**Spring 2020** **Stat 305 (Section 4)** **Final**

**Name:**

*Total points for the exam is 100. Points for individual questions are given at the beginning of each problem. Show all your calculations clearly. Put final answers in the box at the right (except for the diagrams!).*

**1.** [2+2+4+5+4+5+4+4+6=36 points]

The data used in creating the regression analysis JMP output on page 13-14 come from a metal-cutting drilling experiment. The explanatory variables studied were

$x_1$ = natural logarithm of the diameter of the drills
$x_2$ = natural logarithm of the feed rate (rate of drill penetration into the workpiece) of the drills

and the response variable was

$y$ = natural logarithm of the thrust required to rotate the drill on an aluminum alloy

Use the **first** regression analysis output in answering the questions (a)-(d) below.

(a) What fraction of the observed raw variation in $y$ is explained using $x_1$ as a predictor variable?

fraction =

(b) What is the sample correlation between $y$ and $\hat{y}$? (Give a number.)

corr=

(c) Give a 95% two-sided confidence interval for the *increase* in mean value of natural logrithm of thrust ($y$) associated with a 0.5 unit increase in natural logrithm of the drill diameter ($x_1$). (No need to simplify.)

| conf. interval = |
|---|

(d) As it turns out, the data have $\bar{x}_1 = -1.145$ and $\sum(x_{1i} - \bar{x}_1)^2 = 0.475$. Use these facts and give a 90% lower confidence bound for the mean natural logarithm of thrust ($y$) when natural logarithm of drill diameter equals $-1$ (i.e. $x_1 = -1$). (No need to simplify.)

| lower conf. bd = |
|---|

Use the **second** regression analysis output in answering the questions (e)-(i) below. The standard error of the predicted mean response $\hat{\mu}_{y|\boldsymbol{x}}$ for the second regression analysis is in the last column of the data table.

(e) Give the residual corresponding to the second observation ($x_1 = -0.901, x_2 = -5.116$, and $y = 5.927$).

| residual = |
|---|

(f) Give a 90% upper prediction bound for the next natural logarithm of thrust ($y$) when $x_1 = -0.901$ and $x_2 = -5.116$. (No need to simplify.)

| upper pred. bd = |
|---|

(g) Find the value of a $t$-statistic, its degrees of freedom, and the corresponding $p$-value for testing whether the predictor $x_2$ can be dropped from this multiple linear regression model. What is your conclusion? (*Hint: If the slope is zero, the predictor is of no use to predict the response and therefore can be dropped.*)

| Observed $t$ = | df = | $p$-value = |
|---|---|---|

Conclusion (circle only one):

(a) $x_2$ should be dropped

(b) $x_2$ should not be dropped

(h) Give the value of an $F$ statistic, its degrees of freedom, and the corresponding $p$-value for testing $H_0: \ \beta_1 = \beta_2 = 0$ against $H_a: \ $ not $H_0$. What is your conclusion?

| Observed $F$ = | $\text{df}_1$ = | $\text{df}_2$ = | $p$-value = |
|---|---|---|---|

Conclusion (circle only one):

(a) $x_1, \ x_2$ should be dropped

(b) $x_1, \ x_2$ should not be dropped

(i) Fitting the complete second-order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1)^2 + \beta_4 (x_2)^2 + \beta_5 (x_1 x_2) + \epsilon$$

gave $SSR = 1.1093$ and $SSE = 0.01715$. Find the value of an $F$ statistic and its degrees of freedom for testing whether all the second order predictors (i.e. $(x_1)^2, (x_2)^2$ and $(x_1 x_2)$) can be dropped from the complete second-order model involving all 5 predictors. What is your conclusion.?

| Observed $F$ = | $\text{df}_1$ = | $\text{df}_2$ = | $p$-value = |
|---|---|---|---|

Conclusion (circle only one):

(a) $(x_1)^2, (x_2)^2, (x_1 x_2)$ should be dropped

(b) $(x_1)^2, (x_2)^2, (x_1 x_2)$ should not be dropped

**2.** [3+6+4=13 points]

Two independent discrete random variables $X$ and $Y$ can be described using the following probability functions:

| $x$ | $-2$ | 0 | 1 | 5 |
|---|---|---|---|---|
| $f_X(x)$ | 0.1 | 0.4 | 0.2 | 0.3 |

| $y$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $f_Y(y)$ | 0.2 | 0.4 | 0.3 | 0.1 |

(a) Find the cumulative probability function for $X$.

(b) Find the means and standard deviations for $X$ and $Y$ respectively.

| | |
|---|---|
| $\mu_X =$ | $\sigma_X =$ |
| $\mu_Y =$ | $\sigma_Y =$ |

(c) Find the mean and standard deviation for the random variable $3X - 5Y + 2$.

mean = s.d=

**3.** [4+4+3+6+5=22 points]

The spring constants of two types of steel springs are measured. The resulting measurements and some summary statistics are given below.

| Type 1 Springs | Type 2 Springs |
|---|---|
| 1.99, 2.06, 1.99, 1.94, 2.05, 1.88, 2.30 | 2.85, 2.74, 2.74, 2.63, 2.74, 2.80 |
| $\bar{x}_1 = 2.03$ | $\bar{x}_2 = 2.75$ |
| $s_1 = 0.134$ | $s_2 = 0.074$ |

(a) Give a 95% upper prediction bound for the spring contant of the next Type 1 spring. (No need to simplify.)

| upper pred. bd (Type 1)= |
|---|

(b) Give a a 95% lower confidence bound for the mean spring contant for Type 1 springs. (No need to simplify.)

| lower conf. bd (Type 1)= |
|---|

(c) What assumptions have to be made in order to construct the confidence bound in part (b)?

