University of Toronto Scarborough
STAB22 Final Examination

April 2009

For this examination, you are allowed two handwritten letter-sized sheets of notes (both sides) prepared by you, a non-programmable, non-communicating calculator, and writing implements.

This question paper has 17 numbered pages. Before you start, check to see that you have all the pages. You should also have a Scantron sheet on which to enter your answers, and a set of statistical tables. If any of this is missing, speak to an invigilator.

This examination is multiple choice. Each question has equal weight, and there is no penalty for guessing. To ensure that you receive credit for your work on the exam, fill in the bubbles on the Scantron sheet for your correct student number (under "Identification"), your last name, and as much of your first name as fits.

Mark in each case the best answer out of the alternatives given (which means the numerically closest answer if the answer is a number and the answer you obtained is not given.)

If you need paper for rough work, use the back of the sheets of this question paper. The question paper will be collected at the end of the examination, but any writing on it will not be read or marked.

Before you begin, two more things:

- Check that the colour printed on your Scantron sheet matches the colour of your question paper. If it does not, get a new Scantron from an invigilator.

- Complete the signature sheet, but *sign it only when the invigilator collects it*. The signature sheet shows that you were present at the exam.

1. A simple random sample of 70 coffee drinkers was taken. Each sampled coffee drinker was asked to taste two unmarked cups of coffee, one of which is actually Brand A and the other is Brand B, and was asked which one they preferred. 44 coffee drinkers preferred Brand A, and the other 26 preferred Brand B. Use this information to answer this question and the next one.

   The people who commissioned the survey are trying to find out whether a majority of all coffee-drinkers (that is, more than 50% of them) prefer brand A. Carry out a suitable test of significance to assess the evidence. What is the P-value of your test?

   (a) between 0.05 and 0.10
   (b) between 0.025 and 0.05
   (c) greater than 0.10
   (d) between 0.01 and 0.025
   (e) less than 0.01

2. In the situation of Question 1, it turned out that the coffee drinkers had always been given an unmarked cup containing Brand A coffee first, and Brand B coffee second. How do you react to this knowledge?

   (a) The P-value was small, so there will still be good evidence that Brand A is preferred.
   (b) The coffee drinkers received their coffee in unmarked cups, so it doesn't matter which brand is actually tasted first.
   (c) Brand A must have an advantage by being tasted first, so there cannot be a significant difference between brands A and B.
   (d) The P-value was large, so there will still be no evidence that Brand A is preferred.
   (e) A better approach would have been to toss a coin to decide whether each drinker gets Brand A first or Brand B first.

3. A university financial aid office took a simple random sample of students to see how many of them were employed the previous summer. Of the 750 men sampled, 703 had been employed the previous summer; of the 650 women sampled, 592 had been employed the previous summer. Use this information for this question and the next one.

   Test whether the proportion of all male students employed last summer is different from the proportion of all female students employed last summer. What is the P-value of your test?

   (a) between 0.05 and 0.10
   (b) between 0.01 and 0.025
   (c) between 0.025 and 0.05
   (d) less than 0.01
   (e) greater than 0.05

4. From the information given in Question 3, what is the upper limit of a 95% confidence interval for the difference between the proportion of men employed the previous summer and the proportion of women employed the previous summer? (Take the difference as men minus women.)

   (a) 0.05
   (b) -0.10
   (c) 0.10
   (d) 0.00
   (e) -0.05

5. A report states that 3% of all births are "multiple births" (that is, twins, triplets, etc.). A random sample of 5000 births is taken. Assuming the report to be correct, what is the probability that 2.5% or less of the births in the sample are multiple births?

   (a) less than 0.02
   (b) between 0.05 and 0.10
   (c) greater than 0.30
   (d) between 0.02 and 0.05
   (e) between 0.10 and 0.30

6. A population has a mildly non-normal shape with mean 35 and standard deviation 7.5. A sample of size 225 is taken from this population. What is the probability that the sample mean is less than 33.5?

   (a) about 0.001
   (b) 0.05
   (c) 0.42
   (d) more than 0.90
   (e) 0.12

7. A simple random sample of 30 observations is taken from some population. The sample mean is 120. You wish to test the null hypothesis $\mu = 116$ against the alternative $\mu \neq 116$, but you have only the output below.

```
N     Mean  SE Mean         95% CI
30  120.000    1.826  (116.422, 123.578)


N     Mean  SE Mean         99% CI
30  120.000    1.826  (115.297, 124.703)
```

   What can you say about the P-value of your test?

