STATS 101/108 WORKSHOP

SATURDAY 2: CHAPTERS 4 AND 6

SATURDAY 14 AUGUST, 2010

Students **MUST REGISTER** for all workshops with
The Student Learning Centre, 3rd Floor, Information Commons

Student Learning Centre

Topics we teach and can provide advice on include:

- ✓ Essay writing
- ✓ Computer skills
- ✓ Reading and notetaking
- ✓ Memory and concentration
- ✓ Report writing
- ✓ Test and examination skills
- ✓ Thesis and dissertation writing
- ✓ Tutorial skills
- ✓ Research skills
- ✓ Time and stress management
- ✓ Mathematics
- ✓ **Statistics**
- ✓ Oral presentation and seminar skills
- ✓ Language learning
- ✓ Specific learning disabilities
- ✓ Motivation and goal setting
- ✓ Survival skills (in the University system)

Programmes within SLC include:

- Te Puni Wananga
  Maori university tutors committed to enhancing Maori students' success

- Fale Pasifika
  Pacific Island tutors committed to enhancing success for Pacific Island students

- Students with impairments
  Learning assessments are available for students with specific learning disabilities; academic assistance is available for these students and those with mental health impairments.
  If you have any special learning requirements, please feel free to discuss this with Leila in person or via email.

- Academic English Conversation Groups
  Improve your academic English; develop communication skills including critical/creative thinking and clear expression of ideas and opinions. Weekly class held at the SLC on Thursdays, 3-5pm (during semester)

Statistical help available at the SLC

The Student Learning Centre (SLC) offers help for STATS 101/108 by offering:

- one-on-one tutoring help, and
- a number of workshops

One-on-one help

The SLC employs tutors specifically to help students with one-on-one assistance for STATS 101/108. One-on-one tutoring must be booked at SLC reception on the third floor of the Information Commons in person or by calling 373-7599 X 88850. Enquire at the SLC reception for available times.

Note: SLC tutors are not allowed to help students complete their assignments.

SLC STATS 101/108 Workshops

Any questions regarding STATS 101/108 workshops should be forwarded to:

Leila Boyle
SLC Statistics Co-ordinator
l.boyle@auckland.ac.nz

x88850

Workshops are run in a relaxed environment, typically set at a pace for those students that find the Statistics Department's tutorials too fast. All workshops allow plenty of time for questions. In fact, this is encouraged ☺

1) Saturday Workshops x5

These five 3-hour workshops are held on Saturdays throughout the semester to help students with different sections of the course.

2) Computer Workshops: Excel / PASW (SPSS) x 3

These three computer-based workshops introduce students to the skills needed for Excel and PASW (SPSS) use in STATS 101/108 assignments.

3) Pre-test Workshops x 3

These three workshops will cover the basics that you need for the test.

4) Pre-exam Workshops x 6 each repeated twice

These six workshops will cover the basics that you need for the exam.

Note: All workshops concentrate on questions reviewing the basic concepts, rather than questions on finer details. They are designed to assist students to achieve a pass; they are not designed to cover all material.

The timetable for these workshops is available with this handout. Please enrol in each of your preferred classes at the Student Learning Centre by:

- **Going to the SLC in person**
- **Enrolling online at www.slc.auckland.ac.nz (click on "SLC Workshops" then click on "Undergraduate Workshops" then view workshops by "Statistics").**

Useful Websites

- SLC webpage: **www.slc.auckland.ac.nz**
- Online enrolment through the SLC site: **www.slc.auckland.ac.nz**, click on "**SLC Workshops**" and then click on "**Undergraduate Workshops**"
  - For STATS 101/108 workshops, view by "**Statistics**"
  - For SLC workshops aimed at helping students learn basic computing skills (e.g. Excel, Word, PowerPoint), view by "**Computer skills**"
  - For other SLC workshops on skills appropriate for undergraduate students, scroll down the page or view by the appropriate category
- Cecil: **https://cecil.auckland.ac.nz/**
- **Leila's website for STATS 101/108 SLC workshop handouts & information: www.stat.auckland.ac.nz/~leila**

2 qual. var.s

Revision Notes

2-4 Qs in test & exam

Chapter 4 - Probability

Look at blue pages for extra test/exam questions for practice

impossible certain

- A **probability** is a number between 0 and 1 that quantifies uncertainty.
- There are two main sources of probabilities that we will deal with.
  1. Probabilities using a model – some models that may involve equally likely outcomes are *tossing a coin* and *rolling a die*
  2. Probabilities from data → two-way table of counts
- A **random experiment** is an experiment where the outcome cannot be predicted.
- A **sample space** is the collection of all possible outcomes.
- An **event** is a collection of outcomes. An event **occurs** if any outcome making up that event occurs.
- If all **outcomes** are **equally likely:** $\text{pr}(A) = \dfrac{\text{no. of outcomes in } A}{\text{total no. of outcomes}}$
- The **complement** of an event A, denoted $\overline{A}$, occurs if A does not occur. A, and $\overline{A}$ are **mutually exclusive** events, ie they CANNOT occur at the same time.
- General probability rules:
  1. $\text{pr}(S) = 1$
  2. $\text{pr}(\overline{A}) = 1 - \text{pr}(A)$ $\quad \text{pr}(A) = 1 - \text{pr}(\overline{A})$
- **Statistical Independence** – two events (A & B) are statistically independent if knowing whether B has occurred gives no new information about the chances of A occurring.

  memorise

  i.e $\text{pr}(A|B) = \text{pr}(A)$ and $\text{pr}(A \text{ and } B) = \text{pr}(A) \times \text{pr}(B)$

- **Two Types of Test/Exam Questions**
  1. Given a table of numbers/~~proportions~~, find the probability:
     - ☞ Easier question/s (can be between 1 and 3 of this type).
     - ☞ ~~*May want to convert the table into table of probabilities first.*~~

2. Given a short story with proportions, percentages and/or counts about two factors (qualitative variables), find the probability:

- ☞ Harder question/s (can be 1 or 2 of this type).
- ☞ Need to *interpret* the story first, and then construct a table.
- ☞ Use the table to find 1 or 2 probabilities.
- ☞ Steps to constructing a table:

  Step 1: highlight numbers
  Step 2: highlight factors
  Step 3: define factor levels
  Step 4: label table
  Step 5: enter appropriate table total
  Step 6: enter row/column totals from story
  Step 7: enter cell numbers from story
  Step 8: enter remaining numbers by +/-

