UNIVERSITY OF TORONTO
Faculty of Arts and Science

June 2016 EXAMINATIONS

STA302/1001H1F

Duration - 3 hours

Examination Aids: Scientific Calculator

**STA 302/1001**
**Summer 2016**
**Final Exam**
**6/21/2016**
**Time Limit: 3 hours**

Last Name (Print): _____

First Name: _____

Student Number: _____

# Check one: STA302 ☐ STA1001 ☐

This exam contains 19 pages (including this cover page) and 5 problems. Check to see if any pages are missing. Enter all requested information on the top of this page.

- You may *not* use your books or notes on this exam. You may use a scientific calculator and the formulae and tables at the end of the exam. Round the df down to the nearest table entry, if necessary.
- MLR stands for 'Multiple Linear Regression'; MLE for Maximum Likelihood Estimate.
- You are required to show your work on each problem on this exam. Please carry all possible precision through a numerical question, and give your final answer to four (4) decimals, unless they are trailing zeroes or otherwise indicated.
- You may use a benchmark of $\alpha = 5\%$ for all inference, unless otherwise indicated.
- Do not write in the table to the right.

| Problem | Points | Score |
|---|---|---|
| 1 | 20 | |
| 2 | 10 | |
| 3 | 20 | |
| 4 | 30 | |
| 5 | 20 | |
| Total: | 100 | |

1. (20 points) **Multiple Choice** Answer the following questions by circling the *best* answer.

I. Which of the following is a good reason to look at a sequence plot of residuals?

A. To assess whether there is a relationship between the response and predictor(s)
B. To identify influential points
**C. To identify serial correlation**
D. To identify high leverage points

II. Under which marginal distribution of X would residuals have constant variance?

**A. Two equal groups**
B. Three equal groups, equally spaced
C. Uniform distribution
D. Normal distribution

III. Which of the following statements about variable transformations is false?

A. Transformations on X can improve linearity
B. Transformations on Y can improve linearity
**C. Transformations on X can fix variance problems**
D. Transformations on Y can fix variance problems

IV. If you are setting up an experiment whose main purpose is to determine the direction of X's effect on Y, which is the most efficient choice of levels for X?

A. X should have mean $X_h$
B. X should have mean 0
C. X should be uniformly distributed across the domain
**D. X should be composed of two equal-sized groups at the endpoints of the domain**

V. The number of days that a fruit fly lives is best modeled as:

A. A random variable measured with error
**B. A random variable measured without error**
C. A fixed constant measured with error
D. A fixed constant measured without error

VI. A beam is loaded and the time until failure measured. The amount of stress applied to the beam is best modelled as:

A. A random variable measured with error
B. A random variable measured without error
**C. A fixed constant measured with error**
D. A fixed constant measured without error

VII. Which of the following statements about covariance matrices is true?

**A. They are always symmetric**

B. They are always diagonal

C. They are always a constant times the identity matrix

D. They are always multivariate Normal

VIII. Which of the following statements is false when multi-collinearity is present?

A. Parameter estimates **b** have high variance

B. Parameter estimates **b** are highly correlated

**C. Parameter estimates b are no longer unbiased**

D. Parameter estimates **b** follow a Normal distribution

Suppose you have an R data frame `df` with columns [`y, x1, x2, f`]. The variables, `y, x1 and x2` are numerical, and `f` is a factor with two levels (`f1` and `f2`).

IX. Which line of R code properly fits an additive (no interaction) MLR model?

A. `lm(y ~ x1 * x2, data = df)`

B. `lm(y ~ x1:x2, data = df)`

C. `anova(y ~ x1 * x2, data = df)`

**D.** `lm(y ~ x1 + x2, data = df)`

X. From which line could you test whether or not the two factor levels have equal slopes?

A. `lm(y ~ f + x1, data = df)`

**B.** `lm(y ~ f * x1, data = df)`

C. `lm(y ~ f1 * x1 + f2 * x2, data = df)`

D. `lm(y ~ f1 + f2 + x1, data = df)`

2. (10 points) **True or False** Answer the following questions by writing 'T' or 'F' in the blank. Do not write something ambiguous like ∓ or ☋!

_**F**_ Forward and backward selection give the same model if the entry and exit probabilities are equal

_**T**_ Partial F-tests only work when comparing nested models

_**F**_ Choosing between two arbitrary models based on $R^2$ always leads to choosing the bigger model

_**T**_ Adjusted $R^2$ cannot be greater than $R^2$

random error accounts for it

_**F**_ Measurement errors in the Y variable cause a problem for our MLR model.

_**F**_ When trying to determine linearity between X and Y, the best choice for the number of levels of our predictor variable is two

_**T**_ A high Variance Inflation Factor indicates a potential problem with multi-collinearity.

_**T**_ Type I and Type III sums of squares are the same when predictors are independent.

