Homework #2
SOLUTION

Problem 1

(a)

- $\bar{Y}_A = Y_1$

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

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

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

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

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

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

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

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

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

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

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

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

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

(b)

- $\bar{Y}_A = Y_1$

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

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

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

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

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

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

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

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

Ranking

$$\xrightarrow{\text{Better}}$$

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

Problem 2 $\quad Y \sim \lambda e^{-\lambda y} \qquad \mathcal{D} = \{y_i\}_N \overset{iid}{\sim} Y$

$$\ln \mathcal{L}(\hat{\lambda}; \mathcal{D}) = \sum_{i=1}^{N} \ln P(y_i ; \hat{\lambda})$$

$$= \sum_{i=1}^{N} \ln\left(\hat{\lambda} \exp(-\hat{\lambda} y_i)\right)$$

$$= \sum_{i=1}^{N} \left(\ln \hat{\lambda} - \hat{\lambda} y_i\right)$$

$$= N \ln \hat{\lambda} - \hat{\lambda} \sum_{i=1}^{N} y_i$$

$$\frac{d}{d\hat{\lambda}} \ln \mathcal{L}(\hat{\lambda}; \mathcal{D}) = \frac{d}{d\hat{\lambda}} \left(N \ln \hat{\lambda} - \hat{\lambda} \sum_{i=1}^{N} y_i\right)$$

$$= \frac{N}{\hat{\lambda}} - \sum_{i=1}^{N} y_i$$

Stationary point:

$$\frac{N}{\hat{\lambda}} - \sum_{i=1}^{N} y_i = 0 \iff \hat{\lambda} = \frac{N}{\sum_{i=1}^{N} y_i} = \frac{1}{\hat{\mu}_N}$$

The MLE of λ is the inverse of the sample mean. $\blacksquare$

Problem 3

Y ... breaking strength of steel wires.

$\hat{\mu}_{100} = 50 \text{ kN}$

$\hat{\sigma}_{100} = 2 \text{ kN}$

$N = 100$ is "large"

(a)

$$\text{normalize}: \quad t = \frac{\bar{Y}_{100} - \mu_Y}{\sqrt{\hat{S}_N^2 / N}} \sim t(99)$$

- Since $N > 30$, $t(N-1) \overset{\sim}{=} \mathcal{N}(0,1)$
- So we can solve with either z tables or t tables.

z tables

$$\rho = \frac{\sqrt{N}}{\hat{\sigma}_N} \left| \Phi_{\mathcal{N}}^{-1}(0.025) \right| = \frac{2}{\sqrt{100}} \left| -1.959964 \right| \approx 0.392$$

With Python

```
gamma = 0.95
rho = (stdhat/np.sqrt(N))*abs(stats.norm().ppf((1-gamma)/2))
rho
```

```
0.3919927969080108
```

$$\therefore \; I_{N,\gamma} = 50 \pm 0.392 = [49.608, 50.392]$$

t tables

$$\rho \approx \frac{\sqrt{N}}{\hat{\sigma}_N} \left| \Phi^{-1}_{t(99)}(0.025) \right| = \frac{2}{\sqrt{100}} \left| -1.9842 \right| = 0.397$$

With Python

```
gamma = 0.95
rho = (stdhat/np.sqrt(N))*abs(stats.t(df=N-1).ppf((1-gamma)/2))
rho
```

```
0.39684339030173654
```

$$\therefore \; I_{N,\gamma} = 50 \pm 0.397 = [49.603, 50.397]$$

(b)

z:

```
gamma = 0.99
rho = (stdhat/np.sqrt(N))*abs(stats.norm().ppf((1-gamma)/2))
rho
```

```
0.5151658607097801
```

```
[muhat-rho, muhat+rho]
```

```
[49.48483413929022, 50.51516586070978]
```

t:

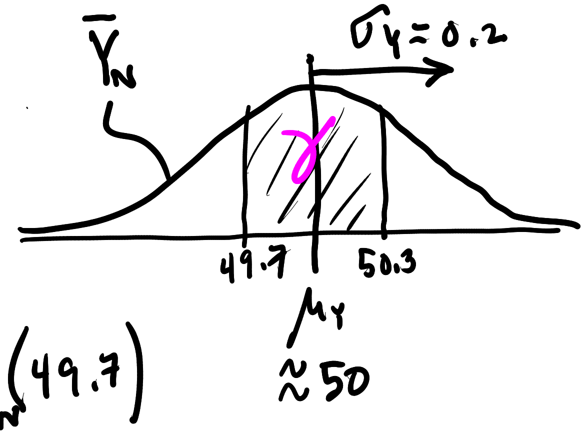
```
gamma = 0.99
rho = (stdhat/np.sqrt(N))*abs(stats.t(df=N-1).ppf((1-gamma)/2))
rho
```

```
0.5252810912770371
```

```
[muhat-rho, muhat+rho]
```

```
[49.47471890872296, 50.52528109127704]
```

