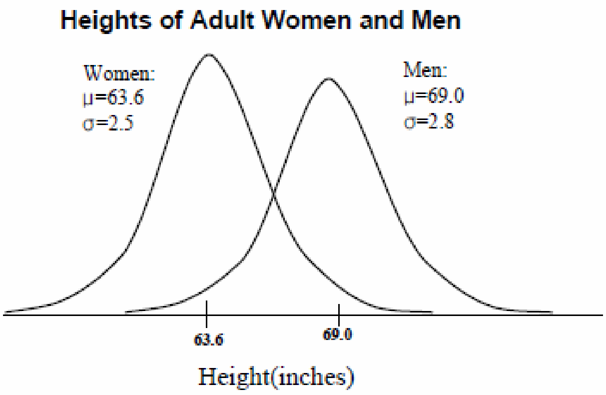
A Introduction to mathematical Statistics **Midterm 2 Solution**

1 The U.S. Army requires that women's heights be between 58 in. and 80 in.
(a) Find the percentage of US women satisfying that requirement assuming that women have heights that are normally distributed with a mean of 63.6 in. and a standard deviation of 2.5 in.
(b) Suppose it is felt that the population standard deviation is still σ = 2.5in, however, the population mean should be re-estimated given the better nutrition and more exercise nowadays. Please (1) first derive the general formula for the 100(1-α)% confidence interval for the population mean μ based on a random sample of size n, when the population is normal and the population standard deviation σ is known; (2) subsequently please derive the 95% confidence interval for the mean height of US women based on a newly obtained random sample of n = 200, and a sample mean of 65 in.

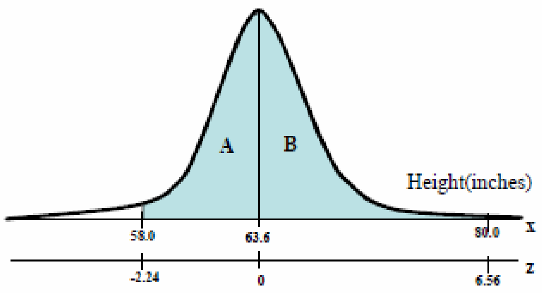
Solution:

(a) In the language of an undergraduate data analysis class, AMS315, we have:



$$z = \frac{x-\mu}{\sigma} = \frac{58.0 - 63.6}{2.5} = -2.24 \qquad z = \frac{80.0-63.6}{2.5} = 6.56$$

Area of regions A and B combined = 0.4875 + 0.4999 = 0.9874.

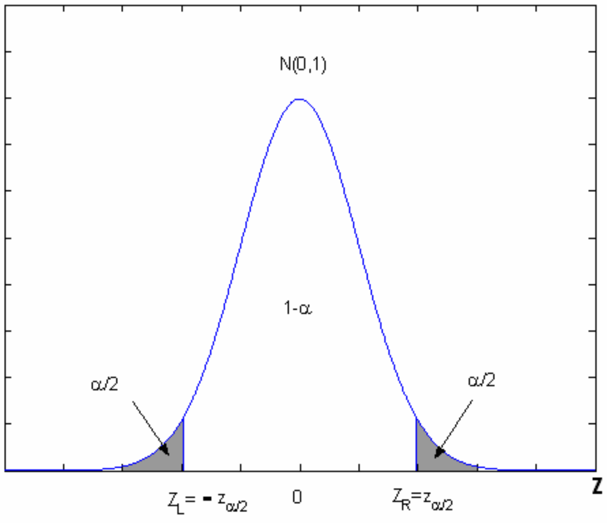


The above solution is acceptable. However, in the language of our mathematical statistics class (AMS312), I will recommend the following solution format:

Let $X \sim N\left(\mu = 53.6, \sigma^2 = (2.5)^2\right)$, and $Z \sim N(0,1)$, the percentage of women eligible for US army is:

$$P(58 \le X \le 80) = P\left(\frac{58-53.6}{2.5} \le \frac{X-53.6}{2.5} \le \frac{80-53.6}{2.5}\right) = P(-2.24 \le Z \le 6.56) \approx 0.4875 + 0.4999 = 0.9874$$

(b. 1)



Let α be any small positive value less than 1 (*usually less than 0.5), in the above figure, we have:

$$P(-Z_{\alpha/2} \le Z \le Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2} \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}) = 1-\alpha$$

$$P(-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \overline{X}-\mu \le Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

$$P(\overline{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

∴ The 100(1-α)% confidence interval for μ (when σ^2 is known and the population is normal) is $\left[\overline{X}-Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}+Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$

(b. 2) We have α = 0.05, $\bar{x} = 65$, $\sigma = 2.5$, n = 200. The 95% CI is

$$65 \pm 1.96\frac{2.5}{\sqrt{200}} = 65 \pm 0.35 \text{ or } (64.65, 65.35)$$

2. A biologist wishes to estimate the effect of an antibiotic on the growth of a particular bacterium by examining the mean amount of bacteria present per plate of culture when a fixed amount of the antibiotic is applied. Previous experimentation with the antibiotic on this type bacterium indicates that the standard deviation of the amount of bacteria present per plate is approximately $12\,cm^2$. (*Assume the population distribution is normal if necessary.)

(a) Please derive the general formula for sample size calculation in this type of scenario based on a maximum error of *E* at a confidence level of 100(1-α) %.

(b) Plug the information given in this particular problem to determine the number of plates necessary to estimate the mean amount of bacteria present within $6\,cm^2$ with a probability of 99%.

