Zeinab Hosseini. 401130596508 HW3

1.

$SE = \frac{\sigma}{\sqrt{n}} = \frac{90000}{\sqrt{81}} = 10000$, $\bar{x} = 800000$

n is smaller than 10 percent of all the housed in NY and also $n \geq 30$. Therefore, we can use the normal distribution and confidence interval.

* اعداد روی بازه ، حدودی هستند نه کاملاً دقیق.

a) For 98% confidence, we have

$z^* = 2.33$ (qnorm (0.01) 2.326348) :

$800000 - 2.33 \times 10000 < \mu < 800000 + 2.33 \times 10000$

$\Rightarrow 776700 < \mu < 823300$

b) For 95% confidence, we have

$z^* = 1.96$ (qnorm (0.025) 1.959964) :

$800000 - 1.96 \times 10000 < \mu < 800000 + 1.96 \times 10000 \Rightarrow$

$\Rightarrow 780400 < \mu < 819600$

c) For 90% " , we have $z^* = 1.64$ (qnorm (0.05) 1.644854)

$800000 - 1.64 \times 10000 < \mu < 800000 + 1.64 \times 10000 \Rightarrow$

$\Rightarrow 783600 < \mu < 816400$

d) For 50% confidence, we have

$z^* = 0.67$ qnorm(0.25) : 0.6744898

$80000 - 0.67 \times 70000 < \mu < 80000 + 0.67 \times 70000 \Rightarrow$

$\Rightarrow 793300 < \mu < 806700$

e) ما برای بازه اطمینان ۹۵ درصدی می‌گیریم ما ۹۵ درصد اطمینان داریم که متوسط خانه در نیویورک به طور متوسط بین ۷۸۰۴۰۰ و ۸۲۹۶۰۰ قرار دارد.

f) برای هر فاصله اطمینانی که درست بگیریم مثلاً گفت ما ۱۰۰ درصد اطمینان داریم اون بازه اطمینان مون مقدار میانگین رو در بر می‌گیره و هر چقدر که confidence level بیشتر باشه، confidence interval مون بیشتر شده یعنی با اطمینان بیشتری می‌تونیم بگیم که اون بازه مقدار میانگین واقعی رو در بر می‌گیره ولی خب دقتش کمتر و برعکس.

g) For 98% confidence, we have $z^* = 2.58$

(qnorm(0.005) 2.575829)

در واقع در نهایت از این فرمول استفاده می‌کنیم: $n = \left(\frac{z^* S}{ME}\right)^2$

margin of error (ME) $= z^* SE = z^* \frac{\sigma}{\sqrt{n}} = 2.58 \frac{90000}{\sqrt{n}} =$

$5000 \Rightarrow \sqrt{n} > 46.44 \rightarrow n > 2156.6736 \rightarrow n = 2157$

h) $\frac{\text{margin of error}_1}{\text{margin of error}_2} = \frac{z^* \frac{\sigma}{\sqrt{n_1}}}{z^* \frac{\sigma}{\sqrt{n_2}}} = \frac{\sqrt{n_2}}{\sqrt{n_1}} \Rightarrow \frac{5000}{2500} = \frac{\sqrt{n_2}}{46.44} \Rightarrow$

$n_2 = 8649$

2.

1. اولاً که در null hypothesis باید علامت تساوی داشته باشیم
نه کوچکتر.

2. مورد بعدی مقداری که در H_0 و H_A قرار می‌دهیم هر دو باید یکسان باشند.

3. در H_A باید از $\bar{x} > 7$ استفاده شود.

4. مورد آخر اینکه تنها این سوال راجع به بچه های دبیرستانی صحبت شده

$$H_0 : x = 7 \text{ hours}$$
$$H_A : x > 7 \text{ hours}$$

---

3. درصورت سوال گفته حداقل 5 سال آموزش نیاز است پس H_0 و H_A میشوند:

$$H_0 : \mu = 5 \text{ years}$$
$$H_A : \mu > 5 \text{ years}$$

ما برای اینکه بتوانیم از Hypothesis test استفاده کنیم باید شرایط استفاده‌اش رو چک کنیم:

Conditions:

1. Independence: Sampled observations must be independent
   - random sample / assignment
   - if Sampling without replacement, $n < 10\%$ of population

2. Sample size / skew : $n \geq 30$ وقتی چولگی زیاد نیست.

برای بخش استقلال، شرایط برقرار است ولی برای مورد 2، n کوچکتر از 30 می باشد
اما استثناً از این مورد چشم پوشی می‌کنیم تا بتوانیم از normal distribution استفاده کنیم.
(در سوالات بعدی فقط به طور کلی شرایط را خیلی ساده بررسی می‌کنیم).