_**T**_ An ANOVA model is a regression model with only indicator variables as predictors

_**F**_ All regression models can be written as General Linear Models.

nonlinear models

3. Universities collect data on performance in school through a student's Grade Point Average (GPA), and can link that score back to their entry scores on the Scholastic Aptitude Test (SAT), in the US at least. Suppose we want to predict the university GPA using the Math and Verbal scores on the SAT. Several predictor models were fit to these data, and selected R code / output is shown on the next few pages. In case you were wondering, the function `solve()` in R finds an inverse, and `t()` the transpose.

   I. (6 points) Some values have been replaced with letters. Fill in those values. You do not need to show any work for this part.

   (A) (C) (E)

   (B) (D) (F)

**Solution:** (A) -3.3128e-03 (C) 0.1211 (E) 0.4414
(B) 0.5482 (D) 0.1077 (F) 1.374

```
> apply(sat, 2, mean)
VERBAL   MATH    GPA
595.65 649.53   2.63

### Model A ###
###############
> anova(modelA)
Analysis of Variance Table

Response: GPA
             Df Sum Sq Mean Sq F value    Pr(>F)
MATH          1  2.529  2.5287  8.4145  0.004148
VERBAL        1  5.249  5.2492 17.4673 4.401e-05
MATH:VERBAL   1  0.341  0.3407  1.1336  0.288318
Residuals   196 58.901  0.3005

> summary(modelA)
Call:
lm(formula = GPA ~ MATH * VERBAL, data = sat)
```

2.63 = -0.00405 * 649.53 + A * 595.65 + (8.516/1000000) * 649.53 * 595.65
A = 0.0033
regression line crosses the mean so…

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.926e+00  3.149e+00   1.247    0.214
MATH        -4.046e-03  4.778e-03  -0.847    0.398
VERBAL             [A]  5.309e-03  -0.624    0.533
MATH:VERBAL  8.516e-06  7.999e-06   1.065    0.288
```

sqrt(MSE) = sqrt(0.3005)

```
Residual standard error:  [B] on 196 degrees of freedom
Multiple R-squared:    [C],  Adjusted R-squared:    [D]
F-statistic: 9.005 on 3 and 196 DF,  p-value: 1.291e-05
```

R^2_adj = 1 - ((58.901 / 196) / ((58.901 + 0.341 + 5.249 + 2.529) / 199)) = 0.108

R^2 = SSreg/SST = 1 - RSS/SST = 1 - 58.901/(58.901 + 0.341 + 5.249 + 2.529) = 0.1211

```
### Model B ###
###############

> anova(modelB)
Analysis of Variance Table

Response: GPA
          Df Sum Sq Mean Sq F value    Pr(>F)
MATH       1  2.529  2.5287  8.4088  0.004158
VERBAL     1  5.249  5.2492 17.4555 4.418e-05
Residuals 197 59.242  0.3007


> summary(modelB)
Call:
lm(formula = GPA ~ MATH + VERBAL, data = sat)

                          E = sqrt(0.1948) = 0.44
Coefficients:             F = 0.606-0 / E = 1.37
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.6062975        [E]     [F]    0.171
MATH        0.0009999  0.0006093   1.641    0.102
VERBAL      0.0023072  0.0005522   4.178 4.42e-05

Residual standard error: 0.5484 on 197 degrees of freedom
Multiple R-squared:  0.1161,Adjusted R-squared:  0.1071
F-statistic: 12.93 on 2 and 197 DF,  p-value: 5.284e-06


> X <- model.matrix(modelB)
> MSE <- anova(modelB)$Mean[3]  for computing variance and covariance matrix for estimated betas
> MSE*solve(t(X) %*% X)


              (Intercept)          MATH        VERBAL
(Intercept)  0.1948393932 -1.861246e-04 -1.216189e-04
MATH        -0.0001861246  3.712995e-07 -9.241260e-08
VERBAL      -0.0001216189 -9.241260e-08  3.049503e-07

