6 Sampling Distributions, Central Limit Theorem, and Estimation

Chapter Preview

This set of notes will examine statistics and their distributions along with the Central Limit Theorem. The Central Limit Theorem is arguably the most important theorem in Statistics and will bridge the gap between our study of Probability and Statistics. We will also examine estimators and their properties.

6.1: Statistics and Their Distributions

Def: A statistic is any quantity whose value can be calculated from sample data.

Def: The random variables $X_1, X_2, \ldots, X_n$ are said to form a random sample of size n if X_i are independent and identically distributed (IID).

In chapter 1, we came across a few statistics. We can measure the center of a sample with the sample mean, $\bar{x}$ and the sample median, $\tilde{x}$. If we want to describe the spread of a sample, there are the sample variance (s^2) and sample standard deviation (s).

In fact, since each random sample could potentially be different, sample statistics have a probability distribution associated with them.

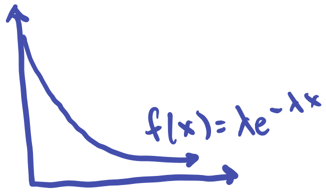
Example: Let's simulate five exponential($\lambda = 3$) random variables a couple times and look at the different sample means.

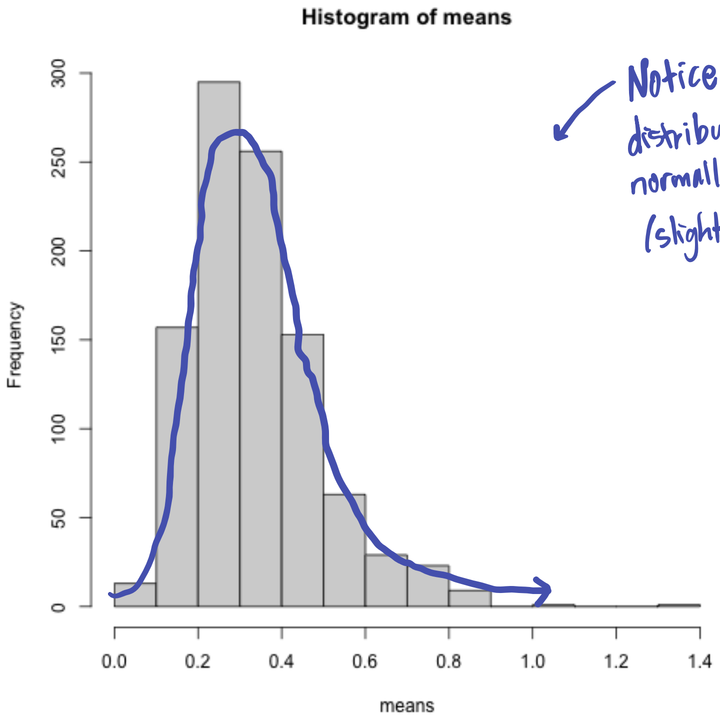
```
sample1 = rexp(n=5,rate=3); sample1

## [1] 0.5274761 0.2818285 0.1090117 0.4920306 0.2566445

mean(sample1)

## [1] 0.3333983

sample2 = rexp(n=5,rate=3); sample2

## [1] 0.07052601 1.36882351 0.14881661 0.04802431 0.28632793

mean(sample2)

## [1] 0.3845037

sample3 = rexp(n=5,rate=3); sample3

## [1] 0.58110200 0.09693622 0.45714040 0.39613499 0.16153774

mean(sample3)

## [1] 0.3385703
```

Example: Let's repeat the previous experiment 1000 times and plot the results.

```
# initialize a vector of length 1000 for the sample means
means = rep(0, 1000)

# collect a sample of five RVs, calculate sample mean, and repeat
for (i in 1:1000) {
    randomsample = rexp(n = 5, rate = 3)
    means[i] = mean(randomsample)
}
mean(means)

## [1] 0.3380338

var(means)

## [1] 0.02367919

hist(means)
```

original data:





Notice this sampling distribution is (approx.) normally dist.
(slightly skewed right)

