Statistics Exam

NAME:____

Part I – Multiple Choice. Each problem is worth 4 points.

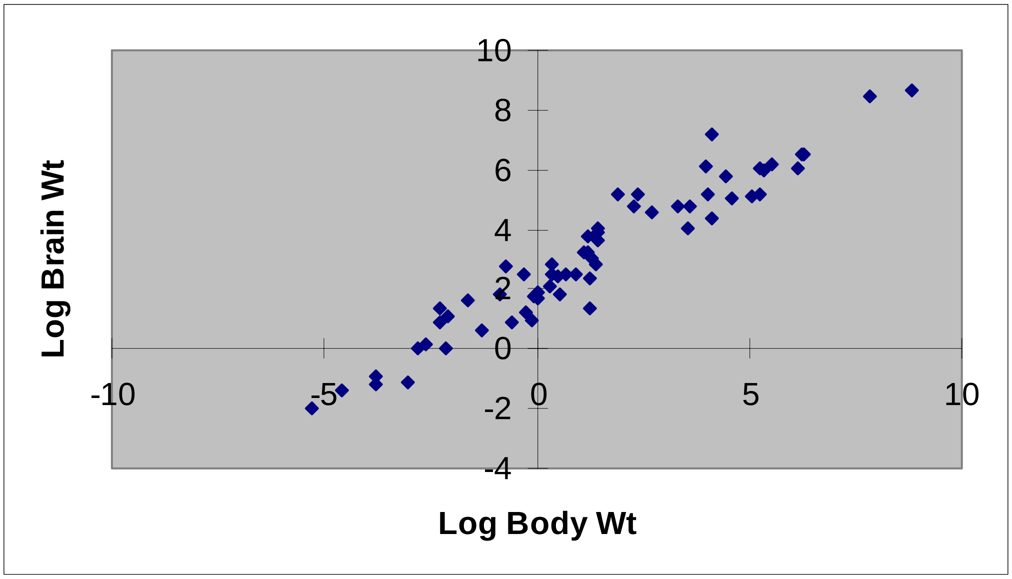
1. Ten pairs of chicks were selected to test the effect of a vitamin supplement on early growth. The chicks in each pair were siblings of high birth weight. One chick in each pair was given the supplement and the other was not. After two weeks, the weight of each chick was recorded. The researcher would like to test the research hypothesis that the supplement increases the growth rate of chicks in the first weeks after hatching against the null hypothesis that it has no effect. The most appropriate test would be
   (a) The normal test
   (b) The paired-sample Student-t test
   (c) The Student-t test for two independent samples
   (d) The chi-squared test of homogeneity
   (e) The F test for one-way analysis of variance.

Answer: b

2. An astronomer collects data on the distance from Earth and the velocity of recession of several dozen galaxies. Which of the following statements is not true?
   (a) This is an example of an observational study and not a designed experiment.
   (b) The astronomer should carefully consider whether this data should be treated as a random sample.
   (c) Techniques of linear regression are useful both for observational studies and designed experiments.
   (d) A significant correlation between these two variables would establish a causal relationship between them.
   (e) Statistics has made many contributions to modern astronomy.

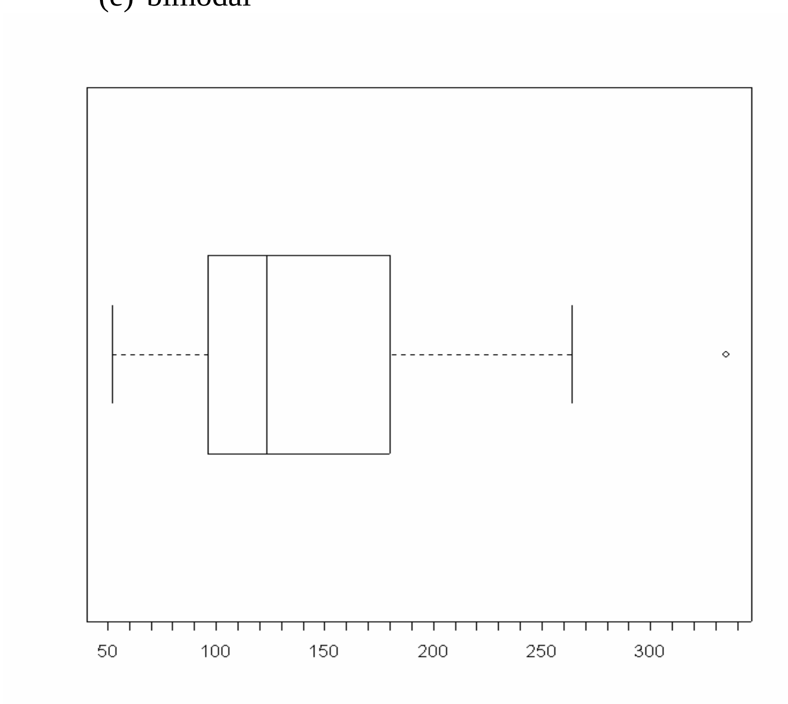
Answer: d

3. The scatter diagram below shows the <u>logarithm</u> of body weight on the horizontal axis and the <u>logarithm</u> of brain weight on the vertical axis for 58 vertebrate species. Based on this data, it would be reasonable to model the relationship between body weight *x* and brain weight *y* as
   (a) linear: $y = a + bx$, where *a* and *b* are constants,
   (b) exponential: $y = ce^{kx}$, where *c* and *k* are constants,
   (c) a power function: $y = ax^r$, where *a* and *r* are constants,
   (d) a reciprocal relationship: $1/y = a + bx$, where a and b are constants,
   (e) a trigonometric function: $y = a\sin(bx + c)$, where *a*, *b*, and *c* are constants.



