STA219 Assignment 8

12312706 Zhou Liangyu

1. (1) $X \sim \mathrm{N}(\mu, 1) \to \overline{X} \sim \mathrm{N}(\mu, \frac{1}{n})$.

   Type I error rate: $\alpha = P(\overline{X} > 2.6|\mu = 2) = P(\sqrt{20}(\overline{X} - 2) > \sqrt{20}(2.6 - 2)) \approx 1 - \Phi(2.68) = 0.0037$.

   Type II error rate: $\beta = P(\overline{X} \le 2.6|\mu = 3) = P(\sqrt{20}(\overline{X} - 3) > \sqrt{20}(2.6 - 3)) \approx \Phi(-1.79) = 0.0367$.

   (2) $\because \beta = P(\overline{X} \le 2.6|\mu = 3) = P(\sqrt{n}(\overline{X} - 3) > \sqrt{n}(2.6 - 3)) = \Phi(-0.4\sqrt{n}) < 0.01$

   $\therefore n > \dfrac{\Phi^{-1}(0.01)^2}{0.16} = \dfrac{(-2.323)^2}{0.16} = 33.73.$

   $\therefore$ The minimum sample size $n = 34$.

2. (1) Let random variable X be the population of the TV time of middle school students, and $E(X) = \mu_X$; $X_1, \dots, X_n$ be a simple random sample from X.

   The testing problem is $H_0 : \mu = 8 \leftrightarrow H_1 : \mu < 8$. The test statistics is $T = \dfrac{\overline{X} - 8}{\frac{S_X}{\sqrt{n}}} \overset{\text{approx.}}{\sim} N(0, 1)$ under H_0.

   For $\alpha = 0.05$, since H_1 is one-sided and left-tail, the rejection region is $\{\mathbf{X} : T < -z_\alpha = -1.645\}$.

   $\because t_{\text{obs}} = \dfrac{6.5 - 8}{\frac{2}{10}} = -7.5 < -1.645$

   $\therefore$ We would reject H_0 at significance level $\alpha = 0.05$.

   (2) $\because p$-value $= P(T < -7.5|\mu = 8) = \Phi(-7.5) < \Phi(-3.49) = 0.0002 < 0.05$

   $\therefore$ The conclusion is the same as in (1), we would reject H_0 at significance level $\alpha = 0.05$.

3. $\because T \overset{\text{approx.}}{\sim} N(0, 1)$ under H_0

   $\therefore T \overset{\text{approx.}}{\sim} N(\dfrac{\mu - 8}{\frac{2}{\sqrt{n}}}, 1) = N(-\dfrac{\sqrt{n}}{4}, 1)$ under the true mean time 7.5.

   $\because$ power $= P(T < -1.645|\mu = 7.5) = P(T + \dfrac{\sqrt{n}}{4} < -1.645 + \dfrac{\sqrt{n}}{4}) = \Phi(-1.645 + \dfrac{\sqrt{n}}{4}) \ge 0.9$

   $\therefore n \ge 16(\Phi^{-1}(0.9) + 1.645)^2 = 16 \times (1.282 + 1.645)^2 = 137.07.$

   $\therefore$ To achieve at least 0.90 power in this significance test, the minimum sample size $n = 138$.

4. (1) The hypothesis testing problem is $H_0 : p_A - p_B = 0 \leftrightarrow H_1 : p_A - p_B \ne 0$.

   $\because \hat{p}_A - \hat{p}_B \overset{\text{approx.}}{\sim} N(p_A - p_B, \dfrac{p_A(1 - p_A)}{n_A} + \dfrac{p_B(1 - p_B)}{n_B})$

   $\therefore$ The test statistics is $T = \dfrac{\hat{p}_A - \hat{p}_B}{\sqrt{\dfrac{\hat{p}_A(1 - \hat{p}_A)}{n_A} + \dfrac{\hat{p}_B(1 - \hat{p}_B)}{n_B}}} \overset{\text{approx.}}{\sim} N(0, 1)$ under H_0.

   Since H_1 is two sided, the rejection region is $\{\mathbf{X} : |T| > z_{\frac{\alpha}{2}}\}$.

