\[ p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n + Z_{\frac{\alpha}{2}}^2}} \quad \text{where} \quad \hat{p} = \frac{Y + 1}{2n + Z_{\frac{\alpha}{2}}^2} \]

\[ (\hat{p}_1 - \hat{p}_2) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1 + 2} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2 + 2}} \quad \text{where} \quad \hat{p}_1 - \hat{p}_2 = \frac{Y_1 + 1}{n_1 + 2} - \frac{Y_2 + 1}{n_2 + 2} \]

\[ \sum^n_{i=1} \frac{(O_i - E_i)^2}{E_i} \]

\[ \sum^n_{i=1} (x_i - \bar{x})(y_i - y) \]

\[ \sum^n_{i=1} (x_i - \bar{x})^2 \]

\[ \bar{y} - b_1 \bar{x} \]

\[ \sqrt{\frac{SS(\text{resid})}{n - 2}} \]

\[ \frac{S_{Y|x}}{\sqrt{\sum^n_{i=1} (x_i - \bar{x})^2}} \]

\[ \frac{\left[ \sum^n_{i=1} (x_i - \bar{x})(y_i - \bar{y}) \right]^2}{\sum^n_{i=1} (x_i - \bar{x})^2 \sum^n_{i=1} (y_i - \bar{y})^2} \]
This exam is worth a total of 120 points.
Part I: Answer eight of the following nine questions. If you complete more than eight, I will grade only the first eight. Five points each.

1) State the definition of a P-value.

2) The Central Limit Theorem says that for any i.i.d. random sample, \( Y_1, Y_2, \ldots, Y_n \) where \( E[Y_i] = \mu \) and \( E[(Y_i-\mu)^2] = \sigma^2 \), then as \( n \to \infty \) the distribution of the sample mean is ______________ with mean, ____, and variance, _______ (note, I’m asking for variance here – not standard deviation).

3) The compound \( m \)-chlorophenylpiperazine (mCPP) is thought to affect appetite and food intake in humans. In a study of the effect of mCPP on weight-loss, eight moderately obese men were given mCPP in a double-blind, placebo controlled experiment. Some of the men took mCPP for two weeks, then took nothing for two weeks (a “washout period”), and then took a placebo for two weeks. The rest of the men took the placebo during the first two weeks, then had a two week washout period, then took mCPP for the final two weeks. The men were asked to rate how hungry they were at the end of each two-week period (hunger rating for mCPP period – hunger rating for placebo period). A QQplot of the differences was constructed (below).

![A QQplot showing the distribution of differences between mCPP and placebo periods.](image)

(Circle the correct answer.) The use of the independent samples t-test / dependent samples t-test / sign test / Wilcoxon-Mann-Whitney test would be appropriate here.
4) State the assumptions we need to check (in terms of the errors) for simple linear regression before using the regression model for inference.

5) (Circle the correct answer.) If the assumptions of a regression model for predicting \( y \) from \( x \) are met, and we do not reject the null hypothesis that \( \beta_1 = 0 \), then we conclude that \( x \) can / cannot be used to predict \( y \).

6) (Circle the correct answer.) If the assumptions of a linear regression model for predicting \( y \) from \( x \) are met and we do reject the null hypothesis that \( \beta_1 = 0 \), then we may / may not conclude that \( x \) causes \( y \).

Questions 7,8, and 9 are on the next page!
A simple linear regression was performed to relate the cocoon temperature (Y in °C) to the outside air temperature (X in °C). Use this portion of DoStat’s output to answer the following 3 questions.

7) What is the r² for this regression?  

8) Interpret the value of the coefficient of determination (r²) in the context of the setting.

9) Referring to the default test DoStat performs for the \( \beta_1 \), fill in the blanks:

\[ H_0: \quad \text{__________________________} \]

\[ H_A: \quad \text{__________________________} \]

\[ P\text{-value: } \quad \text{__________________________} \]
1) It is common folk wisdom that drinking cranberry juice can prevent urinary tract infection in women. In 2001, the British Medical Journal reported the results of a Finnish study in which two groups of 50 women were monitored for these infections over 6 months. One group drank cranberry juice every day and the other group did not drink cranberry juice. At the end of the study, the number of women who had urinary tract infections (one or more) was 8 for the cranberry juice group and 18 for the group that did not drink cranberry juice.