### Model C ###
###############

> anova(modelC)
Analysis of Variance Table

Response: GPA
          Df Sum Sq Mean Sq F value    Pr(>F)
VERBAL     1  6.968  6.9682  22.975 3.216e-06
Residuals 198 60.052  0.3033
```

II. (6 points) Using modelB, find the expected GPA for students who get 700 on Math and 650 on Verbal. You may assume these data are in the range of our dataset. Find a 95% CI for this estimate as well.

**Solution:**
$x_h' = [1 \quad 700 \quad 650]$ (1)

$\widehat{Y_h} = x_h' b = [1 \quad 700 \quad 650][0.6062975 \quad 0.0009999 \quad 0.0023072]' = 2.8059$ (1)

$s^2\{\hat{Y}_h\} = MSE \cdot \mathbf{x_h'}(\mathbf{X'X})^{-1}\mathbf{x_h}$ (1) $= 0.0028433$ (1)

didnt really learnt this...

$s\{\hat{Y}_h\} = 0.05332$ (1)

95% CI for $Y_h$ : $\widehat{Y_h} \pm t_{0.975,150} \cdot s\{\hat{Y}_h\} = 2.8059 \pm 0.105574 = (2.7003, 2.9115)$ (1)

III. (2 points) Compute the extra sum of squares for a model with Math, Verbal and the interaction, compared to a model with just Verbal.

**Solution:** $SSR(Math, Math * Verbal|Verbal) = SSE_{Reduced} - SSE_{Full}$ (1)
$= 60.052 - 58.901 = 1.151$ (1)

RSS_{verbal} - RSS_{math*verbal} = 1.151

IV. (3 points) Perform a partial F-test to test the hypothesis that Math and Math*Verbal are useful predictors above a model that already contains the Verbal score. If you cannot perform this test, state what you are missing in order to do it. If you can, give a test statistic (with df), the most accurate p-value you can, and a conclusion in words.

**Solution:** $F_{obs} = \frac{1.151/2}{0.3005} = 1.9151(2, 196)$ (1)

F^* = (1.151 / 2) / (58.901 / 196) = 1.915

$F_{crit} = 3.09$ (0.5)

F_crit is approximated

F_{2, 196, 0.95 } < 3.09 note 1 tailed test

cant reject, no evidence extra predictors are useful

$\therefore$ Do not reject $H_0$ (0.5): no evidence that extra predictors are useful; (0.5) p > 0.05 (0.5)

V. (2 points) Does knowing a student's Verbal score significantly improve your prediction of their GPA, if you already know their Math score and you assume they **don't** interact? If you cannot perform this test, state what you are missing in order to do it. If you can, give a test statistic and the most accurate p-value you can.

for modelB, where there is no interaction term

we sae t-value = 4.178 for regression coefficients for VERBAL, p-value=4.42x10^-5

**Solution:** Yes (1) $t = 4.178, p = 4.418 \cdot 10^{-5}$ (1)

VI. (1 point) Does knowing a student's Verbal score significantly improve your prediction of their GPA, if you already know their Math score and you assume they **do** interact? If you cannot perform this test, state what you are missing in order to do it. If you can, give a test statistic and the most accurate p-value you can.

**Solution:** Cannot answer. We would need a model with just Math as a predictor to answer this. (1)

cannot answer, need a model with just math as a predictor then use partial f test to see if it is indeed useful

4. Consider the NFL dataset from your three assignments. We are going to try to predict a player's Bench Press score including categorical variables as predictors this time. The R output for a few fitted models is shown below; the data have been pre-formatted as in the assignments. Recall that we created a variable `PosGroup` with three levels: `Big Backs, Linemen` and `Small Backs.`

   I. (6 points) Some values have been replaced with letters. Fill in those values. You do not need to show any work for this part.

   (A) (C) (E)

   (B) (D) (F)

| **Solution:** | (A) 2 | (C) 36.5139 | (E) 23.775 |
|---|---|---|---|
| | (B) 890.35 | (D) 24.3838 | (F) 0.386 |

```
### Model A ###
################


> anova(fitA)
Analysis of Variance Table

Response: Bench
           Df Sum Sq Mean Sq F value    Pr(>F)
PosGroup  [A] 1780.7     [B]     [C] 6.511e-13
Residuals 111 2706.3     [D]
```

A = df_ssreg = p = 2  B = MSreg = ssreg/df_ssreg = 1780.7 /2 = 890.35

C = F-value = msreg / mse = B/D = 890.35/24.38 = 36.5139

df_rss = n - 3 = 111  n = 114

D = mse = rss/df_rss = 2706.3 / 111 = 24.38

```
> summary(fitA)
Call:
lm(formula = Bench ~ PosGroup, data = nfl)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          21.4333     0.9015     [E]  < 2e-16
PosGroupLinemen       3.5474     1.1321   3.134  0.00221
PosGroupSmall Backs  -5.9333     1.2548  -4.728  6.7e-06

Residual standard error: 4.938 on 111 degrees of freedom
Multiple R-squared:  0.3968,Adjusted R-squared:  [F]
F-statistic: [---] on [---] and 111 DF,  p-value: 6.511e-13
```

3 category, use 3-1=2 categories in model

E = 21.433/0.9015 = 23.775

F = R^2_adj = 1 - mse/mst = 1 - (2706.3 / 111) / ((1780.7+2706.3) / 113)=

```
### Model B ###
###############

> anova(fitB)

Analysis of Variance Table

Response: Bench
           Df  Sum Sq Mean Sq F value   Pr(>F)
Wt          1 1857.11 1857.11 82.4526 4.99e-15
PosGroup    2  152.33   76.17  3.3817  0.03756
Residuals 110 2477.57   22.52


> summary(fitB)

Call:
lm(formula = Bench ~ Wt + PosGroup, data = nfl)

