University of Toronto Scarborough
STAB22 Final Examination

December 2011

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 31 numbered pages, with statistical tables at the back. 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. 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.

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.

At the end of the exam, you *must* hand in your Scantron sheet (or you will receive a mark of zero for the examination). You will be graded *only* on what appears on the Scantron sheet. You may take away the question paper after the exam, but whether you do or not, anything written on the question paper will *not* be considered in your grade.

1. When computing a confidence interval for the population mean μ when the population SD σ is known, what value of z^* should be used for an 85% confidence interval? (The formula for the confidence interval is $\bar{x} \pm z^*\sigma/\sqrt{n}$.)

```
(612.tex)  85% in the middle leaves 15% for the ends, or 0.075 for each
 end. Looking up 0.0750 in the body of table A gives z=-1.44, so
 7.5\% is below -1.44 and 7.5% is above z=1.44. Hence z*=1.44.
```

   (a) 1.84
   (b) 1.96
   (c) 1.64
   (d) 1.26
   (e) * 1.44

2. Let $\bar{x}$ be the mean of a random sample of size 4 from a Normally distributed population with mean 10 and standard deviation 20. What can we say about the sampling distribution of $\bar{x}$?

```
(513.tex)  A) will not have a Normal distribution because of the small sample size.

 B) The distribution of will be approximately Normal but not very close to a Normal

 C) will have a Normal distribution with mean 10 and standard deviation 20.

 D) will have a Normal distribution with mean 20 and standard deviation 20.

 E) will have a Normal distribution with mean 10and standard deviation 10.

 Ans E)
 When the population is Normal the sample size doent have to be large
 for the sample mean to have a Normal distribution. Stad dev   =
 20/sqrt(4) = 10
```

   (a) * Normal with mean 10 and standard deviation 10.
   (b) Only approximately Normal because of the small sample size.
   (c) Not Normal because of the small sample size.
   (d) Normal with mean 20 and standard deviation 20.
   (e) Normal with mean 10 and standard deviation 20.

3. A study was carried out on some people who had developed a cold within the previous 24 hours. The people were randomly divided into two groups; the people in the first group had to take zinc lozenges, and the people in the second group had to take placebo lozenges. Everyone was instructed to take the lozenges every 2–3 hours until the cold was gone. The lozenges were designed so that they could not be distinguished by anyone involved in giving the lozenges to the subjects. For each person, the overall duration of cold symptoms was measured.

   What kind of study is this?

   `(311.tex)  the subjects are made to take a particular kind of lozenges, so it's an experiment. The design of the lozenges makes it double-blind.`

   (a) a voluntary-response sample
   (b) * a double-blind experiment
   (c) an observational study
   (d) a stratified sample
   (e) an experiment, but not double-blind

4. The marks in an exam have a Normal distribution with mean 65 and standard deviation 15. A mark of 80 or above qualifies for an A grade. Adam, Bob and Cindy are three students writing this exam. Assume that their marks are independent. What is the probability that that at least one of them will get an A grade?

```
(524.tex)  A) 0.16
 B) 0.05
 C) 0.4
 D) 0.06
 E) 0.5



 Ans  C)

 P(A) = 0.8413 and probability none of them will get an A = 0.8413^3 = 0.595460101
 Or with 68-95-99.5 % rule,, 1- 0.84^3 = 0.407296
```

(a) 0.16

(b) * 0.4

(c) 0.5

(d) 0.05

(e) 0.06

5. A Statistics course has two tutorials (TUT01 and TUT02). TUT01 has 8 men and 12 women. TUT02 has 24 men and an unknown number of women. A student is selected at random from each tutorial. The probability that both these students are of the same gender is 0.44. How many women are in TUT02?

```
(425.tex)  A) 4
 B) 6
 C)24
 D) 30
 E) None of the above

 Ans  B
 If x is the bunber of women in TUT02 and the P(G)=x/(x+24) and P(B) = 24/(x+24). I
 P(GG)+P(BB)=0.6x/(24+x)+0.4*24/(x+24) =0.44 which gives x =6.
 12 also seems it should be a plausible answer.
```

(a) 4

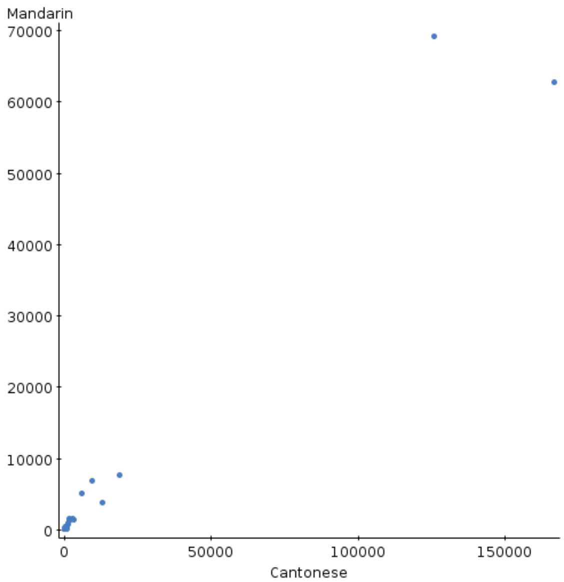
(b) 30

(c) 12

(d) 24

(e) * 6

6. For each of a number of cities across Canada, the number of people with Cantonese as a mother tongue and the number of people with Mandarin as a mother tongue were recorded. A scatterplot of the data is shown below.