   $\because t_{\text{obs}} = \dfrac{0.45 - 0.35}{\sqrt{\dfrac{0.45 \times 0.55}{400} + \dfrac{0.35 \times 0.65}{400}}} \approx 2.90 > z_{\frac{\alpha}{2}} = z_{0.01} = 2.323$

   $\therefore$ There is a significant difference between the support rates of the candidate in town A and town B.

   (2) A 98% CI of p_A is $0.45 \pm 2.323\sqrt{\dfrac{0.45 \times 0.55}{400}} = 0.45 \pm 0.058 = (0.392, 0.508)$.

   A 98% CI of p_B is $0.35 \pm 2.323\sqrt{\dfrac{0.35 \times 0.65}{400}} = 0.35 \pm 0.055 = (0.295, 0.405)$.

A 98% CI of $p_A + p_B$ is $(0.45 - 0.35) \pm 2.323\sqrt{\frac{0.45 \times 0.55}{400} + \frac{0.35 \times 0.65}{400}} = 0.1 \pm 0.080 = (0.020, 0.180)$.

(3) No, we can't make the decision based on whether the 98% confidence intervals of p_A and p_B overlap or not.

The confidence intervals of p_A and p_B is calculated based on their standard error respectively, while the confidence interval of $p_A + p_B$ is calculated based on the combination of the two standard errors.

This may result in the confidence interval of the difference still not including 0.

(4) Yes. If the 98% confidence interval does not contain $\theta_0 = 0$, then a hypothesis test at the level $\alpha = 0.02$ will almost always reject H_0, i.e. the support rates of the candidate in town A and town B are not significantly different.

5. (1) The hypothesis testing problem is $H_0 : \mu_1 - \mu_2 = 0 \leftrightarrow H_1 : \mu_1 - \mu_2 > 0$.

Since $\sigma_1^2 = \sigma_2^2 = 4^2 = 16$, the test statistic is $T = \dfrac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\dfrac{16}{n} + \dfrac{16}{m}}}$.

Since H_1 is one-sided and right-tail, the rejection region is $\{\mathbf{X} : T > z_\alpha = 1.645\}$.

(2) $\because T \overset{\text{approx.}}{\sim} N(0, 1)$ under H_0

$\therefore T \overset{\text{approx.}}{\sim} N(\dfrac{2}{\sqrt{\dfrac{16}{10} + \dfrac{16}{11}}}, 1) = N(1.1443, 1)$ under the true mean difference 2.

Type II error rate: $\beta = P(T \leq 1.645 | \mu_1 - \mu_2 = 0) = P(T - 1.1443 \leq 1.645 - 1.1443) = \Phi(0.5007) = 0.6915$.

(3) $\because$ power $= 1 - \beta = 1 - P(T - \dfrac{2}{\sqrt{\dfrac{16}{n} + \dfrac{16}{m}}} \leq 1.645 - \dfrac{2}{\sqrt{\dfrac{16}{n} + \dfrac{16}{m}}}) = \Phi(-1.645 + \dfrac{2}{\sqrt{\dfrac{16}{n} + \dfrac{16}{m}}}) \geq 0.9$.

$\therefore \dfrac{1}{n} + \dfrac{1}{m} \leq \dfrac{1}{4(\Phi^{-1}(0.9) + 1.645)^2} \approx 0.0292$.

To find the minimum total sample size $N = n + m$, we consider to minimize $n + m$ under a certain value of $\dfrac{1}{n} + \dfrac{1}{m}$.

By the mean inequality, we have $\dfrac{2}{\frac{1}{n} + \frac{1}{m}} \leq \dfrac{n + m}{2}$, the equality holds if and only if $n = m$.

Therefore, when $n = m$, we can find the minimum value of $n + m$:

$\dfrac{1}{n} + \dfrac{1}{m} = \dfrac{1}{2n} \leq 0.0292 \Rightarrow n \geq \dfrac{2}{0.0292} \approx 68.4932 \to N = 2n \geq 136.9864$

$\therefore$ The minimum total sample size $N = 137$.