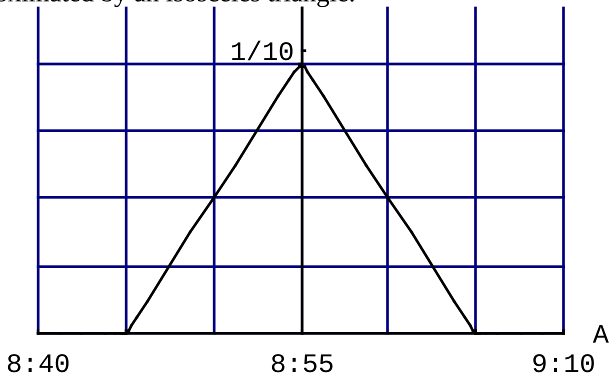
1 Let Y be a random variable whose values are all between 0 and 10, and let F be its cumulative distribution function. What is the numerical value of F(11) — F(—1) ? Why? (Q1&Q2 from MRT2)

2a A person who drives to work tries to reach the office by 9 o'clock. Because of fluctuations in traffic etc., the person actually arrives between 8:45 and 9:05. Suppose that the relative frequencies of the various arrival times ("A") could be approximated by an isosceles triangle.



- Write the appropriate label on the vertical axis.
- If this approximation is valid, what is the probability of arriving at work
  (a) on time?
  (b) at least 10 minutes early?
  (c) not more than 5 minutes early or late?

At the top right is another representation of this pattern.
• Write the appropriate label and scale on the vertical axis
• Give your answers to (a) - (c) in terms of this function.

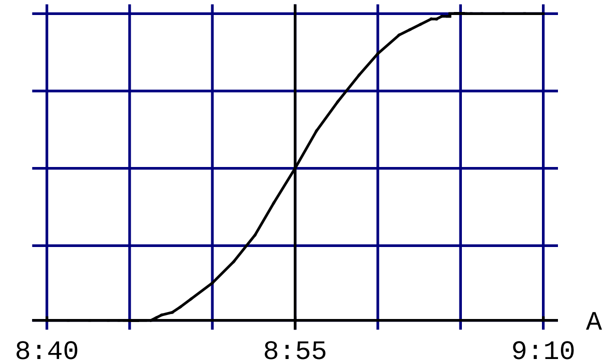
2b Suppose A is (approx.) a normal ("Gaussian") r.v. , with μ = 8:55, and σ= 5 min. (middle panel, next column)

• Under this model, and using the table* on inside front cover of WMS5, calculate probabilities (a)-(c)

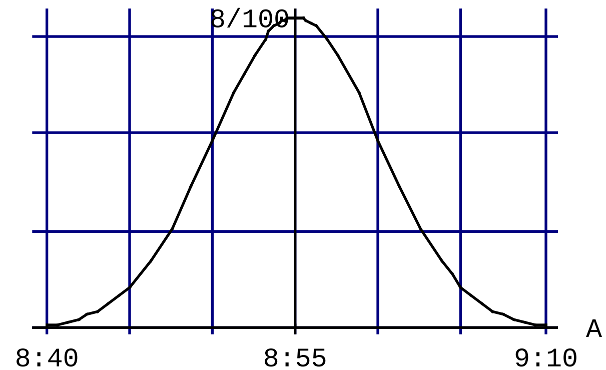
• Put a scale on the vertical axis for the bottom panel. Label the 2 functions shown (2 another representations of the one in the middle panel). How are these functions related to the WMS5 table?

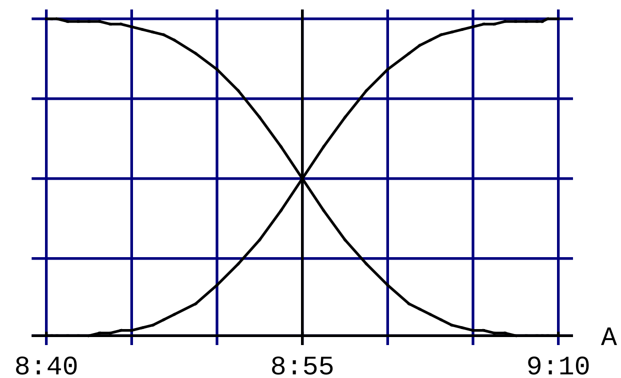
*Excerpts from Table:-
Normal curve areas: standard normal probability (prob) in right-hand tail

z:	-2.0	-1.5	-1.0	-0.5	0	0.5	1.0	1.5	2.0
prob.:	0.097	0.933	0.841	0.691	0.500	0.309	0.159	0.067	0.023



==





2c Say you were given the functional form of p.d.f. for the random variable A with the triangular distribution (panel on the left column of the previous page). How would you calculate the variance, V(A)?

