Sampling Distributions:

\[ X \sim N(\mu, \sigma^2/n) \quad n \geq 20 \]

\[ \hat{p} \sim N\left( p, \frac{p(1-p)}{n} \right) \quad np, n(1-p) > 10 \]

\[ (\bar{X}_1 - \bar{X}_2) \sim N\left( (\mu_1 - \mu_2), \left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right) \right) \quad n_i \geq 20 \]

\[ (\hat{p}_1 - \hat{p}_2) \sim N\left( p_1 - p_2, \left( \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2} \right) \right) \quad n_ip_i, n_i(1-p_i) > 10 \]

\[ Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \]

Central Limit Theorem:

If \( n \) is sufficiently large \( (n > 20) \), the sample means of random samples taken from a population with mean \( \mu \) and standard deviation \( \sigma \) will be approximately normally distributed with mean \( \mu \) and standard deviation \( \sigma/\sqrt{n} \).

95% Confidence Interval, large sample:

sample estimate ±1.96 se(estimate)

Tutorial Exercises

Exercise A

In pairs, discuss the following and write your response:

(a) What is meant by the sampling distribution of \( \bar{X} \)?

(b) How is the mean of the distribution of sample means, \( \bar{X} \), related to the population mean of individual observations, \( X \)?

(c) How is the population standard deviation of \( \bar{X} \) related to the population standard deviation of individual observations, \( X \)?

(d) What happens to the sampling distribution of \( \bar{X} \) if the sample size is increased?
Exercise 16.3

In previous years, exam results give $E(X) = 79$ and $sd(X) = 13$. The number of internal students is 91 and the number of external students is 189.

(a) Use the CLT to give the distribution of the sample mean for both groups (internal and external students).

(b) Find $P(70 < \bar{X} < 80)$ for both groups. Calculate the z-scores and refer to the cdf in Figure 1, below.

(c) Explain the difference between the two.

Exercise B

Suppose that a random variable follows a binomial distribution with probability of success $= 0.2$. Use the normal approximation to the binomial to find $P(\hat{p} > 0.3)$ for 50 trials. Have the conditions necessary for this approximation been met?

You will need to refer to the standardised normal distribution function, figure 1.
Exercise 17.2

Two machines are used to fill a 25kg bag of dog food. Sample information for these 2 machines is:-

<table>
<thead>
<tr>
<th></th>
<th>Machine 1</th>
<th>Machine 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>sample size</td>
<td>81</td>
<td>64</td>
</tr>
<tr>
<td>sample mean</td>
<td>25.5</td>
<td>24</td>
</tr>
<tr>
<td>sample variance</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

(a) Calculate $\mu_{\bar{X}_1 - \bar{X}_2}$ and $\sigma_{\bar{X}_1 - \bar{X}_2}$.

(b) Using the cdf, below, find $P(\mu_1 - \mu_2 \geq 2)$.

Practical Exercises

Exercise: Applets Demonstrating Statistical Concepts


There you will find a number of interactive applets demonstrating important statistical concepts.

(b) Select the Sampling Distribution Simulation, and click on the begin button.

(c) You may need to expand the window. You will see 4 sets of axes. The top one shows the distribution from which the samples are being drawn. The default is a Normal distribution with mean=16 and sd=5.

In the 3rd plot select Mean and sample size, n=2, and in the 4th plot select mean and sample size, n=25. Check the Fit Normal box in both cases.

(d) You can animate a single sample selection by clicking on the animated button a few times until you understand the process. Observe the changing numerical summaries provided to the left of each sampling distribution.

(e) Now you can choose to select 5, 1000 or 10000 samples at a time to see how the sampling distribution of the mean builds up.

(f) Comment on what you have observed. State the sampling distribution of the mean in each case.
Now repeat this for a skewed distribution. What do you observe? Use the Central Limit Theorem to give the sampling distribution of $\bar{X}$.

In your own time you can explore other relevant applets to supplement this week’s learning material:

- Confidence Intervals
- Confidence Intervals on a Proportion
- Normal Approximation to the Binomial - use this to examine the distributions from Exercise 16.1!

Take the time to read the instructions and then you can PLAY! Think about the results and what general rules you can derive from them.

**Exercise 16.1**
Suppose in an experiment where the outcomes are binomial, the population probability of success has been determined in previous long-term studies to be $p = 0.2$. For the new experiment, calculate $P(\hat{p} > 0.3)$ when the sample sizes vary 10, 20, 30, 40, 50 using (a) `pbinom()` (b) `pnorm()` Comment on the reliability of the normal approximation for the various sample sizes.

The following skeleton of an R job may be a starting point. The script file can be found on the website - go to the Tutorial and Practical Exercises link.

```r
options(digits=4)
p <- 0.2
p.test <- 0.3
N <- seq(10,50,10)
cat(" n Pb Pn \n")
for (n in N){
    Pb <- pbinom( )
    Pn <- pnorm( )
cat(n,Pb,Pn,"\n")
}
```

Recall that for binomial data, $E[\hat{\mu}] = np$, $\text{var}[\hat{\mu}] = np(1 - p)$ and for large samples the CLT gives that $\hat{\mu} \sim N(E[\hat{\mu}], \text{var}[\hat{\mu}])$.

Hence $P(\hat{p} > 0.3)$ can be calculated equivalently as $P(\hat{\mu} > n \times 0.3)$ if $n$ is large.

The normal approximation may be couched as either

(a) $\hat{p} \sim N(p, \frac{p(1-p)}{n})$ or (b) $\hat{\mu} \sim N(np, np(1 - p))$

To calculate the binomial probability, you require `pbinom( ,q=n*p.test, )`. In both questions, remember to use `lower.tail=F`. 
Exercise 16.2

The distribution of sample means from Poisson data may also be approximated as normal using the Central Limit Theorem when the sample means are sufficiently large (> 20 say).

If \(X\) is Poisson compute \(P(\bar{X} \leq 42)\) when the Poisson rate parameter \(\lambda = 35, 40, 45, 50\) using the exact Poisson probability and its large sample normal approximation. Comment upon the reliability of the approximation for varying \(\bar{X}\).

Recall from (12.1), (12.2) that if \(X \sim \text{Poisson}(\lambda)\), \(E(X) = \text{var}(X) = \lambda\).

A suggested \texttt{R} strategy follows. The script file can be found on the website - go to the Tutorial and Practical Exercises link.

This exercise is very similar to exercise 16.1 - think carefully about the missing detail in the \texttt{R} code below.

```r
options(digits=2)
lam <- c(35,40,45,50)
Xbar <- 42
cat(" n  Pp  Pn\n")
for (L .... ){
  Pp <- ppois( .... )
  Pn <- pnorm( .... )
}
```

Exercise 17.1

Two instructors offer extra help designed to improve students' scores on an exam. Suppose that for instructors 1 and 2, the percentages of students improving their scores is 85% and 76% respectively. For a random sample of 55 students from instructor 1 and 60 students from instructor 2, compute the probability the difference between percentages of students improving is more than 2.5%.

(a) Write \(n_1 = , n_2 = , \hat{p}_1 = , \hat{p}_2 = .\)
(b) Calculate \(\hat{p}_1 - \hat{p}_2\) and \(\sigma_{\hat{p}_1 - \hat{p}_2}\)
(c) Sketch the shaded area of the normal density corresponding to \((\hat{p}_1 - \hat{p}_2) > 0.025\)
(d) Use \texttt{pnorm()} to find the probability.