The correlation between the number of Cantonese speakers in a city and the number of Mandarin speakers is 0.980. Do you think this correlation is a reasonable summary of the relationship? Why, or why not?

```
(221.tex) %  \begin{minipage}{0.5\linewidth}
%\end{minipage}
%\hfill
%\begin{minipage}{0.5\textwidth}
 the two points top right are definitely influential; the line may
 go a very different place if they are removed, so they may have
 large residuals too. The high correlation is most likely caused by
 those two influential points, and not by the others.
%\end{minipage}
```

(a) Yes, because there are no influential points.

(b) No, because the relationship is clearly curved.

(c) * No, because there are outliers and/or influential points.

(d) Yes, because the relationship is more or less linear.

(e) If the two largest cities (Toronto and Vancouver) were removed, the correlation would be even higher.

7. A social psychologist wants to determine whether restaurant servers will get better tips if they introduce themselves by name to the people they serve. Do you think the social psychologist will use an experiment or an observational study to find out what she wants to know?

```
(312.tex)  experiment can easily be done by randomizing intro or not over
 servers and tables
```

(a) Observational study, because that would allow the researcher to infer cause and effect.

(b) Observational study, because an experiment is difficult to do.

(c) * Experiment, because an experiment can easily be done.

(d) Experiment, because data from an observational study would be worthless.

(e) Observational study, because an experiment would not be ethical.

8. Some people have been complaining that the children's playground at a certain city park is in need of repair. Any repairs will need to be paid for from city taxes. The city will commission a survey on this issue. The survey question is planned to be "the city should allot more funds for the maintenance and repair of children's playgrounds in city parks". What would be the best way to conduct the survey?

```
(321.tex)  a probability sample is best, and followup ensures that the people
 who were actually sampled have their opinions noted (the people who
 don't answer the phone might be in some way different). Using a
 website makes a voluntary-response sample; handing out surveys to
 parents of children actually at playground(s) ignores those parents who
 won't let their children play there (eg. because it is seen as too
 dangerous and needs fixing).
```

(a) * draw a probability sample from all city taxpayers and contact the sampled taxpayers by phone, following up if necessary

(b) hand out surveys to parents of all children at some randomly chosen city parks

(c) use a web site like `surveymonkey.com` to host the survey and advertise it to all city taxpayers.

(d) draw a probability sample from all city taxpayers and contact the sampled taxpayers by phone. If the phone is not answered, ignore this taxpayer and move on to the next.

(e) hand out surveys to parents of all children at the playground in this city park

9. A car rental company records the number of kilometres driven per day by each of its customers, and finds that the number of kilometres driven has a mean of 110 km and a standard deviation of 80 km. Based on this information, what do you think is the shape of the distribution of the number of kilometres driven?

```
(131.tex)  km driven has a lower limit of zero, but if 68-95-99.7 applies,
 there'd be an appreciable prob of getting a value less than 0. So
 not normal. With a lower limit and no upper limit, skewed to right
 is best of the rest.
```

(a) Skewed to the left

(b) * Skewed to the right

(c) Like a normal distribution

(d) Symmetric but not normal

10. In Question 23, you were waiting for a bus at UTSC. The time until departure of the next number 38 bus was a uniformly distributed random variable between 0 and 6 minutes, and the time until departure of the next number 95 bus was uniform between 0 and 10 minutes. Now, you are travelling to Ellesmere and Markham Road, so you can catch either of these buses. You are interested in the chance that the number 38 bus leaves first. This is not possible to calculate using the methods of this course, so you do a simulation instead, using StatCrunch. The first two things you simulate are the times until departure of the next number 38 bus and the next number 95 bus. (These are in the columns labelled "38" and "95" below.) After some more calculation, the worksheet looks as shown below, in part. The "bin column" uses a single cutpoint of zero.

