

Please enter all of your answers on these exam pages. There are 30 questions on this part.
A Defective Question Report, Formula sheet, and Tables will be handed out separately..

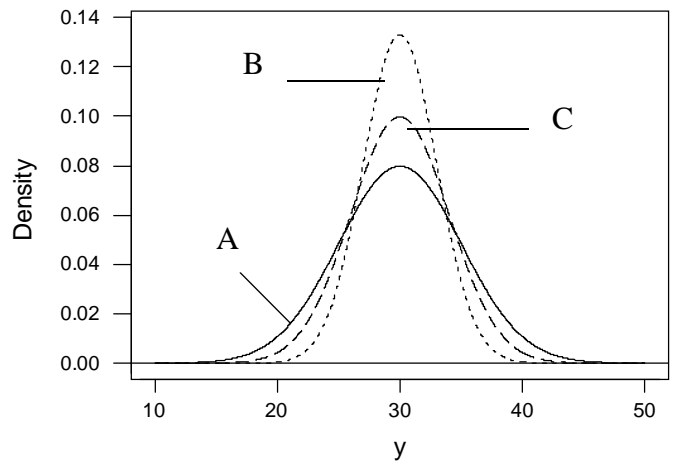
1. The least squares regression line is that line which makes the sum of the vertical distances between the observation pairs and the fitted line as small as possible.
A) True B) False
2. What is the area under the standard normal curve between the z -scores of -0.91 and $+1.3$?
A) 0.0968
B) 0.7218
C) 0.8186
D) 0.9154
E) None of the above.
3. A class has asked their instructor to “grade on the curve.” With this system the instructor is required to give preselected percentages of the various possible grades. In particular, the top 10% of the class **must receive** A’s. If exam scores are normally distributed with mean 82.0 and standard deviation 2.34, what exam score corresponds to the lowest A grade?
A) 70
B) 75
C) 80
D) 85
E) 90
4. Suppose y is a response variable of starting annual salary, x is a continuous predictor variable of years of education ($x = 12$ means high school graduate, $x = 16$ means college graduate, etc.), and z is an indicator variable indicating gender ($z = 1$ means female). Consider the regression model with equation:
 $\hat{y} = b_0 + b_1x + b_2z$. What combination of coefficients represents the predicted starting annual salary for a female college graduate?
A) $b_0 + b_1 + b_2$
B) $16b_1 + b_2$
C) $16b_1$
D) $b_0 + 16b_1$
E) $b_0 + 16b_1 + b_2$
5. In general, the smaller the value of the adjusted R^2 , the better the regression model.
A) True B) False

6. A salary distribution is strongly skewed toward the high values. The mean of the distribution is \$28,000 and the standard deviation is \$10,000. What percentage of the salaries are below \$38,000?
- A) 68%
 B) 82%
 C) 95%
 D) Essentially 100%
 E) Cannot be determined since the distribution of salaries is skewed.
7. A salary distribution is strongly skewed toward the high values. The mean of the distribution is \$28,000 and the standard deviation is \$10,000. A sample of 49 of the salaries is selected randomly and the average of these salaries calculated. If this random sampling were repeated many times, what percentage of the sample averages would be below \$38,000?
- A) 68%
 B) 82%
 C) 95%
 D) Essentially 100%
 E) Cannot be determined since the distribution of salaries is skewed.
8. We discussed a “managed” dice-roll process where we adjusted the current dice total based on the last dice total in an attempt to roll more sevens. This example illustrates:
- A) improving an out-of-control process.
 B) how to find common causes of variation.
 C) how to reduce the variation in an out-of-control process.
 D) how to increase variation in an in-control process.
 E) how to find special causes of variation.
9. **Replication** is important in experimentation since process results vary even under carefully controlled experimental conditions.
- A) True B) False
10. A portion of an ANOVA (analysis-of-variance) table from a multiple regression calculation is shown to the right. What is the value of s (the residual standard deviation) for this regression?

Source	SS	df
Regression	0.427	3
Residual Error		
Total	0.949	93

11. A simple experiment asked students to guess an instructor's age. The class was randomly divided into two equal parts. One half was given the hint that the instructor's oldest son graduated from the University of Iowa in 1985. The other half received no hint. How many factors does this experiment have?
- A) 1
 - B) 2
 - C) 3
 - D) 4
 - E) None of the above.

12. The graph on the right shows three normal curves labelled A, B, and C. One has standard deviation 3, another has standard deviation 4, and another has standard deviation 5. What is the correct order from *smallest* standard deviation to *largest* standard deviation?



