Math 2200 Spring 2016, Exam 1

You may use *any* calculator. You may use ONE "cheat sheet" in the form of a 4" x 6" note card (the medium size of the standard three sizes).

1. This question and the next two pertain to a University of Texas study in which 626 surveyees were categorized in two ways. Each surveyee was placed in category CT if he or she had a tattoo obtained in a commercial tattoo parlor, in category AT if he or she had a tattoo obtained elsewhere (a prison tattoo, for example), or in category NT if he or she had no tattoo. No surveyee fell into both classes CT and AT. Each surveyee was also categorized according to whether he or she had hepatitis C. The joint frequency counts can be found in the following table.

	CT	AT	NT	Total
Has hepatitis C	17	8	18	43
Does not have hepatitis C	35	53	495	583
Total	52	61	513	626

What percentage of the surveyees were tattooed?

A) 12.377 B) 15.214 C) 18.051 D) 20.888 E) 23.725
F) 26.562 G) 29.399 H) 32.236 I) 35.073 J) 37.910

Answer: C) 18.051

Solution

There were $52 + 61$, or 113, tattooed surveyees. The percentage was $113 \times 100/626\%$, or 18.051%.

2. What percentage of the surveyees with hepatitis C were tattooed?

A) 42.929 B) 45.102 C) 47.275 D) 49.448 E) 51.621
F) 53.794 G) 55.967 H) 58.140 I) 60.313 J) 62.486

Answer: H) 58.140

Solution

There were 43 surveyees with hepatitis C. Of these, $17 + 8$, or 25, were tattooed. The percentage was $25 \times 100/43\%$, or 58.140%.

3. What percentage of the surveyees with tattoos had hepatitis C?

A) 20.051 B) 22.124 C) 24.197 D) 26.270 E) 28.343
F) 30.416 G) 32.489 H) 34.562 I) 36.635 J) 38.708

Answer: B) 22.124

Solution

There were $52 + 61$, or 113, tattooed surveyees. Of these, $17 + 8$, or 25, had hepatitis C. The percentage was $25 \times 100/113\%$, or 22.124%.

4. Among other questions in the 2008 General Social Survey, 1993 surveyees were asked to describe their family income with one of the following three levels: Above average, Average, Below average. Surveyees were also asked to describe their level of happiness with one of the following three levels: Not too happy, Pretty happy, Very happy. The following contingency table resulted:

	Not too happy	Pretty happy	Very happy
Above average	26	233	164
Average	117	473	293
Below average	172	383	132

This problem and the two that follow pertain to this contingency table. Suppose that the conditional distributions of the categorical variable Happiness Level were used to determine if the variables Happiness Level and Family Income Level were independent. Which one of the following numbers would be a percentage arising in the determination?

A) 13.250 B) 16.767 C) 20.284 D) 23.801 E) 27.318
F) 30.835 G) 34.352 H) 37.869 I) 41.386 J) 44.903

Answer: A) 13.250

Solution

The conditional distributions of the categorical variable Happiness Level are the rows of the given table. The first row is the conditional distribution of Happiness Level conditioned on the value of Family Income Level being above average. The second row is the conditional distribution of Happiness Level conditioned on the value of Family Income Level being average. The third row is the conditional distribution of Happiness Level conditioned on the value of Family Income Level being below average. To use these conditional distributions to investigate whether the categorical variables Happiness Level and Family Income Level are independent, we calculate row percentages. There are 423 observations in the first row. Multiplying each first row cell entry by 100/423, we obtain 6.146, 55.083, 38.771 for the row percentages. These values are not really close to any of the answer choices, so we proceed to the second row. There are 883 observations in the second row. Multiplying each first row cell entry by 100/883, we obtain 13.250, 53.567, 33.182 for the row percentages. The first of these percentages is an answer choice, so we may stop without considering the third row. At this point we can see, in any event, that the two categorical variables are *not* independent.

5. Consider the conditional distribution of Family Income Level for the value Very Happy of the variable Happiness Level. When this conditional distribution is expressed in terms of relative frequencies, what number is in the middle cell?

A) 0.163 B) 0.205 C) 0.247 D) 0.288 E) 0.330
F) 0.372 G) 0.414 H) 0.455 I) 0.497 J) 0.539

Answer: I) 0.497

Solution

The conditional distribution of Family Income Level for the value Very Happy of the variable Happiness Level is, when written horizontally to save space,

	Above average	Average	Below average
Family Income Level	164	293	132

The sum of the three cell entries is 589. Dividing each cell entry by 589 gives us a distribution of relative frequencies, of which the middle cell entry, 0.497, is our answer:

	Above average	Average	Below average
Family Income Level	0.278	0.497	0.224

6. Which of the following numbers is a frequency count that appears in the marginal distribution of the categorical variable Happiness Level?

   A) 204 B) 241 C) 278 D) 315 E) 352
   F) 389 G) 426 H) 463 I) 500 J) 537

Answer: D) 315

Solution

The frequency of each value of Happiness Level is the total of its joint frequencies with the variable Family Income Level. Thus, the marginal distribution of Happiness Level is obtained by summing each column:$26 + 117 + 172 = 315$, $233 + 473 + 383 = 1089$, $164 + 293 + 132 = 589$.

