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

Name: Solution.

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 = 41.64%

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

corr= 0.6453

$$r = \sqrt{R^2}$$
$$= \sqrt{0.41642} = 0.645323 2$$

(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 = $0.5(0.9934 \pm 2.228(0.3719))$

The quantile used is $t_{n-2,\,1-\alpha/2} = t_{10,\,0.975} = 2.228$.

$0.5(b_1 \pm t_{10,\,0.975}\, se(b_1))$

$= 0.5(0.9934 \pm 2.228(0.3719))$

or

$0.5\left(b_1 \pm t_{10,\,0.975} \dfrac{S_{LF}}{\sqrt{\sum(x_{1i}-\bar{x}_1)^2}}\right)$

$= 0.5\left(0.9934 \pm 2.228 \dfrac{0.2564}{\sqrt{0.475}}\right)$

(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 = $6.0817 - 1.372(0.2564)\sqrt{\frac{1}{12} + \frac{(1.145-1)^2}{0.475}}$

$\hat{\mu}_{y|x} = b_0 + b_1 x = 7.0750973 + 0.9933896 \times (-1)$

$= 6.081708.$

the quantile used is $t_{n-2,\,1-\alpha} = t_{10,\,0.9} = 1.372$.

$\hat{\mu}_{y|x} - t_{10,\,0.9} \cdot S_{LF} \cdot \sqrt{\dfrac{1}{12} + \dfrac{(x-\bar{x}_1)^2}{\sum(x_{1i}-\bar{x}_1)^2}}$

$= 6.081708 - 1.372(0.2564)\sqrt{\dfrac{1}{12} + \dfrac{(-1+1.145)^2}{0.475}}$

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 = 0.019917

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$
$$= 10.26807 + 0.9936207(-0.901) + 0.6776261(-5.116)$$
$$= 5.907083$$
$$e = y - \hat{y}$$
$$= 5.927 - 5.907083 = 0.019917$$

(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 = $5.907083 + 1.383\sqrt{\underline{(0.0653)^2} + (0.03748)^2}$

or $\underline{0.0043}$

$$\hat{\mu}_{y|x} = b_0 + b_1 x_1 + b_2 x_2 = 5.907083.$$

$$t_{n-p,\,1-\alpha} = t_{9,\,0.9} = 1.383.$$

$$\hat{\mu}_{y|x} + t_{n-p,\,1-\alpha}\sqrt{s_{LF}^2 + (se(\hat{\mu}_{y|x}))^2}$$

$$= 5.907083 + 1.383\sqrt{\underbrace{(0.0653)^2}_{\text{or } 0.0043} + (0.03748)^2}$$

(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 = 12.04	df = 12 − 3 = 9	p-value = < 0.0001

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 $F = 127.4438$	$df_1 = 2$	$df_2 = 9$	p-value $= < 0.0001$

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.?

$$F = \frac{(SSR_f - SSR_r)/k}{SSE_f/(n-p)}$$

$$= \frac{(1.1093 - 1.0880309)/3}{0.01715/(12-6)}$$

$$= 2.480362$$

$$p\text{-value} = P(F_{3,6} > 2.480362)$$

$$= 0.158379 \quad (\approx 0.158)$$

Observed $F =$	$df_1 = 3$	$df_2 = 6$	p-value $= 0.158$

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.

$$F_X(x) = \begin{cases} 0, & x < -2 \\ 0.1, & -2 \le x < 0 \\ 0.5, & 0 \le x < 1 \\ 0.7, & 1 \le x < 5 \\ 1, & x \ge 5 \end{cases}$$

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

$\mu_X =$ 1.5	$\sigma_X =$ 2.4187
$\mu_Y =$ 1.3	$\sigma_Y =$ 0.9

$E(X) = \sum x f_X(x)$

$= -2(0.1) + 0(0.4) + 1(0.2) + 5(0.3)$

$= 1.5$.

$E(X^2) = \sum x^2 f_X(x)$

$= (-2)^2(0.1) + 0^2(0.4) + 1^2(0.2) + 5^2(0.3)$

$= 8.1$

$Var(X) = E(X^2) - (E(X))^2 = 8.1 - 1.5^2 = 5.85$.

$SD(X) = \sqrt{5.85} = 2.4187$

$E(Y) = \sum y f_Y(y)$

$= 0(0.2) + 1(0.4) + 2(0.3) + 3(0.1)$

$= 1.3$

$E(Y^2) = \sum y^2 f_Y(y)$

$= 0^2(0.2) + 1^2(0.4) + 2^2(0.3) + 3^2(0.1)$

$= 2.5$.

$Var(Y) = E(Y^2) - (E(Y))^2 = 2.5 - 1.3^2 = 0.81$

$SD(Y) = \sqrt{0.81} = 0.9$

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

mean = 0	s.d= 8.53815

$E(3X-5Y+2)$

$= 3E(X) - 5E(Y) + 2$

$= 3 \times 1.5 - 5 \times 1.3 + 2 = 0$

$Var(3X-5Y+2)$

$= 3^2 Var(X) + (-5)^2 Var(Y)$

$= 9 \times 5.85 + 25 \times 0.81$

$= 72.9$.