• Suppose arrival times are measured as the new random variable A*, where A* is "minutes after 9:00".

• What is V(A*) in terms of V(A)?

• What is V(arrival times, in seconds) in terms of V(A*)?

2d There is a penalty, P, for being late:

$$P = \begin{cases} 0, & \text{if } A \le 9{:}00, \\ (A - 9{:}00)^2 & \text{if } A > 9{:}00 \end{cases}$$

• How would you calculate the expected value of P?

2e Assume the triangular model.

- Suppose that it is 9:00 and the person has not arrived yet. What is the probability that the person will arrive between 9:00 and 9:01; 9:01 and 9:02; etc. ?
- Suppose that it is 9:00 and the person has already arrived. What is the probability that the person arrived between 8:45 and 8:46; 8:46 and 8:47; etc. ?

2f The server "goes down" on average once every 3 hours. There is no predictable pattern to the failure rate: the probability that it will go down in the next δ minutes, given that it has been up and running for the past t minutes, is independent of "up-time" (t), clock time, and load.

- What is the probability of 0,1,2,3, .. crashes between 9:00 and 12:00? (assume that the "down-time" after each crash is negligible) [see exercise 4.108 and 4.109].
- It is 11:10, and the server has been up since 8:45. The person logs in, and wonders what is the probability that the server will crash at least once before 12:00? 1:00? 2:00?
- [More difficult, not explained in the book -- see my notes, especially the "second guess", on the gamma distribution. Is this a fair question for a quiz or the final???] What is the probability model, and associated parameters, for the length of time (from now) until the second-next crash? (again, assume that the down time after each crash is negligible)

  See exercise 4.110 in the textbook. Rather than motivate the gamma as a sum of independently and identically exponentially distributed r.v's, the textbook gives the proof of this fact as an exercise. If you were teaching the gamma distribution, is this how you would do it?

2g [see 4.112] If the durations of meetings between you and your boss follow an exponential distribution with $\beta = 1/12$ hours (i.e., the mean duration = 5 minutes) , is it more likely that it is because the boss...

- sticks to the topic, and doesn't engage in "chit-chat"
- has another meeting
- gets an important (and unplanned!) outside call and terminates the meeting

3 The following are data collected by Statistics Canada at the last census:

Montreal Metropolitan Population by knowledge of official language (percentages are rounded, for sake of exercise)

Total	English only	French only	Both English and French	Neither English nor French
3,287,645	280,205	1,309,150	1,634,785	63,500
100%	8%	40%	50%	2%

- Consider a randomly chosen person. Let E be the variable denoting knowledge(Yes/No) of English, and F knowledge (Yes/No) of French. Use these data to display the joint distribution of E and F.

- Use the data to create (i) the marginal distribution of E (ii) the marginal distribution of F.

- Are E and F independent random variables?

- Compare the probabilities that

  (a) a person who knows French also knows English

  (b) a person who knows English also knows French

- Give **E** the value 1 if the person knows English, and 0 if not. Likewise for **F**.

  (a) What is the value, and meaning, of

  E(**E**)? E(**F**)?

  *(**bold** face for numerical r.v., regular face for expectation)*

  V(**E**)? V(**F**)? [see

  (b) Compute Cov(E,F) [Use the E(**EF**) — E(**E**)E(**F**) version]

- If one randomly chose 2 persons in Montreal, what is the probability that they could understand each other in one or more of the two official languages (sign language and Franglais not allowed!)?

  Hint: Label them P1 and P2, and use the data at the beginning of the question.

The next set of questions goes back to Chapter 3 -- not for this quiz, but a possibility for the final exam!

