Live Session 6

1. Welcome/Intro (including polls)
2. Correlation and Regression – next week
3. Quiz 2 Prep
4. Assignments for next 2 weeks
5. Wrap up and Feedback

Quiz #2 Prep

2-hours timed, multiple choice/drop downs

Open book, open notes, includes:

<u>Chapter 3 </u>–

Describing data numerically (measures of center and spread)

<u>Chapter 6 (skip 6.3) </u>–

Probability Distributions (focus on normal and binomial)

<u>Chapter 8 (skip 8.4)</u>–

Confidence Intervals (for mean, and proportions)

<u>Chapter 9 (skip 9.2 & 9.6, only need p-value method)</u>–

Hypothesis testing (mean and proportion)

<u>Chapter 11 (only 11.2) </u>–

Chi Square test for independence

Probability Distributions: Normal and Binomial

Binomial Probability Example

Suppose we know the population proportion *p* of left-handed students is 0.10, and we have a random sample of 10 students.

What is the probability that there are 2 left-handed students in the sample?

Binomial Probability Example

Suppose we know the population proportion *p* of left-handed students is 0.10, and we have a random sample of 10 students.

What is the probability that there are 2 left-handed students in the sample?

Option 1:

Binomial Table, Table B
N = 10, x = 2, p = 0.10
Probability = 0.1937

Option 2:

Binomial probability formula in Excel
N = 10, x = 2, p = 0.10
=BINOM.DIST(2, 10, 0.1, FALSE)
Probability = 0.1937

There is a 19.37% probability that there are 2 left-handed students in the sample.

Table B Binomial distribution (*continued*)

n	*X*	0.10	0.15	0.20
9	0	0.3874	0.2316	0.1342
	1	0.3874	0.3679	0.3020
	2	0.1722	0.2597	0.3020
	3	0.0446	0.1069	0.1762
	4	0.0074	0.0283	0.0661
	5	0.0008	0.0050	0.0165
	6	0.0001	0.0006	0.0028
	7			0.0003
	8			
	9			
10	0	0.3487	0.1969	0.1074
	1	0.3874	0.3474	0.2684
	2	0.1937	0.2759	0.3020
	3	0.0574	0.1298	0.2013
	4	0.0112	0.0401	0.0881
	5	0.0015	0.0085	0.0264
	6	0.0001	0.0012	0.0055
	7		0.0001	0.0008
	8			0.0001
	9			
	10			

1. What is the probability of 2 students or less?
2. What is the probability of 4 students or more?

Binomial Probability Example

Suppose we know the population proportion *p* of left-handed students is 0.10, and we have a random sample of 10 students.

1. What is the probability of 2 students or less?

Option 1:
Binomial Table, Table B
N = 10, x = 0, 1, and 2 p = 0.10 and add them
Probability = 0.3487+0.3874+0.19371=0.9298

Option 2:
Binomial probability formula in Excel
N = 10, x = 2, p = 0.10
=BINOM.DIST(2, 10, 0.1, TRUE)
Probability = 0.9298

There is a 92.98% probability that there are 2 or less left-handed students in the sample.

2. What is the probability of 4 students or more?

Option 1:
Binomial Table, Table B
N = 10, x = 4,5, 6, 7,8,9,10, p = 0.10
Probability = 0.0112+.0015+.0001=0.0128

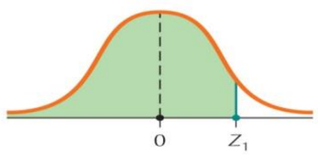
Option 2:
Binomial probability formula in Excel
N = 10, x = 3, p = 0.10
=1-BINOM.DIST(3, 10, 0.1, TRUE)
Probability = 0.0128

There is a 1.28% probability that there are 4 or more left-handed students in the sample.

Normal Probabilities:
Finding Areas Under the Standard Normal Curve

$$Z = \frac{x - \mu}{\sigma}$$

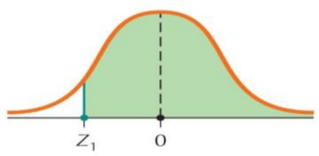
Case 1
Find the area to the left of Z_1.
Step 1 Draw the standard normal curve. Label the Z-value Z_1.
Step 2 Shade in the area to the left of Z_1.



