# HOMEWORK#1

DUE ON 12PM (NOON) MAY 22

Please email your homework (scanned handwritten solution or typed solution) to my email address with subject "HW 1 of NCKU course"

1. If $X_i \sim N(\mu, i * \sigma^2)$ for some unknown $\mu$ and $\sigma^2 > 0$, for $i = 1, \ldots, n$. Define $\bar{X} = \sum X_i/n$ and $S^2 = \sum(X_i - \bar{X})^2/(n-1)$.

   - What is the distribution of $\bar{X}$? Does it converge to $\mu$? (Note that $X_i$'s have different variance, hence Law of Large Number doesn't hold here)
   - Define a weighted average $\bar{X}_w = \sum c_i X_i$ where $c_i = (1/i)/\sum_{i=1}^{n}(1/i)$. What is distribution of $\bar{X}_w = \sum c_i X_i$? What is the its MSE as an estimator of $\mu$?
   - Please compare $\bar{X}$ and $\bar{X}_w$. Which one is a better estimation of $\mu$.
   - Show that $\bar{X}_w$ is independent to $S^2$ (Hint: show that $\bar{X}_w$ is independent to $X_i - \bar{X}$ for any $i$)

2. Let $X_i$ $(i = 1, \ldots n)$ be random variables following $\text{Bin}(m, p_0)$ distribution, where $p_0$ is known but $m$ is an unknown parameter

   - Design a consistent estimator for $m$
   - Define an approximate confidence interval for $m$ based on central limit theory

3. Given **one** observation $X \sim \text{Unif}(\theta, \theta + 1)$, one wants to test $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$

   - Please design a reasonable reject region
   - For your reject region, please derive the corresponding power function $\beta(\theta)$.

4. An experiment tries to compare the typing efficiency of two keyboards denoted by $A$ and $B$. One typist uses the keyboards on six different manuscripts, denoted by 1-6.

   Let $y$ be the amout of time used to type up a manuscript. Note that $y$ depends on keyboard, manuscript, whether the manuscript has already been typed, and experimental error. Let $\mu_A$ and $\mu_B$ denote the effects of keyboard A and B respectively, $\tau_i$ the effect of manuscript $i$ for $i = 1, 2, 3, 4, 5, 6$ and $\epsilon$ denotes the experimental error. Let $\alpha_l$ denote the learning effect. We are interested in estimating the difference between $\mu_B$ and $\mu_A$. Consider a design as follows:

   $$1.A - B; 2.B - A; 3.A - B; 4.B - A; 5.A - B; 6.A - B.$$

   (For the 1st manuscript, the typist tries keyboard A first, then tries keyboard B.) The statistical model for the amount of time for 1st manuscript with keyboard A, denotes by $y_{1A}$ is

   $$y_{1A} = \gamma + \mu_A + \tau_1 + \epsilon_{1A},$$

and the model for the amount of time used for 1st manuscript with keyboard B is

$$y_{1B} = \gamma + \mu_B + \tau_1 + \alpha_l + \epsilon_{1B},$$

where $\gamma$ is some constant value.

- Is $\alpha_l$ positive or negative? Why it is not included in first model?
- Write down the statistical models for the other runs.
- Is $(\sum_{i=1}^{6} y_{iA} - \sum_{i=1}^{6} y_{iB})/6$ a good estimate for $\mu_A - \mu_B$
- Please propose a better design.