(c)



$$P_{\bar{Y}_N}([49.7, 50.3]) = \Phi_{\bar{Y}_N}(50.3) - \Phi_{\bar{Y}_N}(49.7)$$

```python
Ybar = stats.norm(loc=muhat,scale=stdhat/np.sqrt(N))
Ybar.cdf(50.3) - Ybar.cdf(49.7)
```

```
0.8663855974622803
```

With tables: First normalize.

$$P_{\bar{Y}_N}([49.7, 50.3]) = P_{\mathcal{N}}\left(\left[\frac{49.7-\mu_Y}{\hat{\sigma}_N/\sqrt{N}}, \frac{50.3-\mu_Y}{\hat{\sigma}_N/\sqrt{N}}\right]\right)$$

$$= P_{\mathcal{N}}\left(\left[\frac{49.7-50}{0.2}, \frac{50.3-50}{0.2}\right]\right)$$

$$= P_{\mathcal{N}}([-1.5, 1.5])$$

$$= \Phi_{\mathcal{N}}(1.5) - \Phi_{\mathcal{N}}(-1.5)$$

$$= 1 - 2(1 - \Phi_{\mathcal{N}}(1.5))$$

$$= 1 - 2(1 - 0.9331927987)$$

$$= 0.866\ldots$$

(d)

$$N = \left(\frac{\sigma_Y}{\rho} \left| \Phi_{inv}\left(\frac{1-\gamma}{2} \right) \right| \right)^2$$

```
rho = 0.3
gamma = 0.95

( (stdhat/rho)*abs(stats.norm().ppf((1-gamma)/2)) )**2
```

```
170.73150314196113
```

N is at least 171.

(e)

```
rho = 0.3
gamma = 0.99

( (stdhat/rho)*abs(stats.norm().ppf((1-gamma)/2)) )**2
```

```
294.8842933787206
```

N is at least 295.

Problem 4

$N = 8$

$\hat{\mu}_Y = 3410.14\ °C$

$\hat{\sigma}_Y = 1.018\ °C$

(a)

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

```
: gamma = 0.95
  rho = (sigmahat/np.sqrt(N)) * abs( stats.t(df=N-1).ppf((1-gamma)/2) )
  rho
```

```
: 0.8510692980093996
```

```
: [muhat-rho, muhat+rho]
```

```
: [3409.2889307019905, 3410.9910692980093]
```

(b)

```
gamma = 0.98
rho = (sigmahat/np.sqrt(N)) * abs( stats.t(df=N-1).ppf((1-gamma)/2) )
rho
```

```
1.0790147882441998
```

```
[muhat-rho, muhat+rho]
```

```
[3409.0609852117555, 3411.2190147882443]
```

Problem 5

We do not have true standard deviation, so we should use the t distribution.
But on the other hand, N is large so we can use the Gaussian. Either is correct.

(a)

```
gamma = 0.95
rho = (sigmahat/np.sqrt(N)) * abs( stats.t(df=N-1).ppf((1-gamma)/2) )
[muhat-rho, muhat+rho]
```

```
[0.15839046830057163, 0.17160953169942839]
```

```
gamma = 0.95
rho = (sigmahat/np.sqrt(N)) * abs( stats.norm().ppf((1-gamma)/2) )
[muhat-rho, muhat+rho]
```

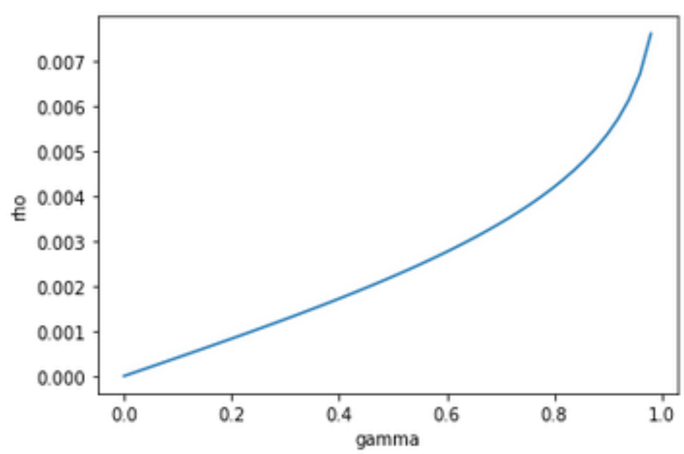
```
[0.15857216803577173, 0.17142783196422828]
```

(b)

```
gamma = np.linspace(0,1)
rho = (sigmahat/np.sqrt(N)) * abs( stats.norm().ppf((1-gamma)/2) )

plt.plot(gamma,rho)
plt.xlabel('gamma')
plt.ylabel('rho')
```

```
Text(0, 0.5, 'rho')
```