Step 3 Use the Z table to find the area to the left of Z_1.

Formula in Excel:
=NORM.S.DIST(Z,TRUE)

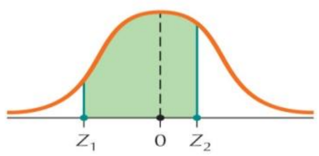
Case 2
Find the area to the right of Z_1.
Step 1 Draw the standard normal curve. Label the Z-value Z_1.
Step 2 Shade in the area to the right of Z_1.



Step 3 Use the Z table to find the area to the left of Z_1. The area to the right of Z_1 is then equal to 1 – (area to the left of Z_1).

Formula in Excel:
=1-NORM.S.DIST(Z,TRUE)

Case 3
Find the area between Z_1 and Z_2.
Step 1 Draw the standard normal curve. Label the Z-values Z_1 and Z_2.
Step 2 Shade in the area between Z_1 and Z_2.



Step 3 Use the Z table to find the area to the left of Z_1 and the area to the left of Z_2. The area between Z_1 and Z_2 is then equal to (area to the left of Z_2) – (area to the left of Z_1).

Formula in Excel:
=NORM.S.DIST(Z2,TRUE) - NORM.S.DIST(Z1,TRUE)

Or calculate the probabilities directly:

Formula in Excel:
= NORM.DIST(x, mean, std dev, TRUE)

Formula in Excel:
= 1-NORM.DIST(x, mean, std dev, TRUE)

Formula in Excel:
= NORM.DIST(x2, mean, std dev, TRUE) - NORM.DIST(x1, mean, std dev, TRUE)

Normal Distribution – Probability example

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of $1000 and a standard deviation of $100. What percent of the supervisors have a weekly income less than $1200?

OPTIONS:

1. Calculate the z value. Look up the value in the z-table (TABLE C)
2. Calculate the z value. Use the formula in Excel: = NORM.S.DIST(z,TRUE)
3. Calculate the % directly using formula in Excel: = NORM.DIST(x, mean, std dev, TRUE)

Normal Distribution – Probability example

The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of \$1000 and a standard deviation of \$100. What percent of the supervisors have a weekly income less than \$1200?

$$Z = \frac{x-\mu}{\sigma} = \frac{1200 - 1000}{100} = 2.0$$

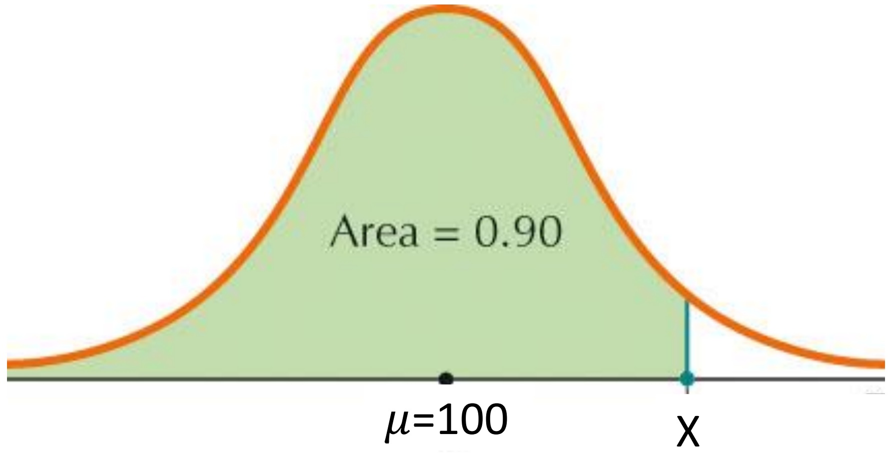
OPTIONS:

1. Calculate the z value. Look up the value in the z-table (TABLE C) = 0.9772
2. Calculate the z value. Use the formula in Excel: = NORM.S.DIST(z,TRUE) = NORM.S.DIST(2,TRUE) = 0.9772
3. Calculate the % directly using formula in Excel: = NORM.DIST(x, mean, std dev, TRUE)
   = NORM.DIST(1200,1000,100,TRUE) = 0.9772

Prep for Quiz #2

Finding *X*-Values for a Given Area

Find the *X*-value with area 0.90 to its left, given that $\mu = 100$, $\sigma = 5$.