- A) A, B, C
- B) B, C, A
- C) C, B, A
- D) A, C, B
- E) B, A, C

13. In a taste test of two recipes for pizza, a coin was flipped to decide which tasters got which recipe. This an example of:

- A) randomization
- B) replication
- C) blinding
- D) stratification
- E) placebo

14. Two variables x and y have an approximate straight-line relationship between them with a correlation coefficient of 0.8. If x were to increase by amount s_x how much would we predict that y would increase?

- A) 0.8
- B) s_y/s_x
- C) $0.8s_x$
- D) $0.8s_y$
- E) $0.8s_y/s_x$

15. In the Physicians' Health Study described in the video, 22,000 male doctors were used as subjects to measure the effect of aspirin in the prevention of heart attacks. By saying that the experiment was **double blinded** we mean that neither the doctors nor the evaluators knew whether the subjects received aspirin or placebo.

- A) True
- B) False

16. A student collects some observational data on her fellow students. For 100 students, she obtains values on two variables: College GPA and Average Weekly Hours Watching Television. She finds the correlation coefficient between the two variables to be +0.95. We can therefore conclude that, for her 100 students, the scatterplot of College GPA versus Average Weekly Hours Watching Television would show a strong upward trend.

- A) True B) False

17. There is a strong positive correlation between the number of highly specialized doctors at University Hospital and the proportion of patients that die at the hospital. This is an example of:

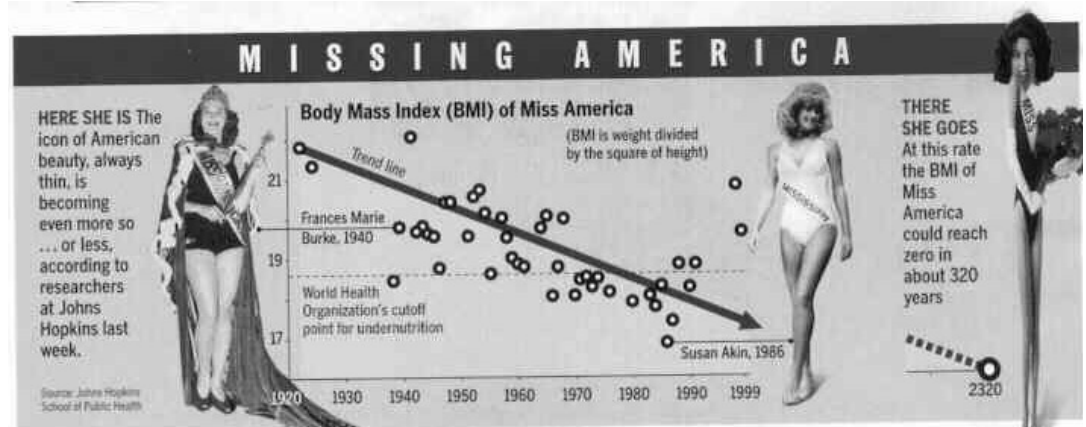
- A) a lurking variable
- B) ecological correlation
- C) cause and effect
- D) consistency
- E) mechanism

18. The graphic displayed below appeared in a recent *Time Magazine* dated April 3, 2000.

The far right says
 “THERE SHE GOES At this rate the BMI of Miss America could reach zero in about 320 years.”

This is an example of:

- A) extrapolation
- B) ecological correlation
- C) cause-and-effect
- D) a lurking variable
- E) None of the above.



19. In the multiple regression example given in the British Steel video, they used two continuous predictor variables (amount of carbon in the steel plate and amount of carbon in the welding wires) to predict the amount of carbon in the final weld. Their regression model is best described as

- A) one line
- B) two parallel lines
- C) one quadratic curve
- D) a plane
- E) None of the above.