StatCrunch Edit Data Stat Graphics Help

Row	38	95	difference	Bin(difference)	var5
1	0.9330059	4.496621	-3.563615	Below 0	
2	0.6775898	8.787588	-8.109999	Below 0	
3	4.1045814	3.2938678	0.81071347	0 or above	
4	1.0094304	4.9581103	-3.94868	Below 0	
5	4.148572	8.051863	-3.9032905	Below 0	
6	5.976367	8.860882	-2.8845143	Below 0	
7	2.717138	6.2869716	-3.5698335	Below 0	
8	2.9964478	2.6268785	0.36956936	0 or above	
9	4.8281407	2.1040385	2.7241025	0 or above	
10	2.4405382	7.0906477	-4.6501093	Below 0	
11	2.7464058	0.45818517	2.2882206	0 or above	
12	2.6572359	9.014571	-6.3573356	Below 0	

Some output was obtained, as below:

```
Summary statistics:
Column     Median     Min         Max      Q1         Q3
38         2.9611802  8.438758E-4 5.997734 1.4819387  4.458607
95         4.9292355  0.015312281 9.938501 2.6126897  7.5193777
difference -1.9829733 -9.351653   5.625704 -4.4648438 0.36462167

Frequency table results for Bin(difference):
Bin(difference) Frequency Relative Frequency
Below 0         701       0.701
0 or above      299       0.299
```

What is your best guess at the probability that the number 38 bus leaves first?

```
(434.tex)  use the bottom table: 701 times out of 1000, the difference "time
 until next 38" minus "time until next 95" is negative. The first
 table is thrown in mainly to confuse, though since Q3 for the
 differences is slightly positive, the probability of the 38 coming
 first is a bit less than 0.75. But we can do better. (The correct
 answer, by integration, is exactly 0.7. I doctored the simulation
 result a little to make it closer to this.)
```

(a) 0.99

(b) * 0.70

(c) 0.60

(d) 0.75

(e) 0.30

11. The probability that a randomly chosen calculus student passes a certain calculus course is 0.70. Use this information for this question and the next two questions.

If 5 students are sampled at random, what is the probabilty that exactly 4 of them pass the course?

```
(522.tex)  this is probability that exactly 1 fails, with P(fail)=0.3, so is
 0.3602 from Table C with n=5, p=0.30, k=1
 normal approx with continuity correction gives 0.3354
```

(a) 0.640

(b) 0.028

(c) * 0.360

(d) impossible to determine from the information in this course

(e) 0.335

12. In Question 11, the probability was 0.7 that a randomly chosen calculus student would pass a certain calculus course. Under these circumstances, suppose a simple random sample of 15 calculus students is taken. What is the probability that 13 or more of these students pass the course?

```
(522.tex)  n=15, p=1-0.7=0.3, 13+ passing means 15-13=2 or fewer failing. Table
 C says 0.0916+0.0305+0.0047=0.1268. normal approx says 0.0795
 (without continuity correction; with cc gives 0.1299, which would be
 the right answer, even though the rule of thumb says the normal
 approx shouldn't work. Here we're in the short tail, which probably helps).
```

(a) 0.000

(b) 0.079

(c) impossible to determine from the information in this course

(d) * 0.127

(e) 0.092

13. In Question 11, the probability was 0.7 that a randomly chosen calculus student would pass a certain calculus course. Under these circumstances, suppose a simple random sample of 150 calculus students is taken. What is the probability that 115 or more of these students pass the course? (You may assume that the total number of students who take this course is over 2000.)

```
(522.tex)  use normal approx to binomial. mean 105, sd 5.612 (n(1-p)=45, so
 rule of thumb is fine).  z=1.78, prob of
 more is 1-0.9626=0.0374. With cont correction is 0.0452
 (z=1.69). exact binomial is 0.0429 (R).
```

(a) 0.21

(b) * 0.04

(c) less than 0.01

(d) impossible to determine from the information in this course

(e) 0.50

14. A test of significance is significant at the 5% level. Which one of the following statements must also be true about this test result? (At most one of the statements below must be true.)

```
(623.tex)
```

(a) The test is not significant at the 10% level.

(b) The test is not significant at the 1% level.

(c) * The test is significant at the 10% level.

(d) None of the other statements must be true.

(e) The test is significant at the 1% level.

15. Researchers speculate that drivers who do not wear a seatbelt are more likely to speed than drivers who do wear one. A random sample of 20 drivers had their speed measured at a certain point, and were then observed to see whether they were wearing a seatbelt. Summary statistics for the drivers are shown below:

```
Summary statistics for speed:
Group by: seatbelt
seatbelt n  Mean      Std. Dev.
n        8  72.5      8.815571
y        12 65.333336 7.487363
```

A test was carried out to test the null hypothesis that the mean speed was equal for the drivers wearing and not wearing seatbelts, against the alternative that drivers not wearing seatbelts travel faster on average. The test statistic was -1.889 (seatbelt mean minus non-seatbelt mean). Using the methods learned in class, what can you say about the P-value for this test statistic?

```
(723.tex)  correct side, use t with 7 df one-sided and ignore minus sign.
```