1. Draw the standard normal curve and label *X*

2. Shade the area to the left of *X* and label with the given area of 0.90.

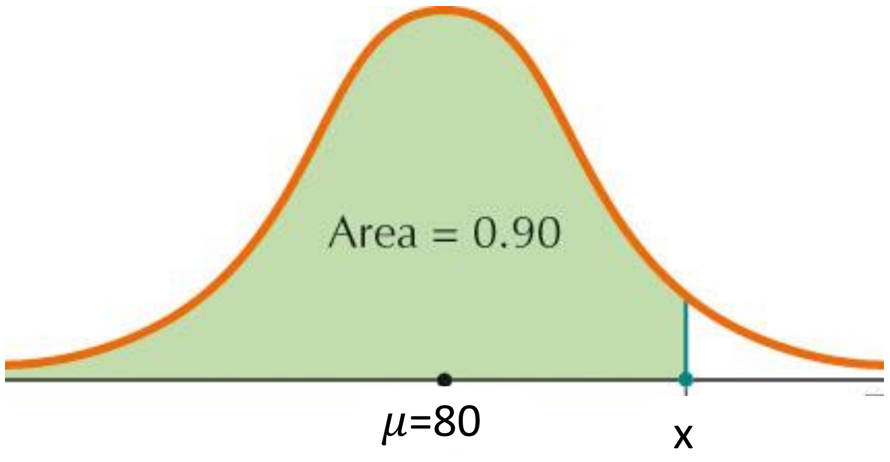
3. Use the Excel formula to solve for X
= NORM.INV(area, μ, σ)
= NORM.INV(0.90, 100, 5)
= 106.4

Example: What weight of 10 year olds at the 90th percentile (greater than 90% of them) if the average weight is 80 lbs and the standard deviation is 7 lbs? Assume a normal distribution

Prep for Quiz #2

Finding *X*-Values for a Given Area

Example: What weight of 10 year olds at the 90th percentile (greater than 90% of them) if the average weight is 80 lbs and the standard deviation is 7 lbs? Assume a normal distribution.



1. Draw the standard normal curve and label *X*

2. Shade the area to the left of *X* and label with the given area of 0.90.

3. Use the Excel formula to solve for X
= NORM.INV(area, μ, σ)
= NORM.INV(0.90, 80, 7)
= 88.97

Hypothesis Testing: Means and Proportions

How do we address hypothesis test questions?

Steps for solving problems:

1. What type of data is this? Discrete or continuous
2. What other relevant information is found in this problem?

List out all of the items provided:

- Sample size (n)
- Means (xbar, μ)
- Number of samples (1 or 2)
- Alpha (α)
- Goals/Targets (μ_o ,p_o)
- Proportions (phat, p)
- P-values (p-value to compare to α)
- Less than, greater than, equal

3. What test/formula do we use?

Purple, orange, green, pink. One tail, two tail.

4. Calculate the values.

Calculate the test statistic. Look up/calculate p-value.

5. Interpret the result. Reject H_o?

Converting Words to Hypotheses

To convert a word problem into two hypotheses, look for key words that can be expressed mathematically.

English words	Symbols	Synonyms
Equal	$=$	Is; is the same as
Not equal	$\neq$	Is different from; has changed from; differs from
Greater than	$>$	Is more than; is larger than; exceeds
Less than	$<$	Is below; is smaller than
At least	$\geq$	Is this much or more; is greater than or equal to
At most	$\leq$	Is this much or less; is less than or equal to

Strategy for Constructing Hypotheses About μ

1. Search the word problem for key words and select the associated symbol.
2. Determine the form of the hypotheses that uses this symbol.
3. Find the value of μ_0 and write your hypotheses in the appropriate form.

