

STAT 3011
Fall 2022
Final Exam (C)
Time Limit: 120 Minutes

Name (Print): _____

Student ID: _____

Instructions:

- Do *not* begin or turn this page until you are instructed.
- Enter all requested information on the top and bottom of this page, and put your initials on the top of every page, in case the pages become separated.
- This exam contains 15 pages (including this cover page and the multiple choice answer sheet). Check to see if any pages are missing. There are 17 multiple-choice problems and 3 short-answer problems.
- The exam is closed book. **Do not** use your books, or any electronic devices on this exam.
- You may use a calculator and two sheets of paper (size A4 or 8.5" by 11") with formulas or other notes on both sides. **Do not** share calculators or notes!
- Show all your work on each problem for full credit except multiple choice problems. The following rules apply:
 - *Organize your work*, in a reasonably neat and coherent way, in the space provided. Work scattered all over the page without a clear order will receive very little credit.
 - *Mysterious or unsupported answers will not receive full credit* for short answer problems. A correct answer, unsupported by calculations, explanation, or algebraic work will not receive full credit; an incorrect answer supported by substantially correct calculations and explanations may still receive partial credit.
 - If you need more space, use the back of the pages; clearly indicate when you have done this.

Honesty Statement and Pledge:

I have not given or received any aid or assistance to or from any other student in this course during the exam period. Everything I have written on this exam represents my own work and knowledge. I sign this knowing that infringements on the University's Academic Honest policy may result in failure or expulsion.

Signed By: _____

Date: _____

Problem 1. (50 points) **Multiple Choice**

Choose ONLY ONE answer for each question. Circle your answers to all questions in the answer sheet provided on page 15. (NO explanation is needed).

1. (3 points) Suppose the University of Minnesota is interested in determining the proportion of students who would be interested in completing summer courses online, compared to on campus. A survey is taken over 100 students who intended to take summer courses. What is the best description of *100 students who intended to take summer courses*?

- (A) Population
- (B) Sample
- (C) Parameter
- (D) Statistic

2. (3 points) Let A and B be any events. Under what condition does the equation hold?

$$P(A \cup B) = P(A) + P(B)$$

- (A) This equation holds when A and B are independent.
- (B) This equation holds when A and B are disjoint.
- (C) This equation never holds.
- (D) This equation always holds.

3. (3 points) Toss a fair coin and roll a fair dice once each. What is the probability that {(the coin lands on a head), or (the dice comes up with a 1), or both}?

- (A) $2 / 3$
- (B) $7 / 12$
- (C) $3 / 4$
- (D) $7 / 8$

4. (3 points) We select 10 balls without replacement from a box, which contains 50 red balls, 30 blue balls, and 20 yellow balls. Let A be the event that the first selected ball is red, B be the event that the first selected ball is not yellow. Let X be the number of times that the selected ball is not yellow. Select the claim that is correct.

- (A) $X \sim Bin(100, 0.2)$
- (B) $X \sim Bin(10, 0.8)$
- (C) $P(A \cap B) = 0.5$
- (D) $P(A \cap B) = 0.4$

5. (3 points) Suppose we have a random variable $X \sim \text{Binm}(3, 0.8)$, what is the probability that X is greater than 2?
- (A) 0.896
 - (B) 0.384
 - (C) 0.512
 - (D) 0.488
6. (3 points) Indicate in which of the following cases the *Central Limit Theorem* will apply to describe the sampling distribution of the sample proportion.
- (A) $n = 100$ and $p = .035$
 - (B) $n = 80$ and $p = .05$
 - (C) $n = 400$ and $p = .28$
 - (D) All of the above because sample size is large.
7. (3 points) Let \bar{x} denote the sample mean from a sample of size $n = 9$ from a population with mean $\mu = 50$ and variance $\sigma^2 = 9$. Which of the following is **true** regarding the sampling distribution of \bar{x} ?
- (A) has a mean of 50 and a standard deviation of 1
 - (B) is normally distributed.
 - (C) Both A and B
 - (D) Neither A nor B
8. (3 points) The 95% confidence interval of a population proportion p is (0.2804, 0.3196). Which of the following statements are correct?
- i. The sample proportion is 0.3
 - ii. The margin of error is 0.0392
 - iii. The standard error is 0.01

Use the following R output if needed.

```
qnorm(0.95)
[1] 1.645
> qnorm(0.975)
[1] 1.96
```

- (A) i. only.
- (B) i. and ii.
- (C) i. and iii.
- (D) i., ii., and iii.