Denote by E_1 whether the first of these randomly chosen persons knows (1) or does not know (0) English. Do the same for E_2, i.e., for the knowledge of English of the second person.

Let $T_E = E_1 + E_2$.

•What is the (i) interpretation (ii) probability distribution of T?

• If we chose 10 persons at random, and compute

$$T_E = E_1 + E_2 + \ \ldots\ + E_{10}$$

$$T_F = F_1 + F_2 + \ \ldots\ + F_{10}$$

- what is the distribution of (i) T_E (ii) T_F ?

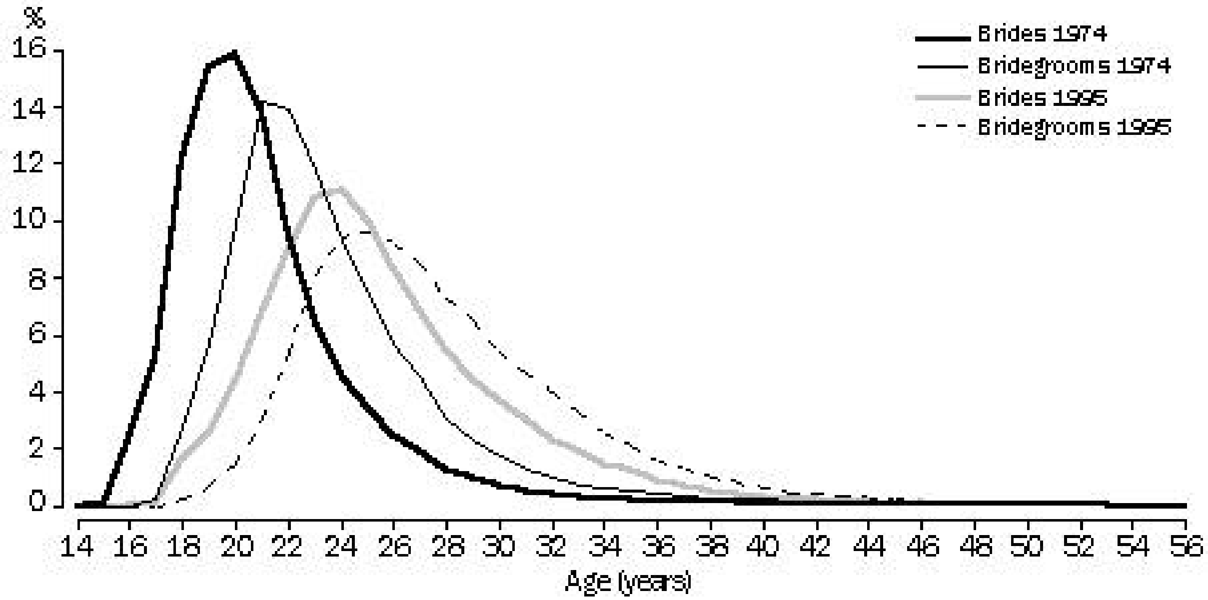
- which one (T_E or T_F) will have the "closer-to-Gaussian" distribution?

AGE DISTRIBUTION OF BRIDEGROOMS AND BRIDES AT FIRST MARRIAGE,

Australia1974 AND 1995

<u>in 1995</u>

Age...	approx SD (Var)
Brides (females)	4.8 (23)
Bridegrooms (males)	5.1 (26)



