Homework 1 Chung Tsz Ting 1155110208

1)

a) $mean = \frac{2+3+6+8+11}{5} = 6$

b) $standard\ deviation = \sqrt{\frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}} = \sqrt{10.8} = 3.286$

c) As all possible samples of size 2 can be drawn with replacement, there are 25 samples, they are as below,

2,2	2,3	2,6	2,8	2,11
3,2	3,3	3,6	3,8	3,11
6,2	6,3	6,6	6,8	6,11
8,2	8,3	8,6	8,8	8,11
11,2	11,3	11,6	11,8	11,11

Table 1: Samples Data

The mean of each sample are,

2	2.5	4	5	6.5
2.5	3	4.5	5.5	7
4	4.5	6	7	8.5
5	5.5	7	8	9.5
6.5	7	8.5	9.5	11

Table 2: Mean of Samples Data

$$Thus,$$

$$mean\ of\ the\ sampling\ distribution\ of\ means = \frac{sum\ of\ each\ sample\ mean}{total\ number\ of\ sample}$$
$$= \frac{150}{25}$$
$$= 6$$

$$which\ is\ the\ same\ as\ the\ mean\ of\ the\ population\ and\ comply\ with\ theorem\ 1.$$

d) Taking the data in table 2 into calculation of formula of $\sqrt{\frac{\sum|x_i-\bar{x}|^2}{n}}$, where x_i is the data in table 2 and $\bar{x}$ is the mean of the data and n is the number of data.

$$standard\ deviation\ of\ the\ sampling\ distribution\ of\ means = 5.4$$
$$= \frac{10.8}{2}$$
$$= \frac{\sigma^2}{n}$$

which is comply with theorem 2.

Calculation procedure of 1(c) and (d):

```
import numpy as np
data=np.asarray([2,2.5,4,5,6.5,
                 2.5,3,4.5,5.5,7,
                 4,4.5,6,7,8.5,
                 5,5.5,7,8,9.5,
                 6.5,7,8.5,9.5,11])
print(data.sum())
print(data.mean())
print(((data-6)**2).sum()/25)
print(data.var())
```

```
150.0
6.0
5.4
5.4
```

2)

a) By theorem 1 and 2, the expected mean of the sampling distribution of means is 68.0 inches and the standard deviation is $\sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{3^2}{25}} = 0.6$ inches.

b) By theorem 1 and 3, the expected mean of the sampling distribution of means is 68.0 inches and the standard deviation is $\sqrt{\frac{\sigma^2(N-n)}{n(N-1)}} = \sqrt{\frac{3^2(3000-25)}{25(3000-1)}} = 0.5976$

3)

a) The Z Test is given by $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, with mean of sample $\bar{X}$, mean of population μ, standard deviation σ and sample size n.
Thus,
66.8 in standard units $= \frac{66.8-68.0}{0.6} = -2,$
68.3 in standard units $= \frac{68.3-68.0}{0.6} = 0.5$
According to the Z score table, -2 is corresponding to 2.28% and 0.5 is corresponding to 69.15%.
Required Probability = (area between z=66.8 and z=68.3)
= (area to the left of z=68.3) - (area to the left of z=66.8)
= 69.15% - 2.28%
= 66.87%
Number of samples expected to find between 66.8 and 68.3 = 80 * 66.87%
= 53.496

b) 66.4 in standard units $= \frac{66.4-68.0}{0.6} = -2.67$
According to the Z score table, -2.67 is corresponding to 0.38%
Required Probability = (area to the left of z=66.4)
=0.38%
Number of samples expected to find between 66.8 and 68.3 = 80 * 0.38%
= 0.304

4)

a) Given the population is binomially distributed, the probability p of getting a head in one toss is 0.5. The probability q of getting a tail is 1-p=0.5.
The mean μ_p is equal to p which is 0.5, the standard deviation $\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.5 \cdot 0.5}{120}} = 0.0456$ with sample size n.