   (a) It is less than 0.01.
   (b) It is greater than 0.95.
   (c) It is less than 0.05.
   (d) It is greater than 0.05.
   (e) It is less than 0.05 but greater than 0.01.

8. It is desired to estimate the mean of a population using a 95% confidence interval. The population is known to have a highly skewed shape. A sample of size 45 is proposed. What do you think of this choice of sample size?

   (a) the sample size may not be large enough to allow the Central Limit Theorem to apply
   (b) the sample size is bigger than 30, so the calculation should be accurate
   (c) the sampling distribution of the sample mean has exactly a normal shape regardless of the shape of the population
   (d) the normal distribution will be an accurate approximation to the t distribution in this case
   (e) The Law of Large Numbers says that the sample mean and population mean will be close, so there is no need to calculate a confidence interval

9. A student's mark on a Psychology exam has a normal distribution has a normal distribution with mean 65 and SD 10. The same student's mark on a Chemistry exam has a normal distribution with mean 60 and SD 15. The two marks are independent of each other. Use this information for this question and the next two.

   What is the probability that the student scores over 80 on both exams?

   (a) between 0.01 and 0.10
   (b) between 0.20 and 0.50
   (c) between 0.10 and 0.20
   (d) less than 0.01
   (e) greater than 0.50

10. What is the probability that the student scores over 80 on exactly one of the two exams?

    (a) between 0.01 and 0.10
    (b) between 0.20 and 0.50
    (c) between 0.10 and 0.20
    (d) less than 0.01
    (e) greater than 0.50

11. What is the probability that the student's mean mark in the two courses is over 80?

    (a) between 0.10 and 0.20
    (b) greater than 0.50
    (c) between 0.01 and 0.10
    (d) between 0.20 and 0.50
   (e) less than 0.01

12. The time it takes (for any student) to complete a STAB22 final exam is a random variable having a normal distribution with mean 160 minutes and standard deviation 15 minutes. Use this information for this question and the next one.

    Anne and Bob are two friends writing this exam. What is the probability that Anne will complete the exam at least ten minutes before Bob completes the exam? (Assume that all students start the exam at the same time and that their completion times are independent.)

    (a) this probability is greater than 0.45 but less than 0.55
    (b) this probability is less than 0.01
    (c) this probability is greater than 0.35 but less than 0.45
    (d) this probability is greater than 0.01 but less than 0.35
    (e) this probability is greater than 0.55

13. Using the information given in Question 12, if 300 students are writing the exam, what is the approximate probability that more than 275 students will complete the exam in less than 180 minutes? Choose the closest value from the options below.

    (a) 0.75

    (b) 0.00

    (c) 1.00

    (d) 0.50

    (e) 0.25

14. In a certain city, the probability that a call to a randomly-chosen telephone number will reach a live person is 0.3. Suppose 10 such calls are made. What is the probability that at least one of them reaches a live person?

    (a) less than 0.10

    (b) between 0.50 and 0.90

    (c) greater than 0.90

    (d) between 0.30 and 0.50

    (e) between 0.10 and 0.30

15. An experiment was conducted to compare a new drug with a standard drug with respect to mean recovery time. It is known that the mean recovery time for the standard drug is 26 days. Following the experiment, which involved 65 patients, a 95% confidence interval estimate was constructed for the mean recovery time (in days) for patients on the new drug. The 95% confidence interval was found to be from 24.6 to 27.8. Which of the following statements is a valid conclusion based on this information?

    (a) The experimenter should reject the claim that the new drug is the same as the standard drug with respect to mean recovery time.

    (b) There is evidence, at the 5% level of significance, that the new drug is better than the standard drug.

    (c) There is insufficient evidence, at the 1% level of significance, to reject the claim that there is no difference between the new drug and the standard drug with respect to mean recovery time.

    (d) There is sufficient evidence, at the 5% level of significance, to reject the claim that there is no difference between the new drug and the standard drug with respect to mean recovery time.

    (e) None of the other statements is a valid conclusion based on the information from this study.