- **Four Types of Probability Calculations**

1. Probability of ***AN EVENT*** *(basic/simple)*

- ☞ pr(*A*) → pr (an event)

rows/s totals/s ÷ GT example 1
or columns/s totals/s ÷ GT

2. Probability of an event <u>AND</u> another event:

- ☞ pr(*A* and *B*) → pr(one event and another event)
- ☞ *Finding* pr(*A*) <u>*and*</u> pr(*B*) (intersection)

cells/s ÷ GT example 3

3. Probability of an event <u>OR</u> another event:

- ☞ pr(*A* or *B*) → pr(one event or another event)
- ☞ <u>*Add*</u> pr(*A*) *to* pr(*B*), then <u>*subtract*</u> pr(*A* and *B*)

[rows/s totals/s + columns/s totals/s
- cells/s] ÷ GT example 4

Use (GT)
GRAND TOTAL (TABLE TOTAL)

4. <u>CONDITIONAL</u> *Probability:*

example 2

☞ Harder to detect but will usually have one of the key words:

- "Given that…"
- "Of those…"
- "Among those…"

} Use **ROW TOTAL/S** *OR* **COLUMN TOTAL/S**

- Look for language that <u>restricts you to part of the table</u> instead of the whole table.

must happen

- pr (one event | another event) = pr($A|B$) = $\dfrac{\text{pr}(A \text{ and } B)}{\text{pr}(B)}$

may happen

"given that"

Examples 1 to **4** refer to the following information.

The following table cross-classifies restaurants surveyed by the business ***outlook*** of the owner(s) and by the ***size*** of the restaurant.

OUTLOOK	**Small**	**<u>SIZE</u>** **Medium**	**Large**	**Total**
Very Unfavourable	46	9	17	72
Unfavourable	34	18	11	63
Neutral	31	21	15	67
Favourable	25	13	11	48
Very Favourable	5	5	3	13
Total	140	66	57	263

Example 1: What percentage of restaurants surveyed were large-sized?

$$pr(\text{large}) = \frac{57}{263} \times 100\%$$

$= 21.67\%$ 2 (4dp)

Example 2: What proportion of restaurant owners of small-sized restaurants had a very favourable business outlook?

$$pr(\text{VF} \mid \text{small}) = \frac{5}{140}$$

$= 0.0357$ (4dp)

Example 3: What is the probability that a restaurant owner, randomly selected from this survey, owned a medium-sized restaurant and at the same time had a very unfavourable business outlook?

$$pr(\text{M } \underline{\&} \text{ VU}) = \frac{9}{263}$$

$= 0.0342$ (4dp)

Example 4: What proportion of restaurant owners had a very favourable business outlook or owned a small-sized restaurant?

$$pr(\text{VF } \underline{\text{or}} \text{ S}) = \frac{(13 + 140 - 5)}{263}$$

$= 0.5627$ (4dp)

Chapter 6 – Continuous Random Variables

3 or 4 Qs in test/exam

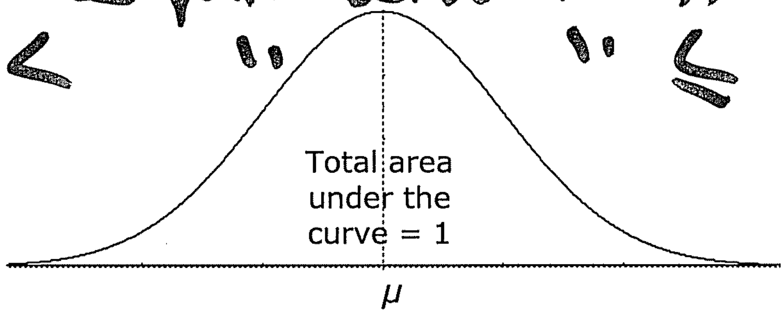
Look at blue pages for good notes and test/exam questions for practice

- A density curve is the probability distribution of a continuous random variable.
- There are no gaps between the values that a continuous random variable can take and therefore, when we calculate probabilities for a continuous random variable it does not matter whether **interval endpoints** are included or excluded

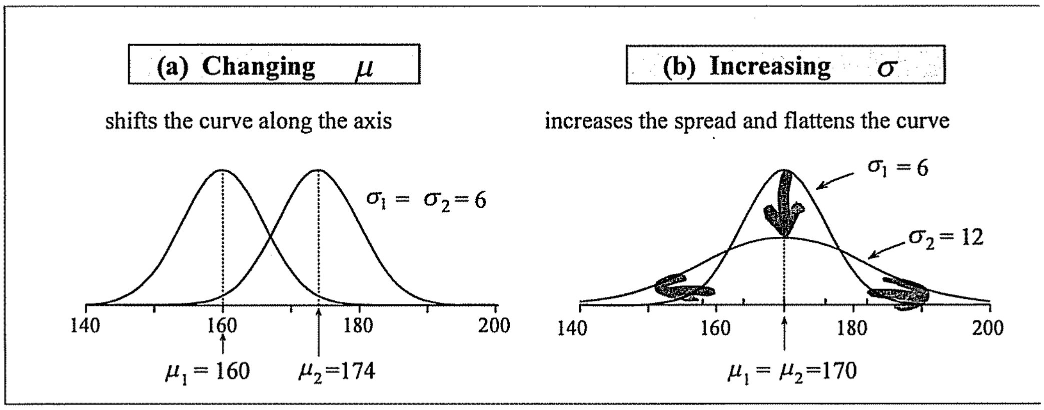
$>$ equivalent to $\geq$
$<$ " " $\leq$

Normal Distribution

- The Normal Distribution has a probability density function curve, which is smooth, **bell-shaped**, and **symmetric**.



- The shape of the curve is solely determined by the parameters μ (mean) and σ (standard deviation).



