Homework #2

SOLUTION

Problem 1

Rubric:
> 2 points for correct procedure and answer.
> 0 points for correct answer but no procedure.
> 1 point for correct procedure but incorrect answer.

(a)

- $\overline{Y}_A = Y_1$

$$\text{Bias}(\overline{Y}_A) = \text{Bias}(Y_1) = E[Y_1] - \mu_Y = \mu_Y - \mu_Y = 0$$

$\therefore$ $\overline{Y}_A$ is unbiased.

- $\overline{Y}_B = \dfrac{Y_1 + 2Y_N}{3}$

$$\text{Bias}(\overline{Y}_B) = E[\overline{Y}_B] - \mu_Y = E\left[\frac{Y_1 + 2Y_N}{3}\right] - \mu_Y$$

$$= \frac{1}{3}E[Y_1] + \frac{2}{3}E[Y_N] - \mu_Y$$

$$= \frac{1}{3}\mu_Y + \frac{2}{3}\mu_Y - \mu_Y = 0$$

$\therefore$ $\overline{Y}_B$ is unbiased.

- $\bar{Y}_C = \frac{1}{M} \sum_{i=1}^{M} Y_i \qquad 1 < M < N$

$$Bias[\bar{Y}_C] = \frac{1}{M} E\left[\sum_{n=1}^{M} Y_N\right] - \mu_Y = \frac{1}{M} \sum_{n=1}^{M} E[Y_N] - \mu_Y$$

$$= \mu_Y - \mu_Y = 0 \qquad \therefore \bar{Y}_C \text{ is unbiased.}$$

- $\bar{Y}_N = \frac{1}{N} \sum_{i=1}^{N} Y_i$

Same argument as $\bar{Y}_C$, but with $M = N$

$\therefore \bar{Y}_N$ is unbiased.

(b)

Rubric:
> 2 points for correct procedure and answer.
> 0 points for correct answer but no procedure.
> 1 point for correct procedure but incorrect answer.

- $\bar{Y}_A = Y_1$

$$Var[\bar{Y}_A] = Var[Y_1] = \sigma_Y^2$$

- $\bar{Y}_B = \frac{Y_1 + 2Y_N}{3}$

$$Var[\bar{Y}_B] = Var\left[\frac{Y_1 + 2Y_N}{3}\right] = \frac{1}{9} Var[Y_1] + \frac{4}{9} Var[Y_N] = \frac{5}{9} \sigma_Y^2$$

- $\bar{Y}_C = \frac{1}{M}\sum_{i=1}^{M} Y_i \qquad 1 < M < N$

$$\mathrm{Var}[\bar{Y}_C] = \frac{1}{M^2}\sum_{i=1}^{M}\mathrm{Var}[Y_i] = \frac{M\sigma_Y^2}{M^2} = \frac{\sigma_Y^2}{M}$$

- $\bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Same argument as $\bar{Y}_C$, but with $M = N$

$$\mathrm{Var}[\bar{Y}_N] = \frac{\sigma_Y^2}{N}$$

c) Ranking

Rubric:
> 1 point for correct answer.

Better $\longrightarrow$

$$\underset{\bar{Y}_A}{\sigma_Y^2} \geq \underset{\bar{Y}_B}{\frac{5}{9}\sigma_Y^2} \geq \underset{\bar{Y}_C}{\frac{1}{M}\sigma_Y^2} \geq \underset{\bar{Y}_N}{\frac{1}{N}\sigma_Y^2}$$

Because $M \geq 2$

Because $M < N$

Problem 2

Y ... yield strength of aluminum alloy.

$$N = 100$$
$$\hat{\mu}_N = 92 \text{ MPa}$$
$$\hat{\sigma}_N = 8 \text{ MPa}$$

(a) N is large so we can assume $\overline{Y}_N$ is Gaussian.

$\gamma = 0.95$

$$\rho = \frac{\hat{\sigma}_N}{\sqrt{N}} \left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

Rubric:
> 1 point correct distribution.
> 1 point correct formula.
> 1 point correct answer.

```
N = 100
muhat = 92
sigmahat = 8
gamma = 0.95
rho = (sigmahat/np.sqrt(N))*abs( stats.norm.ppf((1-gamma)/2) )
print("rho = {:.2f}".format(rho))
print("interval = [{:.2f},{:.2f}]".format(muhat-rho,muhat+rho))
```

```
rho = 1.57
interval = [90.43,93.57]
```

(b)

Rubric:
> 1 point correct formula.
> 1 point correct answer.

```
N = 100
muhat = 92
sigmahat = 8
gamma = 0.99
rho = (sigmahat/np.sqrt(N))*abs( stats.norm.ppf((1-gamma)/2) )
print("rho = {:.2f}".format(rho))
print("interval = [{:.2f},{:.2f}]".format(muhat-rho,muhat+rho))
```

```
rho = 2.06
interval = [89.94,94.06]
```

(c)

Rubric:
> 2 point correct derivation.
> 1 point correct answer.