The Z Test is given by $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, with mean of sample $\bar{X}$, mean of population μ, standard deviation σ and sample size n.
As the data is discrete instead of continuous, calculating the probability of having less than 40%*120=48 tosses and more than 60%*120=72 tosses be head in normal way didn't take the tosses in continue value into consideration, for example, 47.5, 47.6 and 47.67 tosses didn't be taken into consideration.
Thus, (1/2)(1/120)=1/240 should be deduced from the bound of 40% and added to 60%
Thus,
40% - 1/240 in standard units $= \frac{0.4-1/240-0.5}{0.0456} = -2.284$,
60% + 1/240 in standard units $= \frac{0.6+1/240-0.5}{0.0456} = 2.284$.
According to the Z score table, -2.284 is corresponding to 1.13% and 2.284 is corresponding to 98.87%.
Required Probability = (area to the left of z=40%) + (area to the right of z=60%)
= 1.13% + (100%-98.87%)
= 2.26%

b) 5/8 - 1/240 in standard units $= \frac{5/8-1/240-0.5}{0.0456} = 2.650$
According to the Z score table, 2.650 is corresponding to 99.60%.
Required Probability = (area to the right of z=5/8)
= 100% - 99.60%
= 0.4%

5)

a) Unbiased and efficient estimates of the true mean = $\frac{6.33+6.37+6.36+6.32+6.37}{5}$
= 6.35cm
b) Simple variance S^2 = $\frac{(6.33-6.35)^2+(6.37-6.35)^2+(6.36-6.35)^2+(6.32-6.35)^2+(6.37-6.35)^2}{5}$
= 0.00044
Unbiased and efficient estimates of the true variance $\hat{S}^2 = \frac{n}{n-1}S^2$
$= \frac{5}{5-1} \cdot 0.00044$
$= 0.00055\ cm^2$

6)

- $\bar{X} \pm 1.96\sigma_{\bar{X}}$ with sample mean $\bar{X}$ and standard deviation (standard error) $\sigma_{\bar{X}}$ have 95% confidence according to the z score table.
To have the confident of error of estimate not exceed 0.01,
$$1.96\sigma_{\bar{X}} = 0.01$$
$$1.96\frac{\sigma}{\sqrt{n}} = 0.01$$
$$1.96\frac{0.05}{\sqrt{n}} = 0.01$$
$$n = 96.04$$
Therefore, in order to be 95% confident that the error of his estimate won't exceed 0.01, the sample of measurement must be at least 97.

- $\bar{X} \pm 2.58\sigma_{\bar{X}}$ with sample mean $\bar{X}$ and standard deviation (standard error) $\sigma_{\bar{X}}$ have 99% confidence according to the z score table.
  To have the confident of error of estimate not exceed 0.01,
  $$2.58\sigma_{\bar{X}} = 0.01$$
  $$2.58\frac{\sigma}{\sqrt{n}} = 0.01$$
  $$2.58\frac{0.05}{\sqrt{n}} = 0.01$$
  $$n = 166.41$$
  Therefore, in order to be 99% confident that the error of his estimate won't exceed 0.01, the sample of measurement must be at least 167.

7)

a) $\bar{X} \pm 1.96\sigma_{\bar{X}}$ with sample mean $\bar{X}$ and standard deviation (standard error) $\sigma_{\bar{X}}$ have 95% confidence according to the z score table.
Thus, the 95% confidence limits for estimate of the mean of the 200 grades is
$$\bar{X} \pm 1.96\sigma_{\bar{X}} = \bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$
$$= 75 \pm 1.96\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}}$$
$$= 75 \pm 2.407$$

b) The z score of having the mean of all 200 grades be 75 ± 1 is,
$$Z_C\sigma_{\bar{X}} = 1$$
$$Z_C\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 1$$
$$Z_C\frac{10}{\sqrt{50}}\sqrt{\frac{200-50}{200-1}} = 1$$
$$Z_C = 0.814$$
The percentage of area corresponding to the z score of 0.814 is 79.10%. Thus, the degree of confidence is 2*(79.10% - 50%) = 58.2%.

