Section 5.1: Normal Distributions

**To do at home**

Read the book and also listen to the assigned videos.

As you study this chapter, start creating the “flipping study guide” for the chapter; remember that the most effective one contains one topic per page. You will be given a quiz on Wednesday and you will be able to use this study guide if you have it.

1) Go to WILEY HWK - Watch video on learning goal 1 – “Estimate probabilities as areas of a density function”.
   a) What is a density curve?

   Theoretical model of a distribution - smothered out histogram/dot plot

   b) What are the properties of a density curve?

   
   ① Area under it = 1 (accounts for 100% of distribution)
   ② Area of any interval = proportion of distribution in interval

2) Read the solution to example 5.1 on page 325, section 5.1 of the book. Sketch the density curve in this paper, label, and answer the following:
   a. Estimate the proportion is more than 30 minutes?
     \[ 1 - .80 = .20 \]
   b. Estimate the proportion is less than 31 minutes?
     \[ 1 - .05 = .95 \]
   c. Estimate the proportion is below 27?
     \[ .025 \]
   d. Estimate the proportion is between 28 and 29?
     .34

3) Go to WILEY HWK - Watch video on learning goal 2 – Recognize how the mean and standard deviation relate to the center and spread of a normal distribution.

   a. Sketch a normal density curve
   b. What is its shape?
      Bell shaped
   c. Give the parameters that define it
   d. Give the notation used (it’s on the book)

\[ N(\mu, \sigma) \]

4) Copy here Figure 5.4 from the book – page 326. Use the Empirical rule; put the percentages within 1, 2, and 3 standard deviations from the mean. (If you don’t remember it, google it!)
Using the calculator to find areas under the normal curve

Areas = probabilities = proportions

\[ N(0, 1) = \text{z-score distribution} \]

\[ \text{Standard Normal Distribution} \]

\[ \mu = 0 \]
\[ \sigma = 1 \]

To find areas/prob/prop. with calculator
USE normalcdf \( (\text{lower}, \text{upper}, \mu, \sigma) \)

2nd VARS [DISTR]

Area between \( -1 \& 1 \) in \( N(0, 1) \)

\[ \text{normalcdf}(-1, 1, 0, 1) = 0.6827 \]

\[ P(-2 < z < 2) = \text{normalcdf}(-2, 2, 0, 1) = 0.9545 \]

\[ P(-3 < z < 3) = \text{normalcdf}(-3, 3, 0, 1) = 0.9973 \]
Finding Normal Probabilities, proportions, areas under the normal curve

5) Go to WILEY HWK - Watch video on learning goal 3 – Using STATKEY to estimate proportions under a normal curve. (In STATKEY – under THEORETICAL distributions, click on NORMAL)

You will be quizzed on this next class.

a. How do you access it?
b. How do you change the x-value?
c. How do you change the proportion?
d. How do you change the parameters?
e. How do you find an area below, above or between?

Take notes of the examples and the answers; sketch and shade for each example. In class we will learn how to use the calculator to accomplish the same.

6) Example from this video on learning goal 3: “birth weight of babies” - Copy the statement of the problem in this paper and take notes. You will be quizzed on this next class.

a. Here you will learn how to use the standardized values (z-scores)
b. Also how to use the actual parameters given on the problem.

Example: Suppose birth weights are normally distributed with mean \( \mu = 3400 \) grams and std. dev. \( \sigma = 600 \) grams.

What proportion of babies weigh less than 2500 grams? \( 6.7\% \)

Using \( Z \sim N(0,1) \):

\[
Z = \frac{x - \mu}{\sigma} = \frac{2500 - 3400}{600} = -1.5
\]
7) Sketch the normal density curve \( N(5, 2) \) and label one, two and three standard deviations around the mean; then, complete the statements.

   a. 95% of the observations are within \( 1 \) and \( 9 \)
   b. 68% are within \( 3 \) and \( 7 \)
   c. 99.7% are within \( -1 \) and \( 11 \)

8) Women’s heights are normally distributed with a mean of 65.0 inches and standard deviation of 3.5 inches. Men’s heights have the same shape and parameters 70.0 inches and 4.0 inches. Use STATKEY to find the proportions.

   a. Use correct notation to describe the distributions.
   b. Sketch the corresponding density curves.
   c. Find the proportion of women taller than 72 inches
   d. Find the proportion of men shorter than 70 inches
   e. Find the proportion of women with heights between 61.5 and 68.5 inches
   f. Find the proportion of men with heights between 62 and 78 inches

   **Show work below, label a-f**

   (a) \( X \sim N(65.0, 3.5) \)
   \( X \sim N(70.0, 4.0) \)

   (c) \( x(w) = 72 \) in
   \( P(X > 72) = \text{normalcdf}(72, 10^9, 65, 3.5) = 0.023 \)

   (d) \( x(m) = 70 \) in
   \( P(X < 70) = \text{normalcdf}(-10^9, 70, 70, 4) = 0.5 \) (we knew this!)