- The Normal distribution is important because it:
  - fits a lot of data particularly well
  - can be used to approximate other distributions
  - is very important in statistical inference → Ch 8, 9, 10, 12
- If X is a continuous random variable from a Normal distribution then:
  - $E(X) = \mu$ and $sd(X) = \sigma$ — expected value = mean = average
  - Probability distribution function of X is written: $X \sim \text{Normal}(\mu, \sigma)$

"distributed as"

X = random variable

x = a particular number of interest

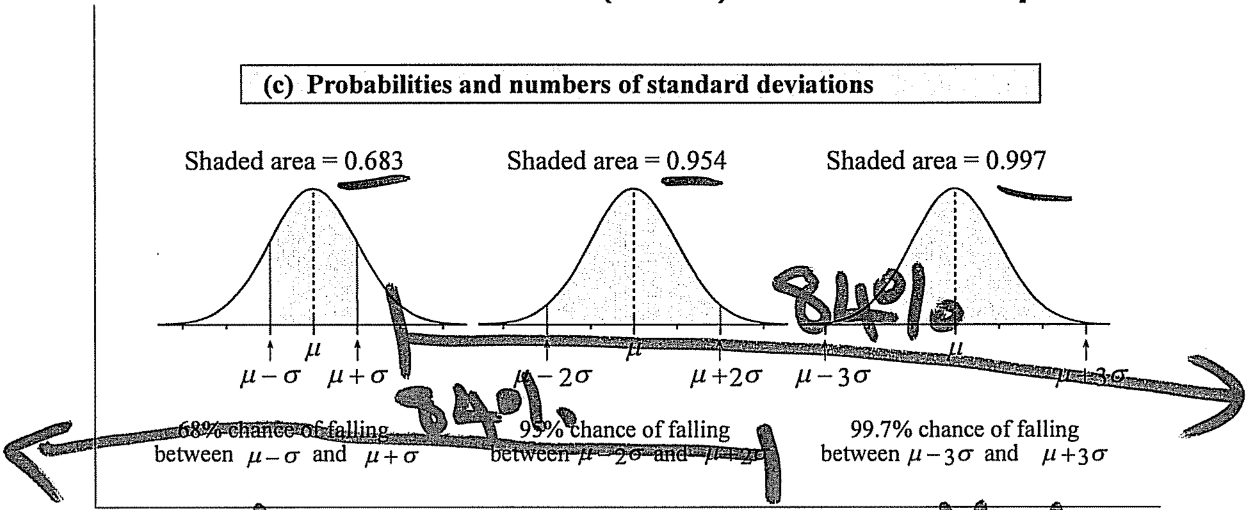
Chapter 6 test/exam questions

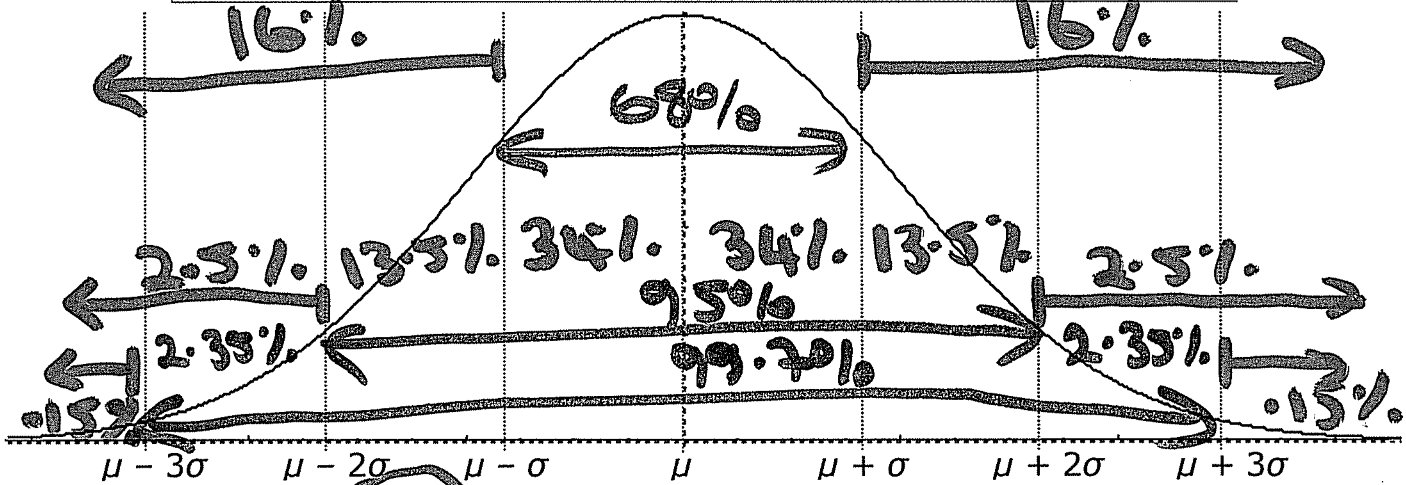
When doing Chapter 6 problems, it is sensible to draw a Normal curve and then mark on it what is known and what is unknown. There are **three (3)** types of Chapter 6 test/exam questions:

1. *True/False (Normal) Chapter 6 problem*

There will be five statements, each about one or the other or both of two different Normal distributions. Use the 68-95-99.7% rule or z-scores to determine whether four of the statements are true or false. The fifth statement will probably be comparing the means (centres/averages) and standard deviations (spread/variability) of the two distributions.

- **68-95-99.7% rule**: A population with a Normal distribution has:
  - ✓ 68% of its observations (values) within 1 $\boldsymbol{\sigma}$ of the $\boldsymbol{\mu}$
  - ✓ 95% of its observations (values) within 2 $\boldsymbol{\sigma}$ of the $\boldsymbol{\mu}$
  - ✓ 99.7% of its observations (values) within 3 $\boldsymbol{\sigma}$ of the $\boldsymbol{\mu}$



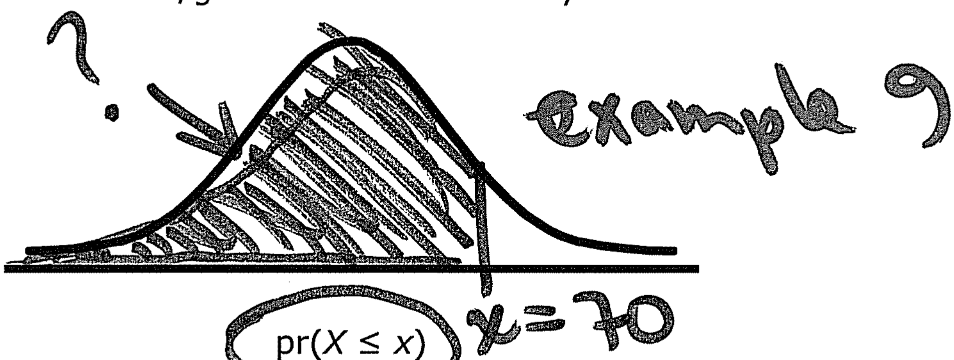


The **z-score**, $z = \dfrac{x - \mu}{\sigma}$, is a standardised number. It represents the number of standard deviations, σ, the value of x is away from the mean, μ. We can use z-scores to compare two or more different Normal distributions.