16. Following the analysis of a well-designed completely randomized experiment it was reported that the observed effect was "statistically significant". Which of the following statements best explains the meaning of the phrase "statistically significant"?

    (a) The observed result made sense to the experimenter since it was what was hoped would happen.

    (b) The observed effect happened because the experiment was properly designed and carried out without bias.

    (c) The experimenter carefully employed the basic principles of experimental design in conducting the study.

    (d) The observed effect was sufficiently large so that it would rarely occur simply by chance.

    (e) The laws of probability say that this observed result would be expected to happen by chance.

17. During a recent year, the average cost of making a movie was $58 million. A simple random sample of 15 action movies was taken during that year; the sampled action movies cost a mean of $62.5 million to make, with a sample standard deviation of $9.5 million. Is this evidence that action movies cost more to make, on average, than movies in general? Use $\alpha = 0.05$.

   (a) The P-value is small, so we can conclude that action movies cost more to make on average.
   (b) The P-value is small, so we have evidence that action movies cost the same to make as movies generally.
   (c) The P-value is not small, so we cannot conclude that action movies cost more to make on average.
   (d) The P-value is not small, so we have evidence that action movies cost the same to make as movies generally.

18. The histogram below shows the scores of 77 players in a golf tournament. Use the histogram for this question and the one following.
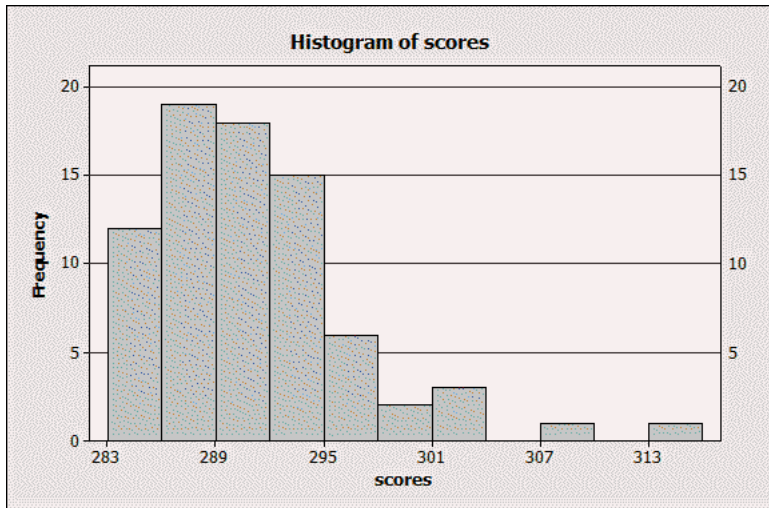
   What is the median score for these golfers?

   (a) between 289 and 292
   (b) between 286 and 289
   (c) cannot tell from a histogram
   (d) between 289 and 292
   (e) greater than 292

19. From the histogram in Question 18, how does the mean score compare with the median?

   (a) cannot estimate mean from a histogram
   (b) the mean score is about the same as the median
   (c) the mean score is bigger than the median
   (d) the mean score is smaller than the median

20. The distribution of the heights of students in a large class is approximately normal. Moreover, the average height is 68 inches, and approximately 16% of the students were taller than 71 inches. Based on the 68-95-99.7 rule, approximately what percentage of students in this class are taller than 65 inches?

    (a) 2.5%
    (b) 84%
    (c) 5%
    (d) 16%
    (e) 95%

21. A company producing a certain item has two machines: Machine A and Machine B. Machine A is an old machine and 10% of the items produced by this machine are defectives. Machine B is a new machine and 5% of the items produced by this machine are defectives. A random sample of two items was selected from each machine for quality control purposes. What is the probability that the sample from machine B will have MORE defective items than the sample from machine A? Choose the closest answer from the options below.

    (a) 0.15
    (b) 0.18
    (c) 0.25
    (d) 0.01
    (e) 0.08

22. In an experiment, the soles of boys' shoes are made of two different synthetic materials, A and B. To see which material is better (lasts longer), measurements were made on the amount of wear of the soles of shoes worn by 10 boys. Each boy wore a special pair of shoes - the left sole was made with material A and the right sole with material B.

    Based on this information, which one of the following statements is true?

