Math 327 – Formulas for exam 1

Equation or Expression	R command(s)
$t^* = t\left(1-\frac{\alpha}{2}; n-2\right)$	`qt(0.975,14)` if $\alpha = 0.05$ and $n = 16$.
$F(1-\alpha; df_1, df_2)$	`qf(0.95,1,14)` if $\alpha = 0.05$, $df_1 = 1$ and $df_2 = 14$.
$P\{\lvert t(df)\rvert > t^*\}$	`2*(1-pt(4.58,14))` if $t^* = 4.58$ and $df = 14$
$P\{F > F^*\}$	`1-pf(16.3,1,14)` if $F^* = 16.3$ with 1 and 14 degrees of freedom
$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$	N/A (theoretical model)
$Y_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j X_{ji} + \varepsilon_i$	N/A (theoretical model)
$\varepsilon_i \sim N(\mu, \sigma^2)$, independent	N/A (theoretical model)
$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i$	`myfit=lm(Y~X)`
$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$	`predict (myfit)`
$SS_{res} = \sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2$	`sum(myfit$residuals^2)`
$MS_{res} = \frac{SS_{res}}{n-p}$ $p =$ # parameters, incl. β_0	`mysumm=summary(myfit)` `mysumm$sigma^2`
$\hat{\beta}_1 \pm t^* se\{\hat{\beta}_1\}$	`confint(myfit)`
$\hat{\beta}_0 \pm t^* se\{\hat{\beta}_0\}$	`confint(myfit)`
$t = \frac{\hat{\beta}_1}{se\{b_1\}}$	`mysumm$coefficients`
$\hat{Y}_h \pm t^* se\{\hat{Y}_h\}$ $se\{\hat{Y}_h\}$ =std. error of mean $\hat{Y}_h$	`predict(fit, data.frame(X=`X_h`), interval="confidence")`
$\hat{Y}_h \pm t\left(1-\frac{\alpha}{2}; n-p\right) se\{\hat{Y}_{h,p}\}$ $se\{\hat{Y}_h\}$ =std. error of individual $\hat{Y}_h$	`predict(fit, data.frame(X=`X_h`), interval="prediction")`
$SS_{Total} = SS_{Model} + SSE$	`anova(fit)` `sum(anova(fit)[,2])`
$R^2 = \frac{SS_{Model}}{SS_{Total}} = 1 - \frac{SSE}{SS_{Toysl}}$	`summary(myfit)$r.squared`
$MS_{Model} = SS_{Model} / (p-1)$ $MSE = SSE / (n-p)$	``anova(fit)$`Sum Sq`/anova(fit)$Df # x=regression or residual``
$F = MS_{Model} / MSE$	``anova(fit)$`F value` `` or `anova(fit)[1,3]/anova(fit)[2,3]`
$r = \pm\sqrt{R^2}$	`cor(X,Y) or cor(Yhat,Y)`

Symbols from Chapters 1-2:

- $X =$ predictor (explanatory) variable
- $Y =$ response variable
- $\mu_Y =$ population mean of the response variable at a particular value of X
  - $Y = f(X) + \epsilon = \mu_Y + \epsilon$
  - $\mu_Y = f(X) = \beta_0 + \beta_1 X$
- $\epsilon =$ population residual = deviation from the mean = observed response – fitted response
- $\beta_0 =$ population y-intercept = expected response when $X = 0$
- $\beta_1 =$ population slope = expected change in Y for a one-unit change in X
- $\hat{y}$ or $\hat{Y} =$ fitted or predicted response value for a particular value of X
- $\hat{\beta}_0 =$ fitted y-intercept = estimated response when $X = 0$
- $\hat{\beta}_1 =$ fitted slope = estimated change in Y for a one-unit change in X
- $residual = y - \hat{y}$
- $\sigma_\epsilon =$ population residual standard deviation
- $\hat{\sigma}_\epsilon =$ estimated residual standard deviation
- $N(0, \sigma_\epsilon) =$ Normal distribution with mean zero and (population) standard deviation, σ_ϵ
- $SSE = \sqrt{\sum_i (y_i - \hat{y}_i)^2}$ = Sum of Squares for Error = Error Sum of Squares = Residual Sum of Squares
- $SE_{\hat{\beta}_1} =$ Standard error of the estimated slope
- $t^* =$ critical value (or quantile) from a t_{n-2} distribution, i.e, a t distribution with $n - 2$ degrees of freedom
- $SSModel = \sum_i (\hat{y}_i - \bar{y})^2 =$ Model Sum of Squares
- $MSModel =$ Model Mean Square (divide by 1 for simple regression
- $SSTotal \ \ = \sum_i (y_i - \bar{y})^2 =$ Total (corrected) Sum of Squares
- $MSE = \frac{SSE}{n-2} = \frac{\sum_i (y_i - \hat{y}_i)^2}{n-2} =$ Mean Squared Error
- $F = \frac{MSModel}{MSE} =$ F-statistic for conducting the F-test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$
- $r =$ linear correlation
- $r^2 = R^2 =$ Coefficient of Determination
- $S_Y =$ Sample standard deviation of the response variable
- $S_X =$ Sample standard deviation of the predictor variable
- $x^* =$ a particular value of X used when predicting a value of Y at x^*
- $SE_{\hat{\mu}} =$ Standard error of an estimated mean response value at $X = x^*$
- $SE_{\hat{y}} =$ Standard error of an estimated individual response value at $X = x^*$