Math 327, Simple Linear Regression, Example R output

This document illustrates what values, as notated in the textbook, are produced by the R functions, summary(), anova(), and confint().

```
> car.fit = lm (travel.time ~ dist.Decorah, data=car.data)
> summary (car.fit)

Call:
lm(formula = travel.time ~ dist.Decorah, data = car.data)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.37660 -0.19209 -0.08742  0.23370  0.34867 
```

Summary statistics of the sample residuals.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.3765987  0.1406053   2.678   0.0253 *  
dist.Decorah 0.0154366  0.0006381  24.190 1.69e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In the table above, we find these values:

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	$\hat{\beta}_0$	$se(\hat{\beta}_0)$	$t = \hat{\beta}_0/se(\hat{\beta}_0)$	$P(T > \|t\| \mid \beta_0 = 0)$
dist.Decorah	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$t = \hat{\beta}_1/se(\hat{\beta}_1)$	$P(T > \|t\| \mid \beta_1 = 0)$

The “Signif. codes” mean that p-values between 0 and 0.001 are marked with ‘***’; p-values between 0.001 and 0.01 are marked with ‘**’; p-values between 0.01 and 0.05 are marked with ‘*’; p-values between 0.05 and 0.1 are marked with ‘.’; and p-values greater than 0.1 are not marked.

```
Residual standard error: 0.2659 on 9 degrees of freedom
Multiple R-squared:  0.9849,	Adjusted R-squared:  0.9832 
F-statistic: 585.2 on 1 and 9 DF,  p-value: 1.687e-09
```

In the three lines above, Residual standard error = $\hat{\sigma} = MS_{res} = SS_{res}/(n-2)$;
degrees of freedom = $n - 2$;
Multiple R-squared = $R^2 = \hat{\beta}_1 SS_{xy}/SS_{xx} = 1 - \frac{SS_{res}}{SS_T}$;
Adjusted R-squared = $R^2_{Adj} = 1 - \frac{SS_{res}/(n-p)}{SS_T/(n-1))}$
F-statistic = MS_{reg}/MS_{res}
p-value = $P(F > F_{stat} \mid \beta_1 = 0)$

```
> anova (car.fit)
Analysis of Variance Table

Response: travel.time
             Df Sum Sq Mean Sq F value    Pr(>F)
dist.Decorah  1 41.362  41.362  585.15 1.687e-09 ***
Residuals     9  0.636   0.071
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the ANOVA table above, we find these values:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dist.Decorah	1	$\hat{\beta}_1 SS_{xy}/SS_{xx}$	$SS_{reg}/1$	MS_{reg}/MS_{res}	$P(F > MS_{reg}/MS_{res} \mid \beta_i = 0)$
Residuals	$n-2$	SS_{res}	$SS_{res}/(n-2)$		

The row with totals is not listed in the anova() function output. Only the "Df" and "Sum Sq" totals are relevant. Total Df $= n-1$, and Total Sum of Squares is $SS_T = \sum_{i=1}^{n}(y_i - \bar{y})^2$.

The confint() function produces confidence intervals for the parameter estimates. The round(..., 3) function rounds the values to 3 decimal places.

```
> round (confint (car.fit), 3)
             2.5 % 97.5 %
(Intercept)  0.059  0.695
dist.Decorah 0.014  0.017
```

The table above has these values:

Intercept: $\hat{\beta}_0 - t_{1-\alpha/2} se(\hat{\beta}_0),\ \ \hat{\beta}_0 + t_{1-\alpha/2} se(\hat{\beta}_0)$

Slope: $\hat{\beta}_1 - t_{1-\alpha/2} se(\hat{\beta}_1),\ \ \hat{\beta}_1 + t_{1-\alpha/2} se(\hat{\beta}_1)$

Interpretation templates:

- Slope:
  - For every one unit increase in the predictor variable, the mean response changes by slope units, p = p-value
  - Or: The mean response {increases | decreases} |slope| y-units per x-unit
  - Or: The mean response changes between LCL and UCL y-units per x-unit with 95% confidence
- Intercept:
  - The mean response when X is zero is between LCL and UCL y-units with 95% confidence
- Predicted responses:
  - Mean: The estimated mean response when X is X-value is between LCL and UCL y-units with 95% confidence
  - Individual: The predicted response for an individual when X is X-value is between LCL and UCL y-units with 95% confidence