    (a) The decision as to whether the left or the right sole is to be made with material A or B should be determined by using a table of the standard normal distribution.
    (b) The type of material is the response variable, with two levels, A and B.
    (c) None of the above other statements is true.
    (d) This is an example of a completely randomized design.
    (e) Increasing the sample size (number of replications) is one way to eliminate the need for any randomization in assigning the materials.

23. The amount of money spent by a customer at a discount store is a random variable with mean $100 and standard deviation $30. What is the approximate probability that a randomly selected group of 50 shoppers will spend a total of more than $5300? Choose the number closest to the answer from the options given below.

    (a) 0.20
    (b) 0.10
    (c) 0.50
    (d) 0.30
    (e) 0.40

24. According to a report, the mean salary for mayors in all Canadian cities is \$120,000. A newspaper believes that this figure is not correct, and wishes to assess the evidence against it. A reporter at the newspaper takes a simple random sample of 10 Canadian cities, and finds that the mean salary for mayors in those cities is \$132,200. Use this information for this question and the following one.

What would a suitable *null hypothesis* for a test of significance say?

(a) The sample mean salary is equal to \$132,200.

(b) The mean salary of all mayors is equal to \$120,000.

(c) The mean salary of all mayors is not equal to \$120,000.

(d) The sample mean salary is not equal to \$132,200.

(e) The mean salary of all mayors is greater than \$120,000.

25. Using the information in Question 24, what would a suitable *alternative hypothesis* for a test of significance say?

(a) The mean salary of all mayors is not equal to \$120,000.

(b) The sample mean salary is not equal to \$132,000.

(c) The mean salary of all mayors is equal to \$120,000.

(d) The sample mean salary is equal to \$132,200.

(e) The mean salary of all mayors is greater than \$120,000.

26. A fair 10-sided die has faces numbered from 1 to 10. The probability of rolling a 1 on any roll is 0.1. If the die is rolled 12 times, what is the value of r such that the probability of rolling r 1's or less is 0.8891?

(a) there is no value of r that works

(b) 2.5

(c) 2

(d) 5

(e) 1

27. In a population, 45% of people have type O blood, 40% have type A, 11% have type B, and 4% have type AB. Use this information for this question and the next one.

Consider an accident victim with type B blood. She can only receive a transfusion from a person with type B or type O blood. There are three people from this population who are willing to donate blood, but they don't know what their blood type is. Assuming that the blood types of the three individuals are independent, what is the probability that at least one of them will be a suitable donor? Choose the closest answer from the following options.

(a) 0.9

(b) 0.6

(c) 0.7

(d) 0.8

(e) 0.5

28. Using the distribution of blood types given in Question 27, what is the probability that both people in a couple will have the SAME blood type if we assume that one partner's blood type does not influence the blood type of the other partner? Choose the closest answer from the following options.

   (a) 0.1
   (b) 0.3
   (c) 0.2
   (d) 0.5
   (e) 0.4

29. A researcher examined the relationship between crimes and other demographic variables for 206 of the most populous counties in the US. The MINITAB output below is a part of the statistical analysis of this data. The variables in the output are:

   CR   Crime rate (Number of crimes divided by total population)
   PVTY   percent below poverty level
   HSC   percent high school graduates

```
Correlations: CR, PVTY, HSC


          CR     PVTY
PVTY   0.573
       0.000


HSC   -0.293  -0.572
       0.000   0.000



Cell Contents: Pearson correlation
               P-Value


Regression Analysis: CR versus PVTY

The regression equation is
CR = 0.0294 + 0.00235 PVTY

Predictor       Coef    SE Coef      T      P
Constant    0.029387   0.001670  17.59  0.000
PVTY       0.0023498  0.0002356   9.97  0.000


S = 0.00896965   R-Sq = 32.8%   R-Sq(adj) = 32.5%
```
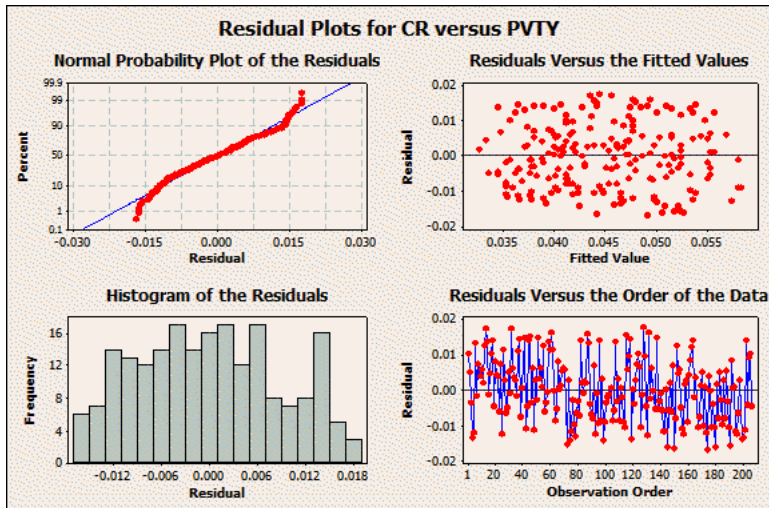
Residual Plots for CR versus PVTY