(a) larger than 0.10

(b) between 0.025 and 0.05

(c) 0.0294

(d) 0.0588

(e) * between 0.05 and 0.10

16. Vehicle speeds at a certain highway location have a mean of 100 kmh and a standard deviation of 10 km/h, with a distribution that is approximately normal. A simple random sample of 25 vehicles is taken. There is a 95% chance that the mean speed of the sampled vehicles is between which of the values below, in km/h?

```
(511.tex)  talking about sample mean, which has an approx normal dist with mean
 100 and sd 10/sqrt(25)=2. So go up and down 2 times this SD (for
 95%) from mean, ie. plus/minus 4.
```

(a) * 96 and 104

(b) 97 and 103

(c) 99.6 and 100.4

(d) 80 and 120

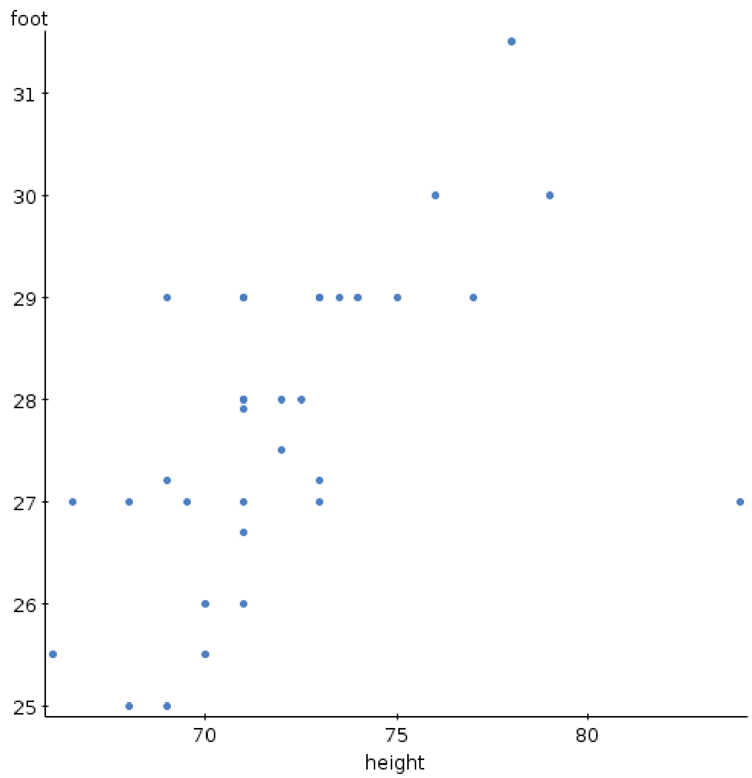
(e) 98 and 102

17. A simple random sample of 64 men has a sample mean foot length of 27.5 cm. Assuming that the standard deviation of foot lengths for all men is 2 cm, obtain a 95% confidence interval for the mean foot length of all men. What is the **upper** limit of this interval, in centimetres?

```
(611.tex)  27.5 + 1.96 2/sqrt(64) = 27.99 cm
```

(a) 28.5
(b) 27.5
(c) 27.0
(d) * 28.0
(e) 26.5

18. For each of the 33 male students in a college class, their height (in inches) and foot length (in centimetres) were measured. None of the men appeared to be unusually tall or short. A scatterplot of the data collected is shown below.



How would you describe what you see on the scatterplot?

```
(232.tex)  the largest height is around 84 inches or 7 feet, which would be
 seen as unusually tall, so that observation would appear to be an
 error. A man with this foot length (27 cm) would be expected to
 have a height around 70 inches (5 ft 10).
```

(a) There is no association.

(b) There is an outlier, and that outlier appears to be a legitimate data value.

(c) There is a curved association.

(d) There is a straight-line association.

(e) * There is an outlier, and that outlier appears to be an error.

19. A fair coin is tossed 400 times. Find the standard deviation of the number of heads that will be obtained.

```
(523.tex)  A) 10
 B) 20
 C) 30
 D) 40
 E) 50

 Ans A
 The number of heads that will be obtained has a Bin (400, 0.5) distribution and so
```

(a) 50

(b) * 10

(c) 20

(d) 40

(e) 30

20. In a game you toss a fair coin once. You win \$20 if a head comes up and you lose \$10 if a tail comes up. Let the random variable X denote the amount of money you win. (Losing \$10 means winning $-$\$10.) What is the standard deviation of X, in dollars?

```
(441.tex)  A) $5
 B) $10
 C) $12.50
 D) $15
 E) $ 17.5

 Ans D

 mean of X = 20*0.5 +(-10)*0.5 = 5
 Var(X) =  (20-5)^2*0.5+(-10-5)^2* 0.5= 225
 Std dev = sqrt(225) = 15

 check: bernoulli for number of heads Y would be mean 0.5, sd
 0.5. Transform Y to X by X=30Y-10. Mean transforms to 30(0.5)-10=5,
 SD transforms to 30(0.5)=15.
```