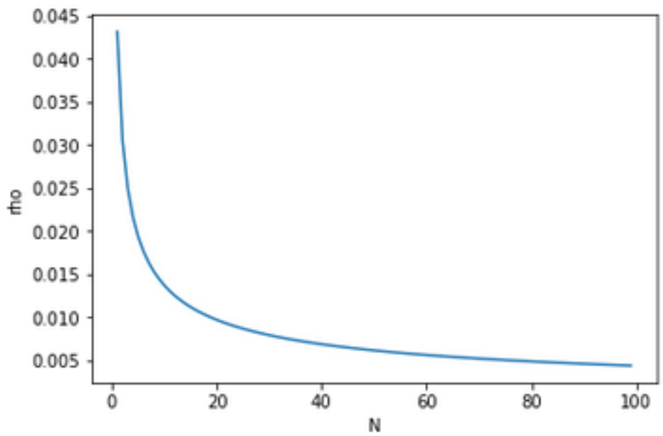


(c)

```
gamma = 0.95
N = np.arange(1,100)
rho = (sigmahat/np.sqrt(N)) * abs( stats.norm().ppf((1-gamma)/2) )

plt.plot(N,rho)
plt.xlabel('N')
plt.ylabel('rho')
```

```
Text(0, 0.5, 'rho')
```



(d) * This was a poorly posed problem. Full points will be awarded to all. See minilecture for explanation.

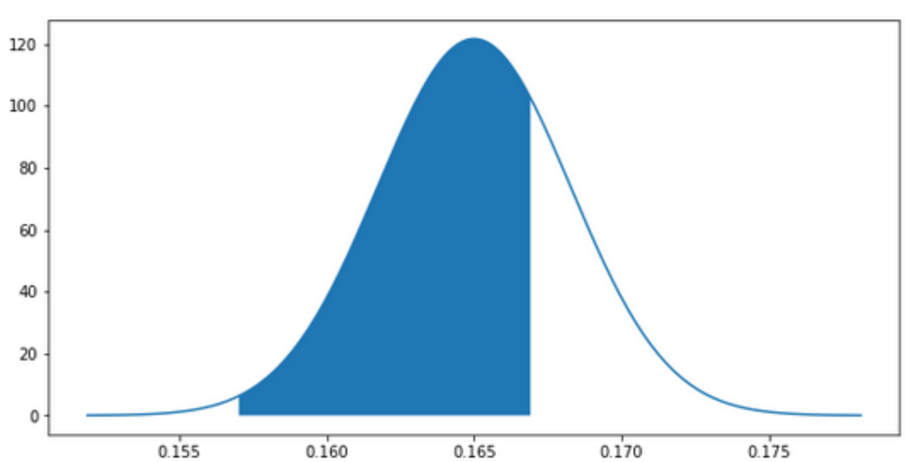
(e)

```
YbarN = stats.norm(loc=muhat, scale=sigmahat/np.sqrt(N))

y = np.linspace(muhat-4*sigmahat/np.sqrt(N),muhat+4*sigmahat/np.sqrt(N),200)

plt.figure(figsize=(10,5))
plt.plot(y,Y.pdf(y))
yint = y[(y>=0.157) & (y<=0.167)]
plt.fill_between(yint, Y.pdf(yint))
```

```
<matplotlib.collections.PolyCollection at 0x7f393397e400>
```