Regression Analysis: CR versus HSC

The regression equation is
CR = 0.0952 − 0.000630 HSC

```
Predictor          Coef     SE Coef       T      P
Constant        0.09521     0.01155    8.24  0.000
HSC          -0.0006303   0.0001442   -4.37  0.000


S = 0.0104614    R-Sq = omitted   R-Sq(adj) = omitted
```

Use this information for this question and the next one.

Which of the following statements is true?

(a) The linear regression of CR on PVTY explains more than 50% of the variability in CR.

(b) Using the regression of CR on PVTY, the residual for an observation with PVTY = 3.7 and CR = 0.05 is greater than 0.005.

(c) The linear regression of CR on HSC explains more than 25% of the variability in CR.

(d) None of the other four statements is true.

(e) The MINITAB output for the regression analysis of CR versus PVTY above indicates the need for a higher order (curved) model rather than a linear model.

30. Using the information in Question 29 above, calculate the slope of the regression line for predicting PVTY from HSC). Choose the closest value from the options below.

(a) −0.5

(b) −0.3

(c) −0.2

(d) −0.4

(e) −0.1

31. A car magazine commissions a study to determine whether single drivers do more driving for pleasure than married drivers. The magazine collects samples of 35 single drivers and 35 married drivers, and records how many kilometres per week of pleasure driving each driver does. Some Minitab output from an analysis is shown below:

```
Two-Sample T-Test and CI: single, married

Two-sample T for single vs married

          N   Mean  StDev  SE Mean
single   35  193.3   27.0      4.6
married  35  190.4   24.7      4.2


Difference = mu (single) - mu (married)
Estimate for difference:  2.82857
95% CI for difference:  (-9.50924, 15.16639)
T-Test of difference = 0 (vs not =): T-Value = 0.46   P-Value = 0.649   DF = 67
```

Using the output as given, what conclusion should the car magazine draw from its study, using $\alpha = 0.05$ if necessary?

(a) There is evidence that the amount of pleasure driving done by single and married people is different on average.

(b) There is no evidence that single people do more pleasure driving than married people, because the P-value is about 0.325.

(c) There is evidence that single people do more pleasure driving than married people, on average.

(d) Single people do about 15 kilometres per week of pleasure driving more than married people do.

(e) There is no evidence that single people do more pleasure driving than married people, because the P-value is about 0.649.

32. A carnival game offers a $100 cash prize for anyone who can break a balloon by throwing a dart at it. It costs $15 for each throw. John is willing to spend up to $45 trying to win. That means he is going to stop if he breaks the balloon on or before the 3rd throw and stop at the 3rd throw even if he failed to break it on the 3rd throw.

The chance that John will break the balloon in any throw is 0.1. Assume that breaking the balloon on any attempt is independent of what happens on the other attempts.

John's net gain is the amount of money he won minus the total cost. Denote this net gain by X. Which of the following numbers is the closest to the mean of X?

