1 Inference for Regression Coefficients

For each regression coefficient in a fitted model, we will perform the following hypothesis test on the corresponding parameter values β_i

H_0 $\beta_i = 0$
H_1 $\beta_i \neq 0$

- If the regression coefficient is significant (indicated by a series of asterisks in the `R` summary), that means that corresponding predictor variable is useful in describing the response variable. If it is not significant - you can try a different model or a different combination of predictor variables (outside the scope of this course).
- Even if an intercept regression coefficient is not significant, it is usually retained in the model (This will happen in the next example). There is a procedure called *Regression Through the Origin*, where the intercept term is removed from the model. It is generally not advised to do this. Amongst other reasons, it is useful in exposing any flaws in the experimental procedure.

1.1 Example

Consider the following experiment.

Concentration	0	5	10	15	20	25	30
Absorbance	0.003	0.127	0.251	0.390	0.498	0.625	0.763

```
# DO A FULL LINEAR REGRESSION ANALYSIS ON THE DATA

>concentration=c(0,5,10,15,20,25,30)
>absorbance=c(0.03,0.127,0.251,0.390,0.498,0.625,0.763)
>regr=lm(absorbance~concentration)
# READ AS; ABSORBANCE DEPENDENT ON CONCENTRATION
>summary(regr)
```

This output from this code is as follows:

```
Call:
lm(formula = absorbance ~ concentration)

...........

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0146429  0.0079787   1.835    0.126
concentration 0.0245857  0.0004426  55.551 3.58e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.01171 on 5 degrees of freedom
Multiple R-squared: 0.9984,     Adjusted R-squared: 0.9981
F-statistic:  3086 on 1 and 5 DF,  p-value: 3.576e-08
```

- Estimation of Slope = 0.0251643
- Estimation of Intercept = 0.0021071
- The regression equation is therefore (to 4 decimal places)

$$\hat{Y} = 0.0021 + 0.0251X$$

  where Y is absorbance and X is concentration.

- The p-value for intercept is 0.126 - implying that it is not significantly different from zero. (**i.e. Not Significant**)

  H_0 $\beta_0 = 0$ - FAIL TO REJECT

  H_1 $\beta_0 \neq 0$

- The p-value for the regression coefficient for the concentratin variable is much less than 5% - implying that it is significantly different from zero (**i.e. Highly Significant**)

  H_0 $\beta_1 = 0$ - REJECT NULL

  H_1 $\beta_1 \neq 0$

2 Confidence Intervals for Regression Coefficients

Confidence Intervals for Regression Coefficients

In the last class we looked how R can be used to determine the estimates and standard errors for the slope and intercept.

The following formulae can be used to compute the confidence intervals for both, for a specified significance level.

for significance level α, the confidence intervals are

$$(1-\alpha) \times 100\% \text{ CI}\left[\widehat{\beta}_0\right] = \widehat{\beta}_0 \pm t_{n-2,1-\alpha/2}\text{SE}\left[\widehat{\beta}_0\right] \, ,$$
$$(1-\alpha) \times 100\% \text{ CI}\left[\widehat{\beta}_1\right] = \widehat{\beta}_1 \pm t_{n-2,1-\alpha/2}\text{SE}\left[\widehat{\beta}_1\right]$$

These calculations provided the basis for end of semester examination questions in previous years, but that will not be the case for this year. To compute the confidence intervals for both estimates, we use the `confint()` command, specifying the name of the fitted model.

- You can be 95% confident that the real, underlying value of the coefficient that you are estimating falls somewhere in that 95% confidence interval.
- We can use confidence intervals as an alternative method of determining the outcome of hypothesis tests.

  $H_0 \;\; \beta_i = 0$

  $H_1 \;\; \beta_i \neq 0$

- If the interval does not contain 0, we reject the null hypothesis. (this will coincide with the $p-$value being 0.05 or less.)
- If the interval does contain 0, we fail to reject the null hypothesis. (this will coincide with the $p-$value greater than 0.05)

2.1 Example 1

Using the data from the previous example.

```
> confint(regr)
                      2.5 %     97.5 %
(Intercept)   -0.005867133 0.03515285
concentration  0.023448025 0.02572340
```

- The 95% confidence interval for the intercept estimate is $(-0.0058, 0.0351)$.
- This confidence interval contains zero.

  H_0: $\beta_0 = 0$ - FAIL TO REJECT NULL

  H_1: $\beta_0 \neq 0$

- The 95% confidence interval for the intercept estimate is $(0.0234, 0.0257)$.
- This confidence interval does not contain zero.

  H_0: $\beta_1 = 0$ - REJECT NULL

  H_1: $\beta_1 \neq 0$

2.2 Example 2

Recall the example used in the previous classes:

```
> Conc=c(0,2,4,6,8,10,12)
> Fluo=c(2.1,5.0,9.0,12.6,17.3,21.0,24.7)
>
> coef(Fit)
(Intercept)        Conc
   1.517857    1.930357

>
> Fit = lm(Fluo ~ Conc)
> confint(Fit)
              2.5 %    97.5 %
(Intercept) 0.75970 2.276014
Conc        1.82522 2.035495
```

- The intercept regression coefficient is 1.5178. The corresponding 95% confidence interval is $(0.7597, 2.2760)$.
- That confidence interval does not contain zero.

  H_0: $\beta_0 = 0$ - REJECT NULL

  H_1: $\beta_0 \neq 0$
- The regression coefficient for concentration is 1.9303. The corresponding 95% confidence interval is $(1.8252, 2.0354)$.
- Again that confidence interval does not containt zero.

  H_0: $\beta_1 = 0$ - REJECT NULL

  H_1: $\beta_1 \neq 0$