Inverse Normal Distribution and Sampling Distributions

Section 4.3 & 4.4

Cathy Poliak, Ph.D.
cathy@math.uh.edu

Department of Mathematics
University of Houston

Lecture 11 - 2311

Outline

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper Set Up

- Fill in all of the proper bubbles.
- Make sure your ID number is correct.
- Make sure the filled in circles are very dark.
- This is popper number 07.

Facts about the Normal distribution

- The curve is symmetric about the mean. That is, 50% of the area under the curve is below the mean. 50% of the area under the curve is above the mean.
- The spread of the curve is determined by the standard deviation.
- The area under the curve is with respect to the number of standard deviations a value is from the mean.
- Total area under the curve is 1.
- Area under the curve is the same a probability within a range of values.
- The normal distribution can be written as $N(\mu, \sigma)$ where we are given the values of μ and σ.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Popper 07 Questions

Let a random variable X have a Normal distribution with mean $\mu = 10$ and standard deviation $\sigma = 2$. For the following questions determine what is the proper way to solve these probabilities.

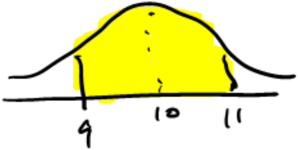
1. $P(X < 7.25)$

a) pnorm(7.25,10,2)
b) 1-pnorm(7.25,10,2)
c) pnorm(7,10,2)
d) dnorm(7.25, 10 , 2)

2. $P(X \geq 5)$

a) pnorm(5, 10, 2)
b) 1 - pnorm(5, 10, 2)
c) 1 - pnorm(4, 10, 2)
d) dnorm(6, 10, 2)

3. $P(9 \leq X \leq 11)$
    a) pnorm(11, 10, 2) - pnorm(8, 10, 2)
    b) pnorm(11, 10, 2) - 1- pnorm(9, 10, 2)
    c) pnrom(11, 10, 2) - pnorm(9, 10, 2)
    d) dnorm(11, 10, 2) - dnorm(9, 10, 2)



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Finding a value when given a proportion

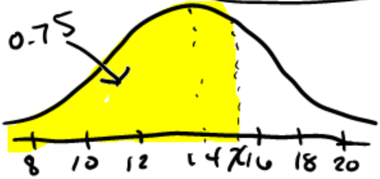
- Called inverse Normal.
- This is working "Backwards" using Z-Table.
- Finding the observed values when given a percent.
- In R: qnorm(proportion,mean,sd).
- In TI-83 or 84: invNorm(proportion,mean,sd).

Example

The average life of a LED light bulb is 14 years with a standard deviation of 2 years. Assume that the lifetime of these bulbs have a Normal distribution.

75th percentile

- What is the third quartile of the life of a LED light bulb?

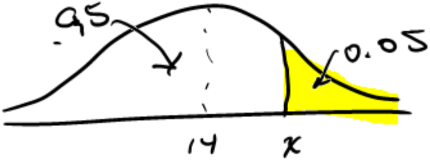


$P(X \le x) = 0.75$ find x

$x = qnorm(0.75, 14, 2)$

$x = 15.35$ years

- 5% of the light bulbs will last longer than what year? That is $P(X \geq x) = 0.05$, what is x?



