生物統計 HW 9 解答

#1:

The distribution of diastolic blood pressures for the population of female diabetics between the ages of 30 and 34 has an unknown mean μ_d and standard deviation $\sigma_d = 9.1$ mm Hg. It may be useful to physicians to know whether the mean of this population is equal to the mean diastolic blood pressure of the general population of females in this age group. 74.4 mm Hg [12].

(a)

What is the null hypothesis of the appropriate test?

Sol):

$H_0 : \mu = 74.4$

(b)

What is the alternative hypothesis?

Sol):

$H_1 : \mu \neq 74.4$

(c)

A sample of ten diabetic women is selected; their mean diastolic blood pressure is $\bar{x}_d = 84$ mm Hg. Using this information, conduct a two-sided test at the $\alpha = 0.05$ level of significance. What is the p-value of the test?

Sol):

$$p-value = P(|Z|>|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}|) = P(|Z|>3.336) = 2\times P(Z>3.336) = (1-0.9996)\times 2 = 0.0008$$

(注意：

$P(|Z|>3.336) = P(Z>3.336)+P(Z<-3.336) = P(Z>3.336)\times 2 = P(Z<-3.336)\times 2$

若是有加絕對值之後，算式就不要再寫乘以 2)

(d)

What conclusion do you draw from the results of the test?

Sol):

$\because$ p-value $= 0.0008 < \alpha = 0.05$, $\therefore$ *reject* H_0 at $\alpha = 0.05$ level

Conclusion: the mean of this population isn't equal to the mean of diastolic blood pressure.

(e)

Would your conclusion have been different if you had chosen $\alpha = 0.01$ instead of $\alpha = 0.05$?

Sol):

$\because$ p-value $= 0.0008 < \alpha = 0.01$, $\therefore$ *reject* H_0 at $\alpha = 0.01$ level

Conclusion: the mean of this population isn't equal to the mean of diastolic blood pressure.

Therefore, the conclusion at $\alpha = 0.01$ is same as the above one.

#2:

E. canis infection is a tick-borne disease of dogs that is sometimes contracted by humans. Among infected humans, the distribution of white blood cell counts has an unknown mean μ and a standard deviation σ. In the general population, the mean white blood cell count is 7250/mm³ [13]. It is believed that persons infected with *E. canis* must on average have lower white blood cell counts.

(a)

What are the null and alternative hypotheses for a one-sided test?

Sol):

H_0: $\mu = 7250$ v.s. H_1: $\mu < 7250$

(p.s.有些同學寫 H_0: $\mu_0 = 7250$ v.s. H_a: $\mu_a < 7250$ ，通常習慣用 μ_0, μ_a 表示已知的值，要檢定的是 μ ，是未知的值。)

(b)

For a sample of 15 infected persons, the mean white blood cell count is $\bar{x} = 4767/\text{mm}^3$ and the standard deviation is $s = 3204/\text{mm}^3$ [14]. Conduct the test at the $\alpha = 0.05$ level.

Sol):

$$\bar{X} = 4767, s = 3204, n = 15, \alpha = 0.05, t_{0.05}(14) \approx 1.7613$$

檢定統計量: $T = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

檢定規則: *reject* H_0 *if* $T < -t_{n-1,\alpha}$

$$\because t = \frac{4767 - 7250}{3204/\sqrt{15}} \approx -3.0014 < -1.7613$$

$\therefore$ *reject* H_0 at $\alpha = 0.05$ level (p.s. 有些同學寫 $p-value=(0.001,0.005)<0.05$，這個寫法不嚴謹，因為 $p-value$ 是一個值，$(0.001,0.005)$ 是一個區間，一個值跟一個區間是無法比較大小的，只能寫 $p-value\in(0.001,0.005)$，"$\in$"在數學上是”屬於”的意思)

(c)

What do you conclude?

Sol):

表示有足夠證據顯示感染 E. canis 的人平均白血球量少於 $7250/mm^3$

(p.s. Conclusion 請用非統計的說法，本次作業中不扣分)

#3:

Body mass index is calculated by dividing a person's weight by the square of his or her height; it is a measure of the extent to which the individual is overweight. For the population of middle-aged men who later develop diabetes mellitus, the distribution of baseline body mass indices is approximately normal with an unknown mean μ and standard deviation σ. A sample of 58 men selected from this group has mean $\bar{x} = 25.0$ kg/m^2 and standard deviation $s = 2.7$ kg/m^2 [15].

(a)

Construct a 95% confidence interval for the population mean μ.

Sol):

$\bar{X} = 25, s = 2.7, n = 58, \alpha = 0.05, t_{57,0.025} \approx 2.0025$

$$100(1\text{-}\alpha)\%\ \text{C.I. of } \mu = (\bar{X} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}})$$

$$95\%\ \text{C.I. of } \mu \approx (25 \pm 2.0025 \times \frac{2.7}{\sqrt{58}}) = (24.2901, 25.7099)$$

(b)

At the 0.05 level of significance, test whether the mean baseline body mass index for the population of middle-aged men who do develop diabetes is equal to 24.0 kg/m^2, the mean for the population of men who do not. What is the p-value of the test?

Sol):

$H_0: \mu = 24$ v.s. $H_1: \mu \neq 24$

檢定統計量: $T = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim t_{n-1}$

$$p-value = P(|T| > |\frac{\bar{x} - \mu_0}{s/\sqrt{n}}|) = P(|T| > \frac{25-24}{2.7/\sqrt{58}}) = 2 \times P(T > 2.8207) \approx 2 \times 0.0033 = 0.0066$$

(c)

What do you conclude?