$n = ۲۰$, $S = ۲,۲$ $\qquad SE = \frac{S}{\sqrt{n}} = \frac{2.2}{\sqrt{20}} = 0.492$

test statistic: $z = \frac{\bar{x} - \mu}{SE} = \frac{4.6 - 5}{0.492} = -0.813$

pnorm(-0.813)

از روی خروجی برای این مقدار z در R مقدار p-value می‌شود:

p-value = 0.208

چون مقدار p-value از $\alpha = 0,05$ بزرگ‌تر است ما fail to reject می‌شویم

و نمی‌توانیم این فرض را رد کنیم.

b) برای محاسبه confidence interval از این فرمول استفاده می‌کنیم:

$\bar{x} \pm \underbrace{z^* SE}_{\text{margin of error}}$ $\qquad \bar{x} = 4,6 \qquad SE = 0.492$

R: qnorm(0,975) = 1.959964

$\Rightarrow 4.6 - 1,959964 \times 0,492 < \mu < 4.6 + 1,959964 \times 0.492$

$\Rightarrow (3.63 , 5.56)$

تفسیر: ما ۹۵ درصد اطمینان داریم که در شهر لوجیک شهروندان به طور متوسط بین 3.63 تا 5.56 سال درس می‌خوانند. همچنین چون $\mu = 5$ در این بازه اطمینان قرار دارد، نمی‌توان آن را رد کرد.

c) همانطور که در بخش b دیدیم، در هر دو روش به پاسخ‌های یکسان رسیدیم. در بخش اول، مقدار p-value چون بیشتر از مقدار α شد پس ما fail to reject می‌شویم و در confidence interval، مقدار $\mu = 5$ در بازه اطمینان قرار می‌گیرد پس مجدداً نمی‌توانیم reject کنیم و به جواب‌های یکسان می‌رسیم.

4. $n = 52$ $\bar{x} = 98.2846$ $S = 0.6824$

a) condition :

1. independence → random sample /assignment ☑
   → $n < 10\%$ of population ☑

2. Sample size /skew : $n \geq 30$ ☑ $(n = 52)$

b) $H_0 : T = 98.6$ Fahrenheit

$H_A : T > 98.6$ Fahrenheit

c)

R : qnorm (0.01) : قدار* 2 می‌شود ، cl: 98 % برای

2.326348

$$SE = \frac{\sigma}{\sqrt{n}} \doteq \frac{0.6824}{\sqrt{52}} = 0.09463185$$

$$98.2846 - 2.33 \times 0.09463185 < \mu <$$

$$98.2846 + 2.33 \times 0.09463185 \Rightarrow$$

$$98.064 < \mu < 98.505$$

قدار ۹۸٫۶ در این بازه نیست ، پس H_0 ، reject می‌شود.

d)

برای آزمون دو طرفه :

$H_0 : \mu = ۹۸٫۶$, $H_A \neq ۹۸٫۶$

$$Z = \frac{۹۸٫۶ - ۹۸٫۲۸۴۶}{۰٫۰۹۴۶۳۱۸۵} = 3.3329$$

قدار p-value برای $P(Z \geq 3.3329)$ می‌شود حدوداً ۰٫۰۰۰۴۲۹

ولی چون آزمون دو طرفه است مقدار فوق را در سمت چپ هم حدود ۰٫۰۰۰۴

که چون این مقدار از $\alpha = ۰٫۰۲$ خیلی کوچک‌تر است ما با اطمینان بالایی این

فرض را رد می‌کنیم.

5. $n = 50$ $\quad S = 5.6$ $\quad \bar{y} = 25.9$

Condition:

1. independence $\rightarrow$ random sample /assignment ☑
   $\searrow$ $n < 10\%$ of population ☑

2. Sample size /skew: $n \geq 30$ ☑ $(n = 50)$

$H_0: \mu \geq 28$ $\rightarrow$ اینجا باید = باشه ولی فقط بیش ایراد نمی‌گیریم.

