

## Estimation and Statistics Lab

### Hypothesis Testing

If some event happens, is it random or was it caused by something else?

Small example: A production line has a typical error rate of 1%, but today a production run of 1200 items resulted in 20 defective items.

Does this result represent a problem with the line? Is it outside the probability of random chance?

Specify the level of random error: typically 5%

Create a confidence interval that shows 95% of the random events.

$XL = \text{mean} - Z * \text{std. dev.}$   $< x < XR = \text{mean} + Z * \text{std. dev}$

$N=1200$ ,  $P=0.01$ ,  $Q=0.99$

Binomial: Variance of  $X$   $V(x) = NPQ$ :  $1200 * .01 * .99 = 11.88$

Std. Dev. =  $\text{Sqrt}(\text{Var}) = 3.4467$ , mean =  $NP = 1200 * 0.01 = 12$

Normal Distribution:  $Z/0.025 = 1.96$

$XL = 12 - 1.96 * 3.4467 = 5.245$ ;  $XR = 12 + 1.96 * 3.4467 = 18.76$

Conclusion: 20 falls outside the confidence interval so it is beyond random chance.

Something is wrong with the production line.

Alternate way to write a test. Compute  $T = (X - \text{exp.}) / \text{std. dev.}$

$T = (20 - 12) / 3.4467 = 2.32$  with 1199 degrees of freedom ( $\sim$ Normal)

$Z/0.025 = 1.96$  and  $2.32 > 1.96$ ,

Reject the null hypothesis that they are the same.

### Regression

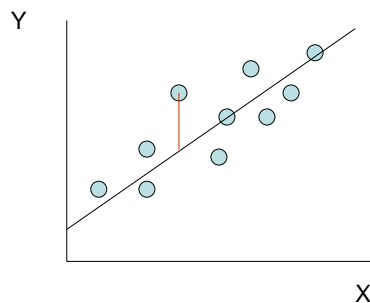
How do independent variables affect a dependent variable?

A search for relationships.

Typically best for linear relationships. Others are possible but risky.

$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots$

Estimate  $b_0$ ,  $b_1$ ,  $b_2$ ,  $b_3$ , ...



Along with their standard deviations to test for significance, different from zero.

Estimated line:  $b_0 = \text{intercept}$ ,  $b_1 = \text{slope}$ .

Found by searching for  $b_0$ , ... that minimize the squared error:

$$\Sigma(Y^{\wedge} - Y)^2$$

$$b_1 = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{\Sigma xy - n\bar{x}\bar{y}}{\Sigma x^2 - n\bar{x}^2}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

$$B = (X'X)^{-1}X'Y$$

The formulas help to understand the process, but let a computer do the work.  
Focus on interpretation:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \dots$$

$$\partial Y / \partial X_1 = b_1, \text{ etc.}$$

Interpretation: An increase in one unit of  $X_1$  results in  $b_1$  change in  $Y$

$$\text{Sales} = 1018 + 12 \text{ Advertising} - 50 \text{ Price}$$

Increasing advertising by one dollar would increase sales by 12

Decreasing price by one dollar would increase sales by 50

Excel has a regression tool.

All data must be numeric with no missing cells.

X column values must be contiguous (next to each other).

Income	Price	Advertising	SalesQty
20000	45	5000	2976
20000	48	5000	2931
20000	50	5000	2968
20000	52	5000	2951
20000	55	5000	2901
20000	60	5000	2813

### Lab Assignment

1. Run the regression for the sample data.
2. Briefly explain the statistical results.
3. For the average values of Income and Advertising, draw the demand curve.
4. Draw a second demand curve for a \$50,000 income.