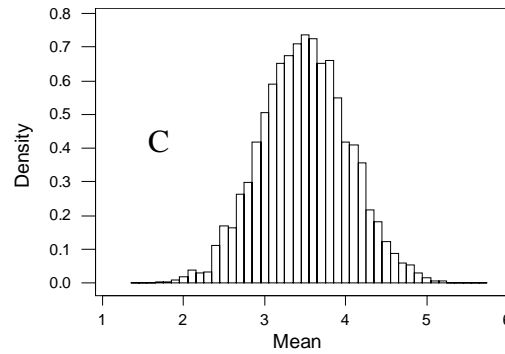
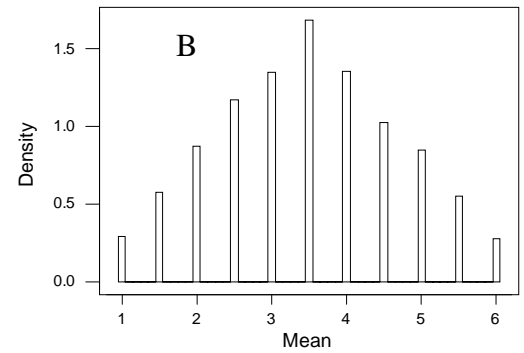
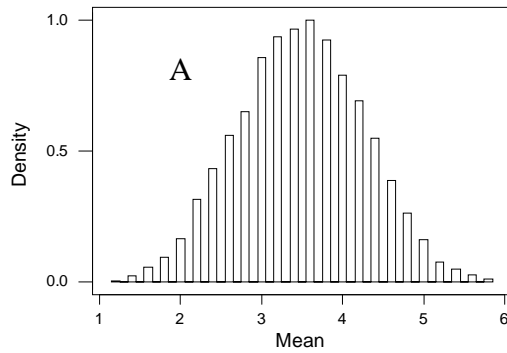
20. We used an example of measuring the variation in Attentiveness in Lecture. In that example “Aaron’s daydreaming about Ames, Iowa” would be considered a **common cause** of variation.

- A) True B) False

21. The results produced by a constant-cause system vary, and may vary over a wide band or a narrow band.

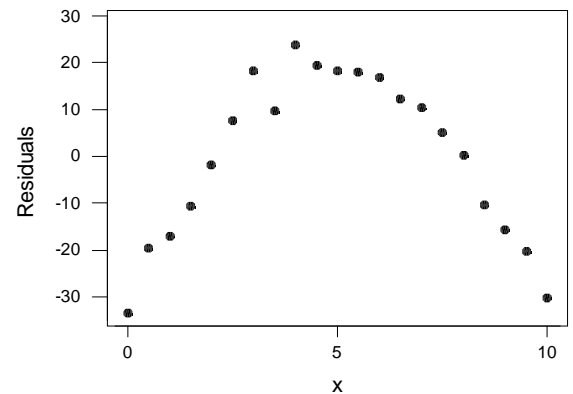
- A) True B) False

22. The three density histograms at the right display the results of finding average (mean) spots on 2, 5, or 10 dice tossed 6000 times. They are labelled A, B, and C. Which of the following is the correct order from 2 to 5 to 10 dice?



- A) A, B, C
- B) C, B, A
- C) B, A, C
- D) A, C, B
- E) B, C, A

23. The graph at the right shows a plot of residuals versus x from a fitted straight-line model. Which of the following best describes what we learn from this graph?



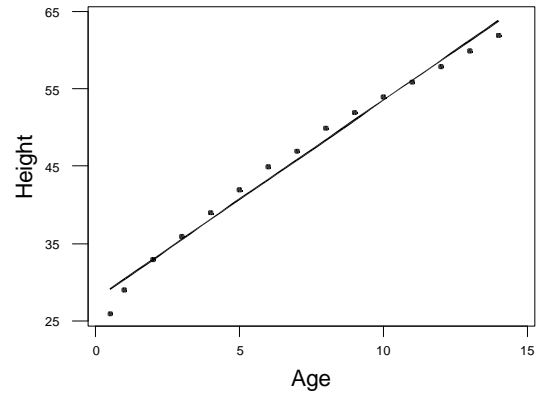
- A) Since the graph is mound-shaped the plot indicates that the straight-line model explains the relationship well.
- B) Since the residuals sum to zero the plot indicates that the straight-line model is a good fit for the relationship.
- C) Since about half of the residuals are positive and half negative the plot indicates that the straight-line model is a good fit for the relationship.
- D) Since the graph shows zero correlation between residuals and x the plot indicates that the straight-line model could be improved.
- E) Since the graph shows a systematic pattern in the residuals—negative then positive then negative—the plot indicates that the straight-line model could be improved—possibly with a quadratic model.

24. A basketball coach has an idea that training in ballet might help her players jump higher. To test this idea she divides her team randomly into two equal groups. One group is required to take six weeks of ballet training before the basketball season begins. The other group receives no ballet training. This study would be classified as

- A) a probability survey
- B) an observational study
- C) an experiment
- D) an observational survey
- E) None of the above.

