GM – Probabilités et Statistique – Corrections 9

Please note : the **reasoning/justifications** for the steps in your solution are also important (not only the final result).

In class

Exercise 1 To carry out a hypothesis test :

1. The parameter of interest is μ = mean life-time of light bulbs produces by the factory
2. $H : \mu = 1600$
   $A : \mu < 1600$ (according to the problem)
3. $T = (\overline{X} - \mu_H)/(S/\sqrt{n})$, thus $t_{obs} = (1570 - 1600)/(85/\sqrt{31}) \approx -1.97$
4. (a) Under H, $T \sim t_{30}$; $t_{30,0.05} = -t_{30,0.95} = -1.697$; (b) $t_{30,0.01} = -t_{30,0.99} = -2.457$
5. (a) Assuming a level $\alpha = 0.05$:
   $t_{obs} = -1.97$ (ou -1.95) $< -1.697 = t_{30,0.05}$, thus we REJECT the NULL hypothesis H.

   (b) Assuming a level $\alpha = 0.01$:
   $t_{obs} = -1.97 > -2.457 = t_{30,0.01}$, thus we DO NOT REJECT the NULL hypothesis H.

Exercise 2 (a) The corresponding confidence interval is $[\overline{Y} \pm t_{n-1,0.95}\, s/\sqrt{n}]$
(b) The length is $\overline{Y} + t_{n-1,0.95}\, s/\sqrt{n} - (\overline{Y} - t_{n-1,0.95}\, s/\sqrt{n}) = 2\, t_{n-1,0.95}\, s/\sqrt{n}$
(c) $t_{9,0.95} = 1.83$, thus for a 90% CI, we have $[\overline{Y} \pm t_{9,0.95}\, s/\sqrt{n}] = [30 \pm 1.83 \times 1.7/\sqrt{10}] \approx [29.02, 30.98]$

Exercise 3 Let F = score final exam, T = score bonus test.
slope $= r s_F/s_M = 1.2$, intercept $= \overline{F} - 1.2\,\overline{T} = -29$ points, thus

pred. Final $= 1.2 \times$ Test $- 29$ points; $RMSE\ (REQM) = s_F\sqrt{1-r^2} = 20\sqrt{1-0.6^2}$ = 16 points.

Exercise 4 pred. height = 0.25 inches per year × ed. + 66.75 inches
If $x = 12$, the regression estimate for height is 0.25*12 + 66.75 = 69.75 inches ;
if $x = 16$, the regression estimate for height is 0.25*16 + 66.75 = 70.75 inches.

However, even though there is an association between height and education, it is clear that going to university does not make you taller. The observed correlation is more easily explained by the association of height with age. In addition, the study is *observational*, so a correlation between height and education could be due to some other factors in the family background.

At home

Exercise 1 The confidence interval is $[\overline{Y} \pm t_{24,0.975}\, s/\sqrt{n}] = [1.6 \pm 2.064 \times 0.3/\sqrt{25}] \approx [1.48, 1.72]$

Exercise 2 (a) 1. The parameter of interest is $\mu_x - \mu_y$ = the difference in mean costs beetween the two concepts

2. $H : \mu_x = \mu_y \Rightarrow \mu_x - \mu_y = 0$
   $A : \mu_x \neq \mu_y \Rightarrow \mu_x - \mu_y \neq 0$
3. $s_p^2 = ((n-1)s_x^2 + (m-1)s_y^2)/(n+m-2) = (11(37.00^2) + 5(36.40^2))/16 = 1355.237$, thus $s_p = \sqrt{1355.237} \approx 36.8$
   $T = (\overline{X} - \overline{Y})/(S_p\sqrt{(n+m)/nm})$, thus $t_{obs} = (400.00 - 327.00)/(36.8\sqrt{(12+6)/(12*6)}) \approx 3.97$
4. Under H (and assuming that $\sigma_x^2 = \sigma_y^2$), $T \sim t_{16} t_{16,0.975} = 2.12 < 3.97\ (= t_{obs})$, thus $p_{obs} \leq \alpha = 0.05$

5. Since $p_{obs} \leq \alpha = 0.05$, thus we REJECT the NULL hypothesis H

(b) $X_1, \ldots, X_n \sim iid\ N(\mu_x, \sigma_x^2)$; $Y_1, \ldots, Y_m \sim iid\ N(\mu_y, \sigma_y^2)$; $\sigma_x^2 = \sigma_y^2$; X and Y independent

Exercise 3 pred. length $= 0.05$ cm per kg $\times$ weight (kg) $+ 439.01$ cm
Si $x = 3$, pred. length = 0.05*3 + 439.01 = 439.16 cm ;
si $x = 5$, pred. length = 0.05*5 + 439.01 = 439.26 cm.

Yes, this is an experimental study. Assuming that the only difference between the measured units is their weight, we can thus deduce causation.