2. *Normal probability problem, i.e. find a probability associated with a number*

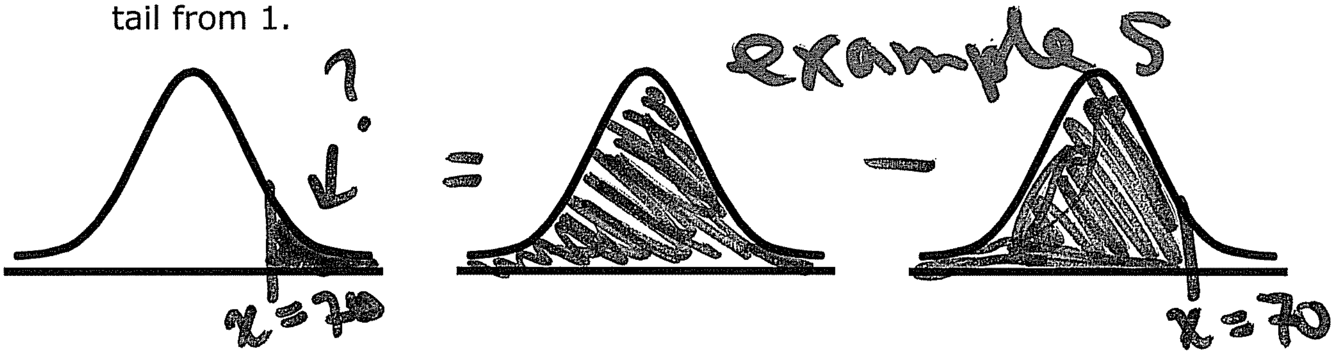
When finding a probability, shade the desired area under the curve and then devise a way to obtain it using lower tail probabilities which is all the computer can give. There are three types of Normal probability problems:

- *Find a lower tail probability (area)*
  The computer can find/give the answer directly.



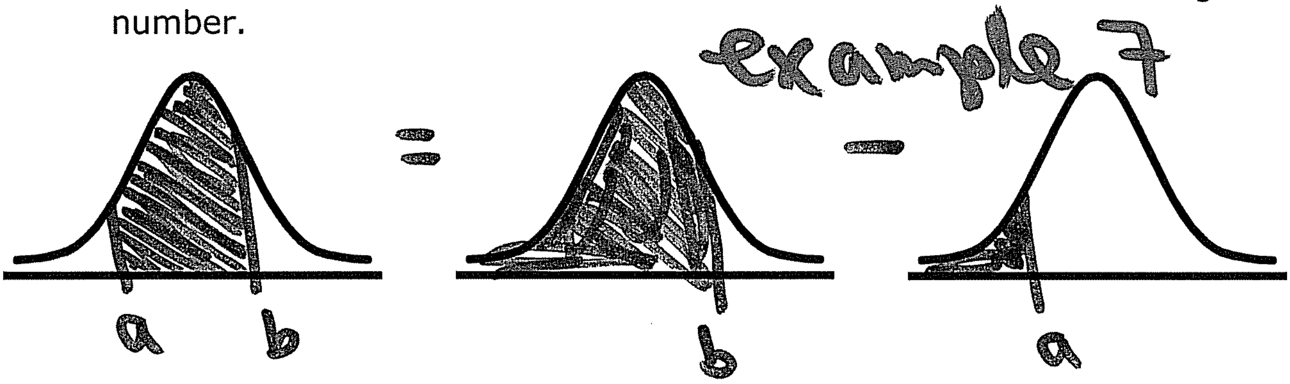
$\text{pr}(X \le x)$

- *Find an upper tail probability (area)*
  The computer cannot find/give the answer directly so subtract the lower tail from 1.



$\text{pr}(X \ge x) \quad = \quad 1 \quad - \quad \text{pr}(X \le x)$

- *Find a probability (area) between two numbers*
  The computer cannot find/give the answer directly so subtract the lower tail beneath the smaller number from the lower tail beneath the larger number.

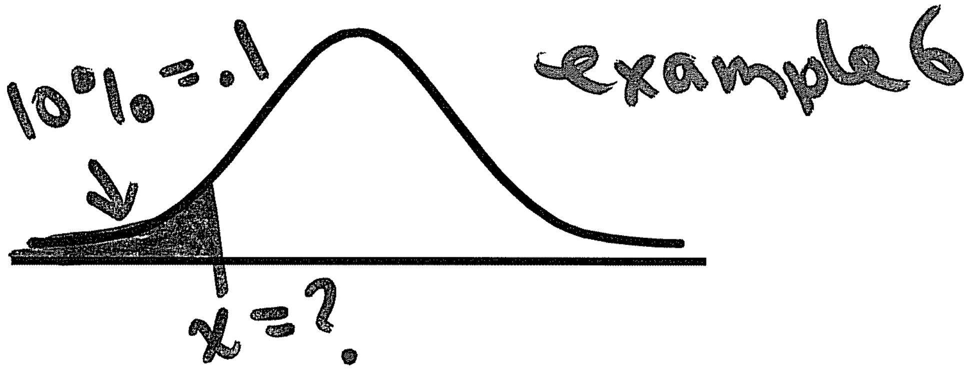


$\text{pr}(a \le X \le b) \quad = \quad \text{pr}(X \le b) \quad - \quad \text{pr}(X \le a)$

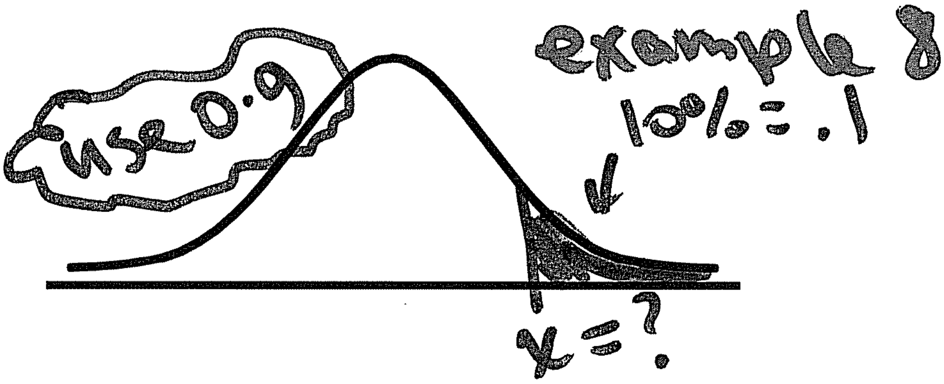
3. *Inverse Normal problem, i.e. find a number associated with a probability*