Answer: c

4. The box and whisker diagram below indicates that the distribution of the data is:
   (a) normal
   (b) heteroscedastic
   (c) skewed
   (d) symmetric
   (e) bimodal



Answer: c

5. Which of the following is not generally true?
   (a) The mean and median are both measures of the center of a distribution.
   (b) The standard deviation and interquartile range are both measures of scale.
   (c) The median is less affected by outliers than is the mean.
   (d) For normal distributions, the interquartile range is about 1.35 times the standard deviation.
   (e) Outliers are usually caused by mistakes and should be removed from the data.

Answer: e

Part II – Free Response. Each problem is worth 8 points. Show your work.

6. It is known that about 20% of all income tax returns have errors. Government officials would like to know that percentage more precisely, to two significant digits. They intend to survey a sample of tax returns. How large a sample should they take to achieve the desired accuracy with 90% confidence?

Answer:

$$n > z^2_{\alpha/2} \frac{\hat{p}(1-\hat{p})}{\varepsilon^2}$$

$$z_{\alpha/2} = 1.645$$

$$\hat{p} = 0.2$$

$$\varepsilon = 0.005$$

$$n > 17319$$

7. Professor A says he deserves a raise because his students have an exceptionally high success rate on the bar exam. The dean compared the success rates of 40 randomly selected students from Professor A's classes and 40 randomly selected students from his own classes. Twenty four of Professor A's group passed the exam and 19 of the students from the dean's classes passed. Find a 95% confidence interval for the difference between the success rates of Professor A and the dean.

Answer:

$$\hat{p}_A - \hat{p}_B \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$$

$$\hat{p}_A = 24/40 = 0.6$$

$$\hat{p}_B = 19/40 = 0.475$$

$$n_A = n_B = 40$$

$$z_{\alpha/2} = 1.96$$

0.125 ± 0.217

8. A random sample of size 16 is taken from the normal distribution with mean 20 and variance 4. Let $\overline{X}$ denote the sample average. What is the probability that $\overline{X}$ is larger than 19?

Answer:

$$P[\overline{X} > 19] = P[\frac{\sqrt{16}(\overline{X} - 20)}{\sqrt{4}} > \frac{4(19-20)}{2}]$$

$$= P[Z > -2] = P[Z < 2] = 0.9772$$

9. Scores on a nationwide professional qualifying exam are normally distributed and have a population mean of 600 and a population standard deviation of 50. An examinee scored 615 on the exam. At what percentile was this examinee's score?

Answer:

$$z = \frac{615 - 600}{50} = 0.3$$

$$P[Z \le z] = P[Z \le 0.3] = 0.618$$

61.8th percentile or 62nd percentile.

10. The stem and leaf diagram below shows exam grades of 69 statistics students. The stem is the first digit of a student's score and the leaf is the second digit. What is the third quartile of grades on this exam? A 1 point error is acceptable.

Stem|Leaves

5 | 778
6 | 1122334
6 | 555666777799999
7 | 001123333344
7 | 5555666777779
8 | 00222244
8 | 5567788999
9 | 1

Answer: 80

11. A six-sided die is tossed 180 times with the following results.

Face	1	2	3	4	5	6
Frequency	39	51	20	22	26	22

Is there strong evidence that the die is not fair? Use a significance level of 10%.

Answer:

$$\chi^2 = \sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{m} \frac{(O_i - np_i)^2}{np_i}$$

$m = 6; n = 180$

$p_1 = p_2 = \cdots = p_6 = 1/6$ (under H_0: The die is fair)

Degrees of freedom for $\chi^2 = m - 1 = 5$.

Critical value: $\chi^2_{0.10}(5) = 9.24$.

Observed value: $\chi^2 = 25.53$.

Very strong evidence that the die is not fair.

12. For June through September of 1999, the rainfall and yield per acre of cotton were measured for 8 farms in Texas. The results are tabulated below. Find a 95% confidence interval for the expected cotton yield of a farm that receives 15 inches of rain.

Rain	Yield
8.2	17.2
6.4	17.6
12.8	19.6
7.6	19.8
13.8	20.0
16.1	23.5
12.4	20.2
12.8	20.2

Answer:

Use calculator to calculate least squares estimates of the slope and intercept, or use formulas 8 and 9. Either way, $\hat{\mu}(x) = 14.59 + 0.46x$, where x is rainfall.