Coefficients:
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)          6.27561    4.83433   1.298  0.19696
Wt                   0.06456    0.02026   3.187  0.00187
PosGroupLinemen     -0.41828    1.65296  -0.253  0.80070
PosGroupSmall Backs -3.61416    1.40860  -2.566  0.01164

Residual standard error: 4.746 on 110 degrees of freedom
Multiple R-squared:  0.4478,Adjusted R-squared:  0.4328
F-statistic: 29.74 on 3 and 110 DF,  p-value: 3.712e-14


### Model C ###
###############

> anova(fitC)

Analysis of Variance Table

Response: Bench
           Df  Sum Sq Mean Sq F value    Pr(>F)
Wt          1 1857.11 1857.11 84.0964 3.842e-15
PosGroup    2  152.33   76.17  3.4491   0.03535
Cone3       1   14.96   14.96  0.6772   0.41237
Overall     1   28.55   28.55  1.2929   0.25806
Shuttle     1   71.17   71.17  3.2229   0.07544
Residuals 107 2362.90   22.08
```

```
> summary(fitC)
Call:
lm(formula = Bench ~ Wt + PosGroup + Cone3 + Overall + Shuttle,
data = nfl)

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         23.538279  13.187847   1.785  0.07712
Wt                   0.079701   0.026728   2.982  0.00355
PosGroupLinemen      0.010462   1.653816   0.006  0.99496
PosGroupSmall Backs -3.649037   1.422982  -2.564  0.01172
Cone3                1.708552   2.783550   0.614  0.54065
Overall             -0.006574   0.006403  -1.027  0.30686
Shuttle             -7.546112   4.203413  -1.795  0.07544