$x = qnorm(0.95, 14, 2)$

$x = qnorm(1-0.05, 14, 2)$

$x = 17.28971$ years

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

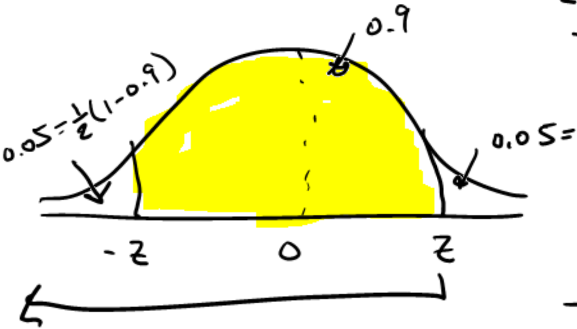
z-score, $\mu = 0$, $\sigma = 1$

What values of $\pm z$ such that

$$P(-z \le Z \le z) = 0.9$$

$-z = \text{qnorm}(0.05)$

$-z = -1.645$



$z = \text{qnorm}(0.9+0.05)$

$z = 1.645$

$$\frac{1+.9}{2} = \frac{1.9}{2} = 0.95$$

$z = \text{qnorm}(1.9/2)$

Popper 07 Questions

Consider one family as a population of five children. We are looking at the ages of these five children: 3, 5, 9, 11, 14.

4. Determine the population mean, μ, age of these children.

   $\frac{3+5+9+11+14}{5}$

   a. 9 b. 10 c. 8.4 d. 11

5. Determine the population standard deviation, σ, of these children.

   $\frac{(3-8.4)^2+(5-8.4)^2+\cdots}{5}$

   a. 10 b. 4 c. 8.4 0

6. Suppose we take a sample of 2 children from this population. What would we **expect** the sample mean, $\bar{x}$ from the 2 children to be?

   a. 2 b. 8.4 c. 4 d. 16

Sampling Distribution of size 2

From the five children, we want to list out all possible pairs of size 2 and determine their mean. Ages are: 3,5, 9, 11, 14

Pairs	Sample mean, $\bar{x}$
(3,5)	4
(3,9)	6
(3,11)	7
(3, 14)	8.5
(5, 9)	7
(5, 11)	8
(5, 14)	9.5
(9, 11)	10
(9, 14)	11.5
(11, 14)	12.5

$E(\bar{x}) = \mu_{\bar{x}}$
$= 8.4$

$\sigma_{\bar{x}} = SD(\bar{x})$
$= 2.82$

The list above is a sampling distribution from a sample of 2 of $\bar{x}$, the possible values of the sample mean. What is the mean of the sample means, $\mu_{\bar{x}}$? What is the standard deviation of the sample means, $\sigma_{\bar{x}}$?

Sampling Distribution of size 3

What about the sampling distribution of size 3 from the family of five?

$\mu_{\bar{x}} = E(\bar{x}) = 8.4$

$\sigma_{\bar{x}} = SD(\bar{x}) = 2.31$

Sets	$\bar{x}$
(3, 5, 9)	5.6667
(3, 5, 11)	6.3333
(3, 5, 14)	7.3333
(3, 9, 11)	7.6667
(3, 9, 14)	8.6667
(3, 11, 14)	9.3333
(5, 9, 11)	8.3333
(5, 9, 14)	9.3333
(5, 11, 14)	10
(9, 11, 14)	11.3333

What is the mean of these means, $\mu_{\bar{x}}$? What is the standard deviation of these means, $\sigma_{\bar{x}}$?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sampling distribution

- When we describe distributions we use three characteristics:
  - Shape
  - Center
  - Spread
- To describe the sampling distribution we can use the same three characteristics.
- This can be shown through histograms or numerical values.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sampling Distribution of $\bar{X}$

- Suppose that $\bar{X}$ is the sample mean of a simple random sample of size n from a large population with mean μ and standard deviation σ.
- $\bar{X}$ is a random variable because every time we take a random sample we will not get the same sample mean $\bar{X}$. Thus we want to know the **distribution** of the sample means $\bar{X}$.
- The center of the sample means (mean of the sample means) $\mu_{\bar{X}}$ is μ. Also called the **expected value**. $E(\bar{x}) = \mu_{\bar{x}} = \mu$
- The spread of the sample means (standard deviation of the sample means) $\sigma_{\bar{X}}$ is $\sigma/\sqrt{n}$.

$$\sigma_{\bar{x}} = SD(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

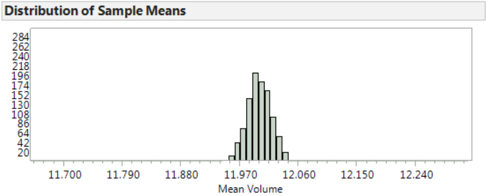
Sampling Distribution Example

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz.

