

STA 3032 - Practice set 3

This practise set is not meant to represent the format or content of the exam, it is simply for practice. You should also practise on textbook problems (assigned and not assigned).

1. Ford wants to compare mean assembly times for Explorers at their 3 assembly plants. They observe random samples of 10 cars at each plant, and obtain the following summary statistics on assembly times (in minutes):

Plant	Mean	Std. Dev.
Atlanta	180	12
Chicago	185	10
Detroit	175	9

- (a) Compute the between plant (Treatment) sum of squares and its degrees of freedom.
 - (b) Compute the within plant (Error) sum of squares and its degrees of freedom.
 - (c) Compute the test statistic and make conclusion.
 - (d) If applicable, use Tukey's multiple comparison procedure with $\alpha = 0.05$ to determine which plant's differ.
2. A study was conducted to relate weight gain in chickens (y) to the amount of the amino acid lysine ingested by the chicken (x). A simple linear regression is fit to the data.

	coefficients	S.E.	t-stat	p-value
Intercept	12.4802	1.2637	9.8762	<0.0000
X	36.8929	7.5640	4.8774	0.0012

Source	df	SS	MS	F	p-value
Regression	1	27.07	27.07	23.79	0.0012
Residual	8	9.10	1.14		
Total	9	36.18			

- (a) Give the fitted equation, and the predicted value for $x = 0.20$.
 - (b) Give a 95% Confidence Interval for the MEAN weight gain of all chickens with $x = 0.20$ (Note: $\bar{x} = 0.16$ and $\sum(x_i - \bar{x})^2 = 0.020$)
 - (c) What proportion of the variation in weight gain is "explained" by lysine intake?
3. A researcher reports that the correlation between length (inches) and weight (pounds) of a sample of 16 male adults of a species is $r = 0.40$. A colleague from Europe transforms the data from length in inches to centimeters (1 inch=2.54 cm) and weight from pounds to kilograms (1 pound=2.2 kg). What is the colleagues estimate of the correlation?

4. A randomized block design is conducted to compare the output of three weaving looms (treatments) for a sample of 10 operators (blocks), where each operator's output is measured on each loom. The Mean Square Error from the ANOVA is $MSE = 500$. Compute Bonferroni's B , i.e. margin of error, the minimum significant difference for concluding that two looms' population means differ if their sample means differ by at least B .
5. Late at night you find the following output in your department's computer lab. The data represent numbers of emigrants from Japanese regions, as well as a set of predictor variables from each region.

	coefficients	S.E.	t-stat	p-value
Intercept	407.070	226.341	1.798	0.079
LANDCULT	-1.685	3.5677	-0.472	0.639
AREAFARM	-2.132	1.056	-2.019	0.050
PIONEERS	175.968	61.222	2.874	0.006

R-square=0.274, R-square(adj)=0.222

Source	df	SS	MS	F	p-value
Regression	3	514814.087	171604.696	5.187	0.004
Residual	41	1356447.158	33084.077		
Total	44	1871261.244			

- (a) How many regions are there in the analysis?
 - (b) Give the test statistic and p-value for testing (H_0) that none of the predictors are associated with the response EMGRANTS.
 - (c) Give the test statistic and p-value for testing whether LANDCULT is associated with EMGRANTS, after controlling for AREAFARM and PIONEERS.
 - (d) What proportion of the variation in EMGRANTS is "explained" by the model?
 - (e) Give the estimated regression equation.
6. A realtor is interested in the determinants of home selling prices in his territory. He takes a random sample of 24 homes that have sold in this area during the past 18 months, observing:
 - selling PRICE (y)
 - AREA (x_1)
 - BEDrooms (x_2)
 - BATHrooms (x_3)
 - POOL dummy ($x_4 = 1$ if Yes, 0 if No)
 - AGE (x_5)

He/She fits the following models (predictor variables to be included in model are given for each model):

Model 1 AREA, BED, BATH, POOL, AGE SSE1 = 250, SSR1 = 450
 Model 2 AREA, BATH, POOL SSE2 = 325, SSR2 = 375

- (a) Test whether neither BED or AGE are associated with PRICE, after adjusting for AREA, BATH, and POOL at the $\alpha = 0.05$ significance level.
- (b) What statement best describes β_4 in Model 1?
 - i. Added value (on average) for a POOL, controlling for AREA, BED, BATH, AGE
 - ii. Effect of increasing AREA by 1 unit, controlling for other factors
 - iii. Effect of increasing BED by 1 unit, controlling for other factors
 - iv. Effect of increasing BATH by 1 unit, controlling for other factors
 - v. Average price for a house with a POOL

7. Yields of entozoic amoebae in 5 treatment conditions. Conditions:

1=None

2=Heat at 70° C for 10min

3=Addition of 10% Formalin

4=Heat followed by Formalin

5=Formalin followed, after 1 hr by heat

Trt 1	Trt 2	Trt3	Trt 4	Trt 5
265	204	191	221	259
292	234	207	205	206
268	197	218	178	179
251	176	201	167	199
245	240	192	224	180
192	190	192	225	146
228	171	214	171	182
291	190	206	214	147
185	222	185	283	182
247	211	163	277	223

Data can also be found at

http://www.stat.ufl.edu/~athienit/STA3032/practice3_entozoic.csv.

Are there differences in the means of the 5 treatments?

8. Endurance Times (in minutes) for nine well-trained cyclists, on each of 4 doses of caffeine (0,5,9,13 mg)

Dose	Cyclists								
	1	2	3	4	5	6	7	8	9
0	36.05	52.47	56.55	45.20	35.25	66.38	40.57	57.15	28.34
5	42.47	85.15	63.20	52.10	66.20	73.25	44.50	57.17	35.05
9	51.50	65.00	73.10	64.40	57.45	76.49	40.55	66.47	33.17
13	37.55	59.30	79.12	58.33	70.54	69.47	46.48	66.35	36.20

Data can also be found at

http://www.stat.ufl.edu/~athienit/STA3032/practice3_endurance.csv

- (a) Which factor is the the treatment and which is the block?
- (b) Test whether the factor populations means are all equal.
 - i. Using parametric technique.
 - ii. (optional) Using non-parametric technique.