Residual standard error: 4.699 on 107 degrees of freedom
Multiple R-squared:  0.4734,Adjusted R-squared:  0.4439
F-statistic: 16.03 on 6 and 107 DF,  p-value: 4.44e-13
```

not covered

II. (4 points) Give a set of two Bonferroni-corrected 98% joint CIs for the true regression slopes $\beta_1$ and $\beta_2$, from modelC (the slopes of `Wt` and `PosGroupLinemen`). The familywise error rate here is 2%.

**Solution:**

$t_{100,0.995} = 2.63$ (1)

$\beta_1 : b_1 \pm 2.63 \cdot s\{b_1\}$ (1) $= 0.079701 \pm 2.63(0.026728) = 0.0797 \pm 0.0703$ (1)

$= (0.0094, 0.1500)$

$\beta_2 : b_2 \pm 2.63 \cdot s\{b_2\} = 0.010462 \pm 2.63(1.653816) = 0.0105 \pm 4.3495$ (1)

$= (-4.34, 4.36)$

III. (3 points) Using modelC, give a 99% CI for $\beta_1$ protected with the Working-Hotelling procedure.

**Solution:**
$F_{crit} = F_{0.99,7,100} = 2.82$ (1)
$W = \sqrt{7 \cdot F_{crit}} = 4.443$ (1)
$\beta_1 : b_1 \pm W \cdot s\{b_1\} = 0.079701 \pm 4.443(0.026728) = 0.079701 \pm 0.11875$ (1)
$= (-0.0391, 0.1985)$

IV. (1 point) What is the average Bench Press score for Big Backs?

**Solution:**
21.4333

when lineman, small backs indicator is zero
equivalent to the value of estimated intercept for
Bench ~ PosGroup

V. (1 point) What is the average Bench Press score for Linemen?.

**Solution:**
24.9807

Bench = 21.433 + 3.5474 lineman - 5.9333 small_back
set lineman to 1 and small_back to 0
Gench = 24.9807

VI. (2 points) Can you test if Big Backs and Linemen are statistically different with respect to their bench press scores, ignoring other predictors? If so, give the test statistic, df, p-value and a conclusion in words. If not, state what you are missing to do the test.

**Solution:**
t value is for posGroupLineman in modelAs
$t_{obs} = 3.134; DF = 111; p = 0.00221$ (1)
$\therefore$ Reject $H_0$ (0.5): Linemen put up more reps on avg. compared to big backs (0.5)

why though .. have to control small back …

VII. (3 points) Maybe linemen are better simply because they are bigger, and not because of their position. Can you test if Big Backs and Linemen are statistically different with respect to their bench press scores, controlling for weight? If so, give the test statistic, df, p-value and a conclusion in words. If not, state what you are missing to do the test.

**Solution:**

add weight Wt to the model, see t value not significant ... for predictor PosGroupLineman

$t_{obs} = -0.253; DF = 110; p = 0.8007$ (2)

$\therefore$ Do not reject $H_0$ (0.5): Linemen and big backs have no difference in score when weight is also considered. (0.5)

VIII. (2 points) Can you test if the relationship between Bench Press score and Weight is the same among all position groups? If so, give the test statistic, df, p-value and a conclusion in words. If not, state what you are missing to do the test.

**Solution:**

You can't do this test (1) without a model that has the interaction between Bench and Wt. (1)

just add interaction term taking into account weight+positiongroup

IX. (2 points) Find the extra sum of squares for adding a player's 3-cone time, overall rank and shuttle time to a model that already contains the player's weight and position group.

**Solution:** RSS_{weigh,position} - RSS_{all} = 2477.57 - 2362.9 = 114.67

$SSR(3cone, overall, shuttle|Wt, PosGroup) = SSE_{Reduced} - SSE_{Full}$ (1)

$= 2477.57 - 2362.9 = 114.67$ (1)

X. (3 points) Can you test if a player's 3-cone time, overall rank and shuttle time are useful predictors of Bench in a model that already contains the player's weight and position group? If so, give the test statistic and df, critical value, the most accurate p-value you can, and a conclusion in words. If not, state what you are missing to do the test.

**Solution:**

$F_{obs} = \frac{114.67/3}{22.08} = 1.73113(3, 107)$ (1)

F_obs = (114.67 / 3) / (2362.9 / 107) = 1.73

F_{3, 107}_crit ~ 2.7

F_obs < F_crit so cannto reject

$F_{crit} = 2.70$(0.5)

$\therefore$ Do not reject $H_0$ (0.5): no evidence that extra predictors are useful;(0.5) p $> 0.05$

(0.5)