(a) $20

(b) $10

(c) $0

(d) −$20

(e) −$10

33. In a study of the effects of college student employment on academic performance, the researchers analyzed the GPAs of a random sample of students who were employed (denoted Emp) and a random sample of students who were not employed (denoted NotEmp). Some MINITAB output obtained from this study is given below. In the questions below, μ_{Emp} denotes the population mean GPA of all students employed and μ_{NotEmp} denotes the population mean GPA of all students not employed.

```
Descriptive Statistics: Emp, NotEmp

Variable   N     Mean    StDev SE Mean  Minimum      Q1  Median      Q3  Maximum
Emp       55   2.8734   0.4294  0.0579   1.9257  2.5987  2.8761  3.2340   3.6547
NotEmp    65   3.0224   0.2717  0.0337   2.3223  2.8569  3.0286  3.2633   3.4923
```


Probability Plot of Emp, NotEmp
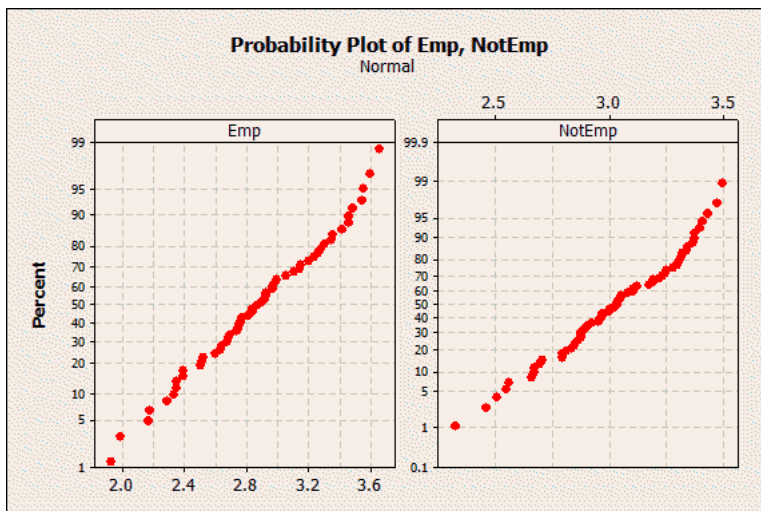
Use this information for this question and the following two questions.

Which of the following statements is true?

(a) Look at the sample of employed students. At least 25% of these students have a GPA equal to or below 2.6000.

(b) None of the other four statements is true.

(c) Look at the distribution of the GPAs in the sample of students who were not employed. This distribution is right skewed.

(d) Look at the sample of students who were employed. In this sample, more than 15 students have a GPA of 3.25 or higher.

(e) Look at the maximum observed GPA of the sample of employed students. According to the $1.5 \times IQR$ criterion, this value is an outlier.

34. Using the information in Question 33 above, which of the following statements is true?

   (a) Consider the 90% confidence interval for μ_{Emp}, the population mean GPA of students who were employed. The margin of error of this confidence interval is greater than 0.11.

   (b) Based on the on the information given in Question 33, we see that t-procedures cannot be applied for this data set.

   (c) Consider the t-test for testing the null hypothesis $\mu_{Emp} = 3.0$ against the alternative hypothesis $\mu_{Emp} < 3.0$. The P-value of this test is less than 0.01.

   (d) None of the other four statements is true.

   (e) Consider the t-statistic for testing the null hypothesis $H_0 : \mu_{Emp} = 3.0$ against the alternative hypothesis $H_a : \mu_{Emp} < 3.0$. This test statistic is between -2.00 and 2.00.

35. Consider the information given in Question 33 above. If we were interested in testing the null hypothesis that the population mean GPAs of the two groups were equal, against the alternative that the means were different, what could we say about the P-value for this test?

   It must be

   (a) between 0.02 and 0.05

   (b) between 0.05 and 0.10

   (c) between 0.01 and 0.02

   (d) greater than 0.10

   (e) less than .01

36. A random variable X might have one of two probability distributions. The null hypothesis is that X has this distribution:

Value	0	1	2	3
Probability	0.50	0.40	0.08	0.02

The alternative hypothesis is that X has this distribution:

Value	0	1	2	3
Probability	0.10	0.30	0.35	0.25

A test of significance is carried out by observing one value of X, and rejecting the null hypothesis in favour of the alternative if the observed value is 2 or larger. Use this information for this question and the following one.

What is the probability of a Type I error for this test?

   (a) 0.60

   (b) cannot determine it, because we need to know the exact value that was observed.

   (c) 0.05

   (d) 0.08

   (e) 0.10

37. In Question 36, a test of significance was described for the distribution of a random variable X. What is the probability of a Type II error for this test?

   (a) 0.35

   (b) 0.40

   (c) 0.60

   (d) 0.05

   (e) 0.10

38. When would you prefer to use t procedures (confidence interval, test of significance) instead of z procedures?

    (a) When you want to obtain a smaller confidence interval for the same sample size.

