**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** ) has a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ relationship with .

The model is:

;

Where, the unknown population parameters are:

 is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_and

 is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

* For any fixed value of the ***x*** (the explanatory variable), the **\_\_\_\_\_\_ \_\_\_\_\_\_\_ variable** ( ) varies according to a **\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_ distribution**.

The repeated responses of are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of each other.

* The **standard deviation** of is the \_\_\_\_\_\_\_\_\_\_ for all values of ( ).

 is also an unknown \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.

**Regression Inference**

**Unbiased Estimators**

**Population Parameters**

* We use the model \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ to estimate the population least square regression line (LSRL)
* The slope of the LSRL is **an \_\_\_\_\_\_\_\_ \_\_\_\_ estimator** of the true **\_\_\_\_\_\_\_\_\_ \_\_\_\_** .
* The intercept of the LSRL is **an \_ \_\_ \_\_\_\_\_\_\_\_\_ estimator** of the true **\_\_\_\_\_\_\_\_\_ \_\_\_\_** .
* The standard error is **an \_\_\_\_\_\_\_\_\_ \_\_\_ estimator** of the **true standard deviation of y** .



Notice degrees of freedom (df) is \_\_\_\_\_\_\_\_\_\_!

We lose \_\_\_\_\_ degrees of freedom for estimating \_\_\_\_\_ parameters .

**Regression Inference**

**Sampling Distribution for**

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



**Notes:**

* **b1 = \_\_\_\_\_\_\_\_ \_ \_\_\_ .**
* **sb1  = 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 ar**e Random**.
**\*Check** that you have an \_\_\_\_\_\_\_\_\_\_\_\_\_ of data.

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

1. **Linear Relationship**

The true relationship is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
**Check** the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ scattering of residual points.

1. **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).

1. **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

**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

****

Notice,

We lose 2 degrees of freedom for \_\_\_\_\_\_\_\_\_\_\_\_\_\_
two parameters!

**Hypothesis Statements for Regression Inference**

**(the slope of the LSRL)**

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

 **means a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_relationship.**

 **means a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_relationship.**

 **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 |

* Must define x and y !!!!
* where

**x= \_\_\_\_\_\_\_\_ \_\_\_\_**

**y= \_\_\_\_\_\_\_\_ \_\_\_\_**

**Recommend write LSRL in context**

* **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:



**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.