XI. (3 points) Can you test if a player's position group is a useful predictor of Bench in a model that already contains the player's weight? If so, give the test statistic and df, critical value, the most accurate p-value you can, and a conclusion in words. If not, state what you are missing to do the test.

**Solution:**

look at anova table for Bench ~ Wt + PosGroup

specifically f-value for PosGroup

$F_{obs} = 3.38; DF = (2, 110)$ (1)

$F_{crit} = 3.09$(0.5); $p = 0.03756$(0.5)

$\therefore$ Reject $H_0$ (0.5): Position Group is a useful predictor;(0.5)

5. Consider the MLR model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with assumption $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$

   (a) (4 points) Show that the least squares estimates of $\boldsymbol{\beta}$ are $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

**Solution:**

$$Q = \mathbf{e}'\mathbf{e} \text{ (1)} = (\mathbf{Y} - \mathbf{X}\mathbf{b})'(\mathbf{Y} - \mathbf{X}\mathbf{b}) = \mathbf{Y}'\mathbf{Y} - 2\mathbf{b}'\mathbf{X}'\mathbf{Y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} \text{ (1)}$$

$$\frac{\partial Q}{\partial \mathbf{b}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\mathbf{b} = 0 \text{ (1)}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \text{ (1)}$$

   (b) (3 points) Derive the distribution of $\mathbf{b}$. You have three things to derive: expected value, variance, and distribution.

**Solution:**

variance hint: use quadratic term

(c) (3 points) Show that $SSR = \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$, where $\mathbf{J}$ is a matrix of 1s, and $\mathbf{H}$ is the hat matrix.

**Solution:**
$SSR = \Sigma\hat{Y}_i - n\bar{Y}^2 = (\mathbf{HY})'\mathbf{HY} - n\frac{(\mathbf{1}'\mathbf{Y})'(\mathbf{1}'\mathbf{Y})}{n^2}$ (1)
$= \mathbf{Y}'\mathbf{H}'\mathbf{HY} - \frac{1}{n}\mathbf{Y}'\mathbf{11}'\mathbf{Y}$(1)
$= \mathbf{Y}'\mathbf{HY} - \frac{1}{n}\mathbf{Y}'\mathbf{JY}$
$= \mathbf{Y}'(\mathbf{H} - \frac{1}{n}\mathbf{J})\mathbf{Y}$ (1)

(d) (3 points) Derive the distribution of the residuals, $\mathbf{e}$. You have three things to derive.

**Solution:**

(e) (3 points) For the usual fixed effects multiple regression model, let $\mathbf{W} = \mathbf{X}'\mathbf{e}$. Simplify this expression for $\mathbf{W}$, and find the probability distribution of $\mathbf{W}$.

**Solution:**

(f) (2 points) Now let $\mathbf{W} = \hat{\mathbf{Y}}'\mathbf{e}$. Simplify this expression for $\mathbf{W}$, and find the probability distribution of $\mathbf{W}$.

**Solution:**

(g) (2 points) What do the results from the last two parts mean, in terms of how we check model assumptions?

**Solution:**

Some formulae:

$$b_1 = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})^2} = \frac{\Sigma X_i Y_i - n\bar{X}\bar{Y}}{\Sigma X_i^2 - n\bar{X}^2} \qquad b_0 = \bar{Y} - b_1\bar{X}$$

$$Var(b_1) = \frac{\sigma^2}{\Sigma(X_i - \bar{X})^2} \qquad Var(b_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma(X_i - \bar{X})^2}\right)$$

$$Var(\hat{Y}_h) = \sigma^2\left(\frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right) \qquad \sigma^2\{pred\} = Var(Y_h - \hat{Y}_h) = \sigma^2\left(1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\Sigma(X_i - \bar{X})^2}\right)$$

$$SSTO = \Sigma(Y_i - \bar{Y})^2 \qquad SSE = \Sigma(Y_i - \hat{Y}_i)^2 \qquad SSR = \Sigma(\hat{Y}_i - \bar{Y})^2 = b_1^2\Sigma(X_i - \bar{X})^2$$

$$r = \frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\Sigma(X_i - \bar{X})^2\Sigma(Y_i - \bar{Y})^2}} \qquad Cov(b_0, b_1) = -\frac{\sigma^2\bar{X}}{\Sigma(X_i - \bar{X})^2}$$

Working-Hotelling coefficient: $W = \sqrt{pF(1-\alpha; p, n-p)}$

$$R^2 = \frac{SSR}{SSTO} \qquad R^2_{adj} = 1 - \frac{(n-1)MSE}{SSTO}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad Cov(\mathbf{b}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{b} = \mathbf{H}\mathbf{Y} \qquad \mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \qquad SSE = \mathbf{Y}'(\mathbf{I} - \mathbf{H})\mathbf{Y}$$

$$SSR = \mathbf{Y}'\Big(\mathbf{H} - \frac{1}{n}\mathbf{J}\Big)\mathbf{Y} \qquad SSTO = \mathbf{Y}'\Big(\mathbf{I} - \frac{1}{n}\mathbf{J}\Big)\mathbf{Y}$$