    (b) When you have a small sample.

    (c) When the population standard deviation is not known.

    (d) When the population distribution is approximately normal.

    (e) When you have a large sample.

39. The costs of major surgery can vary substantially from one place to another. A study of the costs involved in a particular surgery was done in California and Montana. The 95% confidence intervals for the mean costs in the two states were reported to be from $5826.76 to $6173.24 in Montana and from $6061.41 to $6338.59 in California. No other information was given in the report. Assume that these intervals were calculated based on the normal distribution (i. e. assume that both the sample sizes were very large).

    Calculate the upper limit of the 95% confidence interval for the difference between the two population means (i.e. mean cost in California minus mean cost in Montana). Choose the closest answer from the following options.

    (a) $420

    (b) $375

    (c) $340

    (d) $200

    (e) $360

40. A 90% confidence interval for a population mean, based on a sample of size 30, is from 40 to 65. This confidence interval turned out to be too big. Which of the following is a way of making the confidence interval shorter?

    (a) Make the sample size bigger.

    (b) Use the median instead of the mean.

    (c) Make the sample size smaller.

    (d) Choose a higher level of confidence.

41. A magazine is considering the launch of an online edition. In a small survey, the magazine contacted a random sample of 500 current subscribers and asked whether they would be interested in an online edition. Some of these subscribers showed an interest and some did not. Based on the information from this sample, the investigators tested the null hypothesis $H_0 : p = 0.3$ against $H_a : p > 0.3$, where p is the population proportion of subscribers who will be interested in an online edition. The value of the Z-statistic for testing $H_0 : p = 0.3$ against $H_a : p > 0.3$ was 5.86.

    If we were interested in testing $H_0 : p = 0.4$ against $H_a : p > 0.4$ based on the same sample used above, what could we say about the value of the Z-statistic for this test? This value is:

    (a) between 1.5 and 3.0.

    (b) between 3.0 and 4.5.

    (c) greater than 4.5.

    (d) between 0.5 and 1.5

    (e) less than 0.5.

42. A simple random sample of 20 observations is taken from a population of successes and failures where the proportion p of successes is believed to be about 0.8. In the sample, 15 successes are observed. Which of the following statements best applies to this situation? ("Standard procedures" means what we learned in lectures.)

    (a) The probability of observing exactly 15 successes is 0.1746.

    (b) The sample size is too small to use standard procedures to obtain a confidence interval for p.

    (c) There should have been exactly 16 successes observed in the sample.

    (d) A 95% confidence interval for p is from 0.56 to 0.94.

43. A government agency is interested in whether air quality in Ontario cities is changing. The agency made a study: 10 Ontario cities are randomly selected, and the air quality in those cities is recorded on a date in 2007, and also the same date in 2008. A lower figure indicates better air quality. Summaries of the data collected are shown below ("difference" is year 2 minus year 1):

```
Variable      N    Mean    StDev      Q1   Median     Q3
year1        10   28.00    9.98    20.00   29.50   36.00
year2        10   38.90    9.07    29.50   43.00   46.50
difference   10   10.90    9.75     3.50   14.00   17.50
```

Carry out a suitable test of significance to determine whether mean air quality has changed. You may assume normality as necessary to carry out your test. What P-value do you get?

    (a) between 0.005 and 0.01

    (b) between 0.02 and 0.05

    (c) greater than 0.05

    (d) between 0.01 and 0.02

    (e) less than 0.005

44. A random variable X has this distribution:

Value	1	3	5
Probability	0.1	0.6	0.3

X has mean 3.4 and SD 1.2. A sample of 36 values is taken from this distribution. What is the approximate probability that the sample mean is bigger than 3.9?

    (a) 0.38

    (b) 0.30

    (c) 0.25

    (d) 0.006

    (e) 0.34

45. An experimenter has decided to use a small value of α like $\alpha = 0.01$ in her experiment. Which of the following is a consequence of her decision?

    (a) the probability of a Type II error is small.

    (b) the probability of a Type I error is large

    (c) there is a small chance of rejecting the null hypothesis when it is true

    (d) if the null hypothesis is false, the experimenter will easily be able to reject it

    (e) even if the null hypothesis is far from being true, it will be hard to reject it.

