STAT 3004: Solutions of Assignment 1

Cheng Li
28 Sep 2020

---

1 Problem 1

1.1 (7.56)

We will use the paired t test

1.2 (7.57)

We wish to test the hypothesis:

$$H_0 : \mu_d = 0,$$
$$H_1 : \mu_d \neq 0,$$

where $\mu_d =$ mean of 4 year LVM - mean of Baseline LVM. So, we have the following equations,

$$\bar{d} = 18.9g,$$
$$s_d = 26.4g,$$

Thus,

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{10}}} = \frac{18.9}{\frac{26.4}{\sqrt{10}}} = 2.264,$$

The p-value $= 2 * Pr(t_9 > 2.264)$. From the distribution of t_9, we can get the p-value is less than 0.05. Therefore, we can say that there is a significant increase in LVM over 4 years.

1.3 (7.58)

We first find a 95% CI for μ_d given by:

$$\bar{d} \pm \frac{t_{n-1,.975} * s_d}{\sqrt{n}} = 18.9 \pm t_{9,.975} * 8.348 = 18.9 \pm 2.262 * 8.348 = (0, 37.8),$$

Thus, the CI for μ_d is (0, 37.8).

1.4 (7.59)

We use the sample size formula:

$$n = \frac{\sigma^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2},$$

In this case, $\Delta = 10g$, σ is approximated by $s_d = 26.4g$,$z_{1-\alpha/2} = z_{.975} = 1.96$, $z_{1-\beta} = z_{.80} = 0.84$. Thus,

$$n = 54.6 \approx 55.$$

So, we need to study 55 subjects in the main study to achieve 80% power.

2 Problem 2

2.1 (7.12)

We have such a hypothesis:

$$H_0 : p = p_0,$$
$$H_1 : p \neq p_0,$$

where $p_0 = 0.005$, p is the true incidence rate of MI in 2010 among 45-54-year-old men. Since $np_0q_0 = 5000 * 0.005 * 0.995 = 24.88 > 5$, we can use the normal-theory method. So, we can get the test statistics as following:

$$\begin{aligned} z &= \frac{\hat{p} - p_0}{\sqrt{p_0 q_0 / n}} \\ &= \frac{15/5000 - 0.005}{\sqrt{0.005 * 0.995/5000}} \\ &= -2.005 < -1.96 = z_{0.025}, \end{aligned}$$

Therefore, we should reject H_0 at the 5% level.

2.2 (7.13)

The p-value is:

$$\begin{aligned} p &= 2 * \Theta(z) \\ &= 2 * (1 - \Theta(1 - 2.005) \\ &= 0.045. \end{aligned}$$

2.3 (7.14)

From the problem, we can get a new hypothesis:

$$H_0 : p_{death} = p_{(0,death)},$$
$$H_1 : p_{death} \neq p_{(0,death)},$$

where $p_{(0,death)} = 0.25$. Since $n_{MI} p_{(0,death)} q_{(0,death)} = 2.81 < 5$, we must use the exact method to test the hypothesis. Since $\hat{p} = \frac{5}{15} > p_{(0,death)}$, the two-tailed p-value is obtained from

$$\begin{aligned} p &= 2 * \sum_{k=5}^{15} \binom{15}{k} (0.25)^k (0.75)^{(15-k)} \\ &= 2 * (1 - \sum_{k=0}^{4} \binom{15}{k} (0.25)^k (0.75)^{(15-k)}) \\ &= 0.627 > 0.05, \end{aligned}$$

therefore, there is no significant change in the case-fatality rate between 2000 and 2010.

2.4 (7.15)

We can use the power formula in Equation 7.32 using a two-sided formulation whereby

$$Power = \Phi[\sqrt{\frac{p_0q_0}{p_1q_1}}(z_{\alpha/2} + \frac{|p_0 - p_1|\sqrt{n}}{\sqrt{p_0q_0}}],$$

where $p_0 = 0.25, p_1 = 0.2, \alpha = 0.05, n = 50$. So, we have

$$\begin{aligned} Power &= \Phi(-1.238) \\ &= 1 - \Phi(1.238) \\ &= 0.11, \end{aligned}$$

thus, such a study would only have an 11% chance of detecting a significant difference.

2.5 (7.16)

We use the formula in Equation 7.33 using a two-sided formulation whereby

$$n = \frac{p_0q_0(z_{1-\alpha/2} + z_{1-\beta}\sqrt{\frac{p_1q_1}{p_0q_0}})^2}{(p_1 - p_0)^2},$$

where $\beta = 0.1$. So, we have

$$\begin{aligned} n &= \frac{0.1875[1.96 + 1.28(0.9238)]^2}{0.0025} \\ &= 740.6 \approx 741, \end{aligned}$$

thus, we need to study 741 MI case to achieve 92% power.

3 Problem 3

3.1

From the problems, we can get the hypothesis

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = 0, \\ H_1 &: \mu_1 - \mu_2 \neq 0, \end{aligned}$$

and the estimators of sample variance s^2 and sample $\hat{\mu}$ mean

$$\begin{aligned} s^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \\ \hat{\mu} &= \hat{\mu_1} - \hat{\mu_2}, \end{aligned}$$

where $\hat{\mu_1} = 6.56, s_1 = 0.64, \hat{\mu_2} = 6.80\ and\ s_2 = 0.76$.
Therefore, we can get the pivotal function

$$t = \frac{\hat{\mu}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

thus, $t \approx -1.313$. Then, we can get the p-value is 0.1940 larger than 0.05. As a result, we think that there is not a significant difference between the two groups.

3.2

Therefore, the CI is

$$\hat{\mu} \pm t_(63, 0.975)\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_2})},$$

where $t_(63, 0.975) \approx 1.9983$. So, the CI is $[-0.61, 0.13]$.

4 Problem 4

4.1

We should use F-test to compare the standard deviation of diet record vitamin C intake between current smokers vs. nonsmokers. The hpyothesis is

$$H_0 : \sigma_1 = \sigma_2,$$
$$H_1 : \sigma_1 \neq \sigma_2,$$

and the $F = (\frac{s_1}{s_2})^2 = 3.67$, the p-value is 0.002 larger than 0.05, therefore we should reject the null hpyothesis.

4.2

We need to use a two-sample t-test with unequal variance. So, we can get that

$$\begin{aligned} t &= \frac{(\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{306} + \frac{s_2^2}{17}}}, \\ &= 5.07, \end{aligned}$$

and our approximate degrees of freedom is

$$\begin{aligned} d' &= \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1 - 1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2 - 1}}, \\ &= 23.13, \end{aligned}$$

so, we take $d' = 23$, p-value is less than 0.05. Thus, we reject null hypothesis.