A line is fitted to 10 data points \((X_i, Y_i)\). The equation of this fitted line is \(Y=10 + 0.3X\) and the population model that is being estimated is \(Y = \alpha + \beta X + e\) where the errors \(e\) have the usual assumptions, namely \(e \sim N(0, \sigma^2)\) and the \(e\)’s are independent of each other.

The values of \(X\) are 10, 20, 30, 40, 50 and there are two points above each \(X\) giving us our 10 total points. I also computed the sum of the squared residuals (vertical deviations of the points from the line) getting 640 as my error sum of squares.

1. (16 pts.) Compute the average _____ of the 10 \(X\) values and the (corrected) sum of squares of the \(X\)s. \(S_{xx} = \)_______

2. (8 pts.) Compute \(\sum_{i=1}^{10}(X_i - \bar{X})(Y_i - \bar{Y}) = \)_______ (If you can’t do it, insert a guess and use it later when needed)

3. (10 pts.) Compute the error mean square for this regression. \(\text{MSE} = \)_______

4. (8 pts.) We assumed the \(e\)’s have variance \(\sigma^2\).
   Give your best estimate of \(\sigma^2\) _________

5. (16 pts.) Compute, if possible, the regression sum of squares _____ and the (corrected) total sum of squares \(S_{yy} = \)_______ for these data.

6. (24 pts.) Our slope estimate, 0.3, has a standard error. Compute that standard error _____ and the t test _____ for testing \(H_0: \beta = 0\).
   How many degrees of freedom _____ does this t have?

7. (8 pts.) Assuming the p-value for the test in 6 is \(p=0.1720\), do you reject or fail to reject the null hypothesis using the usual 5% level of significance?

8. (10 pts.) Insert the three labels in the picture below to explain what the p-value in question 7 represents.
   Total shaded area is ____.1720 _______
It is easy to see that the mean of X is 30, the deviations are -20, -10, 0, 10, 20 each occurring twice so squaring and summing the 10 deviations we get Sxx = 2000.

Since b is Sxy/Sxx we find Sxy = (2000)(b) = (2000)(0.3) = 600 * (see note below)

MSE = 640/8 = 80
80 is ALSO best estimate $\sigma^2$

SS total = SS regression + SS error = 180 + 640 = 820.

$\sqrt{MSE/Sxx} = \sqrt{80/2000} = 0.2$
$t = 0.3/0.2 = 1.5$ with 8 df.

0.1720 > 0.0500 so we fail to reject H0.

* Note that Sxy would be computed from the data as $\sum (X_i - \bar{X})(Y_i - \bar{Y})$. For the observed points $(X_i, Y_i)$ used in this formula, $Y_i$ is not the predicted value $10+0.3X$. It is only for points Y on the line that $Y=10+0.3X$. Therefore, you would think that substituting the predicted values $10+0.3X$ for $Y_i$ in the Sxy formula would give the wrong answer. Indeed every term in that sum will change, but the sum of them will
actually add up to the right Sxy! While that is not the most logical or quick way to work this one, it is a valid method.

Reason: The difference between Sxy and what is described above is

\[
\sum (X_i - \bar{X})(Y_i - \bar{Y}) - \sum (X_i - \bar{X})(10 + .3X_i - \bar{Y}) = \sum (X_i - \bar{X})(Y_i - \bar{Y} + 0.3\bar{X} - .3X_i)
\]

That is because the intercept (10) is \(\bar{Y} - 0.3\bar{X}\). So now with \(b=0.3\) we have

\[
\sum (X_i - \bar{X})(Y_i - \bar{Y} - 0.3(X_i - \bar{X})) = \sum (X_i - \bar{X})(Y_i - \bar{Y}) - b\sum (X_i - \bar{X})^2 = sxy - bsxx = 0
\]

Here we used the fact that \(b\) is just \(\frac{sxy}{sxx}\). Thus Sxy and what we get by using predicted values in the Sxy formula differ by 0, that is, we have demonstrated that they are always the same once you have summed their terms.