Name: Original

ST 411/511, Fall 2014

FINAL EXAM

- There are 4 questions, worth a total of 50 points. Attempt all questions.
- **Question One** consists of 20 multiple choice questions and must be answered on the scantron sheet. It is recommended you also circle your answers on this exam paper.
- You have 110 minutes to complete the exam.
- Please:

  (i) Do not look at the exam until I tell you to, and

  (ii) stop writing when I announce that the exam is over.

  (violations will be considered acts of academic dishonesty)
- The exam is closed-book but you may have one sheet (8.5 x 11in) double sided of your own notes.
- You may bring a calculator.
- Read the questions carefully.
- **Read the questions carefully.**

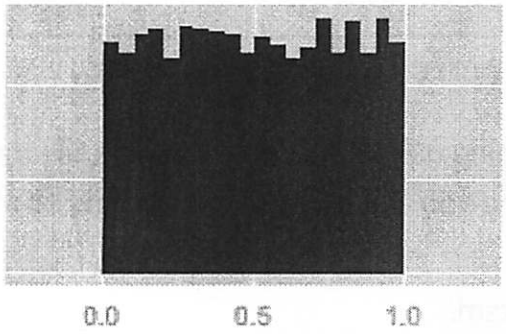
Question	Points	Possibl points
1		20
2		10
3		10
4		10
Total		**50**

Question One – 20 Multiple choice questions

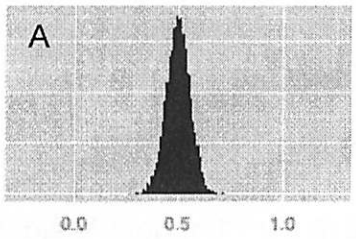
(1 each)

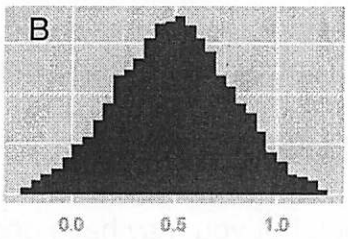
ANSWER ON THE SCANTRON SHEET

For questions 1 and 2, consider the following population histogram:

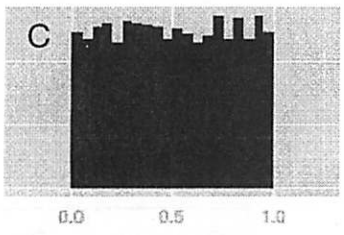


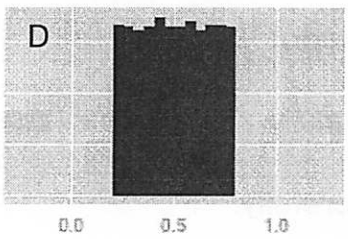
1. Which histogram is most likely to be the **sampling distribution of the sample average** for samples of size 25 from this population?





A





2. Which of the above histograms is most likely to be **a histogram of a single sample** of size 300 from this population?

C

3. A p-value gives: (choose the best answer)

A. the proportion of test-statistics that are greater than zero.
B. the probability that the null hypothesis is true.
C. evidence for the null hypothesis.
D. the probability of a more extreme test-statistic if the null hypothesis was true.
E. the probability the alternative hypothesis is true.

4. Consider the following statistical summary:

"With 95% confidence, the mean weight of male possums is between 30 and 100 grams more than the mean weight of female possums "

What can you say about the p-value you would get from a two-sided t-test of the null hypothesis that the true difference in mean weight between male and female possums is 0 grams? (Choose the best answer)

A. we can say nothing based on the information given
B. it is approximately equal to 0.05
C. it is less than 0.05
D. it is greater than 0.05
E. it is approximately equal to 0.95

5. Researchers are interested in the effect of Seasonal Affective Disorder (SAD) lights on the mood of OSU students.

- They randomly sample 100 OSU students.
- All 100 students are asked to attend a daily 5pm hour long seminar on student success for two weeks in January.
- Fifty of the students are randomly assigned to a seminar room with normal lighting. The other fifty attend the seminar in a room with SAD lights.
- All 100 students conduct a mood evaluation before and after the two weeks of seminar and the researchers record their change in mood score.

What kind of inferences will the researchers be able to make?

A. Both population and causal inferences
B. Population inference
C. Causal inference
D. Neither population nor causal inferences are justified

6. Researchers are interested in the effect of coffee consumption of the amount of sleep students at OSU get.

- They randomly sample 100 OSU students.
- They ask them on average how many cups of coffee they drink a day and on average how many hours of sleep they get at night.

