L2: Licence Sciences et Technologies

UE STA230: final exam, December 17$^{\text{th}}$, 2013

with answers

Duration 2h. Autorized documents: statistics tables and two handwritten A4 sheets (four pages). Calculators are authorized.

Exercise 1 (12 points) : The hemoglobin concentration (HC) in any given population follows a normal distribution $\mathcal{N}(\mu, \sigma^2)$. The results for means and standard deviations reported in the reference cited below are the following (unit: gm/dl).

- Tibetan women: $\mu = 14.2$, $\sigma = 1.1$
- Aymara women: $\mu = 17.8$, $\sigma = 2.1$

1. (a) If the HC of a woman had been measured, what decision rule would you follow to decide on view of her HC, at threshold 1%, between the two hypotheses $\mathcal{H}_0$: "she is Tibetan", against $\mathcal{H}_1$: "she is Aymara".
   reject $\mathcal{H}_0$ if HC > 16.76
   (b) What is the second kind risk of that test? 0.3102
   (c) Mrs T. had her HC measured at 16.5. What is the corresponding p-value? what is your decision? 0.0183 , she is Tibetan
   (d) Mrs T. insists she is Aymara, and wants this to be tested as the null hypothesis. What is her p-value for this other test? what is your decision?
   0.2679 , she is Aymara

2. A sample of 10 Aymara women had their HC measured: the mean of the 10 values was 17.5, and the standard deviation was 1.5.
   (a) At threshold 5%, can it be said that the observed mean is significantly smaller than 17.8? $T = -0.6 > -1.833$, no
   (b) At threshold 5%, can it be said that the observed standard deviation is significantly smaller than 2.1? $T = 5.102 > 3.325$, no

3. Two groups of women had their HC measured. Out of 16 Tibetan women, the mean was 14.5, with standard deviation 1.3. Out of 19 Aymara women, the mean was 17.2, with standard deviation 1.8.
   (a) At threshold 5% can the hypothesis of equality of variances be accepted?
   $T = 1.897 < 2.79$, accept equality of variances
   (b) At threshold 1%, can it be said that Aymara women have a significantly higher HC than Tibetan women on those samples? $T = 4.856 > 2.445$, yes

4. In a larger scale study, Tibetan and Aymara women had their HC measured, but the only information that was retained was the number of those having a low HC (i.e. below 16 gm/dl): 84 among the 92 Tibetan, 22 among the 86 Aymara.

   (a) Using the large numbers approximations, can it be said that the proportion of low HC is larger among Tibetan women? $\boxed{T = 11.85\,,\ \text{yes}}$

   (b) Write a contingency table corresponding to these data.

   (c) Compute the test statistic for the chi-squared test of independence. $\boxed{T = 79.708}$

   (d) What is your conclusion? $\boxed{\text{significant dependence}}$

Exercise 2 (8 points) : For a population of 24 Aymara women, the two variables
X : altitude of residence (in meters),
Y : Hemoglobin Concentration (HC in gm/dl),
have been measured. The following values are given:

$$\overline{x} = 3523.8\,,\ s_x^2 = 92257\,,\ \overline{y} = 17.42\,,\ s_y^2 = 1.095\,,\ c_{xy} = 274.35\,.$$

1. Compute the correlation coefficient of X and Y. $\boxed{r_{xy} = 0.8632}$

2. Give the equation of the regression line of Y onto X. $\boxed{y = 0.002974\,x + 6.9411}$

3. What is the predicted hemoglobin concentration for an Aymara woman living at 3500m? $\boxed{17.35}$

4. Test the pertinence of the regression at threshold 1%. $\boxed{T = 8.02 > 2.508\,;\ \mathcal{H}_0 : a = 0 \text{ rejected, pertinence accepted}}$

5. Give a confidence interval with level 0.99, for the average HC of Aymara women living at 3500m. $\boxed{[17.03; 17.67]}$

6. Mrs. T. lives in a village at 3528m. Give a prediction interval with level 0.99 for her HC. $\boxed{[15.84; 19.02]}$

7. In a previous study, it had been reported that average HC should increase by 0.2 gm/dl, for each 100 m increase in residence altitude. In the linear regression, which value of the slope does this correspond to? Which hypotheses are you testing to confirm what value? $\boxed{\mathcal{H}_0\ :\ a = 0.002 \text{ against } \mathcal{H}_1\ :\ a > 0.002}$

8. Give the value of the test statistic, and an interval containing the p-value. What is your conclusion? $\boxed{T = 2.6257\,;\ \text{p-value} \in [0.005, 0.01]\,;\ \text{significantly higher increase}}$

Reference: C. M. Beall et al. Hemoglobin concentration of high altitude Tibetans and Bolivian Aymara, American Journal of Physical Anthropology, 106:385–400 (1998)