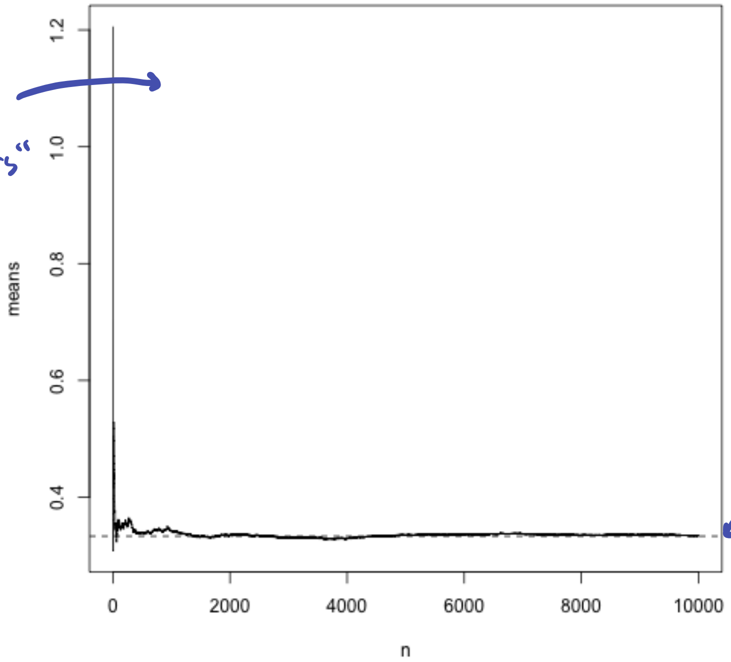
Example: Now, let's look at the sample mean of a random sample as a function of sample size, n.

```r
# generate a sample of 10000 exp(3) RVs
data = rexp(n = 10000, rate = 3)
means = cumsum(data)/seq_along(data)
plot(x = 1:10000, y = means, type = "l", xlab = "n")
abline(h = 1/3, lty = 2)
```

"Law of Large Numbers"

as $n \to \infty$,

$\overline{X}_n = \frac{1}{n} \sum x_i \to \mu$



Recall: $X_i \sim \exp(3)$

$E[X_i] = \frac{1}{\lambda} = \frac{1}{3}$

$\approx \frac{1}{3}$

Theorem: The tendency of $\bar{x}$ to approach μ as $n \to \infty$ is the result of the Law of Large Numbers. Specifically, if $|E[X]| = \mu < \infty$, then $\lim_{n\to\infty} 1/n \sum_{n=1}^{\infty} x_i = \lim_{n\to\infty} \bar{x} = \mu$.

6.2: The Distribution of the Sample Mean

Motivation: Suppose it is claimed that the average lifespan of Galapagos tortoises is 100 years old. To test this claim, a study of 20 Galapagos tortoises finds an average life span of $\bar{x} = 95.9$ years. Is this enough information to reject the claim?

To answer this question, we need to know the distribution of $\bar{X}$.

Theorem: Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ and let $\bar{X} = \frac{1}{n}\sum X_i$. Then:

(i) $E[\bar{X}] = \mu$

(ii) $Var(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$

(iii) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ "standard error of $\bar{X}$"

random variable (sample mean)

$E[X_i] = \mu$

$Var[X_i] = \sigma^2$

Example: Show that $Var(\bar{X}) = \frac{\sigma^2}{n}$.

$$Var(\bar{X}) = Var\left(\tfrac{1}{n}\textstyle\sum X_i\right) = \left(\tfrac{1}{n}\right)^2 Var\left(\textstyle\sum X_i\right) = \left(\tfrac{1}{n}\right)^2 \textstyle\sum Var(X_i) = \frac{1}{n^2}\textstyle\sum \sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} \checkmark$$

$Var(aY) = a^2 Var(Y)$

Since X_i are independent, $Var(X+Y) = Var(X) + Var(Y)$

$\hat{\theta}$ = estimate of θ, where θ is a pop. parameter

Def: If $\hat{\sigma}$ is an estimator for σ, then the standard error of $\hat{\sigma}$ is: $SE(\hat{\sigma}) = \sqrt{Var(\hat{\sigma})}$.

Example: If $X_1, X_2, \ldots, X_{50}$ are iid Uniform[10,20] random variables, then determine $E[\bar{X}]$ and $SE(\bar{X})$.