(d) Give a 95% two sided confidence interval for the difference in mean spring constants for the two types of springs. (No need to simplify.)

| conf. interval = |
|---|

(e) What assumptions have to be made in order to construct the confidence interval in part (d)?

**4.** [5+5+5=15 points]

Jars of a particular type are made in a factory. The jars have weights with mean 120 g and standard deviation 1.6 g.

(a) Assume that the weights are normally distributed and specifications on the weights are 120 g $\pm$ 4 g. What fraction of the weights of jars actually satisfy this specification ?

fraction =

(b) Evaluate the probability that *at most* 11 of the next 12 jars produced are within the specifications of 120 g $\pm$ 4 g.

probability =

(c) Let $\bar{X}$ denote the sample mean weight of 80 jars of this type. Approximate the probability that $\bar{X}$ is bigger than 120.1 g.

probability =

**5.** [7+7=14 points]

An experiment was made in order to measure the compressive strength of 6 different concrete formulas. The data, some summary statistics, and the analysis of variance table are given below.

| Formula 1 | Formula 2 | Formula 3 |
|---|---|---|
| 5659, 6225, 5376<br>$\bar{y}_1 = 5753.33$<br>$s_1 = 432.29$ | 5093, 4386, 4103<br>$\bar{y}_2 = 4527.33$<br>$s_2 = 509.91$ | 3395, 3820, 3112<br>$\bar{y}_3 = 3442.33$<br>$s_3 = 356.37$ |
| **Formula 4** | **Formula 5** | **Formula 6** |
| 2971, 3678, 3325<br>$\bar{y}_4 = 3324.67$<br>$s_4 = 353.50$ | 2122, 1372, 1160<br>$\bar{y}_5 = 1551.33$<br>$s_5 = 505.45$ | 2051 , 2631 , 2490<br>$\bar{y}_6 = 2390.67$<br>$s_6 = 302.49$ |

ANOVA Table

| Source | DF | Sum of Squares | Mean Square | F-ratio |
|---|---|---|---|---|
| Model | 5 | 33584690 | 6716938 | 38.5359 |
| Error | 12 | 2091642 | 174304 | (Prob > F) < 0.0001 |
| Total | 17 | 35676332 | | |

Let $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ and $\mu_6$ be the true compressive strength for concrete formula 1 to formula 6.

(a) Give a 95% two-sided confidence interval for the quantity $\frac{1}{2}(\mu_2 + \mu_3) - \mu_4$. (No need to simplify).

conf. interval =

(b) Assess the strength of the evidence against $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$ in favor of $H_a$ : not $H_0$. (Show all steps!)

# JMP Output

## Data Table

| | x1 | x2 | y | StdErr Pred y |
|---|---|---|---|---|
| 1 | -1.386 | -5.116 | 5.438 | 0.0372993671 |
| 2 | -0.901 | -5.116 | 5.927 | 0.0374781564 |
| 3 | -0.901 | -4.343 | 6.346 | 0.0363956559 |
| 4 | -1.386 | -4.343 | 5.927 | 0.0362003412 |
| 5 | -1.492 | -4.711 | 5.635 | 0.0379006761 |
| 6 | -1.146 | -5.298 | 5.416 | 0.0379368887 |
| 7 | -0.799 | -4.711 | 6.363 | 0.0378322421 |
| 8 | -1.146 | -4.075 | 6.337 | 0.0405601347 |
| 9 | -1.146 | -4.711 | 5.991 | 0.0188611718 |
| 10 | -1.146 | -4.711 | 5.991 | 0.0188611718 |
| 11 | -1.146 | -4.711 | 5.94 | 0.0188611718 |
| 12 | -1.146 | -4.711 | 5.94 | 0.0188611718 |

## Regression 1

**Response y**

**Whole Model** — **x1**

**Summary of Fit**

| | |
|---|---|
| RSquare | 0.416442 |
| RSquare Adj | 0.358087 |
| Root Mean Square Error | 0.256388 |
| Mean of Response | 5.937583 |
| Observations (or Sum Wgts) | 12 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 1 | 0.4691010 | 0.469101 | 7.1363 |
| Error | 10 | 0.6573479 | 0.065735 | Prob > F |
| C. Total | 11 | 1.1264489 | | 0.0234* |

**Parameter Estimates**

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 7.0750973 | 0.4322 | 16.37 | <.0001* |
| x1 | 0.9933896 | 0.371864 | 2.67 | 0.0234* |

# Regression 2

Response y

Whole Model | x1 | x2

### Summary of Fit

| | |
|---|---|
| RSquare | 0.965895 |
| RSquare Adj | 0.958316 |
| Root Mean Square Error | 0.065335 |
| Mean of Response | 5.937583 |
| Observations (or Sum Wgts) | 12 |

### Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Ratio |
|---|---|---|---|---|
| Model | 2 | 1.0880309 | 0.544015 | 127.4438 |
| Error | 9 | 0.0384180 | 0.004269 | Prob > F |
| C. Total | 11 | 1.1264489 | | <.0001* |

### Parameter Estimates

| Term | Estimate | Std Error | t Ratio | Prob>\|t\| |
|---|---|---|---|---|
| Intercept | 10.26907 | 0.287208 | 35.75 | <.0001* |
| x1 | 0.9936207 | 0.094762 | 10.49 | <.0001* |
| x2 | 0.6776261 | 0.056275 | 12.04 | <.0001* |