(a) 5

(b) 10

(c) 12.50

(d) * 15

(e) 17.50

21. You toss 2 fair coins and count the number of heads. Independently, your friend tosses 3 fair coins and counts the number of heads. The winner is the player who gets more heads when they toss their coins. What is the probability that you win?

```
(521.tex)  only possibilities are 1 for you, 0 for him; 2 for you and 0 or 1
 for him. For you: 1: 1/2, 2: 1/4; for him: 0: 1/8, 1: 3/8. Thus
 (1/2)(1/8)+(1/4)(1/8)+(1/4)(3/8)=1/16+1/32+3/32=6/32=3/16. Or
 write out all the equally likely possibilities: HH, HT, TH, TT for
 you, HHH, HHT, HTH, HTT, THH, THT, TTH, TTT for him. Out of these
 4x8=32, only HH-HTT, HH-THT, HH-TTH, HH-TTT, HT-TTT, TH-TTT result
 in wins for you, 6/32=0.19.
```

(a) 0.04
(b) 0.01
(c) 0.50
(d) 0.39
(e) *0.19

22. A sample of size 20 is taken from a population with unknown mean and unknown standard deviation. What value of t^* should be used for a 90% confidence interval for the population mean?

```
(711.tex)  df=20-1=19, t*=1.729
```

(a) * 1.729
(b) 2.093
(c) 1.960
(d) 2.861
(e) 1.645

23. Suppose you arrive at the UTSC bus stop. You have exactly 4 minutes to wait for your bus. While you are waiting, you observe other buses that are leaving. The time until the next number 38 bus leaves is a random variable with a (continuous) uniform distribution between 0 and 6 minutes; independently of that, the time until the next number 95 bus leaves is a random variable with a uniform distribution between 0 and 10 minutes. What is the probability that, while you are waiting, you observe a number 38 bus leaving, but you do not observe a number 95 bus leaving?

```
(433.tex)  P(observe a 38)=4/6; P(do not observe a 95)=1-4/10=6/10;
 independence implies multiplication is ok for "and", so ans
 is 4/6 * 6/10 = 0.40.
```

(a) 0.53
(b) * 0.40
(c) 0.27
(d) 0.20
(e) 0.13

24. A simple random sample of 50 measurements is taken from a slightly skewed population whose standard deviation is known to be 10. We are testing a null hypothesis that the population mean is 60 against the alternative that it is not equal to 60, using a significance level of $\alpha = 0.05$. The sample mean is 63. What do you conclude?

```
(624.tex)  z = (63-60)/(10/sqrt(50))=2.12, prob of greater = 0.0169, P-value
 = 0.0339. The answer would be the same if t were incorrectly used,
 but the only good way to fix that is to have a very small sample
 size. Implication is that a sample size of 50 is good enough for
 CLT with a "slighly skewed" population.
```

(a) * conclude that the population mean is not 60 because the P-value is between 0.025 and 0.05
(b) conclude that the population mean is not 60 because the P-value is between 0.01 and 0.025
(c) conclude that the population mean could be 60 because the P-value is greater than 0.05
(d) conclude that the population mean is not 60 because the P-value is less than 0.01
(e) conclude that the population mean could be 60 because the P-value is less than 0.05

25. The random variable X has a Normal distribution with mean 60 and standard deviation 10. One of the following probabilities is also equal to $P(40 < X \leq 48)$. Which one? (Hint: It is not necessary to use a normal table to answer this question.)

```
(134.tex)  The normal density curve is symmetric about the mean (i.e. 60) and
 (72,80) is the interval symmetrically opposite to (40, 48). Also <
 and <=  give the same probabilities for the normal distribution (in
 fact for all continuous distributions).
```

(a) * $P(72 < X \leq 80)$

(b) $P(64 < X \leq 72)$

(c) $P(50 < X \leq 58)$

(d) $P(80 < X \leq 88)$

(e) $P(56 < X \leq 64)$

26. The salaries paid to the 13 employees of a small market research company are as follows: the five telephone interviewers are each paid $32,000; three administrative assistants are paid $48,000; three data analysts are paid $55,000, one supervisor is paid $65,000 and one senior manager is paid $160,000.