$n = 50$

continuous

↳ see chapter 4

single RV

$$X_i \sim Uniform[a,b] \rightarrow \underbrace{E[X_i] = \frac{a+b}{2} = \frac{10+20}{2} = 15}_{=\mu},\ \underbrace{Var(X_i) = \frac{(b-a)^2}{12} = \frac{100}{12}}_{\sigma^2}$$

$(a=10,\ b=20)$

$E[\bar{X}] = \mu = 15$

$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{100/12}{50} = \frac{1}{6}$

$SE(\bar{X}) = \sqrt{Var(\bar{X})} = \sqrt{1/6}$

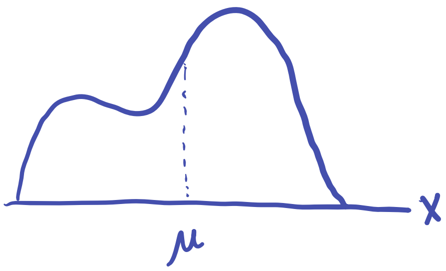
→ these values describe the sampling dist of $\bar{X}$

> Theorem: (Central Limit Theorem) Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. If n is sufficiently large, $\bar{X}$ has approximately a normal distribution with mean $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \sigma^2/n$.

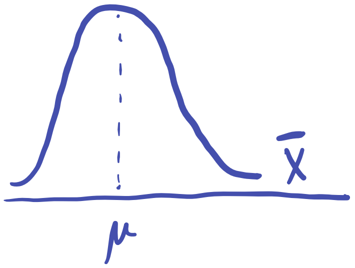
This result is nontrivial and extremely useful! It allows us to determine how likely or unlikely a sample mean is if we are given a hypothesized population mean.

Illustration:

(example)



(dist of individual data)



(sampling dist of $\bar{X}$ - approx normal)

Example: Suppose that $X_1, X_2, \ldots, X_{10}$ is a random sample from a discrete uniform(1,6) random variable. Determine the approximate sampling distribution of $\bar{X}$. (fair six-sided die roll)

(a=1, b=6)

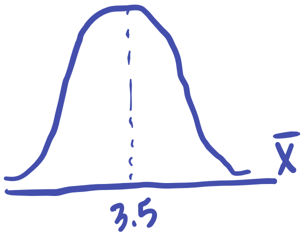
(a=1, b=6)

single die roll

$X_i \sim$ discrete uniform (a, b) (from ch. 3)

↳ $E[X_i] = \frac{a+b}{2} = \frac{1+6}{2} = 3.5$

$Var[X_i] = \frac{(b-a+1)^2 - 1}{12} = \frac{35}{12}$



$E[\bar{X}] = \mu = 3.5$

$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{35/12}{10} = \frac{35}{120}$

$$\longrightarrow \boxed{\bar{X} \,\dot{\sim}\, N(\mu = 3.5, \sigma^2 = \tfrac{35}{120})}$$

" $\dot{\sim}$ " = approximately dist. as

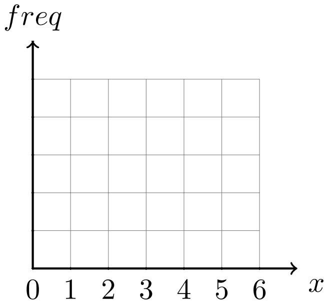
Example: Class exercise. We will collect data based on the following experiments. Plot a histogram for each instance.

(a) Roll your die once and record your number: ________

Sample data

x	1	2	3	4	5	6
freq						

Histogram

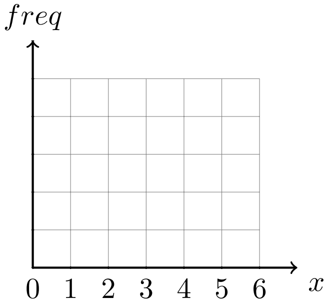


(b) Roll your die twice and record the average of the two die rolls: ________

Sample data

x	(0.5, 1]	(1,1.5]	(1.5,2]	(2,2.5]	(2.5,3]	(3,3.5]
freq						
x	(3.5, 4]	(4,4.5]	(4.5,5]	(5,5.5]	(5.5,6]	
freq						

Histogram

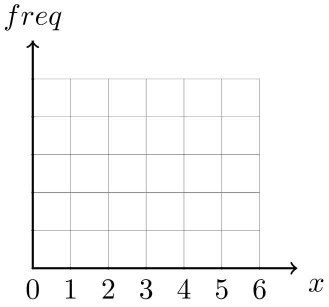


(c) Roll your die ten times and record the average of the ten die rolls: ________

Sample data

x	(0.5, 1]	(1,1.5]	(1.5,2]	(2,2.5]	(2.5,3]	(3,3.5]
freq						
x	(3.5, 4]	(4,4.5]	(4.5,5]	(5,5.5]	(5.5,6]	
freq						