   (e) Women’s height \( N(65, 3.5) \)
   \( P(61.5 < x < 68.5) = \text{normalcdf}(61.5, 68.5, 65, 3.5) = 0.683 \)

   (f) Men’s \( P(62 < x < 78) = \text{normalcdf}(62, 78, 70, 4) = 0.954 \)
   think of empirical rule!
Finding Scores from a given proportion

\[ \text{Inv Norm} (\text{Area to the left, mean, st. dev.}) \]

9) Go to WILEY HWK - Watch video on learning goal 4 – Using technology to find endpoint(s) of intervals with a specified probability for any normal curve. Take notes below.

(a) \( x \) scores on exam \( \sim N \left( \mu = 75, \sigma = 10 \right) \)

What endpoints gives the top 10%?

\[ \text{Inv Norm} \left( 0.90, 75, 10 \right) = 87.816 \]

(b) bottom 20%?

\[ \text{Inv Norm} \left( 0.20, 75, 10 \right) = 66.584 \]

(c) middle 90%?

\[ \text{Inv Norm} \left( 0.05, 75, 10 \right) = 58.551 \]

\[ \text{Inv Norm} \left( 0.95, 75, 10 \right) = 91.449 \]

Converting from a general normal distribution to the standard normal distribution and vice versa

10) Go to WILEY HWK – Watch video on learning goal 5 – Convert in either direction between a general \( N(\mu, \sigma) \) distribution and a standard \( N(0, 1) \) distribution. Write the two formulas given.

\[ \text{From } N(\mu, \sigma) \text{ to } N(0, 1) \text{ use } \frac{X - \mu}{\sigma} \]

\[ \text{From } N(0, 1) \text{ to } N(\mu, \sigma) \text{ use } Y = Z \sigma + \mu \]
Section 5.2: Confidence Intervals and P-values - Using Normal Distributions

• Table 5.2 - Normal percentiles for common confidence levels

<table>
<thead>
<tr>
<th>Confidence level</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
<th>98%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>z*</td>
<td>1.282</td>
<td>1.645</td>
<td>1.960</td>
<td>2.326</td>
<td>2.576</td>
</tr>
</tbody>
</table>

1) Let's check this out with technology (you try STATKEY at home)

2) Let's check this out with STATKEY

• Central Limit Theorem
For random samples with a sufficiently large sample size, the distribution of sample statistics for a mean or a proportion is normally distributed and is centered at the value of the population parameter.

• Confidence Interval based on a Normal Distribution
If the distribution for a statistic follows the shape of a normal distribution with standard error SE, we find a confidence interval for the parameter using

\[ \bar{x} \pm z^* \times SE \]

Where \( z^* \) is chosen so that the area between \(-z^*\) and \(+z^*\) in the standard normal distribution is the desired level of confidence.
3) Example 1: Obesity in America
In Chapter 3, we see that the mean BMI (Body Mass Index) for a large sample of US adults is 27.655. We are told that the standard error for this estimate is 0.009. If we use the normal distribution to find a 99% confidence interval for the mean BMI of US adults:

(a) What is $z^*$? 2.576

(b) Find and interpret the 99% confidence interval.

\[ 3S \pm z^* \text{SE} \]
\[ 27.655 \pm 2.576 (0.009) \]
\[ 27.655 \pm 0.023184 \]
\[ 27.63 \text{ to } 27.68 \]

Note: A BMI > 25 is classified as overweight

⇒ American average BMI in 2010 is considered overweight

4) Example 2: Obesity in America: Exercises vs Non-exercisers
Also in Chapter 3, we see that the difference in mean BMI between non-exercisers (those who said they had not exercised at all in the last 30 days) and exercisers (who said they had exercised at least once in the last 30 days) is $\bar{x}_N - \bar{x}_E = 1.915$, with a standard error for the estimate of SE = 0.16. If we use the normal distribution to find a 90% confidence interval for the difference in mean BMI between the two groups:

(a) What is $z^*$? 1.645

(b) Find and interpret the 90% confidence interval.

\[ 1.915 \pm 1.645 \text{SE} \]
\[ 1.915 \pm 0.2632 \]
\[ 1.94 \text{ to } 1.94 \]

We are 90% confident that people living in the US in 2010 who do not exercise have mean BMI between 1.94 greater than people who do.