Sol):

$\because\ p-value = 0.0066 < 0.05$

$\therefore\ reject\ H_0\ at\ \alpha = 0.05\ level$

表示有足夠證據顯示有糖尿病的中年男性 BMI 值不等於 $24kg/m^2$

(d)

Based on the 95% confidence interval, would you have expected to reject or not to reject the null hypothesis? Why?

Sol):

若 μ_0 落在 95%信賴區間內，則不拒絕 H_0 ，反之則拒絕

$H_0 \ at \ \alpha = 0.05 \ level$

$\because \mu_0 = 24 \notin (24.2901, 25.7099)$

$\therefore reject \ H_0 \ at \ \alpha = 0.05 \ level$

#4:

Data from the Framingham Study allow us to compare the distributions of initial serum cholesterol levels for two populations of males: those who go on to develop coronary heart disease and those who do not. The mean serum cholesterol level of the population of men who do not develop heart disease is $\mu = 219$ mg/100 ml and the standard deviation is $\sigma = 41$ mg/100 ml [18]. Suppose, however, that you do not know the true population mean; instead, you hypothesize that μ is equal to 244 mg/100 ml. This is the mean initial serum cholesterol level of men who eventually develop the disease. Since it is believed that the mean serum cholesterol level for the men who do not develop heart disease cannot be higher than the mean level for men who do, a one-sided test conducted at the $\alpha = 0.05$ level of significance is appropriate.

(a)

What is the probability of making a type I error?

Sol):

H_0: $\mu = 244$ v.s. H_1: $\mu < 244$

$\Pr(reject \ H_0 | \ H_0 \ is \text{ true}) = 0.05$

(b)

If a sample of size 25 is selected from the population of men who do not go on to develop coronary heart disease, what is the probability of making a type II error?

Sol):

$\sigma = 41, n = 25$

檢定統計量: $Z = \dfrac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$

檢定規則: $reject \ H_0 \ if \ \text{Z} < -Z_{0.05} = -1.645$

$$\Pr(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > -1.645) = \Pr(\overline{X} > -1.645 + 244 \times 41/\sqrt{25})$$

$$\Pr(accept\ H_0 | H_1\ is\ \text{true}) = \Pr(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > -Z_\alpha \mid \mu = 219) = \Pr(\overline{X} > \mu_0 - Z_\alpha \frac{\sigma}{\sqrt{n}} \mid \mu = 219)$$

$$= \Pr(Z > -Z_\alpha + \frac{\mu_0 - 219}{\frac{\sigma}{\sqrt{n}}}) = \Pr(Z > -1.645 + \frac{244-219}{\frac{41}{\sqrt{25}}})$$

$$= \Pr(Z > 1.4038) \approx 0.0802$$

(c)

What is the power of the test?

Sol):

$$Power = 1 - \beta = 1 - 0.0802 = 0.9198$$

(d)

How could you increase the power?

Sol):

$$Power = \Pr(Z < -Z_\alpha + \frac{\mu_0 - \mu_a}{\frac{\sigma}{\sqrt{n}}})$$

$If\ -Z_\alpha + \frac{\mu_0 - \mu_a}{\frac{\sigma}{\sqrt{n}}} \uparrow$, then power $\uparrow$

<Method>

i. $(\mu_0 - \mu_a) \uparrow \rightarrow power\ \uparrow$
ii. $\sigma \downarrow \rightarrow power\ \uparrow$
iii. $n \uparrow \rightarrow power\ \uparrow$
iv. $\alpha \uparrow \rightarrow power\ \uparrow$

(e)

You wish to test the null hypothesis

H_0: $\mu \geq 244$ mg/100 ml (而 $\mu_0 = 244$ 同)

against the alternative

H_A: $\mu < 244$ mg/100 ml

at the $\alpha = 0.05$ level of significance. If the true population mean is as low as 219 mg/100 ml, you want to risk only a 5% chance of failing to reject H_0. How large a sample would be required?

Sol):

$$\Pr(\text{reject } H_0 \mid \mu = 219) \approx 0.95$$

$$0.95 \approx \Pr\left(Z < -Z_\alpha + \frac{\mu_0 - \mu_a}{\frac{\sigma}{\sqrt{n}}}\right) = \Pr\left(Z < -1.645 + \frac{244-219}{\frac{41}{\sqrt{n}}}\right)$$

$$-1.645 + \frac{244-219}{\frac{41}{\sqrt{n}}} \approx 1.645$$

$$\sqrt{n} \approx 5.3959 \to n \approx 29.1125$$

$$\therefore n = 29$$

(p.s. 因為改過題目後僅需要接近 0.95，所以寫”29”或”至少 30”的都給對。另外，樣本大小必為正整數，故回答非正整數的都算錯)

(f)

How would the sample size change if you were willing to risk a 10% chance of failing to reject a false null hypothesis?

Sol):

$$\Pr(\text{reject } H_0 \mid \mu = 219) \approx 0.90$$

$$0.90 \approx \Pr\left(Z < -Z_\alpha + \frac{\mu_0 - \mu_a}{\frac{\sigma}{\sqrt{n}}}\right) = \Pr\left(Z < -1.645 + \frac{244-219}{\frac{41}{\sqrt{n}}}\right)$$

$$-1.645 + \frac{244-219}{\frac{41}{\sqrt{n}}} \approx 1.2816$$

$$\sqrt{n} \approx 4.7996 \to n \approx 23.0364$$

$$\therefore n = 23$$

(p.s. 因為改過題目後僅需要接近 0.90，所以”23”或”至少 24”的都給對。另外，樣本大小必為正整數，故回答非正整數的都算錯)