P8130_hw2

xj2249

2019/10/1

Problem1

a)

Let X denote the number of people who develop uveal melanoma in a given year. As we already know, X follows a binomial distribution as followed:

$$X \sim B(8.5 \times 10^6, 5 \times 10^{-6})$$

Therefore

$$P(X = 30) = \binom{8.5 \times 10^6}{30}(5 \times 10^{-6})^{30}(1 - 5 \times 10^{-6})^{(8.5\times10^6-30)} = 0.0093431$$

b)

The population of Asians, non-Hispanic Whites and Black are 1.19×10^6, 3.638×10^6, 2.0655×10^8, and therefore, X_{Asians},$X_{non-HispanicWhites}$ and X_{Black} separately follow binomial distributions as followed:

$$X_{Asians} \sim B(1.19 \times 10^6, 0.39 \times 10^{-6})$$
$$X_{non-HispanicWhites} \sim B(3.638 \times 10^6, 6.02 \times 10^{-6})$$
$$X_{Black} \sim B(2.0655 \times 10^6, 0.31 \times 10^{-6})$$

Therefore

$$P(X_{Asians} = 30) = \binom{1.19 \times 10^6}{30}(0.39 \times 10^{-6})^{30}(1 - 0.39 \times 10^{-6})^{(1.19\times10^6-30)} = 2.3603543 \times 10^{-43}$$

$$P(X_{non-HispanicWhites} = 30) = \binom{3.638 \times 10^6}{30}(6.02 \times 10^{-6})^{30}(1 - 6.02 \times 10^{-6})^{(3.638\times10^6-30)} = 0.0189991$$

$$P(X_{Black} = 30) = \binom{2.0655 \times 10^6}{30}(0.31 \times 10^{-6})^{30}(1 - 0.31 \times 10^{-6})^{(2.0655\times10^6-30)} = 3.0887185 \times 10^{-39}$$

Problem2

a)

Intervention Group

The hypothesis is:

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$t_{stats} = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{-0.76}{1.44/\sqrt{36}} = -3.17$$

Because $|t_{stats}| = 3.17 > t_{35,0.975} = 2.03$, at 5% significance level, we reject H_0 and conclude that there's a difference of BMI between 6-months follow-up and baseline among intervention group.

Control Group

The hypothesis is:

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$t_{stats} = \frac{\bar{d} - 0}{s_d/\sqrt{n}} = \frac{0.28}{0.97/\sqrt{36}} = 1.73$$

Because $|t_{stats}| = 1.73 < t_{35,0.975} = 2.03$, at 5% significance level, we fail to reject H_0 and conclude that there's no enough evidence to support a difference of BMI between 6-months follow-up and baseline among control group.

b)

Now perform a test to compare the BMI absolute changes between the two groups. ### Test for Equality of Variances The hypothesis is:

$$H_0 : \sigma_1 = \sigma_2 \quad vs \quad H_1 : \sigma_1 \neq \sigma_2$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$F_{stats} = \frac{s_1^2}{s_2^2} = \frac{1.44^2}{0.97^2} = 2.204 \sim F_{35,35}$$

Because $F_{stats} = 2.204 > F_{35,35,0.975} = 1.961$, at 5% significance level, we reject H_0 and conclude that there's a difference of variances of BMI change between intervention and control group.

Two-Sample Independent t-test with Unequal Variances

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$t_{stats} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{-0.76 - 0.28}{\sqrt{\frac{1.44^2}{36} + \frac{0.97^2}{36}}} = 3.594$$

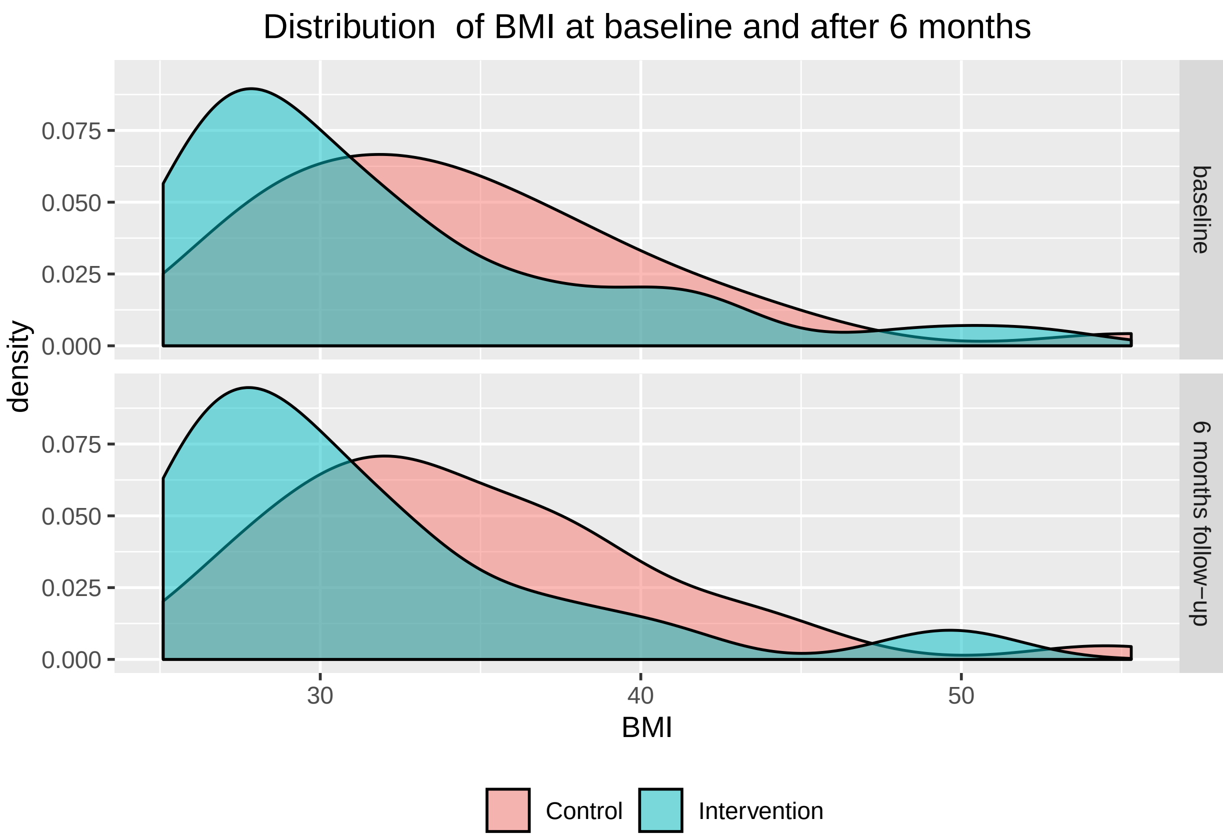
$$d' = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{(\frac{s_1^2}{n_1})^2/(n_1 - 1) + (\frac{s_2^2}{n_1})^2/(n_2 - 1)} = \frac{(\frac{1.44^2}{36} + \frac{0.97^2}{36})^2}{(\frac{1.44^2}{36})^2/(36 - 1) + (\frac{0.97^2}{36})^2/(36 - 1)} = 61.340 = 61$$