$\rho_{max} = 0.5$ MPa

$$\rho = \frac{\hat{\sigma}_N}{\sqrt{N}} \left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right| \leq \rho_{MAX}$$

$$\Leftrightarrow \quad N \geq \underbrace{\left(\frac{\hat{\sigma}_N}{\rho_{max}} \left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right| \right)^2}_{N_{min}} \quad \text{(assuming } N \text{ is large)}$$

```
sigmahat = 8
rhomax = 0.5
gamma = 0.95

Nmin = ((sigmahat/rhomax) * stats.norm.ppf((1-gamma)/2) )**2
Nmin
```

983.4134580976961

$\therefore N \geq 984.$

(d)

Rubric:
> 2 point correct derivation.
> 1 point correct answer.

```
sigmahat = 8
rhomax = 0.5
gamma = 0.99
Nmin = ((sigmahat/rhomax) * stats.norm.ppf((1-gamma)/2) )**2
Nmin
```

1698.5335298614307

$N \geq 1699$

Problem 3

$N = 8$

$\hat{\mu}_N = 78.37\,^\circ C$

$\hat{\sigma}_N = 0.14\,^\circ C$

(a) N is small and σ_Y unknown $\Rightarrow$ t distribution.

$\gamma = 0.95$

$$\rho = \frac{\hat{\sigma}_N}{\sqrt{N}} \left| \Phi^{-1}_{t(\nu)}\left(\frac{1-\gamma}{2}\right) \right|$$

Rubric:
> 1 point correct distribution.
> 1 point correct formula.
> 1 point correct answer.

```
N = 8
muhat = 78.37
sigmahat = 0.14
gamma = 0.98
rho = (sigmahat/np.sqrt(N))*abs( stats.t(df=N-1).ppf((1-gamma)/2) )
print("rho = {:.2f}".format(rho))
print("interval = [{:.2f},{:.2f}]".format(muhat-rho,muhat+rho))
```

```
rho = 0.15
interval = [78.22,78.52]
```

(b) That Y is Gaussian.

Rubric:
> 2 point correct answer.

Problem 4

(a)

Rubric:
> 1 point correct sample mean
> 1 point correct sample standard deviation

```
D = np.loadtxt('survey.txt')
muhat = D.mean()
sigmahat = D.std(ddof=0)   # 0 for biased, 1 for unbiased
N = D.shape[0]

muhat, sigmahat, N
```

```
(0.555, 0.4969657935914704, 400)
```

(b)

Rubric:
> 1 point correct formula.
> 1 point correct answer.

$N = 400 > 30 \Rightarrow$ Use normal distribution

$$\rho = \frac{\hat{\sigma}_N}{\sqrt{N}} \left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

```
gamma = 0.95
rho = (sigmahat/np.sqrt(N))*abs( stats.norm.ppf((1-gamma)/2) )
print("rho = {:.2f}".format(rho))
print("interval = [{:.2f},{:.2f}]".format(muhat-rho,muhat+rho))
```

```
rho = 0.05
interval = [0.51,0.60]
```

(c) $$\sqrt{N}\,\rho = \hat{\sigma}_N \left| \Phi^{-1}_{N}\left(\frac{1-\gamma}{2}\right)\right|$$

Rubric:
> 2 point correct derivation.
> 1 point correct answer.

Assume that the RHS remains unchanged when we increase N, since N is already very large.

So $\sqrt{N}\,\rho$ is constant.

$$\therefore \quad \sqrt{N_1}\,\rho_1 = \sqrt{N_2}\,\rho_2$$

$$\therefore \quad N_2 = N_1\left(\frac{\rho_1}{\rho_2}\right)^2$$

$$= 400\left(\frac{1}{0.5}\right)^2 = 1600.$$

$\therefore$ 1200 additional surveys.