Histogram



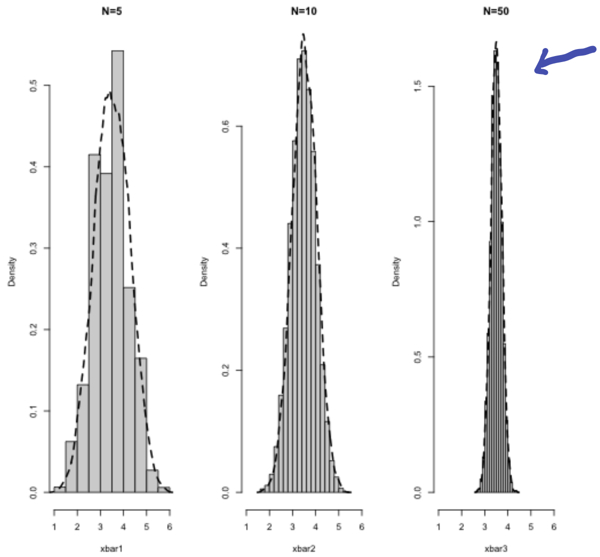
Example: Let's simulate N fair die rolls 10,000 times. Here we will do the simulation for N=5, N=10, and N=50.

```
xbar1 = rep(0, 10000)  # initialize this vector to zeros
xbar2 = rep(0, 10000)  # initialize this vector to zeros
xbar3 = rep(0, 10000)  # initialize this vector to zeros

# 10,000 simulated die rolls for N=5,10,50
for (i in 1:10000) {
    x1 = sample(1:6, 5, replace = T)
    xbar1[i] = mean(x1)
    x2 = sample(1:6, 10, replace = T)
    xbar2[i] = mean(x2)
    x3 = sample(1:6, 50, replace = T)
    xbar3[i] = mean(x3)
}

par(mfrow = c(1, 3))
hist(xbar1, main = "N=5", xlim = c(1, 6), prob = TRUE)
lines(density(xbar1), lwd = 2, lty = 2)
hist(xbar2, main = "N=10", xlim = c(1, 6), prob = TRUE)
lines(density(xbar2), lwd = 2, lty = 2)
hist(xbar3, main = "N=50", xlim = c(1, 6), prob = TRUE)
lines(density(xbar3), lwd = 2, lty = 2)
```

avg 5 die rolls

10

50



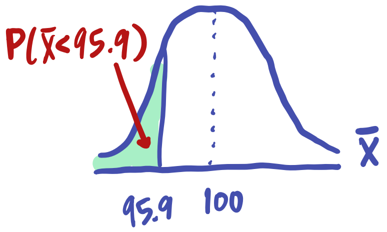
Notice as n increases, the sampling dist. of $\bar{X}$ becomes more "bell-shaped" and the variance decreases

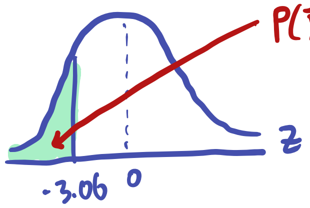
Notice that the variance decreases when N increases. Also notice that the the histograms appear to look more "normal" as N increases.

lifespan $\mu=100$ $\sigma=6$

Example: Suppose the average ~~age~~ of Galapagos tortoises is 100 years and the standard deviation is 6 years. What is the probability that the sample mean of 20 Galapagos tortoises is less than 95.9 years old?

$n=20$ $P(\bar{X}<95.9)$





$P(\bar{X}<95.9)=$ $P(Z<-3.06)=\boxed{0.0011}$

↑calculate the z-score → $z=\dfrac{\bar{X}-\mu}{SE(\bar{X})}=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\dfrac{95.9-100}{\sqrt{1.8}}=-3.06$

characteristics of the samp. dist. of $\bar{X}$

$$\begin{cases} E[\bar{X}]=\mu=100 \\ Var(\bar{X})=\dfrac{\sigma^2}{n}=\dfrac{6^2}{20}=1.8 \\ SE(\bar{X})=\sqrt{Var(\bar{X})}=\sqrt{1.8} \end{cases}$$

$n=50$ (cont.)

Example: Suppose $X_1, X_2, \ldots, X_{50}$ are iid Uniform[10,20] random variables. Determine the sampling distribution of $\bar{X}$ and calculate $P(\bar{X}<14)$.