Prep for Quiz #2

Pop-up polls

1. Provide the null and alternative hypotheses when testing whether the mean exceeds -2

a) H0: $\mu = -2$ vs. Ha: $\mu \neq -2$
b) H0: $\mu \geq -2$ vs. Ha: $\mu < -2$
c) H0: $\mu \leq -2$ vs. Ha: $\mu > -2$

2. The Statistical Abstract of the United States reports that the mean daily number of shares traded on the New York Stock Exchange in 2005 was 1.755 billion. Based on a sample of this year's trading results, a financial analyst would like to test whether the mean number of shares traded will be less than the 2005 level.
Provide the null and alternative hypotheses:

a) H0: $\mu \geq 1.755$billion vs. Ha: $\mu < 1.755$billion
b) H0: $\mu \leq 1.755$billion vs. Ha: $\mu > 1.755$billion
c) H0: $\mu = 1.755$billion vs. Ha: $\mu \neq 1.755$billion

Prep for Quiz #2

1. Provide the null and alternative hypotheses when testing whether the mean exceeds -2

a) H0: $\mu = -2$ vs. Ha: $\mu \neq -2$
b) H0: $\mu \geq -2$ vs. Ha: $\mu < -2$
c) H0: $\mu \leq -2$ vs. Ha: $\mu > -2$

2. The Statistical Abstract of the United States reports that the mean daily number of shares traded on the New York Stock Exchange in 2005 was 1.755 billion. Based on a sample of this year's trading results, a financial analyst would like to test whether the mean number of shares traded will be less than the 2005 level.
Provide the null and alternative hypotheses:

a) H0: $\mu \geq 1.755$billion vs. Ha: $\mu < 1.755$billion
b) H0: $\mu \leq 1.755$billion vs. Ha: $\mu > 1.755$billion
c) H0: $\mu = 1.755$billion vs. Ha: $\mu \neq 1.755$billion

The *p*-Value

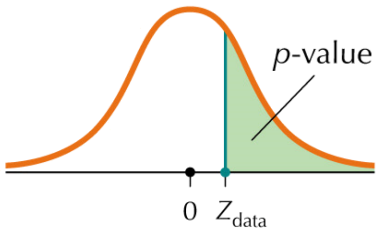
Type of hypothesis test	*p*-Value is tail area associated with Z_{data}

Right-tailed test

$H_0: \mu \le \mu_0$ versus $H_a: \mu > \mu_0$

p-value $= P(Z > Z_{data})$

Area to right of Z_{data}

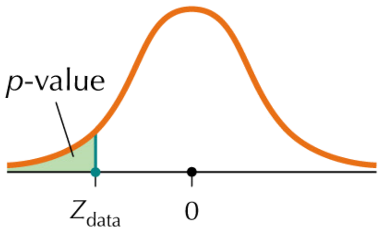


Left-tailed test

$H_0: \mu \ge \mu_0$ versus $H_a: \mu < \mu_0$

p-value $= P(Z < Z_{data})$

Area to left of Z_{data}

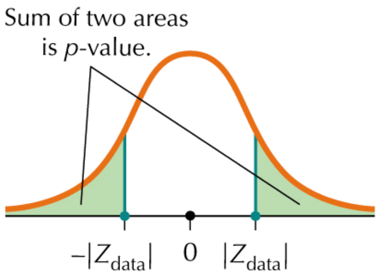


Two-tailed test

$H_0: \mu = \mu_0$ versus $H_a: \mu \ne \mu_0$

$$p\text{-value} = P\left(Z > |Z_{data}|\right) + P\left(Z < -|Z_{data}|\right)$$
$$= 2 \cdot P\left(Z > |Z_{data}|\right)$$

Sum of the two tail areas.