$$\sigma^2\{\hat{Y}_h\} = \sigma^2\mathbf{x}_\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_\mathbf{h} \qquad \sigma^2\{pred\} = \sigma^2\big(1 + \mathbf{x}_\mathbf{h}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_\mathbf{h}\big)$$

Critical values of the $t$ distribution. Upper tail area is the column heading.

| DF | 0.25 | 0.2 | 0.15 | 0.1 | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 5e-04 | 1e-04 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.38 | 1.96 | 3.08 | 6.31 | 12.71 | 31.82 | 63.66 | 318.31 | 636.62 | 3183.10 |
| 2 | 0.82 | 1.06 | 1.39 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 | 22.33 | 31.60 | 70.70 |
| 3 | 0.76 | 0.98 | 1.25 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 | 10.21 | 12.92 | 22.20 |
| 4 | 0.74 | 0.94 | 1.19 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 | 7.17 | 8.61 | 13.03 |
| 5 | 0.73 | 0.92 | 1.16 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 | 5.89 | 6.87 | 9.68 |
| 6 | 0.72 | 0.91 | 1.13 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 | 5.21 | 5.96 | 8.02 |
| 7 | 0.71 | 0.90 | 1.12 | 1.41 | 1.89 | 2.36 | 3.00 | 3.50 | 4.79 | 5.41 | 7.06 |
| 8 | 0.71 | 0.89 | 1.11 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 | 4.50 | 5.04 | 6.44 |
| 9 | 0.70 | 0.88 | 1.10 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 | 4.30 | 4.78 | 6.01 |
| 10 | 0.70 | 0.88 | 1.09 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 | 4.14 | 4.59 | 5.69 |
| 11 | 0.70 | 0.88 | 1.09 | 1.36 | 1.80 | 2.20 | 2.72 | 3.11 | 4.02 | 4.44 | 5.45 |
| 12 | 0.70 | 0.87 | 1.08 | 1.36 | 1.78 | 2.18 | 2.68 | 3.05 | 3.93 | 4.32 | 5.26 |
| 13 | 0.69 | 0.87 | 1.08 | 1.35 | 1.77 | 2.16 | 2.65 | 3.01 | 3.85 | 4.22 | 5.11 |
| 14 | 0.69 | 0.87 | 1.08 | 1.35 | 1.76 | 2.14 | 2.62 | 2.98 | 3.79 | 4.14 | 4.99 |
| 16 | 0.69 | 0.86 | 1.07 | 1.34 | 1.75 | 2.12 | 2.58 | 2.92 | 3.69 | 4.01 | 4.79 |
| 18 | 0.69 | 0.86 | 1.07 | 1.33 | 1.73 | 2.10 | 2.55 | 2.88 | 3.61 | 3.92 | 4.65 |
| 20 | 0.69 | 0.86 | 1.06 | 1.33 | 1.72 | 2.09 | 2.53 | 2.85 | 3.55 | 3.85 | 4.54 |
| 24 | 0.68 | 0.86 | 1.06 | 1.32 | 1.71 | 2.06 | 2.49 | 2.80 | 3.47 | 3.75 | 4.38 |
| 28 | 0.68 | 0.85 | 1.06 | 1.31 | 1.70 | 2.05 | 2.47 | 2.76 | 3.41 | 3.67 | 4.28 |
| 32 | 0.68 | 0.85 | 1.05 | 1.31 | 1.69 | 2.04 | 2.45 | 2.74 | 3.37 | 3.62 | 4.20 |
| 36 | 0.68 | 0.85 | 1.05 | 1.31 | 1.69 | 2.03 | 2.43 | 2.72 | 3.33 | 3.58 | 4.14 |
| 40 | 0.68 | 0.85 | 1.05 | 1.30 | 1.68 | 2.02 | 2.42 | 2.70 | 3.31 | 3.55 | 4.09 |
| 50 | 0.68 | 0.85 | 1.05 | 1.30 | 1.68 | 2.01 | 2.40 | 2.68 | 3.26 | 3.50 | 4.01 |
| 60 | 0.68 | 0.85 | 1.05 | 1.30 | 1.67 | 2.00 | 2.39 | 2.66 | 3.23 | 3.46 | 3.96 |
| 70 | 0.68 | 0.85 | 1.04 | 1.29 | 1.67 | 1.99 | 2.38 | 2.65 | 3.21 | 3.44 | 3.93 |
| 80 | 0.68 | 0.85 | 1.04 | 1.29 | 1.66 | 1.99 | 2.37 | 2.64 | 3.20 | 3.42 | 3.90 |
| 100 | 0.68 | 0.85 | 1.04 | 1.29 | 1.66 | 1.98 | 2.36 | 2.63 | 3.17 | 3.39 | 3.86 |
| 150 | 0.68 | 0.84 | 1.04 | 1.29 | 1.66 | 1.98 | 2.35 | 2.61 | 3.15 | 3.36 | 3.81 |
| 200 | 0.68 | 0.84 | 1.04 | 1.29 | 1.65 | 1.97 | 2.35 | 2.60 | 3.13 | 3.34 | 3.79 |
| 500 | 0.67 | 0.84 | 1.04 | 1.28 | 1.65 | 1.96 | 2.33 | 2.59 | 3.11 | 3.31 | 3.75 |
| 1000 | 0.67 | 0.84 | 1.04 | 1.28 | 1.65 | 1.96 | 2.33 | 2.58 | 3.10 | 3.30 | 3.73 |
| Inf | 0.67 | 0.84 | 1.04 | 1.28 | 1.64 | 1.96 | 2.33 | 2.58 | 3.09 | 3.29 | 3.72 |

*F Distribution Table (Percentiles)*

| | | Numerator df | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 0.95 | 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.46 | 2.42 | 2.40 | 2.37 | 2.35 |
| 0.975 | | 6.12 | 4.69 | 4.08 | 3.73 | 3.50 | 3.34 | 3.22 | 3.12 | 3.05 | 2.99 | 2.93 | 2.89 | 2.85 | 2.82 | 2.79 |
| 0.99 | | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.62 | 3.55 | 3.50 | 3.45 | 3.41 |
| 0.95 | 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.41 | 2.38 | 2.35 | 2.33 | 2.31 |
| 0.975 | | 6.04 | 4.62 | 4.01 | 3.66 | 3.44 | 3.28 | 3.16 | 3.06 | 2.98 | 2.92 | 2.87 | 2.82 | 2.79 | 2.75 | 2.72 |
| 0.99 | | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.52 | 3.46 | 3.40 | 3.35 | 3.31 |
| 0.95 | 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.37 | 2.34 | 2.31 | 2.29 | 2.27 |