Problem 5

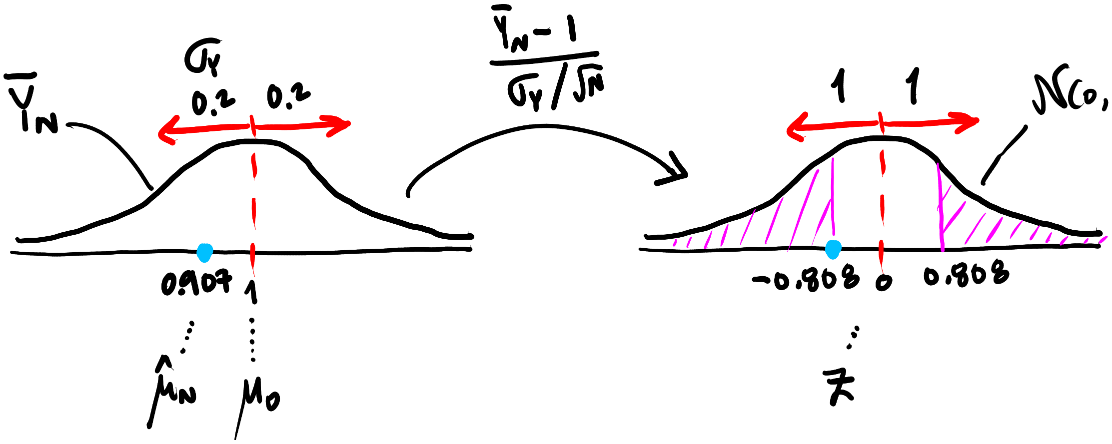
(a) $H_0 : \mu_Y = 1$

$H_1 : \mu_Y \neq 1$

Rubric:
> 1 point correct H0
> 1 point correct H1

(b)

Rubric:
> 1 point correct distribution (vs t dist)
> 1 point correct formulas
> 1 point correct answer

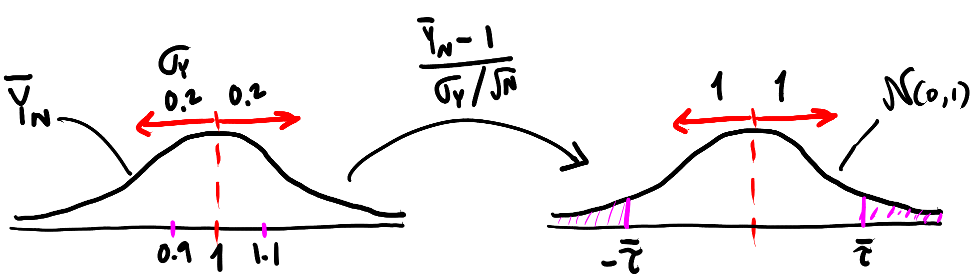



```
mu0 = 1
sigmaY = 0.2
D = np.array([0.92,0.84,0.96])
N = D.shape[0]
muhat = D.mean()

Z = (muhat-mu0) / (sigmaY/np.sqrt(N))
pvalue = 2*stats.norm.cdf(Z)
"{:.3f}, {:.3f}, {:.3f}".format(muhat, Z, pvalue)
```

'0.907, -0.808, 0.419'

(c)



$$\tau = \frac{1.1 - 1}{\sigma_Y/\sqrt{N}} = \left|\Phi_{\mathcal{N}}^{-1}(0.02)\right|$$

Rubric:
> 2 point correct formulas
> 1 point correct answer

$$\therefore \quad N = \left(\frac{\sigma_Y}{0.1}\,\Phi_{\mathcal{N}}^{-1}(0.02)\right)^2 = 16.87 \implies N \geq 17$$

Problem 6

$\alpha = 0.01$

a) $H_0: \mu_Y = 78.5$

$H_1: \mu_Y \neq 78.5$

Rubric:
> 1 point correct H0
> 1 point correct H1

b) t-statistic

Rubric:
> 2 point correct answer

c) $t = \dfrac{\hat{\mu}_N - 78.5}{\hat{\sigma}_N / \sqrt{N}} = -2.62$

$\text{pvalue} = 2\,\Phi_{I_t}(t) = 0.034$

Rubric:
> 1 point correct formula
> 1 point correct answer

d)

Rubric:
> 2 point correct answer

$\text{pvalue} > \alpha \implies$ cannot reject H_0

```python
N = 8
muhat = 78.37
sigmahat = 0.14
t = (muhat-78.5)/(sigmahat/np.sqrt(N))
pvalue = 2*stats.t(df=N-1).cdf(t)

t, pvalue
```

(-2.626396615835656, 0.0340906855788284)

Problem 7

$p \ldots$ proportion of nuts.

a) $\begin{cases} H_0: & p = 0.5 \\ H_1: & p \neq 0.5 \end{cases}$

Rubric:
> 1 point correct H0
> 1 point correct H1

b) $\left.\begin{array}{l} N^+ = 62 \\ N^- = 76 \end{array}\right\} \Rightarrow N = 138$

Rubric:
> 2 point correct answer

N is large, so we can use the Z statistic.
aka unit Gaussian.

c) $\hat{p} = \dfrac{N^+}{N} = 0.231$

Rubric:
> 2 point correct formula
> 1 point correct answer

$$Z = \frac{\hat{p} - 0.5}{\hat{\sigma}_N / \sqrt{N}} = \frac{\hat{p} - 0.5}{\sqrt{\hat{p}(1-\hat{p})} / \sqrt{N}} = -1.198$$

$$\text{pvalue} = 2\,\Phi_N(z) = 0.231$$

d) pvalue $> \alpha \Rightarrow$ cannot reject H_0.

Rubric:
> 2 point correct answer

Enjoy your 1-1 snack!