Statistis STA 2033

Review for Test 2 (H)   Name_____

SHORT ANSWER. Write the word or phrase that best completes each statement or answers the question.

Draw th normal model and use it to answer the question.

1) The amount of Jen's monthly phone bill is normally distributed with a mean of $55 and a standard deviation of $8. Fill in the blanks. 68% of her phone bills are between $___ and $___.

1) _____

In a large class, the professor has each person toss a coin several times and calculate the proportion of his or her tosses that were heads. The students then report their results, and the professor plots a histogram of these several proportions. Use the 68–95–99.7 Rule to provide the appropriate response.

2) If the students toss the coin 80 times each, about 95% should have proportions between what two numbers?

2) _____

Find the specified probability.

3) The number of hours per week that high school seniors spend on computers is normally distributed, with a mean of 5 hours and a standard deviation of 2 hours. 70 students are chosen at random. Let $\bar{y}$ represent the mean number of hours spent on the computer for this group. Find the probability that $\bar{y}$ is between 5.1 and 5.7.

3) _____

Find the specified probability, from a table of Normal probabilities. Assume that the necessary conditions and assumptions are met.

4) Researchers believe that 7% of children have a gene that may be linked to a certain childhood disease. In an effort to track 50 of these children, researchers test 950 newborns for the presence of this gene. What is the probability that they find enough subjects for their study?

4) _____

5) According to Gallup, about 33% of Americans polled said they frequently experience stress in their daily lives. Suppose you are in a class of 45 students.

a. What is the probability that no more than 12 students in the class will say that they frequently experience stress in their daily lives? (Make sure to identify the sampling distribution you use and check all necessary conditions.)
b. If 20 students in the class said they frequently experience stress in their daily lives, would you be surprised? Explain, and use statistics to support your answer.

5) _____

6) The average composite ACT score for Ohio students who took the test in 2003 was 21.4. Assume that the standard deviation is 1.05. In a random sample of 25 students who took the exam in 2003, what is the probability that the average composite ACT score is 22 or more?

6) _____

Find the specified probability, from a table of Normal probabilities. Assume that the necessary conditions and assumptions are met.

7) A summer resort rents rowboats to customers but does not allow more than four people to a boat. Each boat is designed to hold no more than 800 pounds. Suppose the distribution of adult males who rent boats, including their clothes and gear, is normal with a mean of 195 pounds and standard deviation of 10 pounds. If the weights of individual passengers are independent, what is the probability that a group of four adult male passengers will exceed the acceptable weight limit of 800 pounds?

7) _____

Provide an appropriate response.

8) The weights of hens' eggs are normally distributed with a mean of 56 grams and a standard deviation of 4.8 grams. What is the probability that a dozen randomly selected eggs weighs over 690 grams?

8) _____

Answer the question appropriately.

9) A philosophy professor has found a correlation of 0.80 between the number of hours students study for his exams and their exam performance. During the time he collected the data, students studied an average of 10 hours with a standard deviation of 2.5 hours, and scored an average of 80 points with a standard deviation of 7.5 points. Create a linear model to estimate the number of points a student will score on the next exam from the number of hours the student studies.

9) _____

An article in the *Journal of Statistics Education* reported the price of diamonds of different sizes in Singapore dollars (SGD). The following table contains a data set that is consistent with this data, adjusted to US dollars in 2004:

2004 US $	Carat	2004 US $	Carat	2004 US $	Carat
494.82	0.12	688.24	0.15	748.10	0.16
768.03	0.17	944.90	0.18	1076.18	0.19
1105.03	0.20	1071.75	0.21	1289.20	0.23
1508.88	0.25	1504.44	0.26	1597.63	0.27
1826.18	0.28	1908.28	0.29	2038.09	0.32
2096.89	0.33	2409.76	0.35		

10) Make a scatterplot and describe the association between the size of the diamond (carat) and the cost (in US dollars).

10) _____

11) Create a model to predict diamond costs from the size of the diamond.