The rejection rule for performing a hypothesis test using the *p*-value method is:

- **Reject H_0 when the *p*-value ≤ α.**
- **Otherwise, do not reject H_0.**

Choosing the hypothesis test

Continuous

One Sample

One-Sample Hypothesis Tests for Continuous Data (Purple)

Select:	Two-tail test	One-tail test	
	Two-tail	Lower/left-tail	Upper/right-tail
	H_0: $\mu = \mu_0$	H_0: $\mu \geq \mu_0$	H_0: $\mu \leq \mu_0$
	H_a: $\mu \neq \mu_0$	H_a: $\mu < \mu_0$	H_a: $\mu > \mu_0$
Choose:	Sample size		
	Large	Small	
	$n \geq 30$	$n < 30$	
	(or σ known)	(or σ unknown)	
Calculate:	Test statistic		
	$Z = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	$t = \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}$	
	Can replace s with σ if known	$df = n - 1$	
Identify:	p-value		
	Two-tail	Lower/left-tail	Upper/right-tail
	p = 2 × area past Z or t	p = area left of Z or t	p = area right of Z or t

Two Sample

Two-Sample Hypothesis Tests for Continuous Data (Green)

Select:	Two-tail test	One-tail test	
	Two-tail	Lower/left-tail	Upper/right-tail
	H_0: $\mu_1 = \mu_2$	H_0: $\mu_1 \geq \mu_2$	H_0: $\mu_1 \leq \mu_2$
	H_a: $\mu_1 \neq \mu_2$	H_a: $\mu_1 < \mu_2$	H_a: $\mu_1 > \mu_2$
Choose:	Sample size		
	Large	Small	
	$n_1 + n_2 \geq 30$	$n_1 + n_2 < 30$	
	(or σ known)	(or σ unknown)	
Calculate:	Test statistic		
	$Z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$t = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	
		$df = n_1 + n_2 - 2$	
Identify:	p-value		
	Two-tail	Lower/left-tail	Upper/right-tail
	p = 2 × area past Z or t	p = area left of Z or t	p = area right of Z or t

Discrete

One Sample

One-Sample Hypothesis Tests for Discrete Data (Orange)

Select:	Two-tail test	One-tail test	
	Two-tail	Lower/left-tail	Upper/right-tail
	H_0: $p = p_0$	H_0: $p \geq p_0$	H_0: $p \leq p_0$
	H_a: $p \neq p_0$	H_a: $p < p_0$	H_a: $p > p_0$
Choose:	Sample size		
	Must have	Where	
	$np \geq 5$	$p = \frac{X}{n}$	
	$n(1-p) \geq 5$	X = no. of items of interest in sample	
	$n \geq 30$		
Calculate:	Test statistic		
	$Z = \frac{p-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$		
Identify:	p-value		
	Two-tail	Lower/left-tail	Upper/right-tail
	p = 2 × area past Z	p = area left of Z	p = area right of Z

Two Sample

Two-Sample Hypothesis Tests for Discrete Data (Pink)

Select:	Two-tail test	One-tail test	
	Two-tail	Lower/left-tail	Upper/right-tail
	H_0: $p_1 = p_2$	H_0: $p_1 \geq p_2$	H_0: $p_1 \leq p_2$
	H_a: $p_1 \neq p_2$	H_a: $p_1 < p_2$	H_a: $p_1 > p_2$
Choose:	Sample size		
	Must have	Where	
	$n_1 + n_2 \geq 30$	$p_1 = \frac{X_1}{n_1}$ and $p_2 = \frac{X_2}{n_2}$	
		X = no. of items of interest in sample	
Calculate:	Test statistic		
	$Z = \frac{p_1-p_2}{\sqrt{\frac{x_1+x_2}{n_1+n_2}\left[1-\frac{x_1+x_2}{n_1+n_2}\right]\left[\frac{1}{n_1}+\frac{1}{n_2}\right]}}$		
Identify:	p-value		
	Two-tail	Lower/left-tail	Upper/right-tail
	p_1 = 2 × area past Z	p_1 = area left of Z	p_1 = area right of Z

	Calculating the p-value		
Test statistic:	z		t
two tail	if z value is less than 0:	=2*(NORM.S.DIST(z, TRUE))	=T.DIST.2T(t, df)
	if z value is greater than 0:	=2*(1-(NORM.S.DIST(z, TRUE)))	
lower/left tail	=NORM.S.DIST(z, TRUE)		=T.DIST(t, df, TRUE)
upper/right tail	=1-(NORM.S.DIST(z, TRUE))		=T.DIST.RT(t, df)

Prep for Quiz #2

Identify which test to use, calculate the p-value.

