null hypothesis that there is no difference.

Note also the sample \( t = 8.2787 \) is greater than the critical \( t = 1.6477668 \)
(c) Estimate the wage equation: \( \text{Wage} = \beta_0 + \beta_1 X_1 + U \).

Where: Wage is the hourly wage rate and
\( X_1 = \) is dummy variable for females.

Source SS df MS  Number of obs = 526  
Model 828.220467 1 828.220467 F( 1, 524) = 68.54  
Residual 6332.19382 524 12.0843394 Prob > F = 0.0000  
Total 7160.41429 525 13.6388844 R-squared = 0.1157  
Adj R-squared = 0.1140  
Root MSE = 3.4763

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>female</td>
<td>-2.512</td>
<td>.3034092</td>
<td>-8.28</td>
<td>0.000</td>
<td>-3.107878 -1.915782</td>
</tr>
<tr>
<td>_cons</td>
<td>7.09949</td>
<td>.2100082</td>
<td>33.81</td>
<td>0.000</td>
<td>6.686928 7.51205</td>
</tr>
</tbody>
</table>

(i) How much of the variation in wages is explained by gender?
\( R^2 = 0.1140 \)

(ii) Interpret the coefficient on female. Is it statistically significant?
Women make $2.51183 per hour less than men.

(iii) Show your regression equations graphically in a scatter diagram. Describe your graph.
(d) Are your results consistent with the hypothesis test in (b)?

Yes. Note especially the t-test.

(e) Consider the following wage equation:
\[ \text{Wage} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U. \]
Where \( X_1 \) = Experience.
\( X_2 \) = Female.

(i) What are the expected signs of the coefficients?
(ii) Write out the regression equation for men.

For men: \( \text{Wage} = \beta_0 + \beta_1 X_1 + U. \)

(iii) Write out the regression equation for women.

For women: \( \text{Wage} = (\beta_0 + \beta_2) + \beta_1 X_1 + U. \)

(f) Estimate the above equation.

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 526</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>898.161983</td>
<td>2</td>
<td>449.080991</td>
<td>F( 2, 523) = 37.51</td>
</tr>
<tr>
<td>Residual</td>
<td>6262.25231</td>
<td>523</td>
<td>11.9737138</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>7160.41429</td>
<td>525</td>
<td>13.6388844</td>
<td>R-squared = 0.1254</td>
</tr>
</tbody>
</table>

Root MSE = 3.4603

\[ \begin{array}{llllll}
\text{wage} & \text{Coef.} & \text{Std. Err.} & t & P>\text{t} & [95\% \text{ Conf. Interval}] \\
\text{exper} & 0.0269163 & 0.0111369 & 2.42 & 0.016 & 0.0050379 \quad 0.0487948 \\
\text{female} & -2.48142 & 0.3022793 & -8.21 & 0.000 & -3.07525 \quad -1.887589 \\
\_cons & 6.626882 & 0.2862475 & 23.15 & 0.000 & 6.064546 \quad 7.189218 \\
\end{array} \]

(i) How much of the variation in wages is explained by the model.

\( R^2 = 0.1221 \)

(ii) Interpret your estimated coefficients. Are your estimated coefficients significant?
(iii) Show your regression equations graphically in a scatter diagram. Describe your graph.
(g) Construct the interaction variable between female and experience. Interpret this variable.

Consider the regression equation:
\[ \text{Wage} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U. \]
Where \( X_1 = \) Experience.
\( X_2 = \) Female.
\( X_3 = \) Interaction variable.

(i) What are the expected signs of the coefficients?
(ii) Write out the regression equation for men.

For men: \( \text{Wage} = \beta_0 + \beta_1 X_1 + U. \)

(iii) Write out the regression equation for women.

For women: \( \text{Wage} = (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X_1 + U. \)

(h) Estimate the above equation:

- **Source**: SS       df       MS
- **Model**: 971.286312     3  323.762104
- **Residual**: 6189.12798   522   11.856567
- **Total**: 7160.41429   525  13.6388844

Number of obs = 526
F(  3,   522) =   27.31
Prob > F =  0.0000
R-squared =  0.1356
Adj R-squared =  0.1307
Root MSE =  3.4433

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std. Err.</th>
<th>t</th>
<th>P&gt;t</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>wage</td>
<td>.0536048</td>
<td>.0154372</td>
<td>3.47</td>
<td>0.001</td>
<td>(.0232782 .0839314)</td>
</tr>
<tr>
<td>exper</td>
<td>-1.546547</td>
<td>.4818603</td>
<td>-3.21</td>
<td>0.001</td>
<td>(-2.49317 -.5999231)</td>
</tr>
<tr>
<td>female</td>
<td>-.0550699</td>
<td>.022175</td>
<td>-2.48</td>
<td>0.013</td>
<td>(-.098633 -.0115068)</td>
</tr>
<tr>
<td>fexper</td>
<td>6.158275</td>
<td>.3416741</td>
<td>18.02</td>
<td>0.000</td>
<td>(5.48705 6.829501)</td>
</tr>
</tbody>
</table>

(i) How much of the variation in wages is explained by the model?
\[ R^2 = 0.1307 \]

(ii) Interpret your estimated coefficients. Are your estimated coefficients significant?
(iii) Show your regression equations graphically in a scatter diagram. Describe your graph.