-
9. (3 points) In a one-sided hypothesis test with $H_a : p > 0.3$ and 81 observations, the p -value is 0.03. Which of the following R command produces the plausible value of the test statistic in a two-sided hypothesis test with $H_a : p \neq 0.3$ if we use the same sample?
- (A) `qt(0.94, df = 80)`
 - (B) `qt(0.97, df = 81)`
 - (C) `qnorm(0.94)`
 - (D) `qnorm(0.97)`
10. (3 points) Which of the following statements is NOT correct?
- (A) We should always set up the significance level before running the test.
 - (B) When p -value is greater than the significance level, we accept H_0 and conclude that the null hypothesis is true.
 - (C) If we fail to reject the null hypothesis at significance level 0.05, then it is possible that we will reject the null hypothesis at significance level 0.1.
 - (D) When a point estimate and the hypothesized value differ by 0.0001, their difference can still be statistically significant.
11. (3 points) The average time in years to get an undergraduate degree in computer science was compared for men and women. Random samples of 100 men in computer science and 100 women in computer science were taken.
Choose the appropriate parameter(s) for this situation.
- (A) One population proportion p
 - (B) Difference between two population means $\mu_1 - \mu_2$
 - (C) Mean of difference within pairs μ_D
 - (D) None of the above.
12. (3 points) Which of the following is *not* a property of r ?
- (A) r is always between -1 and 1.
 - (B) Switching roles of explanatory variable and response variable changes r .
 - (C) r does not depend on the units of y and x .
 - (D) r has the same sign as the slope of the regression equation.
13. (3 points) Owner of the large chain of coffee shops wanted to test if there are any significant differences in daily average sales (in USD) in three locations (downtown, uptown, and on Campus). Which of the following is the best method to use?
- (A) Chi-square test
 - (B) ANOVA
 - (C) Regression
 - (D) All of the above.

14. (3 points) Which of the following distributions can take on negative values?
- t -distribution with 6 degrees of freedom
 - Chi-squared distribution with 6 degrees of freedom
 - F -distribution with 10 and 45 degrees of freedom
- (A) i only
 (B) i and ii
 (C) i. and iii.
 (D) i., ii., and iii.
15. (3 points) Research found that there is a strong positive association between ice consumption and the number of deaths by drowning on a given day. The best explanation is the existence of what?
- (A) extrapolation
 (B) nonconstant variance
 (C) a lurking variable
 (D) a nonlinear relationship
16. (3 points) Suppose the population regression model is $\mu_x = \alpha + \beta x$ where μ_x is the mean of y given x . Read the following R outputs and choose the statement that is NOT correct.
- ```
> summary(lm(y ~ x))

Call:
lm(formula = y ~ x)

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.1616 0.8943 -0.181 0.857
x -0.5977 0.2828 -2.113 0.040 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.133 on 46 degrees of freedom
Multiple R-squared: 0.08851, Adjusted R-squared: 0.06869

> sd(x)
[1] 1.1
```
- (A) Suppose we want to use the data to test  $H_0 : \beta = 0$  vs.  $H_a : \beta > 0$ , then the p-value is 0.98.  
 (B) Let  $t_{0.025,46}$  be the t-multiplier used to construct a 95% confidence interval and degrees of freedom 46, then we know  $t_{0.025,46} < 2.113$ .  
 (C) Suppose we want to test whether the population mean of  $x$  is greater than 0, then the test statistic follows a  $t$  distribution with degrees of freedom 47.  
 (D) The sample standard deviation of  $y$  is 2.167077.