$SD(3X-5Y+2)$

$= \sqrt{72.9} = 8.53815$

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)= $2.03 + 1.943\,(0.134)\sqrt{1+\frac{1}{7}}$

The quantile used is $t_{n_1-1,\,1-\alpha} = t_{6,0.95} = 1.943$.

$\bar{x}_1 + t_{6,0.95} \cdot s_1 \sqrt{1+\frac{1}{n_1}}$

$= 2.03 + 1.943\,(0.134)\sqrt{1+\frac{1}{7}}$

(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)= $2.03 - 1.943\,(0.134)\cdot\sqrt{\frac{1}{7}}$

$\bar{x}_1 - t_{6,0.95}\, s_1 \sqrt{\frac{1}{n_1}}$

$= 2.03 - 1.943\,(0.134)\cdot\sqrt{\frac{1}{7}}$

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

observations in Type 1 spring sample are iid $N(\mu_1, \sigma^2)$.

(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 = $-0.72 \pm 2.201\,(0.1108)\sqrt{\frac{1}{6}+\frac{1}{7}}$

The quantile used is

$t_{n_1+n_2-2,\,1-\alpha/2} = t_{11,\,0.975} = 2.201$.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{6\,(0.134)^2 + 5\,(0.074)^2}{11}$$

$$= 0.01228327$$

$$S_p = \sqrt{0.1067273} = 0.1108289$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{n_1+n_2-2} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= (2.03 - 2.75) \pm 2.201\,(0.1108)\sqrt{\frac{1}{6}+\frac{1}{7}}$$

$$= -0.72 \pm 2.201\,(0.1108)\sqrt{\frac{1}{6}+\frac{1}{7}}$$

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

- observations from Type 1 spring sample are iid $N(\mu_1, \sigma^2)$
- observations from Type II spring sample are iid $N(\mu_2, \sigma^2)$
- These two samples are independent.

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 =	98.76%

$$P(120-4 \leq X \leq 120+4)$$

$$= P\left(\frac{-4}{1.6} \leq \frac{X-120}{1.6} \leq \frac{4}{1.6}\right)$$

$$= P(-2.5 \leq Z \leq 2.5)$$

$$= 0.9876$$

(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 =	0.13906

Let y be the number of jars within the spec.

$y \sim \text{Binomial}(n=12, p=0.9876)$

$$P(Y \leq 11)$$

$$= 1 - P(Y = 12)$$

$$= 1 - \binom{12}{12}(0.9876)^{12}(1-0.9876)^{0}$$

$$= 0.13906$$

(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 =

$$\bar{X} \sim N\left(120, \frac{1.6^2}{80}\right)$$

$$P(\bar{X} > 120.1)$$

$$= P\left(\frac{\bar{X}-120}{1.6/\sqrt{80}} > \frac{0.1}{1.6/\sqrt{80}}\right)$$

$$= P\left(Z > \frac{0.1}{1.6/\sqrt{80}}\right) = P(Z > 0.559017)$$

$$= \quad 0.2881$$

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

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 μ_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 = $660.16 \pm 2.179\,(417.4973)\sqrt{\frac{1}{2}}$

The quantile used is $t_{n-r,1-\alpha/2} : t_{12,0.975} = 2.179$

$$\hat{\ell} = \frac{1}{2}(\bar{y}_2 + \bar{y}_3) - \bar{y}_4 = \frac{1}{2}(4527.33 + 3442.33) - 3324.67$$

$$= 660.16$$

$$S_p = \sqrt{174304} = 417.4973$$

$$\hat{\ell} \pm t_{n-r}\, S_p \sqrt{\frac{1}{n_2}c_2^2 + \frac{1}{n_3}c_3^2 + \frac{1}{n_4}c_4^2}$$

$$= 660.16 \pm 2.179\,(417.4973)\sqrt{\frac{1}{3}\left(\frac{1}{2}\right)^2 + \frac{1}{3}\left(\frac{1}{2}\right)^2 + \frac{1}{3}(-1)^2}$$

$$= 660.16 \pm 2.179\,(417.4973)\sqrt{\frac{1}{2}}$$

(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!)

1. $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$
   H_a : Not all μ_i's are equal.

2. The test static is
$$F = \frac{SSR/(r-1)}{SSE/(n-r)}$$
   Assuming
   - H_0 is true
   - $Y_{ij} = \mu_i + \varepsilon_{ij}$.
     $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$.

   The test statistic $\sim F_{r-1, n-r} = F_{5,12}$.

3. The observed F is 38.5359.
   The p-value is
   $P(F_{5,12} > 38.5359) < 0.0001$

4. The p-value is very small, we reject H_0 and conclude H_a.

5. There is significant evidence that the means for 6 concrete formulas are not all the same.

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*