What is the median salary of these 13 employees?

```
(128.tex)
 A) $40,000
 B) $48,000
 C) $55,000
 D) $60,000
 E) $105,000

 Ans : Ans B) $ 48 000. The 13 salaries are: 32 32 32 32 32 48 48 48 55 55 55 65 16
 There are 13 values and the median is the 7th  = 48
```

(a) $40,000

(b) $60,000

(c) * $48,000

(d) $55,000

(e) $52,500

27. The random variable X has the probability distribution shown below:

Value	1	2	3	7
Probability	0.65	0.20	0.10	0.05

What is the mean of X?

```
(432.tex)  1(0.65)+2(0.20)+3(0.10)+7(0.05)=1.7
```

(a) * 1.7
(b) 1
(c) 1.5
(d) more than 3
(e) 2.5

28. The random variable Y has the distribution shown below:

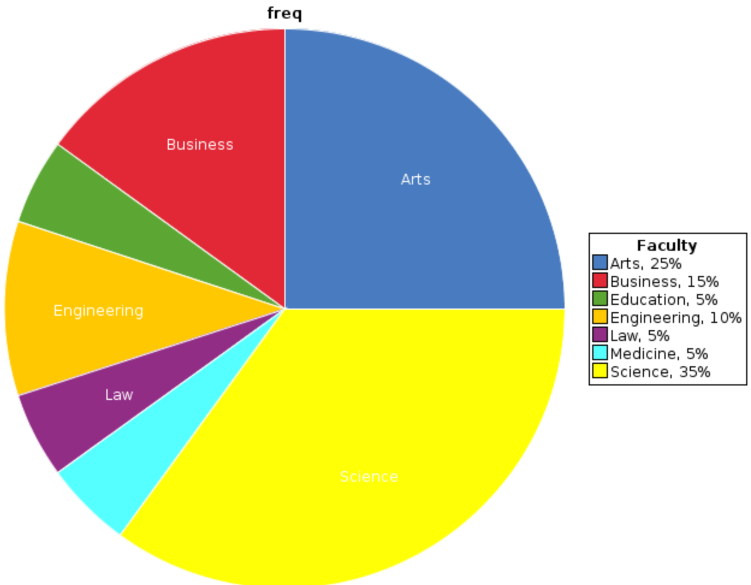
Value	1	4
Probability	0.2	0.8

The mean of Y is 3.4. What is the standard deviation of Y?

```
(424.tex)  variance is (1-3.4)^2(0.2)+(4-3.4)^2(0.8)=1.44, sd 1.2
```

(a) 1.9
(b) 5.0
(c) 2.5
(d) * 1.2
(e) 1.4

29. The pie chart below shows the percentage of students in each faculty at a university.



If there are 3000 students in the faculty of Arts, then **how many** students are there in the faculty of Science?

```
(111.tex)
 Ans E There are 3000 students in the Arts faculty.  That is
 25% of the all students in the university. Thus the number of
 students in the university is $3000* 4 = 12 000 and 35% of them,
 i.e. 12 000*0.35 = 4200 are in Science faculty.
```

(a) 3600

(b) 3800

(c) * 4200

(d) 4000

(e) 3400

30. A simple random sample of 25 observations is taken from a population with mean 80 and SD 20. Assume that this sample size is large enough for the Central Limit Theorem to apply. Use this information for this question and the next one.

    What is the probability that the sample mean is less than 77?

```
(512.tex)  z = (77-80)/(20/sqrt(25))=-0.75, prob 0.2266
```

   (a) 0.44
   (b) * 0.23
   (c) 0.51
   (d) 0.72
   (e) 0.00

31. Question 30 referred to taking simple random samples from a certain population. This time, two simple random samples are drawn. What is the probability that the two sample means will differ by more than 3?

```
(512.tex)  x1bar and x2 bar both N(80,4^2) so difference N with mean 0, SD
 sqrt(16+16)=5.657. P(difference x1bar-x2bar >= 3) uses
 z=(3-0)/5.657=0.53, prob of greater = 0.2979=~0.30. By symmetry,
 P(difference <= -3) also is 0.30, so prob total is 0.60.
```

   (a) 0.45
   (b) * 0.60
   (c) 0.05
   (d) 0.30
   (e) 0.90

32. Should cell phone use be banned by drivers? 188 people took part in a survey. The answers are classified by the gender of the respondent as below:

	Agree	Disagree
Female	68	37
Male	26	46

Use this information for this question and the two following.

What is the conditional proportion of females who agree?

`(251.tex)  68/(68+37)=0.647`

(a) * 0.65

(b) 0.72

(c) 0.55

(d) 0.45

(e) 0.60

33. Refer to the information given in Question 32. What is the marginal proportion of people who agree?

`(251.tex)  (68+26)/(68+37+26+46)=0.531`

(a) 0.60

(b) 0.45

(c) * 0.55

(d) 0.65

(e) 0.72

34. Refer to the information given in Question 32. Is there an association between gender and the response given on the survey?

```
(251.tex)  proportion of females agreeing is 0.65, overall agreement is only
 0.55, so proportion of males agreeing must be less than this.
```

(a) We need to look at a scatterplot to judge association.

(b) No, because the proportions of females and males who agree are very similar.

(c) Association here has nothing to do with what proportions of males and females agree.

(d) Yes, because the proportion of females who agree is lower than the proportion of males who agree.

(e) * Yes, because the proportion of females who agree is higher than the proportion of males who agree.