1. Ho:$\mu \leq 10$ vs Ha:$\mu > 10$, xbar =11, $\sigma=5$, n=25

2. Ho:$\mu \geq 9$ vs Ha:$\mu < 9$, xbar =8, s=4, n=25

3. Ho:$\mu = 100$ vs Ha:$\mu \neq 100$, xbar =94, $\sigma=10$, n=16

Prep for Quiz #2

Identify which test to use, calculate the p-value.

1. Ho:μ=10 vs Ha:μ>10, xbar =11, σ=5, n=25

$$z = \frac{xbar-\mu}{\sigma/\sqrt{n}} = \frac{11-10}{5/\sqrt{25}} = \frac{1}{1} = 1\text{, upper right tail test,}$$

p-value = 1-(NORM.S.DIST(1, TRUE)) = 0.1587

2. Ho:μ=9 vs Ha:μ<9, xbar =8, s=4, n=25

$$t = \frac{xbar-\mu}{s/\sqrt{n}} = \frac{8-9}{4/\sqrt{25}} = \frac{-1}{0.8} = -1.25\text{, lower left tail test,}$$

p-value = T.DIST(-1.25, 24, TRUE)= 0.111676

3. Ho:μ=100 vs Ha:$\mu\neq$100, xbar =94, σ=10, n=16

$$z = \frac{xbar-\mu}{\sigma/\sqrt{n}} = \frac{94-100}{10/\sqrt{16}} = \frac{-6}{2.5} = -2.4\text{, two tail test,}$$

p-value = 2*(NORM.S.DIST(-2.4, TRUE)) = 0.0164

Prep for Quiz #2

1) The distribution of weekly incomes of supervisors at the ABC Company follows the normal distribution, with a mean of \$1000 and a standard deviation of \$100. What percent of the supervisors have a weekly income between \$840 and \$1200?

What type of data is this?
What other relevant information is found in this problem?
How do we find the answer to this question?

Prep for Quiz #2

2) Twenty percent of the employees of ABC Company use direct deposit and have their wages sent directly to the bank. Assume we random sample five employees.

a. What is the probability that all five employees use direct deposit?
b. What is the probability that at least 2 employees use direct deposit?
c. What is the probability that 4 or more employees use direct deposit?

1. What type of data is this?
2. What other relevant information is found in this problem?
3. What test/formula do we use?
4. Calculate the solution.

Prep for Quiz #2

3) Three work shifts producing the same product sorted the finished product into 4 categories based on its quality level and displayed the results in the following table. Determine whether there is dependence between the shift and the quality of the product? (Does product quality depend on the shift that produces it?) Assume alpha= 0.05

	1st Shift	2nd Shift	3rd Shift
Perfect product	185	175	170
Acceptable product	55	60	65
Defective product	15	15	15
Reworked product	10	15	15

1. What type of data is this?
2. What other relevant information is found in this problem?
3. What test/formula do we use?
4. Calculate the solution.

Prep for Quiz #2

4) A bullet manufacturer claims to have produced a projectile having a mean muzzle velocity of more than 3000 feet per second. From a random sample of 60 bullets he calculates a sample mean of 3012 feet per second and a sample standard deviation of 112 feet per second. Does the data from the sample support his claim?

1. What type of data is this?
2. What other relevant information is found in this problem?
3. What test/formula do we use?
4. Calculate the solution.

Next two weeks

1. Project Next Steps – Measure/Analyze Phases

Measure/Analysis tools
Confirm your sample size, discuss your choice
Insights about the problem

2. Coursework BLT's:

6.3 Correlation Video

6.11 Test Your Knowledge: Hand/Foot Exercise

7.8 Test Your Knowledge: Categorical Input Variable

7.9* Relate Regression to Your Project

3. Assignments:

Quiz #2 (covers Chapters **3,6,8,9,11.2**)
3 days after live session 6

Upcoming assignment:

Homework #4: *(worth 5 points)*
Three days after live session 7
LaunchPad Assignments

- **LearningCurve** for Chapter **4**

Reminder: Understanding Variation - *part of HW#4 (page 114-116) week 8*