$H_A: \mu < 28$

$$SE = \frac{S}{\sqrt{n}} = \frac{5.6}{\sqrt{50}} = 0.7919596$$

a) For 90% confidence we have $z^* = 1.64$

R: qnorm(0.05)
1.644854

$$25.9 - 1.64 \times 0.7919596 < \mu < 25.9 + 1.64 \times 0.7919596 \Rightarrow$$

$$\Rightarrow 24.60 < \mu < 27.19$$

Based on the confidence interval, $\mu \geq 28$ is not present in this interval.

Therefore, the H_0 is rejected.

also for p-value: $z = \frac{25.9 - 28}{0.7919596} = -2.65160 \rightarrow$

p-value $= P(z < -2.65160) \approx 0{,}004 \rightarrow$

Because p-value is smaller than α, we can reject H_0.

b) $\mu = 27$ $\quad H_0: \mu = 27$ $\quad H_A: \mu < 27$

β = Type 2 error $= P(\text{Fail to reject} \mid \mu = \mu_a)$

Type 2 error $= P\left(-1{,}74 < \frac{\bar{y} - 28}{0{,}7919596}\right) \rightarrow 28 - 1{,}74 \times 0{,}7919596 = 26.70$

$$P\left(\frac{26.70 - 27}{0.7919596} \leq z\right) = P(-0.378 \leq z) \approx 0.64$$

c) خطای نوع دوم برای وقتی هست که ما رد نکنیم در حالی که باید رد کنیم، اینجا شبیه تست خطای نوع 2 هست چونکه ما رد نکرده ایم.

power = $1-\beta$

مقدار β احتمال خطای نوع دوم را نشان می دهد.

---

6. $\alpha = 0{,}05$ $\quad z^* = 1{,}64$

مثل بخش b سوال قبل محاسبه میشه:

Power = $1-\beta = P(\text{reject } H_0 \mid \mu = \mu_a)$

power: $P\left(\frac{\bar{y}-28}{0{,}7919596} < -1{,}64\right) = P(\bar{y} < 28 - 1.64 \times 0.7919596 = 26.70$

power 27: $P\left(z < \frac{26.7-27}{0.7919596}\right) = P(z < -0.378) = 0.35$

power 26: $P\left(z < \frac{26.7-26}{0.7919596}\right) = P(z < 0.88) = 0.81$

power 25: $P\left(z < \frac{26.7-25}{0.7919596}\right) = P(z < 2.14) = 0.98$

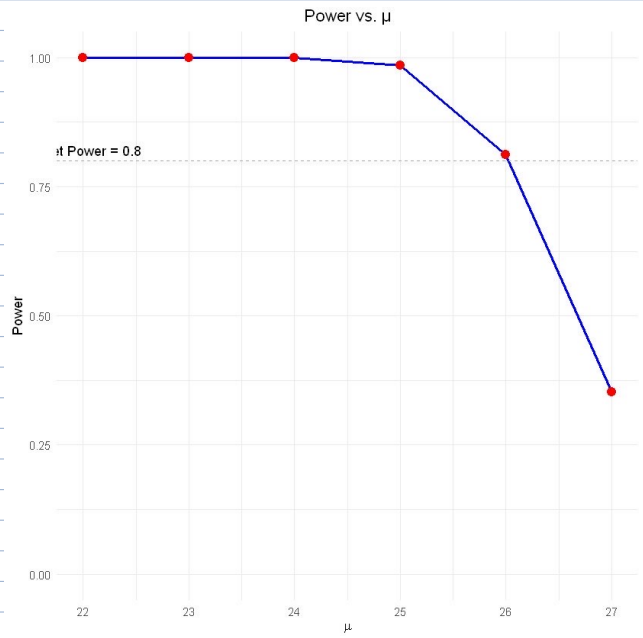
power 24: $P\left(z < \frac{26.7-24}{0.7919596}\right) = P(z < 3.40) = 1$