1a) (20 points) Construct a 95% Agresti-Caffo confidence interval for the difference in the proportion of urinary tract infections for these two groups.

1b) (5 points) Interpret the interval you just computed in part (a).
2) (20 points) Research has indicated that the stress produced by today’s lifestyles results in health problems for a large proportion of society. An article in the *International Journal of Sports Psychology* (July – Sept. 1990) evaluated the relationship between physical fitness and stress. 549 employees of companies participating in the Health Examination Program offered by Health Advancement Services (HAS) were classified into three groups of fitness levels: good, average, and poor. Each person was tested for signs of stress. The following table reports the results.

<table>
<thead>
<tr>
<th></th>
<th>Signs of Stress</th>
<th>No Sign of Stress</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Fitness</td>
<td>38</td>
<td>204</td>
<td>242</td>
</tr>
<tr>
<td>Average Fitness</td>
<td>28</td>
<td>184</td>
<td>212</td>
</tr>
<tr>
<td>Good Fitness</td>
<td>10</td>
<td>85</td>
<td>95</td>
</tr>
<tr>
<td>Total</td>
<td>76</td>
<td>473</td>
<td>549</td>
</tr>
</tbody>
</table>

Test whether there is a significant association between fitness level and stress at the 0.05 significance level. I’ve numbered the steps for you, please write the appropriate step next to the appropriate number.

(1)

(2)

(3)

(4)

(5)

(6)
3) Twenty plots, each of equal area, were randomly chosen in a large field of corn. For each plot, the plant density (number of plants in the plot) and the mean cob weight (g of grain per cob) were observed. The results are given in the table.

<table>
<thead>
<tr>
<th>Plant Density X</th>
<th>Cob Weight Y</th>
<th>Plant Density X</th>
<th>Cob Weight Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>212</td>
<td>173</td>
<td>194</td>
</tr>
<tr>
<td>107</td>
<td>241</td>
<td>124</td>
<td>241</td>
</tr>
<tr>
<td>132</td>
<td>215</td>
<td>157</td>
<td>196</td>
</tr>
<tr>
<td>135</td>
<td>225</td>
<td>184</td>
<td>193</td>
</tr>
<tr>
<td>115</td>
<td>250</td>
<td>112</td>
<td>224</td>
</tr>
<tr>
<td>103</td>
<td>241</td>
<td>80</td>
<td>257</td>
</tr>
<tr>
<td>102</td>
<td>237</td>
<td>165</td>
<td>200</td>
</tr>
<tr>
<td>65</td>
<td>282</td>
<td>160</td>
<td>190</td>
</tr>
<tr>
<td>149</td>
<td>206</td>
<td>157</td>
<td>208</td>
</tr>
<tr>
<td>85</td>
<td>246</td>
<td>119</td>
<td>224</td>
</tr>
</tbody>
</table>

Preliminary calculations yield the following results:

\[
\begin{align*}
\bar{X} &= 139.05 \\
\bar{Y} &= 224.1 \\
\sum (x_i - \bar{x})(y_i - \bar{y}) &= 14,563.1 \\
\sum (x_i - \bar{x})^2 &= 20,209.0 \\
\sum (y_i - \bar{y})^2 &= 14,831.3 \\
S_e(x = 145) &= 1,337.3
\end{align*}
\]

3a) (7 points) Calculate the least-squares regression line using X=plant density as the predictor variable and Y=cob weight as the response.

3b) (7 points) Calculate the residual standard deviation (S_{Y|X}).

3c) (7 points) Give an estimate of the mean and standard deviation of cob weight at a plant density of 145.
3d) (7 points) Calculate a 95% confidence interval for $\beta_1$ (slope of the true line).

3e) (7 points) Interpret the interval you just computed in part (d).