25. One of the following is the equation of the least squares line for the data displayed at the right. Which one is it?

- A) Height = 10.3 + 2.75 Age
- B) Height = 28.8 + 4.01 Age
- C) Height = 37.8 + 2.37 Age
- D) Height = 20.5 + 1.22 Age
- E) Height = 27.8 + 2.57 Age



26. Researchers have observed that drinking red wine, x , seems to lead to fewer men having heart attacks, y . More recently, others have noted that drinking red wine leads to headaches and people with headaches tend to take aspirin. Furthermore, aspirin is known to reduce the chances of having heart attacks. Given these facts, the relationship between x and y would be best described as being due to

- A) cause-and-effect
- B) mechanism
- C) consistency
- D) strong correlation
- E) a lurking variable

27. A set of data pairs has the following summary statistics: $\bar{x} = 50$, $\bar{y} = 10$, $s_x = 4$, $s_y = 12$, and $r = 1/3$. What is the equation for the least squares regression line in **original terms** of x and y ? (Show your work.)

Answer: _____

28. The 381 students took a 35 question exam. The average score was 28.4 correct questions and the standard deviation was 2.6. Class policy is that you must get 88% or more correct to get an A. Here that means you must get 31 or more questions correct. The distribution of scores is approximately a normal curve. About how many students got A's on this test? (Show your work.)

Answer: _____

29. The times necessary to complete service for a class of bank customers is described by a normal distribution with mean 10 minutes and standard deviation 2 minutes. Service times are considered **excessive** if they exceed 15 minutes. Over the long run, what percent of customers will experience **excessive** service times? (Show your work.)

Answer = _____

30. An analysis of 27 pairs of data produced a least squares regression line with equation $\hat{y} = 23 + 3x$. (Show all of your work.)

a) The fitted value when $x = 3$ is _____

b) The residual when $x = 3$ and $y = 31$ is _____

Defective Question Report

Name: _____

Section: _____

If you believe that a test question is defective in some way, please list your complaint here. All complaints will be considered in our interpretation of the test results.

Form: A, B, C, D

Question number: _____ Your answer: _____

Your complaint:

Question number: _____ Your answer: _____

Your complaint:

Question number: _____ Your answer: _____

Your complaint:

Your Computer Number (01-28): _____ Your ID: _____

You are only to use Minitab for this exam. Do not open any other applications. Do not check your email, etc. Keep your eyes on your own screen please! Place all of your answers on this test sheet. **To avoid confusion you may want to close Worksheets and graphs when you go from one dataset to the next.**

31. Open the data worksheet named agehgt_b.mtw from the 22s:008 datasets. This data gives the height of young boys at various ages. Consider a model of the form $\hat{y} = b_0 + b_1 x$ to model y , the height in inches of young boys in terms of x , the age of the boy in years.

- A) What is b_0 ? _____
- B) What is b_1 ? _____
- C) What is the fitted value for $x = 7$? _____
- D) What is the residual for $x = 7$? _____
- E) What is the residual standard deviation? _____
- F) What does the model predict for the height of a 10 year old boy? Show your work.

Ans _____

32. For the same data, now consider a quadratic model of the form $\hat{y} = b_0 + b_1 x + b_2 x^2$.

- A) What is b_0 ? _____
- B) What is b_1 ? _____
- C) What is b_2 ? _____
- D) What is the fitted value for $x = 7$? _____
- E) What is the residual for $x = 7$? _____
- F) What is the residual standard deviation? _____
- G) What does the model predict for the height of a 10 year old boy? Show your work.

Ans _____

33. Which of the models in questions 31 and 32 do you prefer and why?

34. Open the Worksheet named `cars.mtw` from the 22s:008 datasets. These data give values on several variables for 45 cars. We will just look at TripMPG (in miles per gallon for driving on a Trip), Weight (in pounds), and Trans (transmission type with 1=automatic, 0=manual).

a) Use the Weight variable alone to “explain” the response variable, TripMPG.

The least squares regression line has equation: _____

The residual standard deviation for this model is _____ with _____ degrees of freedom.

The coefficient of determination for this model is: $R^2 =$ _____ % and the *adjusted* R^2 is _____ %.

b) Now fit a model that uses both Weight and Transmission Type (Trans) to explain TripMPG.

The least squares regression equation is: _____

The residual standard deviation for this model is _____ with _____ degrees of freedom.

The coefficient of determination for this model is: $R^2 =$ _____ % and the *adjusted* R^2 is _____ %.

How much does the Automatic Transmission “cost us” (in MPG) versus a Manual Transmission for a car of the same Weight? _____