Inferences Concerning the Regression Coefficients (Slopes)

Before a regression equation is used for the purpose of estimation or prediction, we should first determine if a relationship appears to exist between the two variables in the population, or whether the observed relationship in the sample could have occurred by chance.

For the sake of simplicity, we will consider Simple Linear Regression only here, although everything applies to the case of Multiple Linear Regression also.

In the absence of any relationship in the population, the slope of the population regression line would, by definition, be zero: $\boldsymbol{\beta_1 = 0}$.

Therefore the usual null and alternative hypotheses tested is

H_0: $\boldsymbol{\beta_1 = 0}$.
H_1: $\boldsymbol{\beta_1 \neq 0}$.

Such a test is equivalent to a formal test of the linear relationship between the two variables.
If we fail to reject the null hypothesis, we must conclude that the independent variable has no bearing on the value of the dependent variable.

The null hypothesis can also be formulated as a one-tail test, in which case the alternative hypothesis is not simply that the two variables are related, but that the relationship is of a specific type (direct or inverse).

A hypothesized value of the slope is tested by computing a t statistic and using n - 2 degrees of freedom. Two degrees of freedom are lost in the process of inference because two parameter estimates, $\boldsymbol{b_0}$ and $\boldsymbol{b_1}$, are included in the regression equation.

The standard formula is

$$t = \frac{b_1 - (\beta_1)_0}{s_{b_1}}$$

However, when the null hypothesis is that the slope is zero, which generally is the hypothesis, then the formula is simplified and is stated as:

$$t = \frac{b_1}{s_{b_1}}$$

Inferences Concerning the Intercept

The same approach to formal testing is equally applicable to the Intercept.

Therefore the usual null and alternative hypotheses tested is

H_0: $\boldsymbol{\beta_0 = 0}$.
H_1: $\boldsymbol{\beta_0 \neq 0}$.

We will discuss the use of such a test in future classes. It is not usually given as much attention, in general. However it is quite a useful test for chemists.

Performing Inference procedures for intercept and slope.

Recall the example from the last class: Concentration and Fluoresence.

```
> Fit

Call:
lm(formula = Fluo ~ Conc)

Coefficients:
(Intercept)         Conc
      1.518        1.930
```

To compute the p-values for inferences on the intercept and slope, we use the summary command, specifying the regression model we have chosen to use.

The p-value is written as `Pr(>|t|)`. Additionally there is a useful visual aid : the number of asterisks beside the p-value, if the p-value is sufficiently small.

A guide to reading the significance codes is provided in the output.

- Three asterisks indicate a p-value of less than 0.001
- Two asterisks indicate a p-value of less than 0.01
- One asterisk indicate a p-value of less than 0.05

(Remark: we have chosen 0.01 as a threshold for rejecting the null hypothesis. This is an arbitrary level)

```
> summary(Fit)

Call:
lm(formula = Fluo ~ Conc)

Residuals:
       1        2        3        4        5        6        7
 0.58214 -0.37857 -0.23929 -0.50000  0.33929  0.17857  0.01786

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.5179     0.2949   5.146  0.00363 **
Conc          1.9304     0.0409  47.197 8.07e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4328 on 5 degrees of freedom
Multiple R-squared: 0.9978,     Adjusted R-squared: 0.9973
F-statistic:  2228 on 1 and 5 DF,  p-value: 8.066e-08
```

We reject the null hypotheses for both the slope and intercept.

```
> cor(Cheese)

           Taste    Acetic       H2S    Lactic
Taste  1.0000000 0.5495393 0.7557523 0.7042362
Acetic 0.5495393 1.0000000 0.6179559 0.6037826
H2S    0.7557523 0.6179559 1.0000000 0.6448123
Lactic 0.7042362 0.6037826 0.6448123 1.0000000
```

Simple Linear Regression Models

```
> FitA = lm(Taste ~ Acetic, data = Cheese)
> FitB = lm(Taste ~ H2S, data = Cheese)
> FitC = lm(Taste ~ Lactic, data = Cheese)
```

```
> summary(FitA)

Call:
lm(formula = Taste ~ Acetic, data = Cheese)

Residuals:
    Min      1Q  Median      3Q     Max 
-29.642  -7.443   2.082   6.597  26.581 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)  -61.499     24.846  -2.475  0.01964 * 
Acetic        15.648      4.496   3.481  0.00166 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.82 on 28 degrees of freedom
Multiple R-squared: 0.302,	Adjusted R-squared: 0.2771 
F-statistic: 12.11 on 1 and 28 DF,  p-value: 0.001658
```

```
> summary(FitB)

Call:
lm(formula = Taste ~ H2S, data = Cheese)

Residuals:
    Min      1Q  Median      3Q     Max 
-15.426  -7.611  -3.491   6.420  25.687 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -9.7868     5.9579  -1.643    0.112    
H2S           5.7761     0.9458   6.107 1.37e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.83 on 28 degrees of freedom
Multiple R-squared: 0.5712,     Adjusted R-squared: 0.5558 
F-statistic: 37.29 on 1 and 28 DF,  p-value: 1.374e-06
```

```
> summary(FitC)

Call:
lm(formula = Taste ~ Lactic, data = Cheese)

Residuals:
     Min       1Q   Median       3Q      Max 
-19.9439  -8.6839  -0.1095   8.9998  27.4245 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -29.859     10.582  -2.822  0.00869 ** 
Lactic        37.720      7.186   5.249 1.41e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.75 on 28 degrees of freedom
Multiple R-squared: 0.4959,     Adjusted R-squared: 0.4779 
F-statistic: 27.55 on 1 and 28 DF,  p-value: 1.405e-05
```

Regression Models using two or more independent variables.

```
> Fit1 = lm(Taste ~ Acetic + H2S, data = Cheese)
> Fit2 = lm(Taste ~ Acetic + Lactic, data = Cheese)
> Fit3 = lm(Taste ~ H2S + Lactic, data = Cheese)
> Fit4 = lm(Taste ~ Acetic + H2S + Lactic, data = Cheese)
```

Akaike Information Criterion

```
> AIC(FitA)
[1] 246.6389
> AIC(FitB)
[1] 232.0245
> AIC(FitC)
[1] 236.8724
```

For the multiple linear regression models.

```
> AIC(Fit1)
[1] 233.2438
> AIC(Fit2)
[1] 237.3884
> AIC(Fit3)
[1] 227.7838
> AIC(Fit4)
[1] 229.7775
```

Summary of model selection metrics.

Model	**Ind. Variables**	**Multiple R^2**	**Adjusted R^2**	**AIC**
		(highest *)	(highest *)	(lowest *)
FitA	Acetic	0.3020	0.2771	246.6389
FitB	H2S	0.5712	0.5558	232.0245
FitC	Lactic	0.4959	0.4779	236.8724
Fit1	Acetic, H2S	0.5822	0.5512	233.2438
Fit2	Acetic, Lactic	0.5203	0.4847	237.3884
Fit3	H2S, Lactic	0.6517	0.6259 *	227.7838 *
Fit4	All Three	0.6518 *	0.6116	229.7775