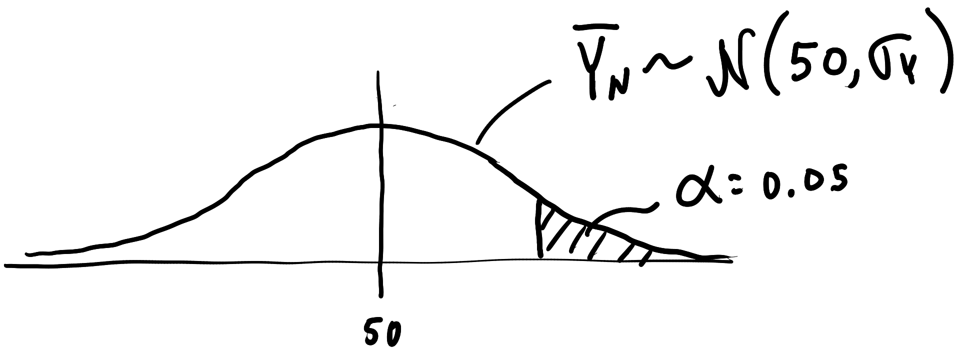
Problem 6

(a) $H_0 : \mu_Y = 50$

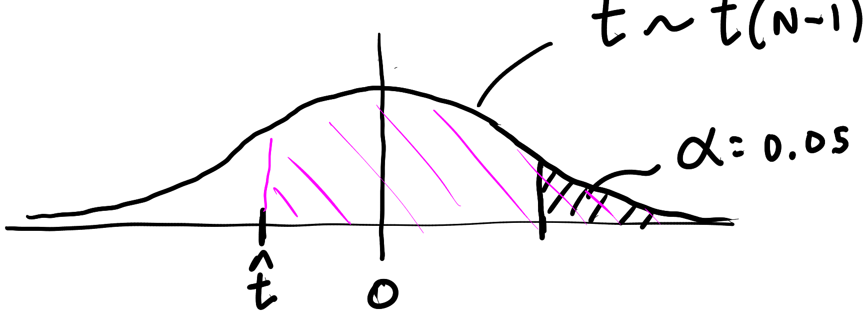
$H_1 : \mu_Y > 50$

(b) $$t = \frac{\bar{Y}_N - 50}{\sqrt{S_N^2 / N}} \sim t(N-1)$$

Assuming H_0 is true:



$$t = \frac{\bar{Y}_N - 50}{\sqrt{S_N^2 / N}}$$



(c) $\hat{t} = \dfrac{\hat{\mu}_N - 50}{\hat{\sigma}_N / \sqrt{N}} = -1.65$

```
y = np.array([48, 42, 58, 45, 51, 39, 45, 58, 43, 50, 43, 43])
N = y.shape[0]
muhat = y.mean()
sigmahat = y.std(ddof=1)
that = (muhat-50)/(sigmahat/np.sqrt(N))
that
```

```
-1.6490831626990412
```

$\text{P-value} = 1 - \phi_{t(v)}(\hat{t}) = 0.93$

```
1 - stats.t(df=N-1).cdf(that)
```

```
0.9363166928058132
```

(d) $\text{Pvalue} > \alpha \implies$ Fail to reject H_0

Problem 7

Not graded. See minilecture video for an explanation.

$$X \sim \mathcal{B}(p) \implies \begin{array}{l} \mu_X = p \\ \sigma_X^2 = p(1-p) \end{array}$$

Sample mean:

$$\bar{X}_N \sim \frac{\mathcal{B}in(N,p)}{N} \implies \begin{array}{l} \mu_{\bar{X}_N} = p \\ \sigma_{\bar{X}_N}^2 = \dfrac{p(1-p)}{N} \end{array}$$

By CLT, because N is large:

$$\tilde{X}_N \approx \mathcal{N}\left(p, \frac{\sqrt{p(1-p)}}{\sqrt{N}}\right)$$

"approximately distributed as"

(a) $\gamma = 0.98$

$\rho_{max} = 0.05$

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

$$= \frac{\sqrt{p(1-p)}}{\sqrt{N}} \left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|$$

We want $\rho \leq 0.05$

$$\therefore \quad \frac{\sqrt{p(1-p)}}{\sqrt{N}} \underbrace{\left| \Phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right|}_{= 2.326} \leq 0.05$$

$$\therefore \quad N \geq \frac{p(1-p)\,(2.326)^2}{(0.05)^2} = 2164.76 \cdot p(1-p)$$

But $p(1-p)$ has a maximum value of 0.25 on $p \in [0,1]$.

So, prior to collecting data, we must assume to worst.

$$N \geq 2164.76 \times 0.25 = 541.18$$

$$\therefore \underline{N \geq 542}$$

(b) $\mathcal{D} = \{\underbrace{1, 1, \ldots, 1}_{\text{30 failures caused by stress corrosion cracking}}, \underbrace{0, 0, \ldots, 0}_{\text{170 failures have other causes.}}\}$

$$\hat{p} = \frac{30}{200}$$

$$\hat{\sigma}_N^2 = \hat{p}(1-\hat{p}) = 0.15 \times 0.85 = 0.1275$$

$$\bar{X}_N \sim \mathcal{N}(\hat{p}, \hat{\sigma}_N / \sqrt{N})$$

$$\therefore \quad \rho = \frac{\hat{\sigma}_N}{\sqrt{N}} \left| \phi_N^{-1}\left(\frac{1-\gamma}{2}\right) \right| = 0.0587$$

```
N = 200
phat = 30/200
sigmahat = np.sqrt(phat*(1-phat))
gamma = 0.98
rho = (sigmahat/np.sqrt(N)) * np.abs( stats.norm.ppf((1-gamma)/2))
rho
```

```
0.05873740460555898
```

$$\therefore \quad I_{\gamma,N} = 0.15 \pm 0.0587$$

(c) $N \geq 2164.76 \cdot \hat{p}(1-\hat{p}) = 2164.76 \cdot 0.15 \cdot 0.85 = 276$

```
gamma = 0.98
rhomax = 0.05
phat = 30/200
phat*(1-phat) * ( stats.norm.ppf((1-gamma)/2) / rhomax )**2
```

```
276.00661598377127
```

$N = 276$ or 277