What kind of inferences will the researchers be able to make (assuming the students are honest and accurate)?

A. Both population and causal inferences
B. Population inference
C. Causal inference
D. Neither population nor causal inferences are justified

7. The two sample t-test is resistant to outliers.

A. TRUE
B. FALSE

8. There are occasions when both the paired *t*-test and the Wilcoxon Signed Rank test are appropriate.

A. TRUE
B. FALSE

9. Researchers are interested in whether your hand span is different between your dominant and non dominant hand.

- They randomly sample 20 people.
- For each person they measure their hand span on their dominant hand and their non-dominant hand.

To compare the hand span between the dominant and non-dominant hands the researchers should conduct a:

A. a paired t-test
B. a two sample t-test

10. In context of comparing the mean response between two samples, which of these statements **is not** a valid reason for trying a log transform of the response?

A. The distribution of the response is severely right skewed.
B. A lot of observed responses are zero.
C. The spread of response for the group with the higher average is larger than the group with the lower average.
D. The observed responses range over several orders of magnitude

11. An observational study is conducted to compare the income between graduates of OSU and graduates of UO. 1000 graduates from 2011 are randomly sampled from each school and surveyed to obtain their yearly income in dollars.

The sample average of the **log** income for OSU students is found to be 10.9.
The sample average of the **log** income for UO students is found to be 10.4

The best interpretation of this result is:
10.9 - 10.4 = 0.5

A. The mean income for OSU students is 0.5 dollars more than the mean income for UO students.
B. The median income for OSU students is exp(0.5) times the median income for UO students.
C. The mean income for OSU students is 0.5 times the mean income for UO students.
D. Going to OSU increases your income by exp(0.5) times the income you would have received if you went to UO.

12. In a one-way ANOVA, a large F-statistic is obtained when the within group variation is large compared to the between-group variation.

A. TRUE
B. FALSE

13. A study is designed to compare the fitness levels of OSU students from four groups: those that walk, bike, bus or drive to campus. The response of interest is the time taken to complete a 1 mile run. Let μ_{walk}, μ_{bike}, μ_{bus} and μ_{drive} represent the mean response for four populations respectively.

Which of the following is a linear combination that compares the average of the mean fitness levels for the walk and bike populations to the average of the mean fitness levels for the bus and drive populations?

A. $\frac{1}{2}(\mu_{walk} + \mu_{bike}) + \frac{1}{2}(\mu_{bus} + \mu_{drive})$
B. $\frac{1}{2}(\mu_{walk} + \mu_{bike}) - \frac{1}{2}(\mu_{bus} + \mu_{drive})$
C. $\frac{1}{2}(\mu_{walk} - \mu_{bike}) + \frac{1}{2}(\mu_{bus} - \mu_{drive})$
D. $\mu_{walk} - \mu_{bike} + \mu_{bus} - \mu_{drive}$

14. In simple linear regression, the standard error on the predicted response at X_0, is larger the further X_0 is from $\bar{X}$.

A. TRUE
B. FALSE

15. In simple linear regression, a confidence interval on the slope is wider for larger samples.

A. TRUE
B. FALSE

16. The null hypothesis in a regression ANOVA is equivalent to the null hypothesis that the slope in the regression equation is zero.

A. TRUE
B. FALSE

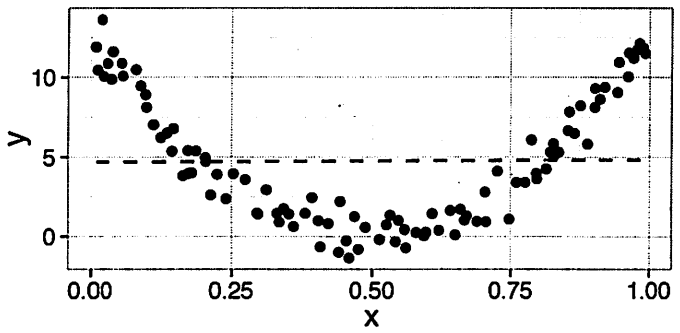
17. The following is part of the output from a simple linear regression model with explanatory variable X.

```
Coefficients:
             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)    -0.05       0.20    -0.25     0.803
x               1.00       0.02    50.00    <2e-16 ***
```

The value for the t-statistic for testing the null hypothesis that the slope is equal to one (β_1=1) is:

A. -0.25
B. 50.00
C. 0
D. Not calculable from this output

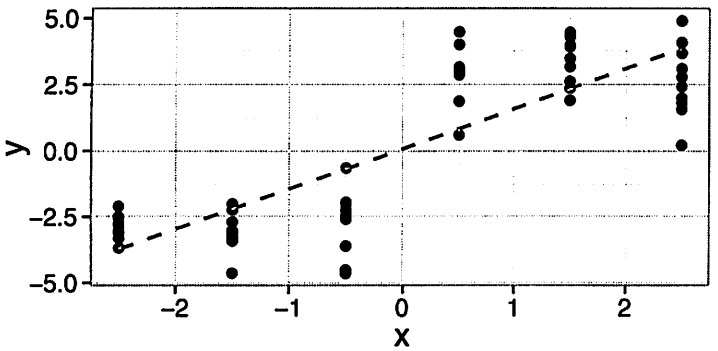
18. Consider the following plot of data from an experiment, where Y is the response, and X the explanatory variable. The dashed line is the fitted simple linear regression line.



Would you expect to reject the null hypothesis in a **regression ANOVA F-test**?

A. Yes, reject the null.
B. No, fail to reject the null.

19. Consider the following plot of data from an experiment, where Y is the response, and X the explanatory variable. The dashed line is the fitted simple linear regression line.

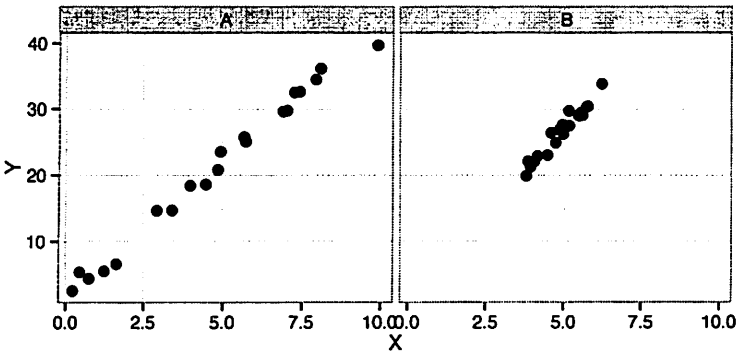


Would you expect to reject the null hypothesis in a **lack of fit F-test**?

A. Yes, reject the null.
B. No, fail to reject the null.

20. The following two plots come from datasets that both satisfy the same regression model for the mean response:

$$\mu(Y \mid X) = 2 + 4X$$



Is the standard deviation of the explanatory values, s_x, for A **smaller**, **larger** or **about the same size** as B?

A. Smaller
B. Larger
C. About the same size

Question Two – Two group comparisons

a) A local large running group is interested in a new training program for runners. Twenty members are randomly selected to implement the new training program. For each runner, they record the runner's time to run 10km before the new training program and after six weeks of the new training program.

Below you will find R output for both a paired t-test and a two sample t-test, comparing the before measurements to the after measurements.

```
> t.test(minutes ~ training, data = run_m, paired = TRUE)

        Paired t-test

data:  minutes by training
t = 17.4761, df = 19, p-value = 3.642e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 4.282944 5.448419
sample estimates:
mean of the differences
                4.865681

> t.test(minutes ~ training, data = run_m, var.equal = TRUE)

    Two Sample t-test

data:  minutes by training
t = 1.61, df = 38, p-value = 0.1157
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.252458 10.983820
sample estimates:
mean in group before  mean in group after
            62.38725              57.52156
```

i) Is the most appropriate analysis a paired t-test or a two sample t-test? Justify your answer.

Paired, since each runner has two measurements

(2)

ii) Based on your answer above, write a statistical summary that includes interpretation of the test result, a point estimate and a confidence interval.

There is convincing evidence the mean difference in run time is not equal to zero (paired t-test pvalue < 0.0001).

It is estimated the time after the training program is 4.87 minutes less on average than the time before training.

→ or phrase mean difference

With 95% confidence, the time after training is between 4.29 and 5.45 minutes less on average than before training. (3)

b)

i) State the assumptions of a two sample t-test.

Normal populations

Equal population SDs

Independence

(3)

ii) Name an alternative to the two sample t-test.

Wilcoxon Rank Sum OR Welch's test OR Randomization test

(1)

iii) What advantages does the test you named above it have over the two sample t-test?

No Normality assumption

No Normality assumption

Resistant

No equal SD assumption (1)

Total points for Question 2 out of (10)

Question Three - Multiple Group Comparisons

The following partially completed one-way ANOVA table is from a randomized experiment comparing heights of seedlings (in cm) treated with different fertilizers.