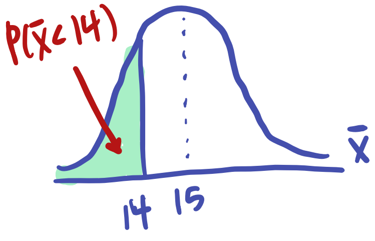
From chapter 4: $X_i \sim Uniform[10,20] \rightarrow E[X_i]=15$

$(a=10, b=20)$ $Var[X_i]=\dfrac{100}{12}$

characteristics of sampling dist of $\bar{X}$

$$\begin{cases} E[\bar{X}]=\mu=15 \\ Var(\bar{X})=\dfrac{\sigma^2}{n}=\dfrac{100/12}{50}=\dfrac{1}{6} \end{cases}$$

→ Our (approx.) sampling dist of $\bar{X}$ is: $\bar{X} \overset{\cdot}{\sim} N(\mu=15, \sigma^2=\frac{1}{6})$



→ calculate the z-score:

$$z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}=\frac{14-15}{\sqrt{1/6}}=-\sqrt{6}$$

$$\boxed{P(\bar{X}<14)=P(Z<-\sqrt{6})=0.0071}$$

θ = parameter (likely unknown)

$\hat{\theta}$ = point estimator of θ (calculated from the sample data)

6.3: Several General Concepts of Point Estimation

Def: A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ and is often denoted with a hat (i.e., $\hat{\theta}$). A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic, $\hat{\theta}$ is called the point estimator of θ.

Example: ① $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a point estimator of μ.

② $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a point estimator of σ^2.

Def: The bias of a point estimator $\hat{\theta}$ for a parameter θ is defined as $Bias[\hat{\theta}] = E[\hat{\theta}] - \theta$.

Def: A statistic $\hat{\theta}$ is called an unbiased estimator of the parameter θ if $E[\hat{\theta}] = \theta$.

γ is a biased est of σ^2.

Note: $\bar{X}$ is an unbiased estimator for μ since $E[\bar{X}] = \mu$

ex) S^2 is an unbiased est of σ^2 since $E[S^2] = \sigma^2$, $\gamma = \frac{1}{n}\sum(X_i - \bar{X})^2$, $E[\gamma] \neq \sigma^2$

Note: From the theorem above, note that as $n \to \infty$, $Var(\bar{X}) \to 0$. This means that we can estimate μ as precisely as we would like as long as we make n large.

Example: Suppose $X_1, X_2, \ldots, X_n$ are exponential($\lambda = 2$) random variables. Is $\bar{X}$ an unbiased estimator for λ? How about if $X_1, X_2, \ldots, X_n$ are Poisson($\lambda = 2$) random variables?

(a) $X_1, \ldots, X_n \overset{iid}{\sim} \exp(\lambda = 2)$, chapter 4: $X \sim \exp(\lambda)$, then $E[X] = \frac{1}{\lambda}$

$$E[\bar{X}] = E\left[\frac{1}{n}\sum X_i\right] = \frac{1}{n}E\left[\sum X_i\right] = \frac{1}{n}\sum \underbrace{E[X_i]}_{=1/\lambda} = \cancel{\frac{1}{n}} \cdot \cancel{n} \cdot \frac{1}{\lambda} = \frac{1}{\lambda}$$

→ $\bar{X}$ is not an unbiased estimator of λ, however,

$\bar{X}$ is an unbiased estimator of $\frac{1}{\lambda}$.

(b) $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda = 2)$, chapter 3: $X \sim \text{Poisson}(\lambda)$, $E[X] = \lambda$

$$E[\bar{X}] = \frac{1}{n}\sum \underbrace{E[X_i]}_{=\lambda} = \frac{1}{n} \cdot n\lambda = \lambda \ ✓$$

→ $\bar{X}$ is an unbiased est. of λ since $E[\bar{X}] = \lambda$ for Poisson.

Def: The mean squared error (MSE) of an estimator $\hat{\theta}$ of a parameter θ is defined by $E[(\hat{\theta}-\theta)^2]$.

Typically, unbiased estimators are preferred though some advanced methods consider biased estimators that minimize MSE.

Theorem: Suppose $X_1, X_2, \dots, X_n$ is a random sample from a distribution with mean μ and variance σ^2. Then:
(i) $\hat{\theta} = \bar{X}$ is an unbiased estimator for μ.
(ii) $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator for σ^2.

Example: Suppose $X_1, X_2, \dots, X_n$ are a random sample of Binomial(n,p) random variables. Derive $E[\bar{X}]$ using the Binomial distribution and compare it to the result from the theorem above. Also, find $Var(\bar{X})$.

$X_1, X_2, \dots, X_n \overset{iid}{\sim} \text{Binomial}(n,p)$, chapter 3: $E[X_i] = np$, $Var(X_i) = np(1-p)$

$$E[\bar{X}] = \frac{1}{n}\sum \underbrace{E[X_i]}_{=np} = \cancel{\frac{1}{n}}\cdot\cancel{n}\cdot np = np$$

→ Since $E[\bar{X}] = np = \mu$, $\bar{X}$ is an unbiased est. of μ.

$$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{\cancel{n}p(1-p)}{\cancel{n}} = p(1-p) = Var(\bar{X})$$

Def: The standard error of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$. It is the magnitude of a typical or representative deviation between an estimate and the value of θ. If the standard error is a function of unknown parameters, then we can estimate it with the estimated standard error which is denoted $s_{\hat{\theta}}$.

Example: Find the standard error for $\bar{X}$ in the example above.

$$\sigma_{\bar{X}} = SE[\bar{X}] = \sqrt{Var(\bar{X})} = \sqrt{p(1-p)}$$

for binomial

6.4: Methods of Point Estimation

There are a variety of ways to obtain a point estimate. These methods include:
(i) Method of Moments
(ii) Maximum Likelihood Estimation
(iii) Bayesian Estimation

While we won't cover how these estimation methods are completed, we will compare different estimators. Each of these methods could produce a different point estimate.

Example: Suppose X_1, X_2, X_3 are a random sample of $\mathcal{N}(\mu = 10, \sigma^2 = 4)$. Four estimators are given below. Determine if each estimator is unbiased and calculate its standard error. Which of the four estimators is preferred? for estimating μ?

(a) $Y_1 = \frac{X_1+X_2+X_3}{3}$ $\rightarrow E[Y_1] = E\left[\frac{X_1+X_2+X_3}{3}\right] = \frac{1}{3}(E[X_1]+E[X_2]+E[X_3]) = \frac{1}{3}(10+10+10) = \underline{10 = \mu}$

$\rightarrow$ Y_1 is an unbiased est. for μ.

$$Var(Y_1) = Var\left(\frac{X_1+X_2+X_3}{3}\right) = \left(\frac{1}{3}\right)^2(Var X_1 + Var X_2 + Var X_3)$$
$$= \frac{1}{9}(4+4+4) = 4/3$$
$$\rightarrow SE(Y_1) = \sqrt{Var(Y_1)} = \underline{\sqrt{4/3} = SE(Y_1)}$$

(b) $Y_2 = \frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{4}$ $\rightarrow E[Y_2] = \frac{1}{2}E[Y_1] + \frac{1}{4}E[Y_2] + \frac{1}{4}E[Y_3] = \frac{1}{2}(10) + \frac{1}{4}(10) + \frac{1}{4}(10) = \underline{10 = \mu}$

$$Var(Y_2) = Var\left(\frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{4}\right) = \left(\frac{1}{2}\right)^2 Var X_1 + \left(\frac{1}{4}\right)^2 Var X_2 + \left(\frac{1}{4}\right)^2 Var X_3$$
$$= \frac{1}{4}(4) + \frac{1}{16}(4) + \frac{1}{16}(4) = 3/2$$

↳ Y_2 is an unbiased est for μ.

$$\rightarrow \underline{SE(Y_2) = \sqrt{3/2}}$$

(c) $Y_3 = X_1$ $\rightarrow E[Y_3] = E[X_1] = \underline{10 = \mu}$ $\rightarrow$ Y_3 is an unbiased est. for μ.

$$Var(Y_3) = Var(X_1) = 4 \rightarrow \underline{SE(Y_3) = \sqrt{4} = 2}$$

(d) $Y_4 = \frac{X_1+X_2+X_3}{2}$ $\rightarrow E[Y_4] = \frac{1}{2}(E[X_1]+E[X_2]+E[X_3]) = \frac{1}{2}(10+10+10) = \underline{15 \neq \mu}$

$\rightarrow$ Y_4 is a biased est. for μ

$$Var(Y_4) = \left(\frac{1}{2}\right)^2(Var X_1 + Var X_2 + Var X_3)$$
$$= \frac{1}{4}(4+4+4) = 3 \rightarrow \underline{SE(Y_4) = \sqrt{3}}$$

Conclusion: Y_1 is the preferred estimator for μ since:
① Y_1 is an unbiased est. for μ
② Y_1 has the lowest std error of these four estimators.