8)

a) For the confidence limits of 50%, the percentage of area under the normal curve would be 75% and its corresponding z score is 0.67.
Thus, the probable error of the mean = $0.67\sigma_{\bar{X}}$
$$= 0.67\frac{\hat{S}}{\sqrt{n}}$$
$$= 0.67\frac{1}{\sqrt{n}}\frac{S\cdot\sqrt{n}}{\sqrt{n-1}}$$
$$= 0.67\frac{1}{\sqrt{50}}\frac{0.5\cdot\sqrt{50}}{\sqrt{50-1}}$$
$$= 0.0479$$

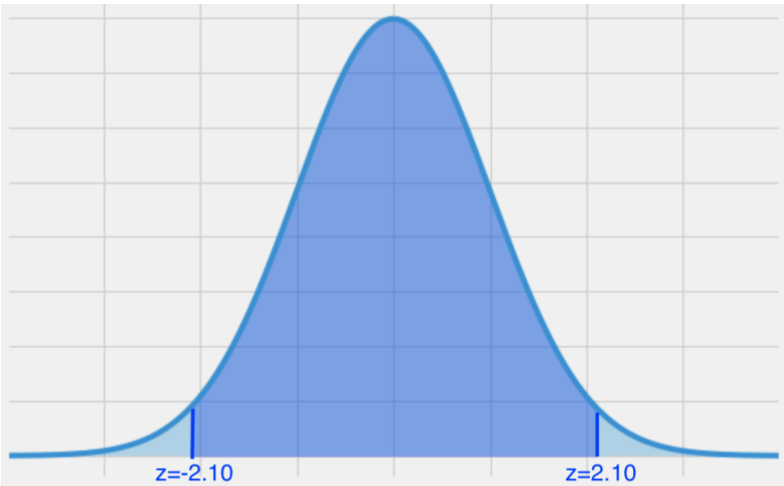
b) the 50% confidence limits = 18.2 ± 0.0479 V

9)

a) Given the population is binomially distributed, the probability p of getting a head in one toss is 0.5. The probability q of getting a tail is 1-p=0.5.
The mean μ_p is equal to np which is 100 · 0.5 = 50, the standard deviation $\sigma_p = n\sqrt{\frac{pq}{n}} = 100\sqrt{\frac{0.5 \cdot 0.5}{100}} = 5$ with sample size n.
The Z Test is given by $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, with mean of sample $\bar{X}$, mean of population μ, standard deviation σ and sample size n.
As the data is discrete instead of continuous, calculating the probability of having 40 to 60 head in 100 tosses in normal way didn't take the tosses in continue value into consideration.
Thus, 39.5 and 60.5 should be the range be placed into the formula above instead.
Thus,
39.5 in standard units $= \frac{39.5-50}{5} = -2.1$,
60.5 in standard units $= \frac{60.5-50}{5} = 2.1$.
According to the Z score table, -2.1 is corresponding to 1.79% and 2.1 is corresponding to 98.21%.
Required Probability = (area between z=-2.1 and z=2.1)
= (area to the left of z=2.1) - (area to the left of z=-2.1)
= 98.21% - 1.79%
= 96.42%

b) The probability of rejecting the hypothesis when it is actually correct = 1 - 96.42%
= 3.58%

c)



The z score is calculated based on part (a). The area in the left of z=-2.10 and in the right of z=2.10 (3.58% of the total area) shaded in light blue color is the area that reject the hypothesis (probability of Type 1 error of the decision rule). The area in between (96.42% of the total area) shaded in deep blue is the area the accept the hypothesis.

d) If the sample of 100 tosses yielded 53 heads, the hypothesis of the coin is fair is accepted because it lies between 40 and 60 heads. For the yielding of 60 heads, the hypothesis is also accept because of 60 is inclusive in the range of 40 to 60 heads in 100 tosses.

e) Yes, the conclusion in part (d) could be wrong (i.e. the coin is unfair) which is the Type 2 error in decision rule that the hypothesis is accepted when it happens to be false (i.e. false positive).