35. A newly-designed highway sign is being examined. 27 drivers, of various ages, were tested to see at what maximum distance (in feet) each driver could clearly read the sign. Each driver's age was also recorded. Some summary statistics are shown below:

Column	n	Mean	Std. Dev.	Median	Range	Min	Max	Q1	Q3
age	27	51.33	21.944	55	64	18	82	28	71
distance	27	423.33	82.927	420	310	280	590	360	460

The correlation between age and distance is -0.796. Calculate the **intercept** of the least-squares regression line for predicting distance from age.

```
(231.tex)  slope = (-0.796)*(82.927/21.944)=-3.008
 intercept=423.33-(-3.008)(51.33)=577.74
```

(a) 275

(b) * 575

(c) 350

(d) 500

(e) 425

36. The null hypothesis $H_0 : \mu = 15$ is being tested against the alternative $H_a : \mu \neq 15$. The P-value for the test is 0.12. What can you say about the 90% and 95% confidence intervals for the population mean μ?

```
(621.tex)  P-value is greater than both 0.10 and 0.05, so in both cases 15 is
 a plausible value for the mean, so is in both intervals.
```

(a) The 95% confidence interval contains 15 but the 90% confidence interval does not.

(b) The P-value of the test tells us nothing about whether 15 is inside or outside a confidence interval.

(c) The 90% confidence interval contains 15 but the 95% confidence interval does not.

(d) Neither confidence interval contains 15.

(e) * Both confidence intervals contain 15.

37. The distribution of the grades in an exam has a Normal distribution. Approximately 2.5% of the students scored 50 or below. 16% of them scored 85 and above. What can you say about the mean grade of all the students?

```
(133.tex)  A) 50
 B) 62
 C) 67.5
 D) 73
 E) 77


 Ans D

 50 is two standard deviations below the mean and 85is one standard
 deviations above the mean and so 85 -50 = 3 SD and SD =35/3 = 11.67
 and so mean = 50 + 2* 11.67 = 73.3
 or: bottom 2.5% goes with mark 50 or z=-1.96; top 16% goes with mark
 85 or z=0.99. Hence mu - 1.96 sigma = 50 and mu+0.99 sigma=85, so
 that 2.95 sigma = 35 or sigma = 11.86, mu= 50+1.96 sigma =
 73.25. But the 68-95-99.7 way is easier. (However, can't legislate
 doing it that way.)
```

(a) less than 60

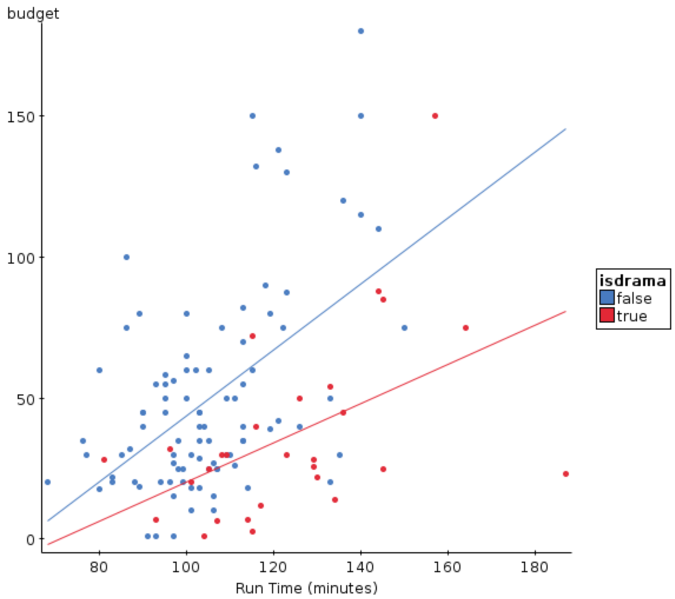
(b) * between 70 and 75

(c) between 65 and 70

(d) greater than 75

(e) between 60 and 65

38. A study was made of movies in 2005. For each movie, the following were recorded: its budget (the amount it cost to make, in millions of dollars), its running time (from start to finish, in minutes), and the genre (drama, comedy, action, etc.). The scatterplot below shows the budget vs. running length for all movies. Two regression lines are shown on the scatterplot, for predicting budget from running length. The lower line is for movies of the "drama" genre, while the upper line is for movies of all other genres. Which of the statements below do you most strongly agree with?




```
(234.tex)  since the colours won't come
 through in the final version, I have
 to phrase the question as shown.
 Since the lower line is dramas,
 dramas cost less to make, and
 because the lines get further apart,
 the difference gets larger with
 dramas continuing to cost less to
 make. In B27 terms, there is an
 interaction between length and genre.
%\begin{minipage}{0.5\linewidth}
%\end{minipage}
%\hfill
%\begin{minipage}{0.5\linewidth}
%\end{minipage}
```

(a) * Dramas cost less to make than other movies of the same length, and that difference increases as the movies get longer.