DISTRIBUTION OF THE DIFFERENCE IN AGE BETWEEN COUPLES AT FIRST MARRIAGE(a),

1974 AND 1995 (note **bold** line denotes 1995, whereas it denoted 1974 in panel above)

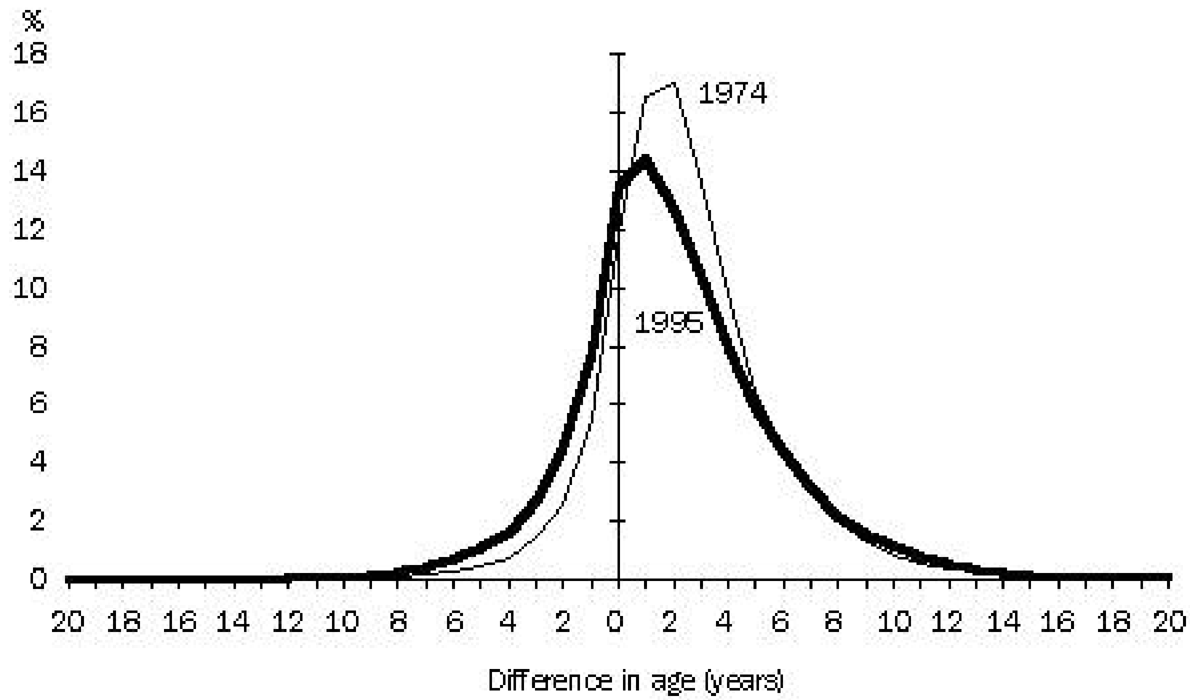
<u>in 1995</u>

Difference < 0:
Bridegroom younger than bride

Difference < 0:
Bridegroom older than bride

Age Difference(M—F)...

approx. SD (Var)
5 (25)



4 The foregoing data on age at first marriage were reported by The Australian Bureau of Statistics

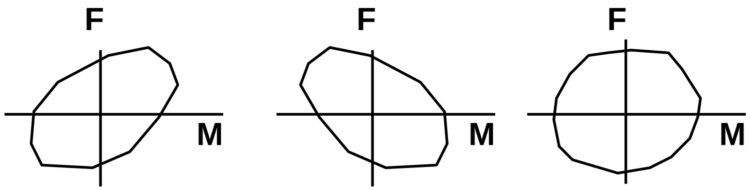
• The three measures of central tendency, the mean, the median, and the mode are always in either dictionary, or reverse dictionary, order. Which order are they in here?

• Which are more variable, the ages in 1974 or those in 1995?

For the rest of this exercise, focus on the 1995 data.

• Are the two 1995 distributions in the top panel, (i) joint (ii) marginal or (iii) conditional distributions of age at marriage?

• (From your own experience and observations in Canada--with similar patterns) Which of the following do you think most resembles the joint distribution of Age of the Bridegroom (M) and the Bride (F)?



Consider the variance of the differences (Male age—Female age) in relation to the variance of the Male ages and the variance of the Female ages.

Recall that the variance of a linear combination $U = a_1 Y_1 + a_2 Y_2$ of two random variables Y_1 and Y_2 is

$$\text{Var}(U) = a_1^2 \text{Var}(Y_1) + a_2^2 \text{Var}(Y_2) + 2\, a_1\, a_2 \text{Covar}(Y_1, Y_2)$$

From this formula, and the variances given in the figure, deduce the covariance, Covar(Age of Male, Age of Female).

From this, deduce the approx. correlation of the Age of the Male and the Age of the Female. Does that fit with your guess above?

$$\text{correlation coefficient} = \frac{\text{Covar}(Y_1, Y_2)}{\text{SD}(Y_1)\ \text{SD}(Y_2)}$$