| 0.975 | | 5.98 | 4.56 | 3.95 | 3.61 | 3.38 | 3.22 | 3.10 | 3.01 | 2.93 | 2.87 | 2.81 | 2.77 | 2.73 | 2.70 | 2.67 |
| 0.99 | | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.43 | 3.37 | 3.32 | 3.27 | 3.23 |
| 0.95 | 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.74 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.34 | 2.31 | 2.28 | 2.26 | 2.23 |
| 0.975 | | 5.92 | 4.51 | 3.90 | 3.56 | 3.33 | 3.17 | 3.05 | 2.96 | 2.88 | 2.82 | 2.76 | 2.72 | 2.68 | 2.65 | 2.62 |
| 0.99 | | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.36 | 3.30 | 3.24 | 3.19 | 3.15 |
| 0.95 | 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.31 | 2.28 | 2.25 | 2.22 | 2.20 |
| 0.975 | | 5.87 | 4.46 | 3.86 | 3.51 | 3.29 | 3.13 | 3.01 | 2.91 | 2.84 | 2.77 | 2.72 | 2.68 | 2.64 | 2.60 | 2.57 |
| 0.99 | | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.29 | 3.23 | 3.18 | 3.13 | 3.09 |
| 0.95 | 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.26 | 2.23 | 2.20 | 2.17 | 2.15 |
| 0.975 | | 5.79 | 4.38 | 3.78 | 3.44 | 3.22 | 3.05 | 2.93 | 2.84 | 2.76 | 2.70 | 2.65 | 2.60 | 2.56 | 2.53 | 2.50 |
| 0.99 | | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.18 | 3.12 | 3.07 | 3.02 | 2.98 |
| 0.95 | 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.22 | 2.18 | 2.15 | 2.13 | 2.11 |
| 0.975 | | 5.72 | 4.32 | 3.72 | 3.38 | 3.15 | 2.99 | 2.87 | 2.78 | 2.70 | 2.64 | 2.59 | 2.54 | 2.50 | 2.47 | 2.44 |
| 0.99 | | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.09 | 3.03 | 2.98 | 2.93 | 2.89 |
| 0.95 | 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.18 | 2.15 | 2.12 | 2.09 | 2.07 |
| 0.975 | | 5.66 | 4.27 | 3.67 | 3.33 | 3.10 | 2.94 | 2.82 | 2.73 | 2.65 | 2.59 | 2.54 | 2.49 | 2.45 | 2.42 | 2.39 |
| 0.99 | | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 3.02 | 2.96 | 2.90 | 2.86 | 2.81 |
| 0.95 | 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2.15 | 2.12 | 2.09 | 2.06 | 2.04 |
| 0.975 | | 5.61 | 4.22 | 3.63 | 3.29 | 3.06 | 2.90 | 2.78 | 2.69 | 2.61 | 2.55 | 2.49 | 2.45 | 2.41 | 2.37 | 2.34 |
| 0.99 | | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.96 | 2.90 | 2.84 | 2.79 | 2.75 |
| 0.95 | 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.13 | 2.09 | 2.06 | 2.04 | 2.01 |
| 0.975 | | 5.57 | 4.18 | 3.59 | 3.25 | 3.03 | 2.87 | 2.75 | 2.65 | 2.57 | 2.51 | 2.46 | 2.41 | 2.37 | 2.34 | 2.31 |
| 0.99 | | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.91 | 2.84 | 2.79 | 2.74 | 2.70 |
| 0.95 | 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.04 | 2.00 | 1.97 | 1.95 | 1.92 |
| 0.975 | | 5.42 | 4.05 | 3.46 | 3.13 | 2.90 | 2.74 | 2.62 | 2.53 | 2.45 | 2.39 | 2.33 | 2.29 | 2.25 | 2.21 | 2.18 |
| 0.99 | | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.73 | 2.66 | 2.61 | 2.56 | 2.52 |
| 0.95 | 50 | 4.03 | 3.18 | 2.79 | 2.56 | 2.40 | 2.29 | 2.20 | 2.13 | 2.07 | 2.03 | 1.99 | 1.95 | 1.92 | 1.89 | 1.87 |
| 0.975 | | 5.34 | 3.97 | 3.39 | 3.05 | 2.83 | 2.67 | 2.55 | 2.46 | 2.38 | 2.32 | 2.26 | 2.22 | 2.18 | 2.14 | 2.11 |
| 0.99 | | 7.17 | 5.06 | 4.20 | 3.72 | 3.41 | 3.19 | 3.02 | 2.89 | 2.78 | 2.70 | 2.63 | 2.56 | 2.51 | 2.46 | 2.42 |
| 0.95 | 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.95 | 1.92 | 1.89 | 1.86 | 1.84 |
| 0.975 | | 5.29 | 3.93 | 3.34 | 3.01 | 2.79 | 2.63 | 2.51 | 2.41 | 2.33 | 2.27 | 2.22 | 2.17 | 2.13 | 2.09 | 2.06 |
| 0.99 | | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 2.82 | 2.72 | 2.63 | 2.56 | 2.50 | 2.44 | 2.39 | 2.35 |
| 0.95 | 80 | 3.96 | 3.11 | 2.72 | 2.49 | 2.33 | 2.21 | 2.13 | 2.06 | 2.00 | 1.95 | 1.91 | 1.88 | 1.84 | 1.82 | 1.79 |
| 0.975 | | 5.22 | 3.86 | 3.28 | 2.95 | 2.73 | 2.57 | 2.45 | 2.35 | 2.28 | 2.21 | 2.16 | 2.11 | 2.07 | 2.03 | 2.00 |
| 0.99 | | 6.96 | 4.88 | 4.04 | 3.56 | 3.26 | 3.04 | 2.87 | 2.74 | 2.64 | 2.55 | 2.48 | 2.42 | 2.36 | 2.31 | 2.27 |
| 0.95 | 100 | 3.94 | 3.09 | 2.70 | 2.46 | 2.31 | 2.19 | 2.10 | 2.03 | 1.97 | 1.93 | 1.89 | 1.85 | 1.82 | 1.79 | 1.77 |
| 0.975 | | 5.18 | 3.83 | 3.25 | 2.92 | 2.70 | 2.54 | 2.42 | 2.32 | 2.24 | 2.18 | 2.12 | 2.08 | 2.04 | 2.00 | 1.97 |
| 0.99 | | 6.90 | 4.82 | 3.98 | 3.51 | 3.21 | 2.99 | 2.82 | 2.69 | 2.59 | 2.50 | 2.43 | 2.37 | 2.31 | 2.27 | 2.22 |