Practical "Hypothesis testing"
P2

Read the paper provided for the discussion (Du Prel JB et al., 2009. Confidence Interval or P-Value? *Dtsch Arztebl Int* 106:335-339). In groups of four or five, discuss the following questions:

1. Consider an example from Gardner and Altman (1986) where samples of 100 diabetic and 100 non-diabetic men of a certain age are compared in a study. A difference of 6.0 mm Hg is found between their mean systolic blood pressures, and the standard error of this difference between sample means is 2.5 mm Hg.

   (a) A two independent samples two-tailed t-test is used and the associated p-value is 0.02. Based on the p-value, would you say this might be due to a real difference, or that it might be due to chance? Does the p-value provide information about the exact size and direction of the difference?

   (b) The best approximation to the population mean difference is provided by the difference of the sample means, 6.0. How precise is this? Explain why this can be shown by a confidence interval.

   (c) Looking at the confidence interval shown in Figure 1, explain, as in 1a), the difference in mean systolic blood pressure. Is it possible that this interval does not contain the true mean difference?

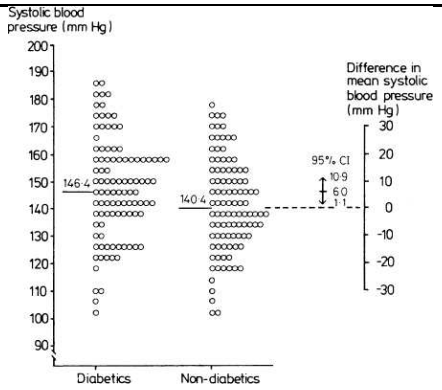

Figure 1: Systolic blood pressure in 100 diabetics and 100 non-diabetics with mean levels of 146.4 and 140.4 mm Hg, respectively. The difference between the sample means of 6.0 mm Hg is shown to the right together with the 95% confidence interval from 1.1 to 10.9 mm Hg.

   (d) If a systolic blood pressure difference of at least 4 mm is defined as clinically relevant, to which of the four examples in Figure 2 of Du Prel et al. (2009) does this example correspond?

   (e) Would it be equally helpful to report confidence intervals for the mean systolic blood pressure of diabetics and non-diabetics separately instead of the confidence interval for the mean difference?

2. Suppose that the same analysis is done using samples half the size, that however have the same observed mean difference (6.0) and standard deviation (do not confound with standard error) of the difference in sample means as in 1a. Similarly to Figure 1, means and confidence interval are displayed in Figure 2.

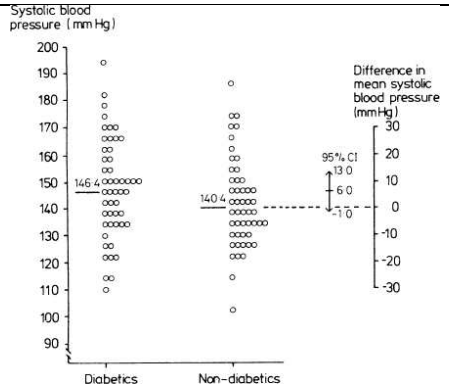

Figure 2: Systolic blood pressure in 50 diabetics and 50 non-diabetics with mean levels of 146.4 and 140.4 mm Hg, respectively. The difference between the sample means of 6.0 mm Hg is shown to the right together with the 95% confidence interval from -1.0 to 13.0 mm Hg.

(a) Explain why the confidence interval is now wider. Does this affect your conclusion about the difference in means?

(b) As before, if a systolic blood pressure difference of at least 4 mm is defined as clinically relevant, to which of the four examples in Figure 2 of Du Prel et al. (2009) does this example correspond?

(c) Does a p-value > 0.05 imply that there is no difference in mean systolic blood pressure between diabetics and non-diabetics? What is the risk of simply distinguishing between significant and non-significant?

(d) Explain what "publication bias" is.

3. Researchers plan to conduct a study on the quality of life (QoL) after breast cancer surgery to see whether women who undergo breast conserving surgery are more satisfied than women who undergo mastectomy. It is known that women after mastectomy score their QoL on average as 15 and it is expected that women after breast conserving surgery will score their QoL on average as 25. The standard deviation in both groups is about 13. How many women should be included in each group, assuming both groups have similar number of women, to reach the power level of 90%?