	Sum of squares	d.f.	Mean sum of squares	F-stat	p-value
Between groups	i) 40	4	10	2.5	0.20
Within groups	140	ii) 35	4		
Total	**180**	39			

a) and b) Find the values that have been left blank in the above one way ANOVA table.

(2)

c) How many fertilizers and how many seedlings were in the study?

5 fertilizers 40 seedlings

(1)

d) Based on the p-value given, write a one sentence summary of the results of the one way ANOVA.

There is no evidence against all fertilizers having the same mean seedling height (one way ANOVA pvalue = 0.20) or treatment effect.

(2)

e) One of the fertilizers used in the study is the current standard. The researchers are interested in finding which of the other fertilizers gives the biggest improvement over the current standard and reporting a confidence interval for the improvement. Which multiple comparison adjustment procedure would be most appropriate?

Dunnett

(1)

f) You find out that there was another factor involved in the study. The seedlings were distributed between four growing plots each located on a different farm. Does this raise any concerns for the above one way ANOVA? Justify your answer.

Yes, violates independence.

(2)

g) The one way ANOVA is robust to violation of one of it's assumptions. Which assumption, and under what conditions?

Normality

Large sample sizes

(2)

Total points for Question 3 out of (10)

Question Four - Simple Linear Regression

The dataset queen provides observations of the time spent at sources of pollen in seconds (`Duration`) and the proportions of pollen removed (`Removed`) by bumblebee queens pollinating a species of lily. The following R output comes from fitting a simple linear regression of `Removed` on `Duration`.

```
Call:
lm(formula = Removed ~ Duration, data = queen)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.130031   0.063335   2.053   0.0486 *
Duration    0.020547   0.004075   5.043  1.9e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1408 on 31 degrees of freedom
Multiple R-squared: 0.4506,   Adjusted R-squared: 0.4329
F-statistic: 25.43 on 1 and 31 DF,  p-value: 1.9e-05
```

a) Find a 95% confidence interval for the slope parameter.

$0.021 \pm 2.04\,(0.004)$

```
qt(0.975, 33) = 2.03
qt(0.975, 31) = 2.04
qt(0.975, 29) = 2.05
```

$(0.013,\ 0.029)$

(1)

b) Assuming the regression model is appropriate, write a three sentence statistical summary interpreting the **slope** estimate in the context of the study. Include a statement of evidence, point estimate and confidence interval.

For every one second increase in duration, the predicted average increase in proportion of pollen removed is .021.

There is strong evidence of a linear relationship b/w duration (seconds) and proportion of pollen removed ($\hat{\beta}_1 = .021$, $t_{31} = 5.043$, 2-sided p-value = .000019).

With 95% confidence, the average increase in proportion of pollen removed is b/w .013 and .029 for every one second increase in duration.

(3)

c) The R^2 is reported to be 0.45. How should this number be interpreted?

45% of the variation in the proportion of pollen removed is explained by the regression model on duration.

(1)

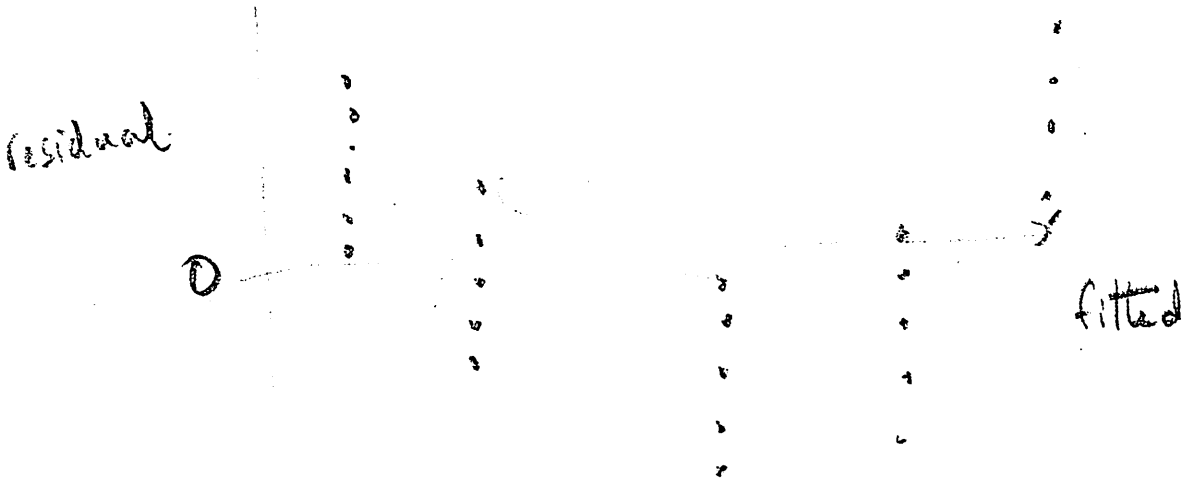
d) The researchers are interested in providing a range for the proportion of pollen a queen who spends 10 seconds at the source might remove. Would it be **more appropriate** to report a 95% confidence interval on the mean response, or a 95% prediction interval? Justify your answer.