$\hat{\mu}(15) = 14.59 + 0.46(15) = 21.49$

$$s^2 = \frac{SS(resid)}{n-2} = 1.3806$$

$\bar{x} = 11.26$

$S_{xx} = 82.299$

$n = 8$

$t_{\alpha/2}(n-2) = t_{.025}(6) = 2.45$

$$\hat{\mu}(x) \pm t_{\alpha/2}(n-2)\, s\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} = 21.49 \pm 1.56$$

13. A binomial experiment is based on 100 trials and an unknown success probability p. Let X be the number of successes. The null hypothesis is H_0: p = 0.5 and the alternative hypothesis is H_1: p = 0.6. H_1 is accepted if X > 58. Find the probabilities of type 1 and type 2 error. Use the normal approximation and the continuity correction.

Answer:

$$P[Type\,1\,error] = P_{H_0}[X > 58.5] = P[\frac{X-50}{5} > 1.7] = P[Z > 1.7] = 0.045$$

$$P[Type\,2\,error] = P_{H_1}[X < 58.5] = P[\frac{X-60}{\sqrt{100(.6)(.4)}} < \frac{58.5-60}{4.90}] = P[Z < -.31] = 0.378$$

14. An assembly line produces widgets with a mean weight of 10 and a standard deviation of 0.200. A new process supposedly will produce widgets with the same mean and a smaller standard deviation. A sample of 20 widgets produced by the new method has a sample standard deviation of 0.126. At a significance level of 10%, can we conclude that the new process is less variable than the old? State appropriate null and alternative hypotheses.

Answer: Assume a normal distribution.

$H_0 : \sigma^2 = (0.2)^2 = \sigma_0^2$

$H_1 : \sigma^2 < (0.2)^2$

Reject $H_0 : \sigma^2 = \sigma_0^2$ in favor of $H_1 : \sigma^2 < \sigma_0^2$ if $s^2 < \sigma_0^2 \dfrac{\chi^2_{1-\alpha}(n-1)}{n-1}$

$n = 20$.

Critical value: $\chi^2_{1-\alpha}(n-1) = \chi^2_{0.90}(19) = 27.20$

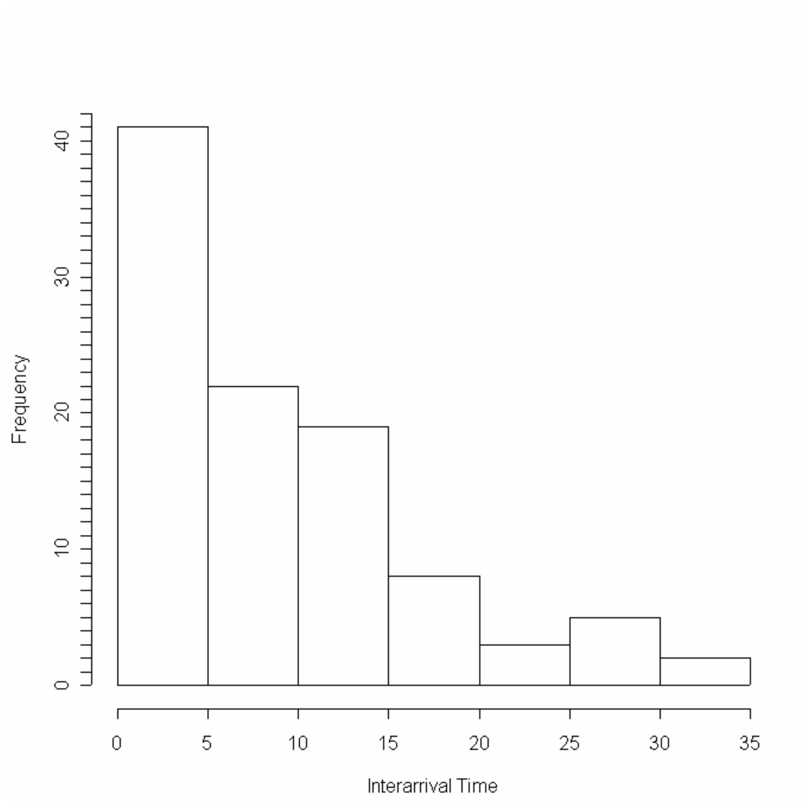
Reject H_0 and accept H_1 if the observed value of s^2 is < 0.057.

Since s^2 (observed) = 0.126^2 = 0.0159, accept H_1.

15. The figure below is a histogram of 100 times between successive customer arrivals in a store. Estimate the mean of the data.

Answer: Use the midpoint of each class interval to represent all the observations in that interval.

$$\bar{x} \approx (41(2.5) + 22(7.5) + 19(12.5) + 8(17.5) + 3(22.5) + 5(27.5) + 2(32.5))/100 = 9.15$$



Frequency
0
10
20
30
40
0
5
10
15
20
25
30
35
Interarrival Time