This type of problem occurs when we know the probability (e.g. the highest 10% in the class) and we need to find out the number associated with it, x (e.g. the mark). There are three types of inverse Normal problems:

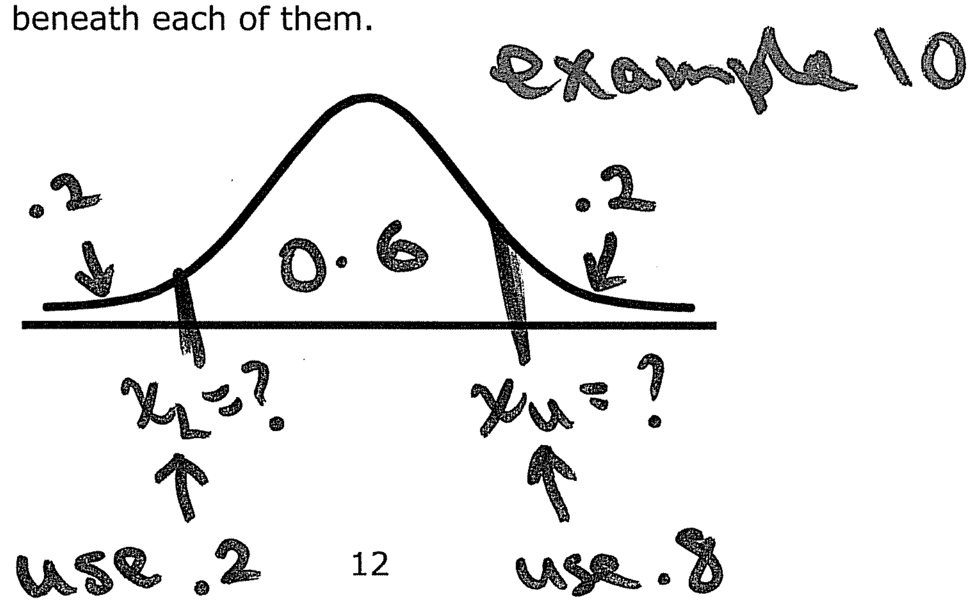
- *Given a lower tail probability, find the number associated with it*
  The computer can find/give the answer directly.



- *Given an upper tail probability, find the number associated with it*
  The computer cannot find/give the answer directly so subtract the upper tail probability from 1 & use the lower tail probability to find the answer.



- *Given a central area/probability, find the two numbers associated with it (the lower limit and the upper limit)*
  The computer can give the two limits as long as you use the lower tails/areas beneath each of them.



Examples 5 to **10** are about the following information.

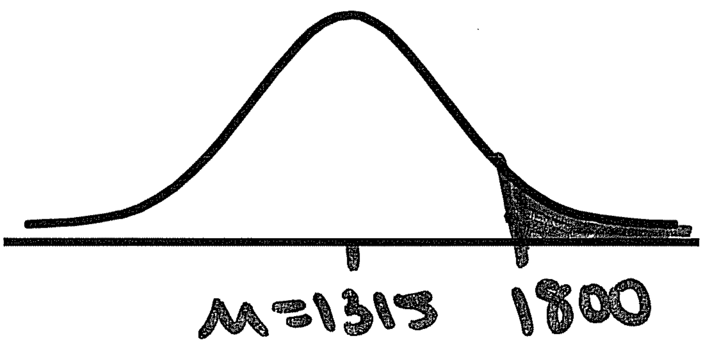
The annual rainfall recorded at the Mt. Albert weather station from 1949 to 1990 is approximately Normal with mean μ =1315 mm and standard deviation σ = 225 mm.

Normal with mean = 1315.000 and standard deviation = 225.0000

x	Pr(X ≤ x)
830	0.0156
1000	0.0808
1178	0.2713
1400	0.6472
1500	0.7945
1800	0.9844

Pr(X ≤ x)	x
0.025	874.008
0.050	944.908
0.100	1026.651
0.125	1056.171
0.250	1163.240
0.500	1315.000

Example 5: Assuming the same distribution holds for this year, ~~what is~~ the probability that this year's annual rainfall will exceed 1800 mm? [The tables given are for a Normal(1315, 225) distribution].

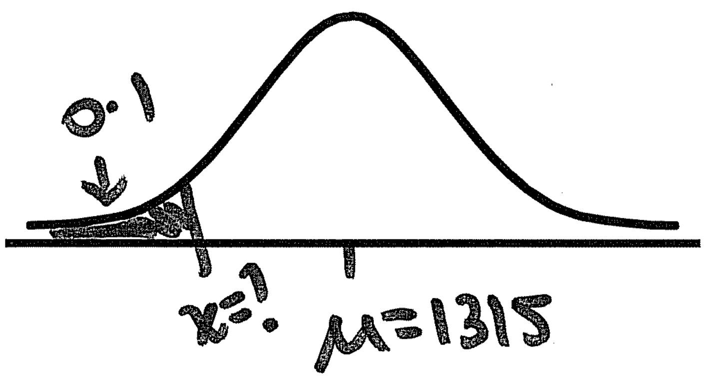


$1 - .9844$

$= .0156$ (4dp)

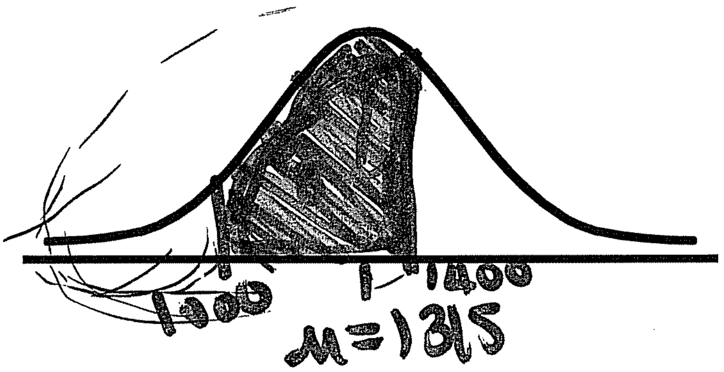
Example 6: What annual rainfall would the driest 10% of years lie below?

