

MA 180/418 Final Exam (Sample)

Student Name (PRINT):

Student Signature:

Use pencil, so that you can erase and rewrite if necessary.

The test consists of 20 questions.

Questions 1 through 10 are multiple-choice and worth 2.5 points each.

Questions 11 through 20 are computational and worth 7.5 points each.

For questions 1 through 10, circle the correct answer (**a**, **b**, **c**, or **d**) after each question.
Each question is 5 points. No partial credit.

Q1 What is the **standard normal distribution**?

- (a) The standard normal distribution is any bell-shaped probability distribution.
- (b) The standard normal distribution is a bell-shaped probability distribution with $\mu > 0$ and $\sigma = 1$.
- (c) The standard normal distribution is any bell-shaped probability distribution with $\sigma \neq 1$.
- (d) The standard normal distribution is a bell-shaped probability distribution with $\mu = 0$ and $\sigma = 1$.

Correct answer (circle one): (a) (b) (c) (d)

Q2 Under what circumstances could you use the **normal approximation to the binomial distribution**? Here we assume that the sample is a simple random sample of size n from a population in which the proportion of successes is p .

- (a) $np \geq 5$ and $nq \geq 5$.
- (b) $np \geq 5$ and $nq < 5$.
- (c) The sample standard deviation is less than 2.
- (d) The sample mean is at least n .

Correct answer (circle one): (a) (b) (c) (d)

Q3 What is a **critical value**?

- (a) A critical value is a calculated error margin of less than 5%.
- (b) A critical value is the degree of confidence when constructing a confidence interval.
- (c) A critical value is the length of a confidence interval.
- (d) A critical value is the number on the borderline separating sample statistics that are likely to occur from those that are unlikely to occur.

Correct answer (circle one): (a) (b) (c) (d)

Q4 In a test of hypothesis, what is the **critical region** (or **rejection region**)?

- (a) The critical region (or rejection region) is the set of negative z-scores.
- (b) The critical region (or rejection region) is the set of positive t-values.
- (c) The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the alternative hypothesis.
- (d) The critical region (or rejection region) is the set of all values of the test statistic that cause us to reject the null hypothesis.

Correct answer (circle one): (a) (b) (c) (d)

Q5 What is the **P-value** for a hypothesis test?

- (a) The P-value is two times the level of significance, α . That is, $P = 2\alpha$.
- (b) The P-value is the computed test statistic.
- (c) The P-value is the same as the critical value. For matched-pair (dependent) samples you construct a test of hypothesis.
- (d) The P-value is the probability of getting a value of the test statistic that is at least as extreme as the one representing the sample data, assuming the null hypothesis is true.

Correct answer (circle one): (a) (b) (c) (d)

Q6 What is the best description for **linear correlation coefficient**?

- (a) The linear correlation coefficient is the probability of a Type-I error when constructing a prediction interval.
- (b) The linear correlation coefficient times two is the coefficient of determination.
- (c) The linear correlation coefficient measures the strength of the linear correlation between two variables.
- (d) The linear correlation coefficient is the P-value for a multiple regression equation.

Correct answer (circle one): (a) (b) (c) (d)

Q7 What is a **regression equation**?

- (a) A regression equation is the inverse of a confidence interval.
- (b) A regression equation is a best-fit algebraic statement of the relationship between two variables.
- (c) A regression equation is what you get when the null hypothesis has been rejected.
- (d) A regression equation is used to calculate the P-value of a t-test (σ unknown).

Correct answer (circle one): (a) (b) (c) (d)

Q8 What is the meaning of the term **prediction interval**?

- (a) A prediction interval is an interval estimate of a predicted value of the dependent variable given some value of the independent variable.
- (b) A prediction interval is the interval between α and β in a test of hypothesis.
- (c) A prediction interval is like a confidence interval, only with a 100% confidence level.
- (d) A prediction interval is created when the sample size is at least 30 and the population distributions are not known to be normal.

Correct answer (circle one): (a) (b) (c) (d)

Q9 Which of the following is the most accurate statement about the relationship between **causality** and **correlation**?

- (a) Correlation implies causality.
- (b) Causality is the square of the correlation coefficient.
- (c) If two variables have high correlation, one causes the other or vice-versa.
- (d) You can have correlation without causality.

