Final Exam - Part I

Started: Apr 25 at 2pm

Quiz Instructions

This exam covers all materials in this course.

There are two types of problems---

True/false problems: Read each statement and decide whether it is true or false.

Multiple-choice problems: Select the answer that is the BEST in each case (may not necessarily be the perfect one).

Question 1 2 pts

A couple has two children. The probability that one is a girl and one is a boy if we know the younger is a girl is

- ○ 1/3
- ◉ 1/2
- ○ 1/4
- ○ 2/3

Question 2 2 pts

Two disjoint events must be independent of each other.

- ○ True
- ◉ False

Question 3 2 pts

A shipment of 20 laptop computers to a retail outlet contains 3 that are defective. A school makes a random purchase of 2 of these computers. The number of defective computers purchased by the school can be modeled by

- ○ Binomial distribution
- ○ Bernoulli distribution
- ◉ Hypergeometric distribution
- ○ Negative Binomial distribution

Question 4 2 pts

The time to failure of a component in an industrial system can be modeled by

- ○ Poisson distribution
- ◉ Exponential distribution
- ○ Gamma distribution
- ○ Uniform distribution

Question 5 2 pts

Which of the following are among the things you should consider when you make a plan for data analysis?

(1) What questions to address by the analysis

(2) What methods/techniques to use

(3) How data will be presented

(4) What to learn from the analysis

- ◉ All of them
- ○ (1), (2) and (3)
- ○ (1), (2) and (4)
- ○ (1) and (2)

Question 6 2 pts

If event A and event B are independent, then they must be disjoint.

- ○ True
- ◉ False

Question 7 2 pts

Z is the standard normal variable. The area between -1.97 and 0.86 under the PDF curve of the standard normal distribution is

- ○ 0.80511
- ○ 0.97558
- ◉ 0.78069
- ○ 0.02442

Question 8 2 pts

According to *Chemical Engineering Progress* (November 1990), approximately 30% of all pipework failures in chemical plants are caused by operator error. The number of pipework failures caused by operator error among the next 20 pipework failures can be modeled by

- (◉) Binomial distribution
- ( ) Poisson distribution
- ( ) Hypergeometric distribution
- ( ) Negative Binomial distribution

Question 9 2 pts

Z is the standard normal variable. $P(Z < 0) =$

- ( ) 0.32119
- (◉) 0.5
- ( ) 0
- ( ) 0.76573

Question 10 2 pts

Three different awards will be given to the staff of a mechanical engineering department. There are 5 staff people who are qualified for those awards. If each person can receive at most one award, the number of ways to assign the awards is

- ( ) 60
- (◉) 120
- ( ) 10
- ( ) None of the above

Question 11 2 pts

Roll a fair die until we see an even number. When we repeat this many times, the average number of rolls is

- (◉) 2
- ( ) 10
- ( ) 1
- ( ) 6

Question 12 2 pts

In time series data analysis, which of the following is not a purpose of differencing?

- ○ Eliminating linear trend
- ○ Eliminating seasonality
- ◉ Eliminating noises
- ○ Eliminating quadratic trend

Question 13 2 pts

Which of the following assumptions in linear regression is wrong?

- ○ The response y and the predictor x has a linear relationship.
- ○ The response y follows a normal distribution with mean 0 and a constant variance.
- ○ The response y has a constant variance for all values of x.
- ◉ Observations are independent.

Question 14 2 pts

For a Gamma distribution with Mean = 8 and Variance = 32, its parameters (α, β) are

- ◉ (2, 4)
- ○ (8, 4)
- ○ (4, 2)
- ○ (4, 4)

Question 15 2 pts

The sample mean is an unbiased estimator of the population mean because the expected value of the sample mean is equal to the population mean.

- ◉ True
- ○ False

Question 16 2 pts

Suppose the 95% confidence interval for the population mean μ is (1.5, 3.8) based on a sample of size 20. This means that the probability that μ falls in this interval is 0.95, that is, $P(1.5 \leq \mu \leq 3.8) = 0.95$.

- ○ True
- ◉ False

Question 17 2 pts

Which of the following statements about design of experiments is false?

- ○ With experimental error, an influential variable may be found to be not influential.
- ○ The first industrial applications of experimental design appeared in the 1930's.
- ◉ An observational study should be conducted rather than an experimental study if human subjects are involved.
- ○ In a manufacturing process, operator is an uncontrollable factor in experimental design.

Question 18 2 pts

For two independent random variables X and Y, $Var[X+Y]$ is equal to $Var[X]+Var[Y]$.

- ◉ True
- ○ False

Question 19 2 pts

Which of the following is not a correct set-up of hypotheses in hypothesis testing?

A. $H_0: \mu = 1$ vs. $H_1: \mu \neq 1$

B. $H_0: \mu = 1$ vs. $H_1: \mu > 1$

C. $H_0: \mu < 1$ vs. $H_1: \mu \geq 1$

D. $H_0: \mu \leq 1$ vs. $H_1: \mu > 1$

- ○ A
- ○ B
- ◉ C
- ○ D

Question 20 2 pts

When the population variance is known, which of the following should be used to find the confidence interval for the population mean?

- ○ t interval
- ◉ Z interval
- ○ ChiSquare interval
- ○ F interval

Question 21 2 pts

In a data table, each row represents one attribute of the object of interest and each column represents one instance of that object.

- ◉ True
- ○ False

Question 22 2 pts

When present a data analysis study, one should be willing to discuss limitations. Which of the following is not a limitation of a study?

- ○ A small sample size was used.
- ○ A low response rate on a survey
- ◉ Qualitative data were used.
- ○ A simple linear regression was used.

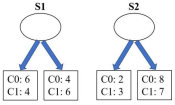
Question 23 2 pts

A doctor collected a dataset on patients who received his treatment during the past five years. The dataset contains two variables: each patient's age and outcome of treatment (1 = death, 0 = survival). He constructed a logistic regression model using the data, with $b_0 = -3.68, b_1 = 0.077$. What is the probability of death for a 50 years old patient based on this model?

- ◉ 0.54
- ○ 0.24
- ○ 0.05
- ○ 0.85

Question 24 2 pts

There are two ways (S1 and S2) to split a node, as shown in the figure below. Before the splitting, there are 10 records of class 0 and 10 records of class 1. The class distribution after splitting is given in the figure. Using Gini index as performance measure, which of the following statements is correct?



- S2 has a higher Gini index than S1, so S2 is a better split.
- S2 has a lower Gini index than S1, so S2 is a better split.
- S1 has a higher Gini index than S2, so S1 is a better split.
- (selected) S1 has a lower Gini index than S2, so S1 is a better split.

Question 25 2 pts

Which of the following is not among common objectives of designed experiments?

- (selected) Achieve desired nominal values of the influential variables
- Minimize the effects of uncontrollable variables on the response
- Find the most influential variables on the response
- Minimize variability of the response

Quiz saved at 2:38pm Submit Quiz