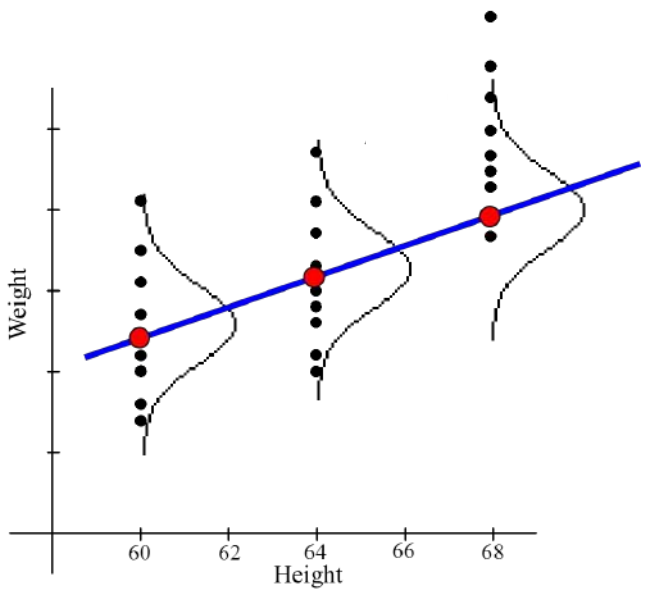
12.1 Regression Inference KEY POINTS

Regression Model Example – Refer to the graph

- **How much would an adult female weigh if she were 5 feet?**
  - **What would you expect the distribution of weights to be?**

- **What would you expect the distribution of weights to be for other heights?**



- Where would you expect the **TRUE LSRL** to be?

- What is the population regression model equation?

- What about the **standard deviations** of all these normal distributions?

Now the Regression Model:

- The **mean response** (μ_y) has a _____ relationship with x.

  The model is:

  $$\mu_y = \alpha + \beta x;$$

  Where, the unknown population parameters are:

  β is the _____and

  α is the _____.

- For any fixed value of the ***x*** (the explanatory variable), the _____ **variable** (y) varies according to a _____ **distribution**.

  The repeated responses of y are _____ of each other.

- The **standard deviation** of y (σ_y) is the _____ for all values of (x).

  σ_y is also an unknown _____

  _____.

Regression Inference
Unbiased Estimators

Population Parameters

- We use the model _____ to estimate the population least square regression line (LSRL) $\mu_y = \alpha + \beta x$.

- The slope b of the LSRL is **an _____ _____ estimator** of the true _____ β.

- The intercept a of the LSRL is **an _____ _____ estimator** of the true _____ α.

- The standard error s is **an _____ _____ estimator** of the **true standard deviation of y** (σ_y).

$$s = \sqrt{\frac{\sum (y - \hat{y})^2}{(n-2)}} = \sqrt{\frac{\sum residuals^2}{n-2}}$$

Notice degrees of freedom (df) is _____!
We lose _____ degrees of freedom for estimating _____ parameters $(\alpha \text{ and } \beta)$.

Regression Inference
Sampling Distribution for β

- Suppose you took many samples of the same size from this population & calculated the LSRL for each, **what shape will this sampling distribution have?**

_____

- The **mean** of the sampling distribution:

$$\mu_b = \beta$$

- The **standard deviation** of the sampling distribution:

$$s_{b_1} = \frac{\sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}}{\sqrt{\sum (x_i - \bar{x}_i)^2}} = \frac{s}{\sqrt{\sum (x_i - \bar{x}_i)^2}}$$

Notes:

- b_1 = _____ _____ _____ _____.
- s_{b1} = SE(b) (the S.D. of the slope)
  - Rarely do we need to calculate.
  - Typically it will be given in computer output.
- s = the S.D. of our LSRL model

Regression Inference

Conditions for Regression Inference on Slope

1. **Random and Independent Observations**

   The observations are **Random**.
   ***Check** that you have an _____ of data.

   The observations are **Independent**
   ***Check** the 10% condition when sampling without replacement.