Correct answer (circle one): (a) (b) (c) (d)

Q10 What is the meaning of the term **coefficient of determination**?

- (a) The coefficient of determination is the amount of the variation in y that is explained by the regression line.
- (b) The coefficient of determination is the square root of the linear correlation coefficient.
- (c) The coefficient of determination is the sum of all the predicted y -values.
- (d) The coefficient of determination is the constant term in the regression equation.

Correct answer (circle one): (a) (b) (c) (d)

For questions 11 through 20, write your answer in the space provided. Show your work. Each question is worth 7.5 points.

Q11 In the accompanying table, the random variable x represents the number of televisions in a household in the US. Find the mean and standard deviation of the random variable x . Write down the corresponding formulas.

x	$P(x)$
0	0.02
1	0.13
2	0.28
3	0.27
4	0.17
5	0.13

Q12 For women, systolic blood pressures are normally distributed with a mean of 115.1 and a standard deviation of 23.2. Hypertension is commonly defined as a systolic blood pressure above 140.

(a) If a woman is randomly selected, find the probability that her systolic blood pressure is greater than 140.

(b) If 4 women are randomly selected, find the probability that their mean systolic blood pressure is greater than 140.

Q13 For a random sample of 100 people, the mean body temperature is $\bar{x} = 98.25^\circ\text{F}$ and the standard deviation is $s = 0.65^\circ\text{F}$.

(a) What is the best point estimate of the mean body temperature for the population?

(b) Construct a 90% confidence interval estimate of the mean body temperature for the population.

Q14 Weights (in grams) of 91 randomly selected M&M candies have standard deviation of 0.0512g. Construct a 95% confidence interval estimate of the standard deviation of weights of all M&Ms.

Q15 In a poll of 725 randomly selected adults, 573 said that it is morally wrong to not report all income on tax return. Use a 0.02 significance level to test the claim that 75% of adults say that it is morally wrong to not report all income on tax return. Use traditional method, draw a diagram, find critical value(s). State the final conclusion. *For an extra credit, find the P-value of the test.*

Q16 A statistician collected a random sample of the cents portion from 81 checks and from 99 credit card charges. The cents portions of the checks have a mean of 24.2 cents and a standard deviation of 33.0 cents. The cents portions of the credit card charges have a mean of 44.1 cents and a standard deviation of 32.2 cents. Construct a 90% confidence interval for the difference between the mean of the cents portion from checks and the mean of the cents portion from credit card charges.

Q17 Listed below are body temperatures (in F°) of six subjects measured at 8:00AM and then again at 6:00PM. Use a 0.05% significance level to test for a difference between the morning and evening measurements.

8:00 AM	97.1	96.2	97.7	96.4	98.3	99.9
6:00 PM	98.9	97.1	99.1	96.8	99.2	100.1

Compute the test statistic, draw a diagram, find critical value(s). State the final conclusion. *For an extra credit, find the P-value of the test by calculator or find an interval for the P-value by using a table.*

Q18 Use a significance level of $\alpha = 0.05$ to test the claim that weights of babies born to mothers given placebos vary more than weights of babies born to mothers given zinc supplements. Sample results are summarized below:

Placebo group:	$n = 16$	$\bar{x} = 3014$	$s = 721$
Treatment group:	$n = 16$	$\bar{x} = 3103$	$s = 662$

Compute the test statistic, draw a diagram, find critical value(s). State the final conclusion.

Q19 Listed below are the measurements of the blood pressure of five subjects taken on their right arm and left arm. Use the pressure in the right arm as the x variable and the pressure in the left arm as the y variable.

Right arm:	103	111	99	92	105
Left arm:	128	139	121	109	126

(a) Find the linear correlation coefficient. Use 0.05 significance level to test the hypothesis that there is a linear correlation between x and y

(b) Find the equation of the regression line

(c) Find the best predicted blood pressure in the left arm given that the blood pressure in the right arm is 100. (Show which formula you use and explain why.)

Q20 Use the data from Question 19 (above) to do the following. In each case write the formula that you use (you can copy the formula from the tear-out card).

(a) Find the coefficient of determination r^2

(b) Find the standard error of estimate s_e

(c) Find the total variation

(d) Find the explained variation

(e) Find the unexplained variation

[Bonus] Find a 95% prediction interval for the estimate of the blood pressure in the left arm given that the blood pressure in the right arm is 100.