Power 23: $P\left(z < \frac{26.7-23}{0.7919596}\right) = P(z < 4.67) = 1$

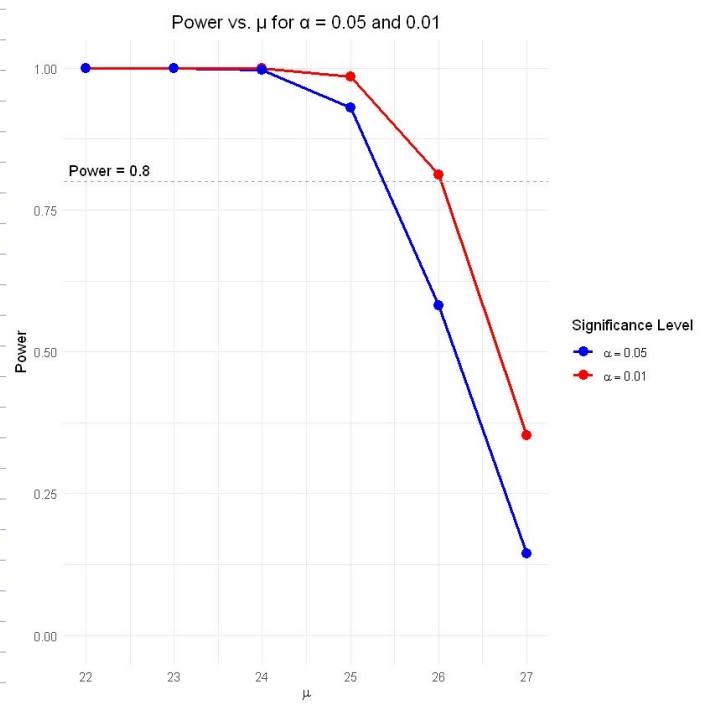
power 22: $P\left(z < \frac{26.7-22}{0.7919596}\right) = P(z < 5.11) = 1$

a.

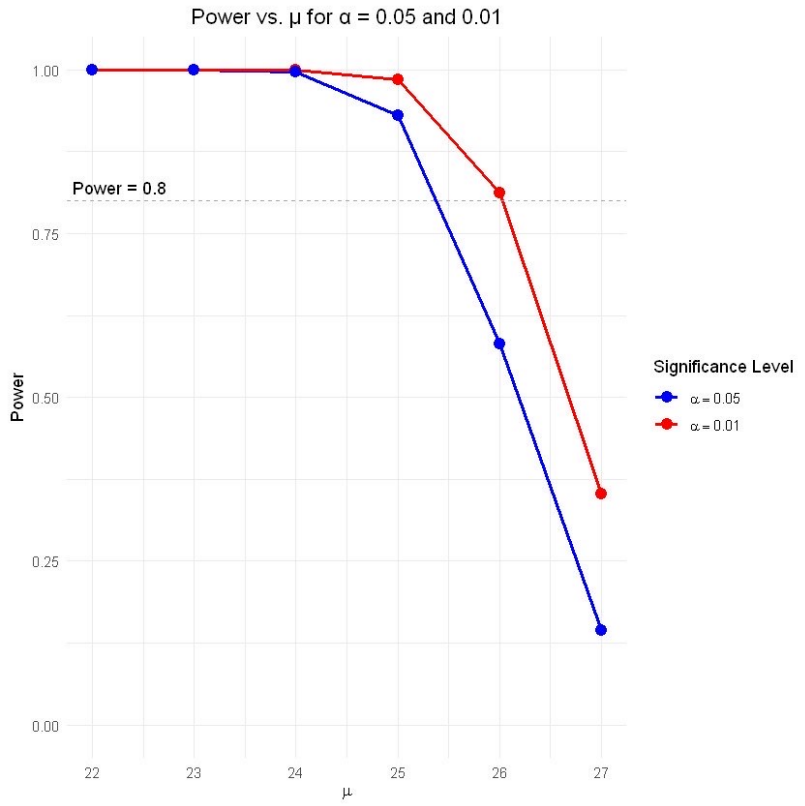
power همون طوری که قبل تر توضیح دادیم یعنی وقتی که باید رد کلیت بدیم، رد کلیت بکنیم،
برای این کنش می‌بینیم که هر چقدر فاصله‌ی بین μ_0 و μ بیشتر باشد احتمال آن افزایش پیدا می‌کند
و برای مقادیری با اختلاف کمتر زیاد این مقدار برابر با یک می‌شود. (effect size : $\mu_0 - \mu$) در واقع هر چه
این مقدار بیشتر باشد احتمال همون بیشتر در هر چقدر کمتر باشد، احتمال رد کلیت کردن H_0 هم کمتر میشه.
(احتمال رد کلیت کردن H_0 در واقع منظور وقتی گفتم احتمال)



b) $\alpha = 0.01$



c) $n = 20$, " " " smaller slope.



7. لطفا نوتبوکی که همراه فایل pdf ارسال شده رو چک کنید
پاسخ این سوال در فایل نوتبوک می باشد.