46. A nutrition laboratory tested a random sample of 50 "reduced sodium" hot dogs. The mean sodium content of the sample was 309mg. Let μ denote the population mean sodium content. A test was carried out to test the null hypothesis $H_0 : \mu = 300$ against $H_a : \mu > 300$. The P-value of this test was 0.038. A test was also carried out of the null hypothesis $H_0 : \mu = 298$ against $H_a : \mu > 298$ (using the same sample); the P-value was 0.015. Assume that the data satisfy all assumptions required for the tests.

If we calculate a 95% confidence interval for μ using this sample, what can we say about its margin of error? Choose the correct range for this margin of error from the following alternatives.

(a) between 12.00mg and 16.00mg

(b) between 8.00mg and 12.00mg

(c) greater than 16.00mg

(d) between 4.00mg and 8.00mg

(e) less than 4.00mg

47. A random variable X has this probability distribution:

Value	1	3	5
Probability	0.1	0.3	0.6

What is the variance of X?

(a) 2.8

(b) 3.3

(c) 1.8

(d) 1.2

(e) 4.0

48. Let Z be a standard normal random variable and let T be a random variable with a t distribution with 5 degrees of freedom. Which of the following events is most likely (i.e. has the highest probability) to occur?

(a) Z and T are both greater than 2.0

(b) Z is less than -2.0

(c) T is greater than 2.0

(d) Z is greater than 2.0

(e) Z is greater than or equal to 2.0

49. A 99% confidence interval for a population mean goes from 27 to 33. The interval was based on a sample size of 50. The interval was calculated using a known population standard deviation, but unfortunately the value has been lost. What is the population standard deviation?

(a) 10.8

(b) 24.6

(c) 8.2

(d) 1.2

(e) 16.4

50. An insurance company is trying to estimate the mean number of sick days that food service workers take per year. The company wishes to use a 90% confidence interval. The standard deviation of the number of sick days is known to be 2.5. How big a sample does the insurance company need to take if they wish to estimate the mean to within 0.5 days?

    (a) 42
    (b) 30
    (c) 8
    (d) 68

51. An association of Christmas tree growers in Indiana commissioned a survey, in which a simple random sample of all Indiana residents was taken. One of the questions in the survey was "Did you have a Christmas tree during the last holiday season"? 400 people were surveyed, and 331 answered "yes". What is the upper limit of a 99% confidence interval for the proportion of all Indiana residents who had a Christmas tree during the last holiday season?

    (a) 0.83
    (b) 0.86
    (c) 0.80
    (d) 0.88
    (e) 0.78

52. A regression was carried out to predict a variable y from a variable x. The residuals from the regression were plotted against x, as shown below:
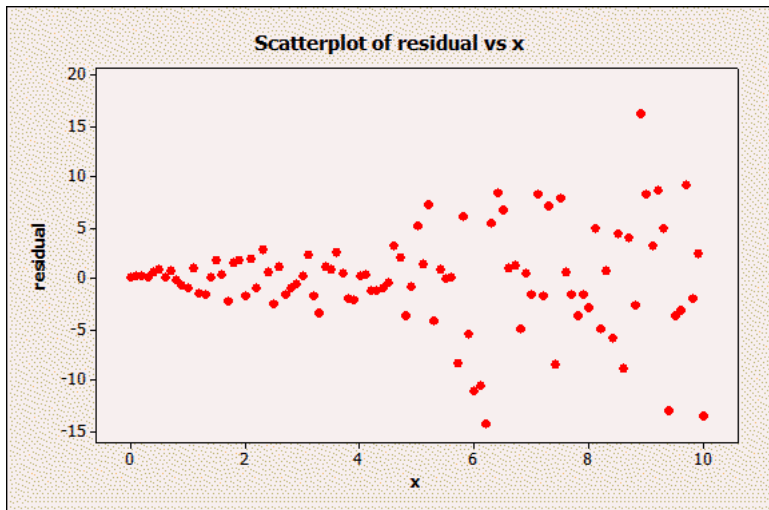
What do you learn from the plot?

    (a) A curve should be fitted rather than a straight line.
    (b) The residuals have a straight-line relationship with x.
    (c) A straight-line relationship describes the relationship well.
    (d) The predictions become less accurate as x increases.
    (e) The correlation between x and y is low.