11) _____

12) Do you think a linear model is appropriate here? Explain.

12) _____

13) Interpret the slope of your model in context.

13) _____

14) Interpret the intercept of your model in context.

14) _____

15) What is the correlation between cost and size?

15) _____

16) Explain the meaning of R^2 in the context of this problem.

16) _____

17) Would it be better for a customer buying a diamond to have a negative residual or a positive residual from this model? Explain.

17) _____

1)
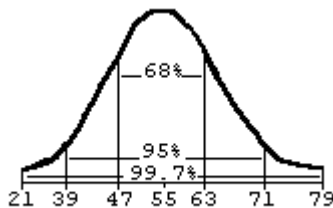
   47, 63

2) 0.39 and 0.62

3) 0.336

4) 0.9821

5) a. We want to find the probability that no more than 12 students in the class will say that they frequently experience stress. This is the same as asking the probability of finding less than 26.7% of "stressed" students in a class of 45 students.

Check the conditions:

1. 10% condition: 45 students is less than 10% of all students who could take the class

2. Success/failure cond.: $np = 45(0.33) = 14.85$, $nq = 45(0.67) = 30.15$, which both exceed 10

We can use the $N(0.33, \sqrt{\frac{(0.33)(0.67)}{45}} = 0.070)$ to model the sampling distribution.

We need to standardize the 26.7% and then find the probability of getting a z–score less than or equal to the one we find: $z = \dfrac{0.267 - 0.33}{0.070} = -0.90$

$P(\hat{p} < 0.267) = P(z < -0.90) = 0.1841$, so the probability is about 18.4% that no more than 12 students will say that they frequently experience stress in their daily lives.

b. From part a, we can use N(0.33, 0.070) to model the sampling distribution. Twenty students is about 44.4% of the class. This is about 1.63 standard deviations above what we would expect, which is not a surprising result.

6) Check the conditions:

1. Random sampling condition: We have been told that this is a random sample.

2. Independence assumption: It's reasonable to think that the scores of the 25 students are mutually independent.

3. 10% condition: 25 students is certainly less than 10% of all students who took the exam.

We're assuming that the model for composite ACT scores has mean $\mu = 21.4$ and standard deviation $\sigma = 1.05$. Since the sample size is large enough and the distribution of ACT scores is most likely unimodal and symmetric, CLT allows us to describe the sampling distribution of $\bar{y}$ with a Normal model with mean 21.4 and $SD = (\bar{y}) = \dfrac{1.05}{\sqrt{25}} = 0.21$.

An average score of 22 is $z = \dfrac{22 - 21.4}{0.21} = 2.86$ SDs above the mean.

$P(\bar{x} > 22) = P(Z > 2.86) = 0.0021$, so the probability that the average composite ACT score for a sample of 25 randomly selected students is 22 or more is 0.0021.

7) 0.023

8) $P\left(\bar{y} > \dfrac{690}{12}\right) = P\left(z > \dfrac{57.5 - 56}{4.8/\sqrt{12}}\right) = P(z > 1.08) = 14\%$
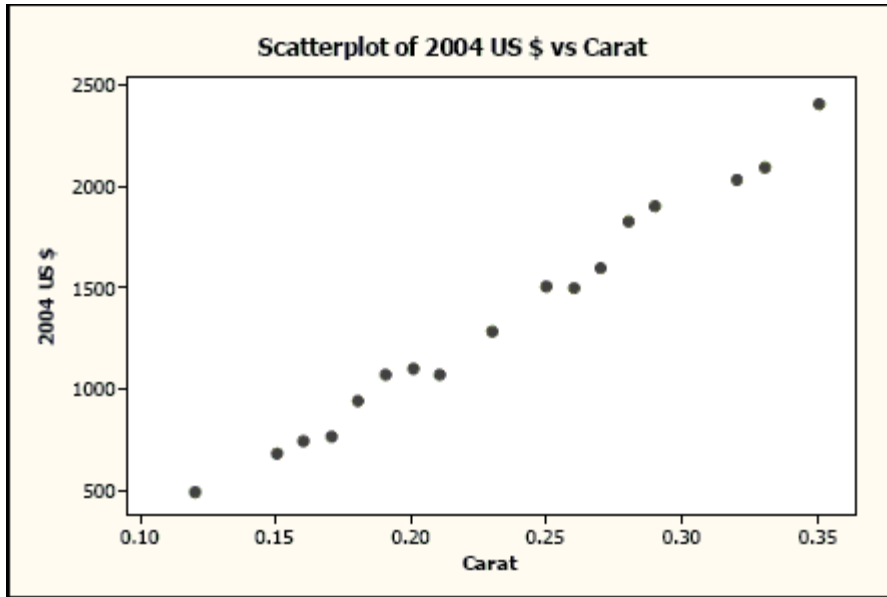
9) Explanatory variable: number of hours spent studying