- We take a sample of 25 cans and find the mean amount $\bar{X}$ in these 25 cans. What would we expect the mean to be? Would the sample mean be exactly that value? If not how far off could the sample mean be?

$E(\bar{x}) = \mu_{\bar{x}} = 12$

$SD(\bar{x}) = \sigma_{\bar{x}} = \frac{0.09}{\sqrt{25}} = 0.018$



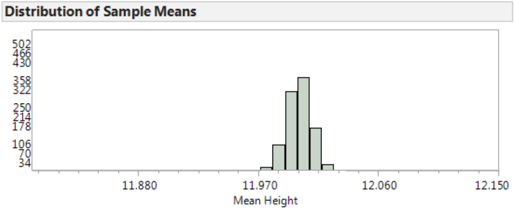
Means Summary Table	
Mean of Sample Means:	12.0001
Std Dev of Sample Means:	0.01786
No. of Sample Means:	1000

Sampling Distribution Example

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz.

- We take a sample of 100 cans and find the mean amount $\bar{X}$ in these 100 cans. What would we expect the mean to be? Would the sample mean be exactly that value? If not how far off could the sample mean be?

$$\mu_{\bar{x}} = E(\bar{x}) = 12 \qquad \sigma_{\bar{x}} = SD(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009$$



Means Summary Table	
Mean of Sample Means:	12.0002
Std Dev of Sample Means:	0.00895
No. of Sample Means:	1000

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Shape of the Sample Mean Distribution

- If a population has a Normal distribution, then the sample mean $\bar{X}$ of n independent observations also has a Normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.
- **Central limit theorem**: For *any* population, when n is large ($n > 30$), the sampling distribution of the sample mean $\bar{X}$ is approximately a Normal distribution with mean μ and standard deviation $\sigma/\sqrt{n}$.

Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 4 cans are examined, describe the distribution of the sample means $\bar{X}$.

- Center: $\mu_{\bar{X}} = \mu = 12$
- Spread: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{4}} = 0.045$
- Shape: Unknown because we do not know the original distribution and the sample size is small.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example: Amount of Pepsi

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 100 cans are examined, describe the distribution of the sample means $\bar{X}$.

- Center: $\mu_{\bar{X}} = \mu = 12$
- Spread: $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.09}{\sqrt{100}} = 0.009$
- Shape: Normal because we have a large sample thus we can apply the **Central Limit Theorem.**

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Finding Probabilities

Assume that cans of Pepsi are filled so that the actual amount have a mean $\mu = 12$ oz and a standard deviation $\sigma = 0.09$ oz. Suppose that a random sample of 36 cans are examined, determine the **probability** that a sample of 36 cans **will have a sample mean amount**, $\bar{X}$ of at least 12.01 oz.

- To find this probability we need to first describe the distribution:
  - Shape: Normal because of the **Central Limit Theorem**
  - Center: $E[\bar{X}] = \mu_{\bar{x}} = \mu = 12$
  - Spread: $SD[\bar{X}] = \sigma_{\bar{x}} = \sigma/\sqrt{n} = 0.09/\sqrt{36} = 0.015$ this is the standard deviation we use.

n=30

$P(\bar{X} \geq 12.01) = 1 -$ pnorm(12.01, 12, 0.09/sqrt(36))

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Finding Probabilities

1. Put the question into a probability statement:

$$P(\bar{X} \geq 12.01)$$

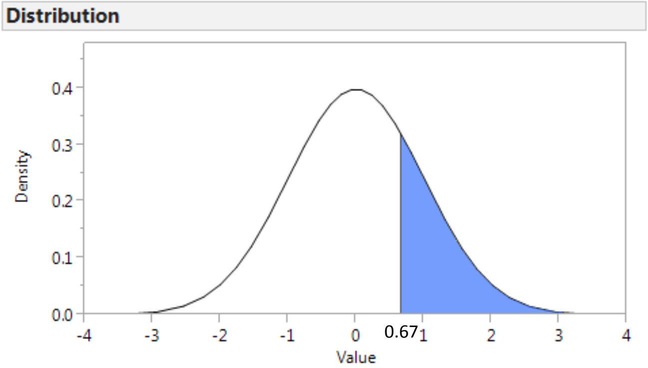
2. Convert to z-score:

$$P\left(\frac{(\bar{X} - \mu_{\bar{X}})}{\sigma_{\bar{X}}} \geq \frac{(12.01 - 12)}{0.015}\right) = P(Z \geq 0.67)$$

Notice: we are using the standard deviation divided by the square root of the sample size in the denominator.

Finding Probabilities

3. Draw and shade desired area.



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

4. Use Z-table or R to determine probability:
5. R: 1 - pnorm(12.01, 12, 0.015) = 0.2524925

$$\begin{aligned} P(\bar{X} \geq 12.01) &= P(Z \geq 0.67) \\ &= 1 - P(Z < 0.67) \\ &= 1 - 0.7486 \\ &= 0.2514 \end{aligned}$$

6. **Answer**: The probability that 36 cans will have a mean of at least 12.01 ounces is **0.2514.**

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Notes about finding probabilities for $\bar{X}$

- We have a sample size n. Thus the standard deviation changes by that value $\mathrm{SD}(\bar{X}) = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.
- The mean stays the same. $\mathrm{mean}(\bar{X}) = \mu_{\bar{X}} = \mu$.
- If we know that the original distribution is Normal **or** we have a large enough sample ($n > 30$). We can use the Normal distributions to find the probabilities.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example Sampling Distribution

$n=4 \quad \mu = 4.7 \quad \sigma = 0.4$

An orange juice producer buys all his oranges from a large orange grove. The amount of juice squeezed from each of these oranges is approximately normally distributed, with a mean of 4.70 ounces and a standard deviation of 0.40 ounce. Suppose we take a random sample of 4 oranges and determine the mean of this sample, $\bar{X}$.

1. What is the shape of the sampling distribution of $\bar{X}$.

   a. Normal b. Uniform c. Skewed left d. Skewed right

2. What is the mean of the sampling distribution of $\bar{X}$.

   $\mu_{\bar{x}} = E(\bar{x}) = 4.7$

   a. 4.7 b. 1.175 c. 2.35 d. 18.8

3. What is the standard deviation of the sampling distribution of $\bar{X}$.

   $\sigma_{\bar{x}} = SD(\bar{x}) = \frac{0.4}{\sqrt{4}}$

   a. 0.40 b. 0.10 c. 0.20 d. 1.60

4. What is the probability that the sample mean of the 4 oranges will be at 4.5 or less?

   $P(\bar{x} \le 4.5) = \text{pnorm}(4.5, 4.7, 0.2)$

   a. 0.1587 b. 0.3085 c. 0.8413 d. 0.6915

Sample Proportions

- The population proportion is p a parameter. In some cases we do not know the population proportion, thus we use the sample proportion, $\hat{p}$ to estimate p.
- The sample proportion is calculated by: $\hat{p} = \frac{X}{n}$
- X = the number of observations of interest in the sample or the number of "successes" in the sample.
- n = the sample size or number of observations.

$X \sim Bin(n, p)$ $E(X) = np$ $Var(X) = np(1-p)$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

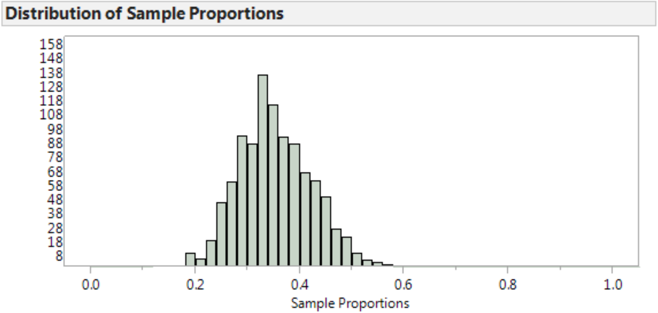
Example

$p = 0.34 \quad n = 50$

- According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes.
- A sample of 50 taxpayers was selected what do we expect the sample proportion $\hat{p}$ to be? $E(\hat{p}) = 0.34$
- Of we take other samples will the sample proportions always be the same value?
- If not what would $\hat{p}$ be off by?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sample Distribution of $n = 50$.

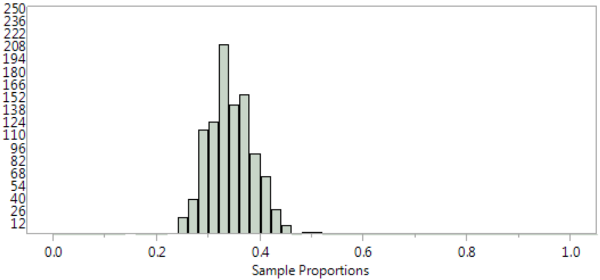


Summary Statistics	
Mean of Sample Proportions	0.34244
Std Dev of Sample Proportions	0.06987
No.of Samples	1000

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Sample Distribution of $n = 125$

Distribution of Sample Proportions



Summary Statistics

Mean of Sample Proportions	0.34082
Std Dev of Sample Proportions	0.04375
No.of Samples	1000

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Shape of the distribution of $\hat{p}$

- Notice from the previous histograms that it appears to have a **Normal** distribution.
- We can use the Normal distribution as long as $np \geq 10$ the number of successes are at least 10 and $n(1-p) \geq 10$ the number of failures are at least 10.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Center of the distribution of $\hat{p}$

- The center is the mean (expected value): $\mu_{\hat{p}} = p$ the proportion of success.
- $\hat{p} = \frac{X}{n}$ where X is the number of **successes** out of n observations. Thus X has a binomial distribution with parameters n and p.
- The mean of X is:

$$\mu_X = np$$

- Thus by rule 1b for means, the mean of $\hat{p}$ is:

$$\mu_{\hat{p}} = \mu_{\frac{X}{n}} = \frac{\mu_X}{n} = \frac{np}{n} = p$$

$E(\hat{p}) = p$

$E(\hat{p}) = E(\frac{x}{n}) = \frac{1}{n} E(x) = \frac{1}{n}(np) = p$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Spread of the distribution of $\hat{p}$

- The spread is the standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- The standard deviation of X is:

$$\sigma_X = \sqrt{np(1-p)}$$

- By rule 1b for standard deviation, the standard deviation of $\hat{p}$ is:

$$\sigma_{\hat{p}} = \sigma_{\frac{X}{n}} = \frac{\sigma_X}{n} = \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{np(1-p)}{n^2}} = \sqrt{\frac{p(1-p)}{n}}$$

$$Var(\hat{p}) = Var\left(\frac{X}{n}\right) = \frac{1}{n^2} Var(X) = \frac{1}{n^2}(np(1-p))$$

$$Var(\hat{p}) = \frac{p(1-p)}{n} \qquad SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Assumptions

- The sampled values must be random and independent of each other. This can be tested by **10% Condition**: The sample size must be no larger than 10% of the population.
- The sample size, n must be large enough. This can be be tested by **Success** / **Failure Condition**: The sample size has to be big enough so that both np and $n(1-p)$ at least 10.

$$N\left(E(\hat{p}) = p,\ SD(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}\right)$$

Example for distribution of $\hat{p}$

According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the distribution of $\hat{p}$, the sample proportion of the 125 taxpayers that used computer software to do their taxes?