2. **Linear Relationship**

   The true relationship is _____.
   Check the _____
   for _____ scattering of residual points.

3. **Constant Standard Deviation**

   The standard deviation of the

   _____ variable (y) is constant.

   Check _____ , **WITH THE LSRL INCLUDED IN THE PLOT,** to see if the points are _____ spaced across LSRL (points above and below the LSRL).

4. **Responses Vary Normally**

   Since samples will typically have fewer than 30 observations (aka CLT cannot be applied), the responses (response variable, y) must vary normally about the _____ regression line.

   Check the _____ of _____ for symmetry and no outliers.

Regression Inference

Confidence Interval Formula – (for the slope of the LSRL)

Identify the…

- point estimate
- critical value
- degrees of freedom
- standard deviation of what?

  _____
- margin of error

$$b \pm t^* * s_{b_1}$$

Regression Inference

Hypothesis Test Formula

Here is the test statistic value we calculate for the slope of the least squares regression line:

Identify the…

- test statistic,
- sample statistic
- population parameter
- standard deviation of the statistic

$$t = \frac{b - \beta}{s_{b_1}}$$

Notice, $df = n - 2$

We lose 2 degrees of freedom for _____ two parameters!

Hypothesis Statements for Regression Inference (the slope of the LSRL)

H_0: $\beta = 0$

- This implies that there is no _____ between x and y; **or**
- that x should not be used to _____ y.

H_a: $\beta > 0$ **means a _____relationship.**

H_a: $\beta < 0$ **means a _____relationship.**

H_a: $\beta \neq 0$ **means _____relationship.**

You must always define the population parameter!!!!

Where, is the true _____ of the LSRL ... (in context)

Regression Inference

EXAMPLE 1"Weight vs. Body Fat"

It is difficult to accurately determine a person's body fat percentage without immersing him or her in water. Researchers hoping to find ways to make a good estimate immersed 20 randomly selected male subjects, and then measured their weights.

Find the LSRL, correlation coefficient, and coefficient of determination.

Weight (lb)	Body Fat (%)
175	6
181	21
200	15
159	6
196	22
192	31
205	32
173	21
187	25
188	30
188	10
240	20
175	22
168	9
246	38
160	10
215	27
159	12
146	10
219	28

LSRL:

$\hat{y} =$

- Must define x and y !!!!
- where

x= _____ _____ ____

y= _____ _____ ____

Recommend write LSRL in context

Corrleation Coefficent:

$r =$

Coefficient of Determination:

$r^2 =$

Rewrite LSRL:

$$\textit{body fat} = \_\_\_\_\_ + \_\_\_\_\_ \textit{(weight)}$$

- **Example 2:**

Explain the meaning of the slope in context of the problem.

Does the y-intercept of the LSRL have meaning in context of the problem?

Explain the meaning of the correlation coefficient in context of the problem.

Explain the meaning of the coefficient of determination in context of the problem.

- **Example 3:**

Estimate α, β and σ for the problem.

Create a scatterplot for the data.

Create a residual plot for the data.

Example 4: **(Significance Test)** Is there sufficient evidence that weight can be used to predict body fat?

Conditions – You must sketch the graphs to support conditions

- **Example 5:** **(Confidence Interval)** Find a 95% confidence interval for the true slope of the LSRL.

- **Example 6:** **(Computer generated data)** Here is the computer generated result from the data:

Sample size: $n = 20$ $R - square = 43.83\ \%$ $s =$

Parameter	Estimate	Std. Err.
Intercept	-27.376263	11.547428
Weight	0.24987414	0.060653996

What are the degrees of freedom? _____

What is the correlation coefficient?

- **How do you determine the sign?**

What does the value of s represent?

What is the equation of the LSRL?

What does the value 0.060653996 represent?
IMPORTANT!!!

It is difficult to accurately determine a person's body fat percentage without immersing him or her in water. Researchers hoping to find ways to make a good estimate immersed 20 randomly selected male subjects, and then measured their weights.