Question: If we test the hypotheses $\mu_N = \mu_E$ vs $\mu_N > \mu_E$, what will the conclusion be? $R_H$ or $DNR_H$
Hypothesis Test based on a Normal Distribution

When the distribution of the statistic under $H_0$ is normal, we compute a standardized test statistic using

$$ z = \frac{\text{Sample Statistic} - \text{Null Parameter}}{SE} $$

The p-value for the test is the probability a standard normal value is beyond this standardized test statistic, depending on the direction of the alternative hypothesis.

You will be using

(a) The calculator and normalcdf (lower, upper, mean, standard error) to find the p-value.
(b) or STATKEY

You will then use methods of chapter 4 to decide whether to reject $H_0$ or do not reject $H_0$.

5) Example 3: Is Divorce Morally Acceptable?

In a study introduced in Chapter 4, we learn that 67% of women in a random sample view divorce as morally acceptable. Does this provide evidence that more than 60% of women view divorce as morally acceptable? The standard error for the estimate assuming the null hypothesis is true is 0.021.

(a) What are the null and alternative hypotheses for this test?

$$ H_0: p = 0.60 \quad H_a: p > 0.60 $$

(b) What is the test statistic?

$$ z = \frac{\hat{p} - p}{SE} = \frac{0.67 - 0.60}{0.021} = 3.33 $$

(c) Use the normal distribution to find the p-value.

$$ p-value = P(\hat{p} > 0.67) = P(x > 3.33) = 1 - \text{normalcdf}(3.33, \infty, 0, 1) = 0.00043 $$

(d) What is the conclusion of the test?

There is very strong evidence that women view divorce as morally acceptable.

6) Example 4: Do Men and Women Differ in Opinions about Divorce?

In the same study described above, we find that 71% of men view divorce as morally acceptable. Use this and the information in the previous example to test whether there is a significant difference between men and women in how they view divorce. The standard error for the difference in proportions under the null hypothesis that the proportions are equal is 0.029.

(a) What are the null and alternative hypotheses for this test?

$$ H_0: p_m = p_w \quad H_a: p_m \neq p_w $$

(b) What is the test statistic?

$$ z = \frac{(0.71 - 0.67)}{0.029} = 1.379 $$

(c) Use the normal distribution to find the p-value.

$$ p-value = 2 \times [1 - \text{normalcdf}(1.379, 0, 1)] = 2 \times 0.8339 = 1.66 $$

(d) What is the conclusion of the test?

$p = 1.66 > 0.05 \Rightarrow$ we do not find evidence of a difference between the views of men and women that view divorce as morally acceptable.
7) **Quick Self-Quiz: Confidence Intervals using the Normal Distribution**

In a recent survey of 1000 US adults conducted in January 2013, 57% said they dine out at least once per week. The standard error for this estimate is 0.016. Use the normal distribution to find a 95% confidence interval for the proportion of US adults who dine out at least once per week. Interpret your answer.

\[ \hat{p} \pm z \times SE \]
\[ 0.57 \pm 1.96 \times 0.016 \]
\[ 0.57 \pm 0.03136 \]
\[ 0.539 \text{ to } 0.601 \]

*We are 95% confident that the proportion of all US adults who dine out at least once per week is between 0.539 to 0.601.*

8) **Quick Self-Quiz: Hypothesis Tests using the Normal Distribution**

A sample of baseball games shows that the mean length of the games is 179.83 minutes. (The data is given in Baseball Times). The standard error is 3.75. Does this sample provide evidence that the mean length of time for baseball games is more than 170 minutes? Use the normal distribution and show all details of the test.

**Given**
\( \bar{x} = 179.83 \text{ min} \)
\( SE = 3.75 \text{ min} \)

**Ho**: \( \mu = 170 \)
**Ha**: \( \mu > 170 \)

**Test Statistic**
\[ Z = \frac{\bar{x} - \text{mean}}{SE} = \frac{179.83 - 170}{3.75} \]
\[ Z = 2.621 \]

**P-value**
\( P(Z > 2.621) = 0.004 < 0.05 \)

**RHo & S Ha**
We have strong evidence that the average length of a baseball game is $> 170$ minutes.