Because $t_{stats} = 3.594 > t_{61,0.975} = 2$, at 5% significance level, we reject H_0 and conclude that there's a difference of BMI change between intervention and control group.

c)

The assumption is that the distribution of BMI is normal.

i) Check the normality



As we can see in the plot above, the distribution of BMI is right-skewed. In another word, the normality assumption is (very likely) invalid.

ii) Effect of non-normality and remedies

Our tests are based on the normality assumption, which decide the underlying distribution of BMI. Therefore, non-normality invalidate our tests and may distort the truth. Fortunately, we do have alternatives,such as **Non-parametric test** and **Transformation**.

Problem3

Let X denote the number of restaurants that close by the end of 2019. As we know,

$$X \sim B(20, 0.60)$$

Therefore, use exact method:

$$P(X \geq 10) = 1 - P(X < 10) = 1 - F(9) = 0.872$$

- Poisson approximation: since n=20 is not large and p=0.60 is not small, it's inappropriate to use poisson approximation to binomial.
- Normal approximation: since n(1-p) = 8 < 10, it's also inappropriate use normal approximation.
- According to the result, the probability that more than 10 restaurants will close by the end of 2019 is 87.2%.

Problem4

a)

Test for normality

According to *Shapiro-Wilk test*, the distributions of increase of sleep (in hours) in both group are normal.

Test for Equality of Variances

The hypothesis is:

$$H_0 : \sigma_1 = \sigma_2 \quad vs \quad H_1 : \sigma_1 \neq \sigma_2$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$F_{stats} = \frac{s_1^2}{s_2^2} = \frac{3.20}{4.01} = 0.798 \sim F_{9,9}$$

Because $F_{9,9,0.025} = 0.248 < F_{stats} = 0.798 < F_{9,9,0.975} = 4.026$, at 5% significance level, we fail to reject H_0 and conclude that there's no enough evidence to support a difference of variances of sleep time increase between groups.

Two-Sample Independent t-test with Equal Variances

$$H_0 : \mu_{drug1} = \mu_{drug1} \quad vs \quad H_1 : \mu_{drug1} < \mu_{drug2}$$

With the significance level $\alpha = 0.05$, compute the test statistic:

$$t_{stats} = \frac{\bar{X_1} - \bar{X_2}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{0.75 - 2.33}{\sqrt{3.605(\frac{1}{10} + \frac{1}{10})}} = -1.861$$

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)} = \frac{(10 - 1) \times 3.20 + (10 - 1) \times 4.01}{(10 + 10 - 2)} = 3.605$$

Because $t_{stats} = -1.861 < t_{18,1-0.95} = -1.67$, at 5% significance level, we reject H_0 and conclude drug2 has a better efficacy than drug1 in increasing sleep time.

b)

According to the formula

$$95\%CI = (\bar{X} - t_{1-0.05/2,9}S/\sqrt{n}, \bar{X} + t_{1-0.05/2,9}S/\sqrt{n})$$

* For drug1:

$$95\%CI_{drug1} = (0.75 - 2.26 \times 1.79/\sqrt{10}, 0.75 + 2.26 \times 1.79/\sqrt{10}) = (-0.529, 2.029)$$

* For drug2:

$$95\%CI_{drug2} = (2.33 - 2.26 \times 2.00/\sqrt{10}, 2.33 + 2.26 \times 2.00/\sqrt{10}) = (0.898, 3.762)$$

c)

According to the formula

$$Power = \Phi(Z_\alpha + \frac{|\Delta|}{\sigma/\sqrt{n}}) = \Phi(-1.64 + \frac{1.58}{\sqrt{3.605}/\sqrt{10}}) = 0.839$$

- The posterior power is 0.839, meaning

d) PROs and CONs of using a posteriori/post-hoc power analysis

- PROs: When the test result of a difference turns out to be non-significant, using a posteriori/post-hoc power analysis helps us to interpret the result. For instance, if the difference is true, but the power is not enough high, then we will get a non-significant difference, which is actually due to insufficient power to detect rather than no true difference.
- CONs: