STAT 41600/41700

Project # 05

Task 1: --

The purpose of this task is to introduce you to the term ***Sampling Distribution*** of a statistic. We consider a random sample of n independent $x_1, x_1, ..., x_n$ observations drawn from a large population with (population) mean $E(X) = \mu$ and (population) variance $V(X) = \sigma^2$**.** Usually, these two (population) parameters μ and σ^2 are unknown to us and we use the n observations in the sample to ***estimate*** these parameters. A natural ***estimator*** for the population mean μ is the sample mean

$$\overline{x}_n = \frac{1}{n}\sum_{1}^{n} x_i$$

and a natural ***estimator*** of the population variance σ^2 is the sample variance

$$S_n^2 = \frac{1}{n-1}\sum_{1}^{n}(x_i - \overline{x}_n)^2 \textbf{.}$$

However, since the values of these two statistics $\overline{x}_n$, and S_n^2 depend upon the observed in values in the sample, these sample statistics are actually, also random variables. The distribution of their values (varying from one sample to another) is called ***Sampling Distribution***.

A. You will now use R to generate (simulate), the sampling distribution of the statistic $\overline{x}_n$ (the sample average) when sampling from the <u>*Exponential distribution*</u>.

- Generate $M = 10000$ different random samples, each with $n = 16$ observations, from the Exponential distribution, $X \sim Exp(\lambda = 0.1)$**,** which has mean $\mu = 10$ and a standard deviation $\sigma = 10$ **);**

a) Create a table to record the Mxn sample values

```
➢ M<-10000
➢ n<-16
➢ XSamples<-matrix(rep(0,n*M), M, n)
```

b) Generate a different random sample in each row of the matrix

```
➢ for(j in 1:M){XSamples[j , ]<- rexp(n, rate=0.1)}
```

c) Calculate the sample mean of each sample (row)

- Xbars<-apply(XSamples, 1, mean)

d) The question however is what are the mean and the standard deviation of this (sampling) distribution? Get some summary statistics about this distribution.

- mean(Xbars)
- var(Xbars)
- sd(Xbars)
- summary(Xbars)

e) Let us use denote by $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ the men and the standard deviation of the sampling distribution of $\bar{x}_n$**,** by part e) above**,** for a sample of size $n = 16$**;**

$\mu_{\bar{x}} \approx$ __________ **and** $\sigma_{\bar{x}} \approx$ __________

f) To study the overall shape of this **sampling distribution** of $\bar{x}_n$**,** obtain the histogram of all these $M = 10000$ sample averages $\bar{x}_n$ you simulated in each one of the previous step. For example, when $\lambda = 0.1$, and $n = 16$

- hist(Xbars, nclass=30, freq=F, main="Sampling Distribution of Xbar when n=16 and la mbda=0.1")

g) From this histogram you can see that this sampling distribution is relatively symmetric, almost 'bell-shaped'. Add 'an approximated' density curve to the histogram above.

- dens<-density(Xbars)
- lines(dens$x, dens$y, col=4)

h) Add to the histogram the density curve of the Normal distribution with mean $\mu \equiv \mu_{\bar{x}}$ and standard deviation $\sigma \equiv \sigma_{\bar{x}}$

- start<-mean(Xbars)-4*sd(Xbars)
- end<-mean(Xbars)+4*sd(Xbars)
- xx<-seq(start, end, length=200)
- lines(xx, dnorm(xx, mean(Xbars), sd(Xbars), col=2, lwd=2)
-

Complete and Submit the next page

Results of Project 4 **NAME:** ______________________________

i) Repeat a)-h) above for $n = 32,\ 64,\ 128$, and $\lambda = 0.1,\ 1$, obtain the 6 histogram plots (nicely labeled and graphed) and summarize the results in the table below.

Sampling Distribution of $\overline{x}_n$ when sampling from the Exponential Distribution $Exp(\lambda)$

$\lambda =$	$\mu =$	$\sigma =$	$n =$	$\mu_{\overline{x}} =$	$\sigma_{\overline{x}} =$
0.1	10	10	32		
			64		
			128		
1	1	1	32		
			64		
			128		

Can you conclude that Sampling Distribution of $\overline{x}_n$ **is approximately the**

$$N(\mu_{\overline{x}}, \sigma_{\overline{x}}^2) \ \textbf{distribution with}\ \mu_{\overline{x}} = \mu \ \textbf{and}\ \sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

Answer: ____