1. Check if we can use the Normal distribution.
   - $p = 0.34$, $n = 125$
   - $np = 125(0.34) = 42.5$
   - $n(1-p) = 125(1-0.34) = 125(0.66) = 82.5$
   - Both np and $n(1-p)$ are greater than 10 so we can use the Normal distribution.
2. The mean is: $\mu_{\hat{p}} = p = 0.34$. If we take a sample we "expect" 34% to have used computer software to do their taxes.
3. The standard deviation is:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.34(1-0.34)}{125}} = 0.0424$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example continued

According to the National Retail Federation, 34% of taxpayers used computer software to do their taxes. A sample of 125 taxpayers was selected. What is the probability that between 28% and 40% of the taxpayers from the sample of 125 used computer software to do their taxes?

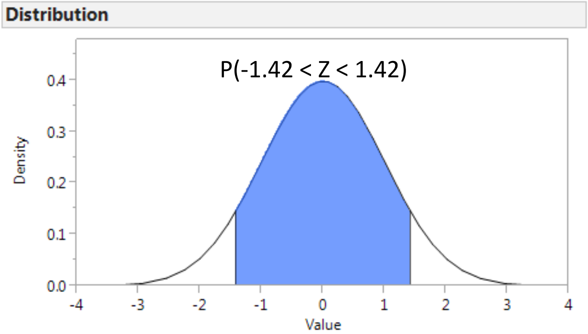
1. We want: $P(0.28 < \hat{p} < 0.40)$
2. We know that $\hat{p}$ has a **Normal** distribution with mean, $\mu_{\hat{p}} = 0.34$ and standard deviation, $\sigma_{\hat{p}} = 0.0424$. Thus we can apply the same rules as we did for the Normal distribution to find probability.
3. Convert to z: $z = \frac{\text{value} - \text{mean}}{\text{SD}}$

$$\begin{aligned} P(0.28 < \hat{p} < 0.40) &= P\left(\frac{(0.28 - 0.34)}{0.0424} < \frac{(\hat{p} - \mu_{\hat{p}})}{\sigma_{\hat{p}}} < \frac{(0.40 - 0.34)}{0.0424}\right) \\ &= P(-1.42 < Z < 1.42) \end{aligned}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example Continued

4. Draw and shade



UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example Continued

5. Use Z-table or R to determine probability:
6. R: pnorm(0.4,0.34,0.0424) - pnorm(0.28,0.34,0.0424) = 0.8429592

$$\begin{aligned} P(0.28 < \hat{p} < 0.40) &= P(-1.42 < Z < 1.42) \\ &= 0.9222 - 0.0778 = 0.8444 \end{aligned}$$

7. Answer: The probability that between 28% and 40% of the taxpayers from the sample of 125 used computer software to do their taxes is 0.8444.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Example $n = 140 \quad p = 0.69$

The Social Media and Personal Responsibility Survey in 2010 found the 69% of parents are "friends" with their children on Facebook. A random sample of 140 parents was selected and we determined the proportion of parents from this sample, $\hat{p}$ that are "friends" with their children on Facebook.

1. What is the shape of the sampling distribution of $\hat{p}$.

   $np = 140(0.69) = 96.6 > 10$

   $n(1-p) = 140(0.31) = 43.4 > 10$ } Normal dist.

   ($0.31 \leftarrow 1 - .69$)

2. What is the mean of the sampling distribution of $\hat{p}$.

   $E(\hat{p}) = 0.69$

3. What is the standard deviation of the sampling distribution of $\hat{p}$.

   $SD(\hat{p}) = \sqrt{\frac{0.69(1-0.69)}{140}} = 0.0391$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Probability of Proportions

4. What is the probability that the sample proportion of 140 parents is greater than 72%?

$$P(\hat{p} > 0.72) = 1 - \text{pnorm}(0.72, 0.69, 0.0391)$$
$$= 0.2215$$