(b) Dramas cost less to make than other movies of the same length, and that difference decreases as the movies get longer.

(c) Dramas cost more to make than other movies of the same length, and that difference increases as the movies get longer.
(d) Dramas cost more to make than other movies of the same length, and that difference decreases as the movies get longer.
(e) The difference in budget between dramas and other movies of the same length remains constant as the movies get longer.

39. A *98%* confidence interval was calculated for a population mean, based on a simple random sample of size 25. The population standard deviation was not known. The interval was from 54 to 66. What must the standard deviation of this sample have been?

```
(713.tex)  A) 5
 B) 6
 C) 10
 D) 12
 E) 14
```

```
Ans: margin of error  = (66-54)/2 = 6. The t table value for df = 25-1 = 2.492 (fo
i.e. 2.492 *( s/5) = 6 and so s = 12.03852327
```

(a) * 12
(b) 14
(c) 10
(d) 6
(e) 5

40. A sample of 20 measurements is taken from an approximately normal population whose standard deviation is known. A 95% confidence interval for the population mean goes from 64.3 to 72.1. What is the test statistic for testing $H_0 : \mu = 75$ against $H_a : \mu \neq 75$?

```
(622.tex)  interval is 7.8 long, so margin of error is 3.9, and sample mean is
 middle of interval, 68.2. Margin is 1.96 sigma/sqrt(20)=3.9, so
 sigma=8.899. z = (68.2-75)/(8.899/sqrt(20)=-3.42.
```

(a) 2.0
(b) 0
(c) -2.0
(d) 3.5
(e) * -3.5

41. Bob and Carol go swimming every morning. They each swim a fixed distance. The time each person takes to complete the swim has a normal distribution, and the times are independent of each other. Bob has a mean time of 10 minutes, with a standard deviation of 1 minute, while Carol has a mean time of 11 minutes with a standard deviation of 0.5 minutes. What is the probability that, on a randomly chosen morning, Carol will complete her swim more quickly (in fewer minutes) than Bob does?

```
(431.tex)  Difference C-B has mean 11-10=1, variance 1^2+0.5^2=1.25, so SD
 sqrt(1.25)=1.118$ minutes. Prob of difference being less than 0 uses
 z=(0-1)/1.118; 0.1855
```

(a) 0.5
(b) 0.7
(c) * 0.2
(d) 0.1
(e) 0.3

42. How does a systematic sample differ from a simple random sample?

```
(322.tex)  the key thing is that independence fails. It is usually *more*
 convenient to take than a simple random sample. If parts of the
 population differ from each other, a stratified sample is the thing
 to use.
```

(a) * Knowing about one item in the sample tells you about which other items will be in the sample.

(b) If parts of the population differ from each other, a systematic sample will give more accurate results.

(c) A systematic sample is less convenient to take than a simple random sample.

(d) Each item in the population has the same chance to be in the sample.

43. A simple random sample of students was taken from Penn State University. For each student, their sex was recorded and their pulse rate measured. The results were as follows:

Sex	n	sample mean	sample SD
Women	35	76.9	11.6
Men	57	70.42	9.95

It is desired to see whether there is evidence of a difference in mean pulse rates between women and men. Calculate the test statistic for assessing this evidence. Assess the difference as women minus men. What do you get?

```
(721.tex)  statcrunch says 2.74, not pooled. 2.84 is pooled.
```

(a) greater than 3

(b) 2.84

(c) 2.50

(d) * 2.74

(e) less than 2

44. Pulse rates of women have a normal distribution with mean 75 and standard deviation 8. Use this information for this question and the next three questions.

What proportion of women have pulse rates less than 71?

```
(132.tex)  z=(71-75)/8=-0.5, prop=0.3085
```

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

45. Using the information in Question 44, what proportion of women have pulse rates greater than 85?

```
(132.tex)  z=(85-75)/8=1.25, prop less is 0.8944, greater is
 0.1056. (Sanity-check: answer should be less than 0.5 since 85 is
 greater than mean.)
```

(a) 0.70
(b) 0.90
(c) * 0.10
(d) 0.30
(e) 0.50

46. Using the information in Question 44, what proportion of women have pulse rates between 59 and 95?

```
(132.tex)  for 59, z=(59-75)/8=-2; for 95 z=(95-75)/8=2.5. Proportion between
 is 0.9938-0.0228=0.9710.
```

(a) 0.99
(b) 0.85
(c) 0.95
(d) less than 0.50
(e) * 0.97

47. Using the information in Question 44, what pulse rate x is such that only 5% of women have a higher pulse rate than x?

```
(132.tex)  5 percent more is 95 percent less, which goes with
 z=1.645. Unstandardize to get pulse rate of 75+8*1.645=88.15.
```

(a) * 88
(b) 80
(c) 103
(d) 91
(e) 99