17. (2 points) Did you circle multiple choice answers on page 15?

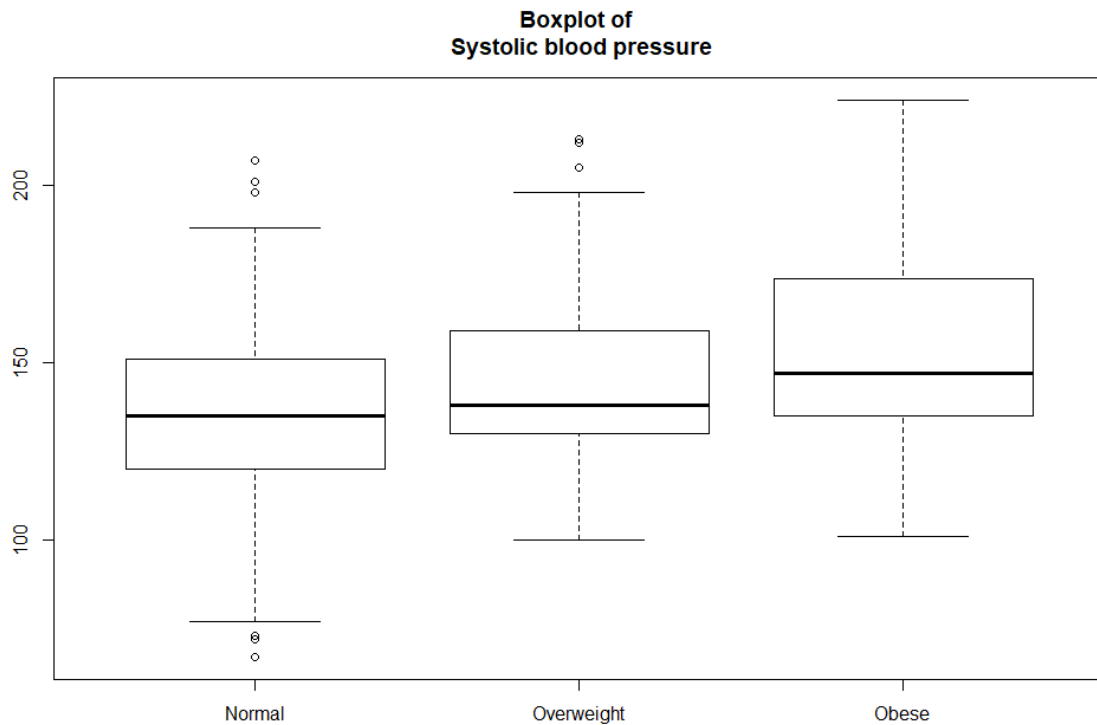
- (A) Yes, I did.
- (B) I will now.
- (C) I will now.
- (D) I will now.

**Problem 2.** (20 points) Be sure to show all work for full credit.

A person's systolic blood pressure can be a signal of serious issues in their cardiovascular system. We want to test whether there are significant differences between average systolic blood pressure based on weight. The dataset `dat` contains blood pressure and weight category information from a random sample of 500 adults.

Two variables are named :

- **SystolicBP:** Systolic blood pressure (measured in millimeters of mercury (mmHg))
- **Overwt:** 0 for normal, 1 for overweight, 2 for obese.



1. (4 points) State the assumptions of hypothesis testing. Determine whether each assumption is met or not. Briefly explain.

2. (5 points) State the null and alternative hypotheses. Remember to define parameters of interest.

3. (3 points) Fill in the blank of ANOVA table below. (a), (b), (c)

```
> summary(aov(dat$SystolicBP~dat$Overwt))
 Df Sum Sq Mean Sq F value Pr(>F)
dat$Overwt (a)___ 27801 (b)_____ 19.018 1.1e-08 ***
Residuals (c)___ 363274 730.93

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. (3 points) Based on R output above, use  $p$ -value and draw a conclusion in the context of the problem. If you don't know the  $p$ -value, make up a value and draw a conclusion. Use  $\alpha = 0.05$



5. (5 points) Determine which group means are statistically different, if any, at  $\alpha = 0.05$  using Tukey's HSD multiple comparisons. Interpret *one of* the confidence interval(s) that are statistically significant at  $\alpha = 0.05$ .