.1



x = 1027mm

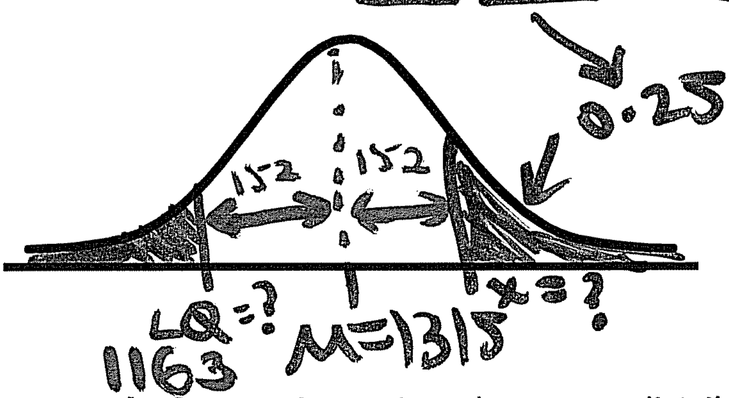
Example 7: The probability that this year's annual rainfall will be between 1000mm and 1400mm is:



.6472 − .0808

= .5664 (4dp)

Example 8: The upper quartile for the annual rainfall is:

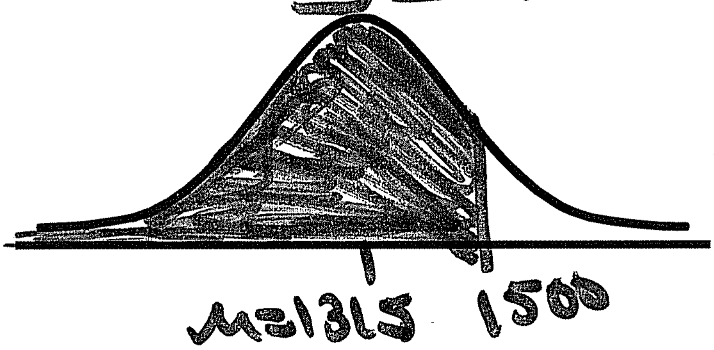


LQ = 1163mm

UQ = 1315 + 152

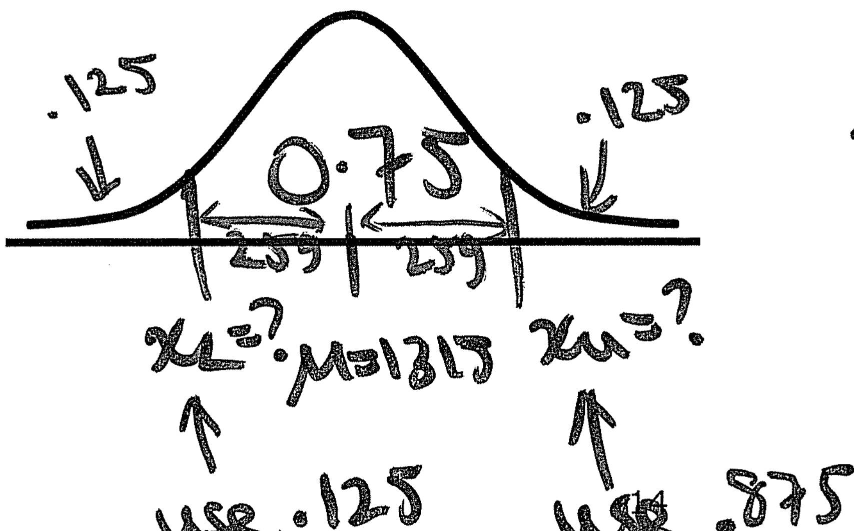
= 1467mm

Example 9: Assuming the same distribution holds for this year, what is the probability that this year's annual rainfall will be no more than 1500 mm?



.7945

Example 10: The central 75% of annual rainfall is between:



x_L = 1056 mm

x_u = 1315 + 259

= 1574 mm

Chapters 4 & 6 – Questions

Questions 1 and 2 are about the following information.

The data in the following table was drawn from a study of 507 Swedish women which was reported in the *Journal of Consulting and Clinical Psychology* (December 1989). Each woman had an aggressiveness score recorded at age 10 by their teacher and then the number of convictions for law breaking until they turned 27 was recorded 17 years later.

Number of Crimes by Age 27	**Aggressiveness Score at Age 10** Low	Medium	High	Total
0	152	291	21	464
1	6	17	3	26
2 or more	3	11	3	17
Total	161	319	27	507

1. If one of the 507 women in the study is chosen at random, the probability that she is convicted of at least one crime by the age of 27 is:

   (1) 0.018
   (2) 0.051
   (3) 0.034
   (4) 0.056
   (5) 0.085

2. Given that the women has a high score for aggressiveness by age 10, the conditional probability that she is convicted of at least one crime by the age of 27 is:

   (1) 0.085
   (2) 0.222
   (3) 0.140
   (4) 0.630
   (5) 0.286

   $Pr(\geq 1 \mid H) = \frac{(3+3)}{27}$
   $= .222$

3. Roulette is a gambling game. A wheel with 37 slots is spun. The slots are numbered from 0 to 36. The slot numbered 0 is coloured green, the even numbered slots (2, 4, 6, ...) are coloured black, and the odd numbered slots (1, 3, 5, ...) are red. While the wheel is spinning a ball is dropped onto it. When the wheel comes to rest, the ball is in one of the 37 slots. The probability that the ball lands in a red slot is:

   (1) $\frac{1}{36}$
   (2) $\frac{18}{37}$
   (3) $\frac{17}{36}$
   (4) $\frac{18}{36}$
   (5) $\frac{1}{37}$

   0 — 1
   2, 4, 6, ... — 18
   1, 3, 5, ... — 18

   18/37

Questions 4 and **5** refer to the following information.

The following table gives the distribution of the number of rooms per dwelling in the greater Auckland area, (that are controlled by the Auckland Regional Council – the residential rateable base of 278,105 dwellings). The entries in the table are the probabilities of a randomly chosen dwelling having a particular number of rooms.