Response variable: score on exam

slope: 2.40   intercept: 56   Model: $\widehat{Score} = 56 + 2.4(Hours)$

10) There is a strong, positive, linear association between the size of the diamond and its cost. The cost of a diamond increases with size.


Scatterplot of 2004 US $ vs Carat

11) The regression equation is
2004 US $ = –559 + 8225 Carat

Predictor	Coef	SE Coef	T	P
Constant	–558.52	57.88	–9.65	0.000
Carat	8225.1	239.1	34.40	0.000

S = 64.9355   R–Sq = 98.7%   R–Sq(adj) = 98.7%

Predicted cost = –558.52 + 8225.1(carat)

12) A linear model is appropriate for this problem. The residual plot shows no obvious pattern.

Residuals Versus the Fitted Values
(response is 2004 US $)
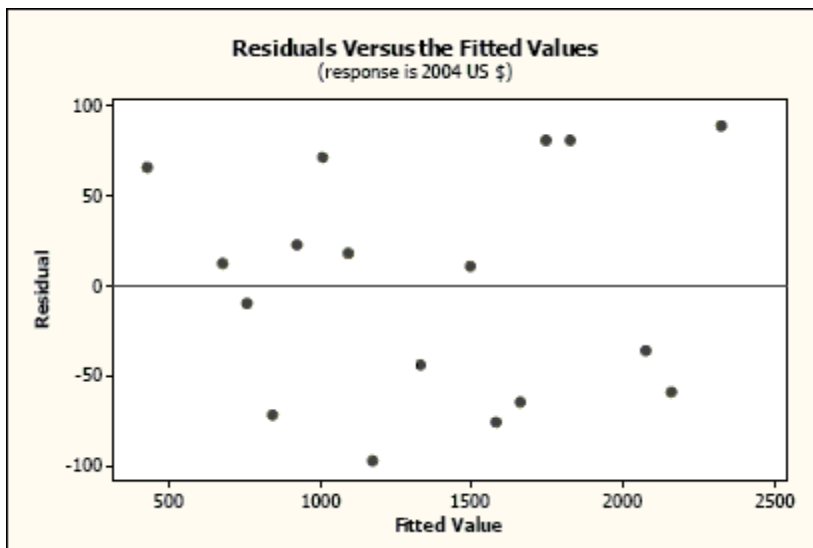
13) The slope of the model is 8225.1. The model predicts that for each additional carat, the cost of the diamond will increase by $8225.10, on average. This can also be interpreted as for each additional 0.01 carat, the cost of the diamond will increase by $82.251, on average.

14) The intercept of the model is –558.52. The model predicts that a diamond of 0 carats costs –$558.52. This is not realistic.

15) The correlation, r, is $r = \sqrt{0.987} = 0.993$. Since the scatterplot shows a positive relationship, the positive value must be used.

16) $R^2 = 0.987$. So 98.7% of the variation in diamond prices can be accounted for by the variation in the size of the diamond.

17) It would be better for customers to have a negative residual from this model, since a negative residual would indicate that the actual cost of the diamond was less than the model predicted it to be.