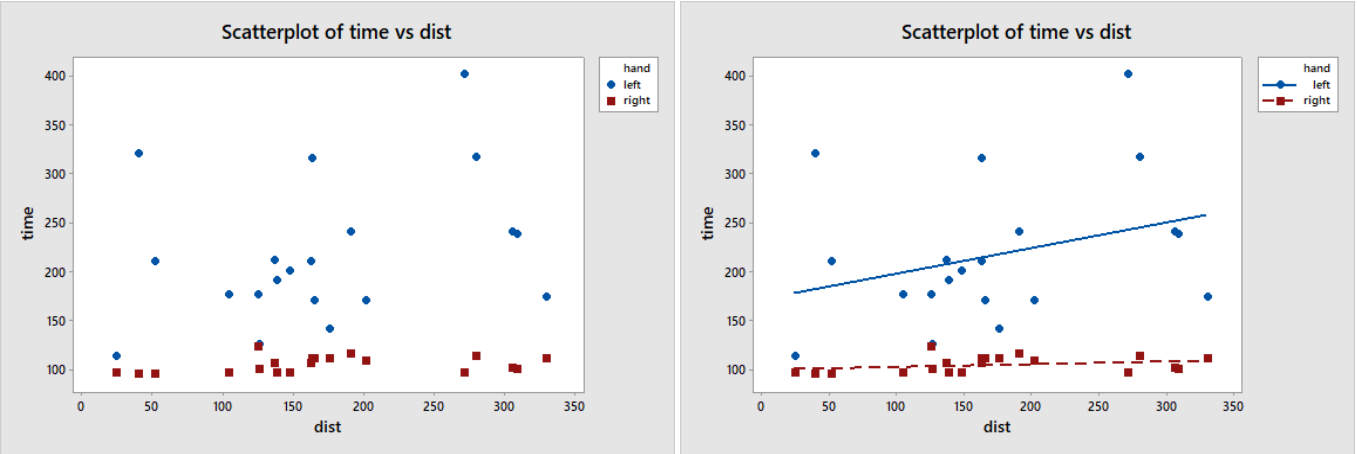
Supplementary exercise 2.59 of IPS7e

Data: Data on reaction time (with a computer mouse) and distance on the computer screen for 40 trials carried out by one subject, 20 trials with each hand. The reaction time is a response variable, whereas the distances between successive points on the screen are controlled in the trial (determined by the software). The hand used is an explanatory variable.

(a) The requested scatterplots are shown below in two versions, with and without overlaid least-squared regression lines. Note that these are obtained from different submenus of Minitab's `Graph-Scatterplot` menu.



(b) For the interpretation of (left) graph we note the following:

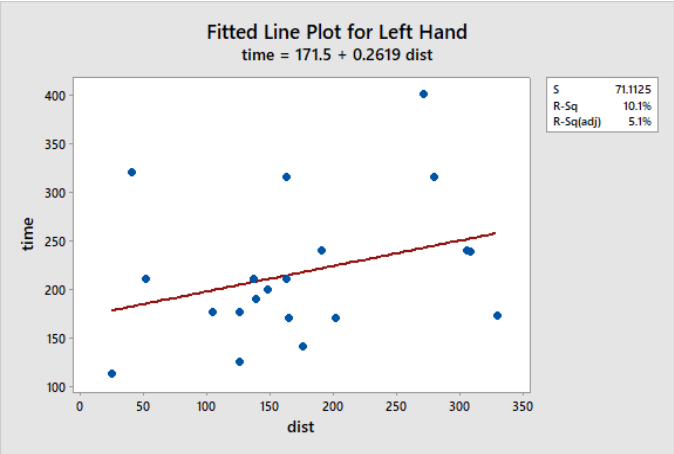
- The right-hand points lie below the left-hand points. This means that the right-hand times are shorter, so it seems reasonable to infer that the subject is right-handed.
- The left-hand points show a wide scatter with perhaps a slight tendency towards larger reaction times with larger distances.
- The right-hand points are squeezed together at the bottom of the plot, so it is difficult to see any patterns; however, there does not seem to be any clear dependence of reaction times on distance.

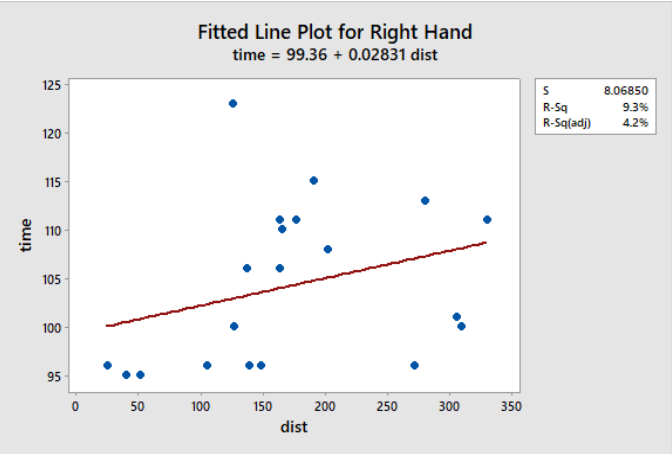
(c) The right graph shows that the slopes are quite different, steeper for the left hand. The two regression lines can (inside Minitab) be read of the plot by moving the pointer to each of the lines. Realistically it is more practical to perform separate analyses for the left-hand and right-hand data, and for this we split the Minitab worksheet by the variable hand (using the `Data-Split Worksheet` menu). Next we fit for both the left-hand and right-hand data the linear regression model,

$$\text{time}_i = \beta_0 + \beta_1 \text{dist}_i + \varepsilon_i,$$

where the errors (ε_i) are i.i.d. from $\text{N}(0, \sigma)$, and all model parameters (β_0, β_1, σ) are different for left and right hands.

The fitted line plots (on the next page) for the two subsets look quite similar when viewed on their own scale: the points are scattered widely around a line with a slightly increasing slope. None of the separate regressions correspond to a significant association between distance and time ($P = 0.17$ and $P = 0.19$ for left-hand and right-hand data, respectively). The P-values were read off the corresponding ANOVA tables (not shown).





Looking at the estimated coefficients $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$ and the computed R^2, we see that the slope for the left hand line is indeed steeper and also that the points have a much larger spread about the line, both by a factor of about 10 relative to the right hand. The intercept for the left-hand line is about twice that of the right-hand line, corresponding to the reaction time for a point that requires no movement of the mouse. The two regressions have similar strength of association when assessed by the P-value for the test for no association (this comparison makes sense because the two subsets are of the same size) and also by the R^2, the proportion of variance explained (the numerical measure requested). To describe the R^2 as a measure of the "success of the regression" is in my view poor terminology. Another potential measure of the predictive ability of the regression is the spread about the regression line, but this measure is difficult to compare between the two datasets because the left-hand values by themselves are far more variable than the right-hand values.

(d) Together with the analyses for (c) we could also request 4-in-1 panels of (standardized) residual plots, shown below. The lower right corner shows a plot of the residuals against observation order. No obvious increasing or decreasing trends are seen in these plots, neither for the left nor for the right hand. We therefore conclude that none of the suggested systematic trends are visible in the data.