Number of Rooms	Owner Occupied	Rented	Total
2 or fewer	0.004	0.003	0.007
3	0.022	0.006	0.028
4	0.078	0.061	0.139
5	0.191	0.106	0.297
6	0.230	0.024	0.254
7	0.139	0.012	0.151
8 or more	0.111	0.013	0.124
Total	0.775	0.225	1.000

4. The probability of randomly choosing an owner occupied dwelling with 5 or more rooms is:

   (1) 0.191 + 0.230 + 0.139 + 0.111
   (2) 0.004 + 0.022 + 0.078 + 0.191
   (3) 0.191
   (4) 0.297 + 0.254 + 0.151 + 0.124
   (5) 0.004 + 0.022 + 0.078

5. Given that a randomly chosen dwelling is rented, the probability that it contains 7 or more rooms is:

   (1) 0.003 + 0.006 + 0.106 + 0.024 + 0.139 + 0.111
   (2) 0.225×0.124
   (3) 0.225 + 0.124 − 0.025
   (4) $\dfrac{0.013+0.012}{0.124}$
   (5) $\dfrac{0.013+0.012}{0.225}$

5E-03 5×10^{-3} .005

6. In a country with a stable population, 0.8% of all the adult males develop lung cancer and 30% of all adult males are smokers. Studies of lung cancer patients have shown that 56.3% of adult male lung cancer patients are smokers. The probability of an adult male non-smoker developing lung cancer is approximately:

   (1) 0.003
   (2) 0.015
   (3) 0.050
   (4) 0.005
   (5) 0.437

	S	$\bar{S}$	
C	$\frac{56.3}{100} \times 80 = 45$	35	$\frac{.8}{100} \times 10{,}000 = 80$
$\bar{C}$	2955	6965	9920
	$\frac{30}{100} \times 10{,}000 = 3000$	7000	10,000

$Pr(C \mid \bar{S}) = \frac{35}{7000} = .005$

Questions 7 to **10** refer to the following information.

In a large American work place with over 6,000 employees, investigators wanted to evaluate the attitude of the employees to a no-smoking policy. Four months after a smoking ban had been implemented, a random sample of 687 employees were mailed a questionnaire which was to be answered anonymously. Out of these 687 employees, 434 (63%) returned the questionnaire. Distributions of age, gender and length of employment for the respondents were similar to those for the entire employee population. Respondents also provided information on their own smoking status and approval of the smoking ban. These results are shown in the table below.

	Approval of Smoking Ban			
Smoking Status	**Approve**	**Do Not Approve**	**Not Sure**	**Total**
Never Smoked	237	3	10	250
Ex-Smoker	106	4	7	117
Current Smoker	24	32	11	67
Total	367	39	28	434

7. Which one of the following statements is true?

    (1) The distribution of age, gender and length of employment for the respondents were similar to those of the entire population. This means that the response rate of 63% does not contribute to the non-sampling error (by way of non-response bias).

    (2) Results obtained from the sample should not be used as reliable estimates for the population parameters because the period between the date of the implementation of the smoking ban and the date of the survey (4 months) was too short for a person to form an opinion on this issue.

    (3) The reason for mailing the questionnaires was to make this a double-blind survey.

    (4) One of the reasons for the anonymity would have been to reduce the non-sampling error, which may have arisen from the behavioural considerations.

    (5) Results obtained from the sample should not be used as reliable estimates for the population parameters because a sample size of 434 out of a population of over 6,000 (less than 10%) is far too small.

8. The proportion of respondents who were current smokers and did not approve of the smoking ban is:

(1)	0.4776	(4)	0.0553
(2)	0.0253	(5)	0.0737
(3)	0.8205		

9. The proportion of ex-smokers who approved of the smoking ban is:

(1) 0.2442
(2) 0.9060
(3) 0.2888
(4) 0.7558
(5) 0.0940

10. Which one of the following statements is **false**?

T (1) Age of the respondent, classified in age groups – under 20 years, 20–39 years, 40-59 years, over 60 – is a qualitative ordinal variable.

T (2) Approval of the smoking ban, with levels as shown in the table above is a qualitative variable.

(3) Gender of the respondent, coded male = 1, female = 2, is a ~~quantitative~~ qual. variable.

T (4) Smoking status is a qualitative variable.

T (5) Length of employment of the respondent to the nearest year is a quantitative variable.

11. The NZ Herald, 9 April 1996, reported that prisoners in New Zealand would be tested for drug taking using a urine test. Suppose that the probability of having a positive test given that the person had taken drugs is 0.95 and that the probability of having a positive test given that the person had not taken drugs is 0.05. A pilot study at the Christchurch prison revealed that 18% of prisoners were on drugs. Assuming that this figure applies to all New Zealand prisoners, the probability that a New Zealand prisoner with a positive test is taking drugs is:

(1) 0.171
(2) 0.041
(3) 0.212
(4) 0.950
(5) 0.807

test	D	$\bar{D}$	
+ve	$.95 \times 1800 = 1{,}710$	$.05 \times 8{,}200 = 410$	2120
-ve	90	7,790	7880
	$\frac{18}{100} \times 10{,}000 = 1{,}800$	8,200	10,000

$$Pr(D|+ve) = \frac{1710}{2120} = .807$$

12. By examining the past records of a company's account balances, an auditor finds that 20% of all the account balances are in error. i.e. the balance is incorrect. Of those account balances in error, 40% are regarded as unusual values based on historical figures. Of all account balances, 15% have unusual values. If the ~~value for~~ a randomly selected account balance appears unusual, then the probability that this account balance is in error is:

error

(1) ~~0.08~~
(2) 0.533
(3) ~~1.000~~
(4) 0.4
(5) 0.267

	E	$\bar{E}$	
unusual U	$\frac{40}{100} \times 2000 = 800$	700	$\frac{15}{100} \times 10000 = 1500$
$\bar{U}$	1200	7,300	8,500
	$\frac{20}{100} \times 10000 = 2000$	8000	10,000

$$Pr(E|U) = \frac{800}{1500}$$

Questions 13 and **14** refer to the ~~following information~~.

A survey was taken of 901 people, and they were classified by their income, and how they rated their job satisfaction.

	Job Satisfaction				
Income (US$)	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied	Total
< 6000	20	24	80	82	206
6000 – 15, 000	22	38	104	125	289
15,000 – 25,000	13	28	81	113	235
>25,000	7	18	54	92	171
Total	62	108	319	412	901

13. What is the probability that a randomly chosen person *has income under $6,000* or is *dissatisfied*?

(1) $\dfrac{20+24}{62+108}$

(2) $\dfrac{20+24}{901}$

(3) $\dfrac{206+62+108}{901}$

(4) $\dfrac{206+62+108-20-24}{901}$

(5) $\dfrac{20+24}{206}$

14. What is the probability that a randomly chosen person has an *income under $15,000* and is *dissatisfied*?

(1) $\frac{206}{206+289+235}$

(2) $\frac{20+24+22+38}{206+289}$

(3) $\frac{20+24+22+38}{62+108}$

(4) $\frac{20}{20+22}$

(5) $\frac{20+24+22+38}{901}$

Questions 15 and **16** are about the following information.