Prediction, since they are interested in response (proportion of pollen) not mean response.

(1)

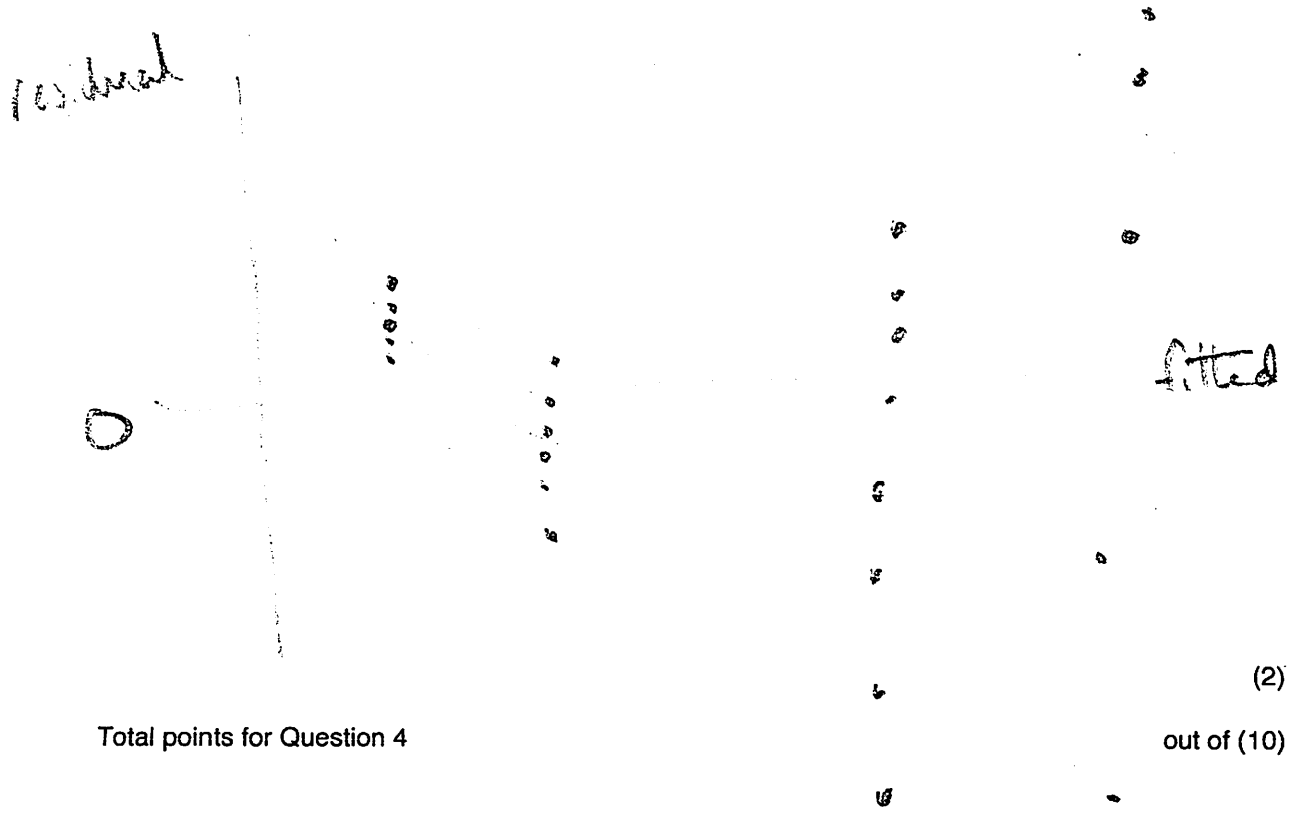
Question Four is continued on the next page...

e) Sketch an example of a residual plot that gives evidence of a violation of the linearity assumption, but shows **no evidence** against the constant spread assumption. Don't forget to label your axes.



(2)

f) Sketch an example of a residual plot that gives evidence of a violation of the linearity assumption, and **also shows evidence** against the constant spread assumption. Don't forget to label your axes.



(2)

Total points for Question 4 out of (10)