Solution: This is a sample size determination problem for the inference of μ based on the maximum error E.

(a) P.Q. $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$

$$P(|\bar{X} - \mu| \le E) = 1 - \alpha$$

$$P(-\frac{E}{\sigma/n} \le \frac{\bar{X} - \mu}{\sigma/n} \le \frac{E}{\sigma/n}) = 1 - \alpha$$

$$\frac{E}{\sigma/n} = Z_{\alpha/2} \Rightarrow n = \left(\frac{Z_{\alpha/2}\sigma}{E}\right)^2$$

(b)We just need to plug in the numbers:

E=6, $\alpha = 1 - 0.99 = 0.01$, $\sigma = 12$, $Z_{\alpha/2} = Z_{0.05} = 2.575$, $n = (\frac{2.575*12}{6})^2 = 27$

3. $Y_1, \cdots, Y_n \overset{i.i.d.}{\sim} U[0,\theta]$

(a) Find the MOME for θ

(b) Find the MLE for θ

Solution:

(a) $f(y) = \frac{1}{\theta}$, $0 \le y \le \theta$

$$E(Y) = \int_0^\theta y \cdot \frac{1}{\theta} dy = [\frac{y^2}{2\theta}]_0^\theta = \frac{1}{2\theta}[\theta^2 - 0] = \frac{\theta}{2}$$

$$E(Y) = \bar{Y}$$

$$\frac{\theta}{2} = \bar{Y} \Rightarrow \hat{\theta}_1 = 2\bar{Y}$$

(b) $L = \prod_{i=1}^{n} f(y_i) = (\frac{1}{\theta})^n$, $0 \le y_1, \cdots, y_n \le \theta$

$\ln L = -n \log \theta$, $0 \le y_1, \cdots, y_n \le \theta$

$\frac{d \ln L}{d\theta} = \frac{-n}{\theta} = 0 \Rightarrow \hat{\theta} = \pm\infty$? This is not good.

So, $0 \le y_1, \cdots, y_n \le \theta \Rightarrow 0 \le y_{(1)}, \cdots, y_{(n)} \le \theta$ $(\because \theta \ge Y_{(n)})$

$\therefore$ L is maximized when $\theta = Y_{(n)}$

$\therefore$ The MLE for θ is $\hat{\theta}_2 = Y_{(n)}$

4. In an upcoming election, in order to estimate the percent of votes candidate A will receive with a margin of error = 2%, how many voters should be polled? Please first derive the sample size calculation formula based on the maximum error E at the significance level of $(1-\alpha)$

Solution: First, we derive the general formula for sample size calculates for 1 population proportion *p* using the sample proportion $\hat{p}$ based on the ***maximum error E***.

Definition. $P(|\hat{p} - p| \le E) = 1 - \alpha$

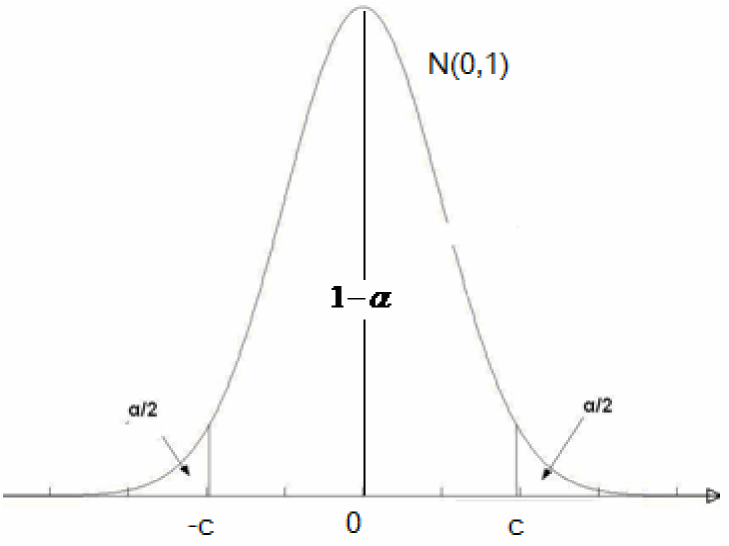
We want to estimate p **<u>within</u>** E with a probability of $(1-\alpha)$.

Derive the general formula for n

$P(|\hat{p}-p| \le E) = 1-\alpha$

$P(-E \le \hat{p}-p \le E) = 1-\alpha$

$$P\left(-\frac{E}{\sqrt{\frac{\hat{p}(\hat{p}-p)}{n}}} \le \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(\hat{p}-p)}{n}}} \le \frac{E}{\sqrt{\frac{\hat{p}(\hat{p}-p)}{n}}}\right) = 1-\alpha \text{ and } \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(\hat{p}-p)}{n}}} \dot{\sim} N(0,1)$$



$$c = Z_{\alpha/2} = \frac{E}{\sqrt{\frac{\hat{p}(\hat{p}-p)}{n}}}$$

$$\therefore n = \frac{(Z_{\alpha/2})^2\,\hat{p}(1-\hat{p})}{E^2} \le \frac{(Z_{\alpha/2})^2}{4\cdot E^2}$$

Next, in order to estimate the percent of votes Candidate A will receive with a margin of error = 2%, we calculate how many voters should be polled.

The margin of error is indeed the maximum error when $\alpha = 0.05$

Here margin of error = E = 0.02, and $\alpha = 0.05$.

Therefore we have: $n = \frac{(Z_{0.025})^2}{4E^2} = \frac{(1.96)^2}{4(0.02)^2} = 2401$