A European study on the transmission of the HIV virus involved 470 heterosexual couples. Originally only one of the partners in each couple was infected with the virus. There were 293 couples that always used condoms. From this group, 3 of the non-infected partners became infected with the virus. Of the 177 couples who did not always use a condom, 20 of the non-infected partners became infected with the virus.

15. What proportion of couples in this study always used condoms?

(1) $\frac{290}{470}$

(2) $\frac{157}{177}$

(3) $\frac{23}{470}$

(4) $\frac{3}{293}$

(5) $\frac{293}{470}$

	C	$\bar{C}$	
HIV+	3	20	23
HIV-	290	157	447
	293	177	470

16. If a non-infected partner became infected, what is the probability that he/she was one of a couple that always used condoms?

(1) $\frac{23}{470}$

(2) $\frac{23}{293}$

(3) $\frac{3}{23}$

(4) $\frac{3}{20}$

(5) $\frac{3}{293}$

17. A study involved psychiatric hospitals in New York and London. Two hundred patients, admitted to hospitals in each city during a certain period of time, were selected at random for a comparison of diagnoses by psychiatrists. Assume the percentages are true for all patients admitted to the hospitals.

Diagnoses of 200 patients by Hospital Psychiatrists

	New York	London
Schizophrenia	75%	35%
Other diagnoses	25%	65%

×3

Which one of the following statements about the above table is **true**?

T (1) Psychiatrists in New York are 3 times more likely to diagnose schizophrenia than any other disorder. 75% 25%

(2) Conclusions from this table cannot be trusted because of the problems associated with transferring findings.

(3) Other diagnoses are roughly 3 times more common in New York hospitals than in London hospitals. 65% × 3 ≠ 25%

(4) Schizophrenia is approximately twice as common in ~~London~~ hospitals as in ~~New York~~ hospitals.

(5) London psychiatrists are 35% more likely to diagnose schizophrenia than any other disorder.

Questions 18 to **21** refer to the following information.

The height of adult women born in New Zealand is approximately Normally distributed with a mean of 161.8cm and a standard deviation of 8.4cm.

```
Normal with mean = 161.800 and standard deviation = 8.400

        x          pr( X <= x)
     145.000          0.0228
     149.400          0.0699
     156.000          0.2449
     159.000          0.3694
     161.800          0.5000
     167.000          0.7321
     174.200          0.9301
     178.000          0.9731
```

18. The proportion of New Zealand women who are at most 156 cm is:
(1) 0.3694
(2) 0.2449
(3) 0.2280
(4) 0.7551
(5) 0.6306

19. The proportion of New Zealand women who are taller than 174.2cm, the mean adult male height, is approximately:
(1) 0.440
(2) 0.429
(3) 0.070
(4) 0.930
(5) 0.154

20. We would expect the proportion of New Zealand women whose height was between 145 and 167cm to be:
(1) 0.0709
(2) 0.7093
(3) 0.0041
(4) 0.6622
(5) 0.0800

21. We would expect approximately 25% of New Zealand women to be shorter than:
(1) 156cm
(2) 145cm
(3) 178cm
(4) 167cm
(5) 159cm

Questions 22 and **23** refer to the following information.

NZ students who wish to enrol for an undergraduate degree in the United States must sit the American Scholastic Aptitude Test (SAT), which includes a test on verbal ability. Each year, the scores for the verbal test are approximately Normally distributed with a mean of 430 and a standard deviation of 100. Use the following probability table from *Excel* to answer the questions:

x	$\Pr(X \le x)$	x	$\Pr(X \le x)$
302	0.1003	550	0.8849
420	0.4602	558	0.8997
430	0.5000	570	0.9192
443	0.5517	650	0.9861
530	0.8413	850	1.0000

22. In a given year, the proportion of scores from 430 to 550 (inclusive) is approximately:

(1) 0.385
(2) 0.115
(3) 0.885
(4) 0.074
(5) 0.615

23. Some American universities have a fixed number of places for overseas students. To be accepted into the History department of a certain university, a New Zealand student must be placed in the top 10% in the verbal ability test. To the nearest whole number, a New Zealand student wishing to be accepted into this History department must therefore score at least:

(1) 650
(2) 558
(3) 302
(4) 443
(5) 850

Questions 24 & 25 refer to the following information.

The mean wingspan of the adult female albatross is 303 cm (Sorensen J.H., *The Royal Albatross*). Assume that the wingspan for the adult female is well modelled by a Normal distribution with a standard deviation of 13 cm. Use the computer output given below to answer the questions.

```
        Normal with mean = 303 and standard deviation = 13
X              P( X <= x)                 300.41         0.4210
76.00          0.0000                     301.73         0.4611
278.00         0.0272                     304.27         0.5389
279.93         0.0380                     311.77         0.7500
294.00         0.2444                     312.00         0.7556
294.23         0.2500                     326.00         0.9616
300.00         0.4087
```

24. The proportion of adult female royal albatrosses with a wingspan larger than 326 cm (the mean adult male's wingspan) is approximately:

(1) 0.962
(2) 0.038
(3) 0.421
(4) 0.539
(5) 0.461

25. Approximately 25% of adult female royal albatrosses have a wingspan less than:

(1) 300 cm
(2) 312 cm
(3) 294 cm
(4) 278 cm
(5) 76 cm

26. Which of the following statements is **false**?

(1) The shape of the Normal distribution curve is a symmetric bell-shape.
(2) 68% of data from a Normal distribution lie within 1 standard deviation of the mean
(3) 99.7% of data from a Normal distribution lie within 2 standard deviation of the mean
(4) When calculating probabilities we usually devise a way of obtaining it via the lower tail probabilities.
(5) The z-score is how many standard deviations X is away from the mean.

27. If $x = 12.5$ is an observation from a random variable X, where X is distributed as Normal($\mu_X = 15$, $\sigma_X = 10$), the z-score for X is:

(1) -25
(2) -0.25
(3) 0.994
(4) -10
(5) 40

ANSWERS

1.	**(5)**	2.	**(2)**	3.	**(2)**	4.	**(1)**	5.	**(5)**	6.	**(4)**
7.	**(4)**	8.	**(5)**	9.	**(2)**	10.	**(3)**	11.	**(5)**	12.	**(2)**
13.	**(4)**	14.	**(5)**	15.	**(5)**	16.	**(3)**	17.	**(1)**	18.	**(2)**
19.	**(3)**	20.	**(2)**	21.	**(1)**	22.	**(1)**	23.	**(2)**	24.	**(2)**
25.	**(3)**	26.	**(3)**	27.	**(2)**						