```
> TukeyHSD(m)
 Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = dat$SystolicBP ~ dat$Overwt)

 diff lwr upr p adj
1-0 8.051464 0.3927115 15.71022 0.0366867
2-0 16.865865 10.4316024 23.30013 0.0000000
2-1 8.814400 1.2740746 16.35473 0.0170703
```

**Problem 3.** (8 points) Be sure to show all work for full credit.

Is there an association between blood type and likely to be bitten by mosquito? That is what a medical entomologist in Japan wanted to know. He exposed 180 people to mosquitoes, and recorded their blood type and whether or not the mosquitoes bit their subjects. He is interested in determining whether there is an association between blood type and mosquito bites at a significance level  $\alpha = 0.05$ .

The contingency table of blood type and being bitten is given in the following table:

| Mosquito<br>Bites | Blood Type |    |    |    | Total |
|-------------------|------------|----|----|----|-------|
|                   | A          | B  | AB | O  |       |
| Yes               | 38         | 10 | 16 | 48 | 112   |
| No                | 41         | 9  | 8  | 10 | 68    |
| Total             | 79         | 19 | 24 | 58 | 180   |

1. (2 points) The partial table of the expected cell counts is shown in the following. Fill in a) b) c) and d). Round your answer to three decimal places.

| Mosquito<br>Bites | Blood Type |        |         |    | Total |
|-------------------|------------|--------|---------|----|-------|
|                   | A          | B      | AB      | O  |       |
| Yes               | a)_____    | 11.822 | b)_____ |    | 112   |
| No                | c)_____    |        | d)_____ |    | 68    |
| Total             | 79         | 19     | 24      | 58 | 180   |

2. (2 points) Write down the formula for computing the test statistic of the chi-squared test from the observed and expected cell counts. Fill in this formula with data from the single cell, **blood type B, and Yes, only**. No calculation is needed.

3. (2 points) What is the distribution of the test statistic if  $H_0$  is true?

4. (2 points) Compare the risk of mosquito bite for blood type A's to blood type B's. Use relative risk and interpret. Below is a copy of the table from the previous page.

| Mosquito<br>Bites | Blood Type |    |    |    | Total |
|-------------------|------------|----|----|----|-------|
|                   | A          | B  | AB | O  |       |
| Yes               | 38         | 10 | 16 | 48 | 112   |
| No                | 41         | 9  | 8  | 10 | 68    |
| Total             | 79         | 19 | 24 | 58 | 180   |

**Problem 4.** (22 points) Be sure to show all work for full credit.

Is there a relationship between the weight of a mountain bike and its price? A lighter bike is often preferred, but do lighter bikes tend to be more expensive? The data set `mtb` contains two variables `Price` (\$) and `Weight` (measured in LB, pound) for 12 brands.

Use the following R commands to answer questions.

Call:

```
lm(formula = mtb$Price ~ mtb$Weight)
```

Residuals:

```
 Min 1Q Median 3Q Max
-340.4 -262.5 -147.7 403.4 458.2
```

Coefficients:

```
 Estimate Std. Error t value Pr(>|t|)
(Intercept) 1895.91 1187.19 1.597 0.141
mtb$Weight -40.45 37.76 -1.071 0.309
```

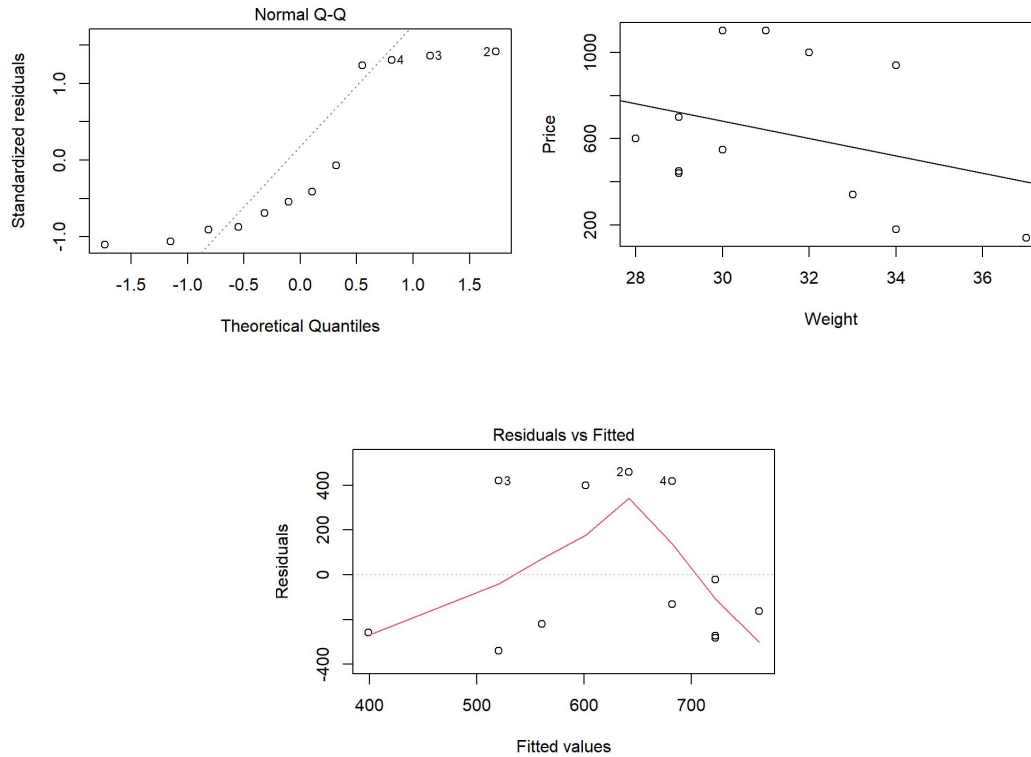
Residual standard error: 339.1 on 10 degrees of freedom

Multiple R-squared: 0.103, Adjusted R-squared: 0.01326

F-statistic: 1.148 on 1 and 10 DF, p-value: 0.3092

1. (5 points) Based on the R outputs provided above, a) write down the estimate regression equation. Remember to use correct notations. b) Interpret the slope in the context of the problem.





5. (6 points) Based on R output provided, conduct a hypothesis test for  $\beta \neq 0$ . Use  $\alpha = 0.05$ .

- Hypotheses
- Test statistic
- P-value
- Conclusion and interpretation.

Name: \_\_\_\_\_

**Lecture Section:** 001      006      0011      016      021  
**Lecture time:** 9:05 am   8:00 am   10:10 am   11:15 am   12:20 pm  
**(Circle One)**      Zhang      Yang      Park      Park      Park

| Question | Answer |   |   |   |
|----------|--------|---|---|---|
| 1        | A      | B | C | D |
| 2        | A      | B | C | D |
| 3        | A      | B | C | D |
| 4        | A      | B | C | D |
| 5        | A      | B | C | D |
| 6        | A      | B | C | D |
| 7        | A      | B | C | D |
| 8        | A      | B | C | D |
| 9        | A      | B | C | D |
| 10       | A      | B | C | D |
| 11       | A      | B | C | D |
| 12       | A      | B | C | D |
| 13       | A      | B | C | D |
| 14       | A      | B | C | D |
| 15       | A      | B | C | D |
| 16       | A      | B | C | D |
| 17       | A      | B | C | D |

Please do NOT write in the following table. This is for grading purpose only!

|          |   |    |     |    |     |
|----------|---|----|-----|----|-----|
| Question | I | II | III | IV | 100 |
| Score    |   |    |     |    |     |
| Total    |   |    |     |    | 100 |