	Not too happy	Pretty happy	Very happy
Happiness Level	315	1089	589

Any one of these cell entries would anser the question, but only 315 is an answer choice.

7. A statistics class is divided into two sections. The following (incomplete) table is the contingency table for the two categorical variables Section and Gender:

	F	M	Total
Section 1	66		
Section 2			171
Total			285

What number of female students in Section 2 provides the *strongest* evidence for the independence of the variables Section and Gender?

   A) 93 B) 94 C) 95 D) 96 E) 97
   F) 98 G) 99 H) 100 I) 101 J) 102

Answer: G) 99

Solution

We first complete the marginal column with the missing cell entry, $285 - 171$, or 114, in the first row. This allows us to calculate the missing cell entry, $114 - 66$, or 48, in the second column of the first row. Our contingency table is now

	F	M	Total
Section 1	66	48	114
Section 2			171
Total			285

The two categorical variables are independent if the rows of the table are (nearly) proportional (so that if the frequencies in each row are replaced by their row relative frequencies, then rows that result are (nearly) identical). Let r_1 denote the missing cell entry under 66 and let r_2 denote the missing cell entry under 48. Then, for some unknown proportionality constant λ, which we will soon eliminate, we have $r_1 = 66\,\lambda$ and $r_2 = 48\,\lambda$. It follows that $r_2 = (48/66)\,r_1$, or $r_2 = (8/11)\,r_1$. Additionally, because the second row total is 171, we have $r_1 + r_2 = 171$. Substituting for r_2 in this last equation, we obtain $r_1 + (8/11)\,r_1 = 171$, or $(19/11)\,r_1 = 171$, or $r_1 = (11/19)\,171$, or $r_1 = 11\,(171/19)$, or $r_1 = 11 \times 9 = 99$.

8. This problem and the next pertain to he following incomplete table, which provides data for the results of an exam taken by two sections of a calculus class at First President University. The variables N_F and N_M represent the numbers of females and males, respectively, per group (Section1, Section 2, or Sections 1 and 2 combined). The variable A_F (respectively A_M) represents the average score attained by the female (respectively male) students per group (Section1, Section 2, or Sections 1 and 2 combined).

	N_F	A_F	N_M	A_M
Section 1	40	76		75
Section 2	80	70	33	68
Sections 1 & 2	120			

What was the value of A_F in the last row?

A) 70.5 B) 71.0 C) 71.5 D) 72.0 E) 72.5
F) 73.0 G) 73.5 H) 74.0 I) 74.5 J) 75.0

Answer: D) 72.0

Solution

The number of points obtained by the female students in Section 1 was 40×76, or 3040. The number of points obtained by the female students in Section 2 was 80×70, or 5600. The total number of points obtained by the $40 + 80$ female students was $3040 + 5600$, or 8640. Their average was $8640/120$, or 72.

9. According to the data of the preceding problem, the inequality $A_F > A_M$ is true for each of Sections 1 and 2. However, the unspecified value of N_M for Section 1 was such that, had it been filled in so that the table could have been completed, the values A_F and A_M for Sections 1 and 2 combined would have satisfied the reverse inequality $A_M > A_F$. What is the smallest value the variable N_M might have had for Section 1?

A) 42 B) 45 C) 48 D) 51 E) 54
F) 57 G) 60 H) 63 I) 66 J) 69

Answer: B) 45

Solution

Let n be the value of N_M for Section 1. Then the total number of points obtained by the male students in the class was $n \times 75 + 33 \times 68$ and the value of A_M for the combination of both sections was $(n \times 75 + 33 \times 68)/(n + 33)$. The given inequality, $A_M > A_F$ for combined sections 1 and 2, together with the value of $A_F = 72$ for the combined sections (as found in the preceding problem) leads to the inequality

$$\frac{n \times 75 + 33 \times 68}{n + 33} > 72,$$

which gives $n \times 75 + 33 \times 68 > 72\,(n + 33)$, or $(75 - 72)\,n > 72 \times 33 - 68 \times 33$, or $3\,n > 4 \times 33$, or $n > 4 \times 11$, or $n > 44$, or $n \geq 45$.

10. This problem and the next three pertain to the following sorted observations, which represent the number of milligrams of sodium per serving for 24 types of breakfast cereal: 0, 35, 50, 55, 70, 100, 130, 140, 140, 150, 160, 180, 180, 180, 190, 200, 200, 200, 210, 210, 220, 290, 320, 340. Determine the range of the given data and bin the data so that the class width is equal to $1 + \frac{1}{10}$ range. What is the largest upper class limit if the least lower class limit is 0 and the rightmost bin is not empty?

    A) 340 B) 341 C) 342 D) 343 E) 344
    F) 345 G) 346 H) 348 I) 350 J) 351

Answer: I) 350

Solution

The range is $340-0$, or 340. The class width is $1+340/10$, or 35. If the least lower class limit is 0 and the rightmost bin is not empty, then the bins are $[0, 35)$, $[35, 70)$, $[70, 105)$, $[105, 140)$, $[140, 175)$, $[175, 210)$, $[210, 245)$, $[245, 280)$ The largest upper class limit is 350.

11. In this problem we will use the term "mode" in the sense that has been adapted to histograms, but not in its loosest sense. When the data in problem 10 is binned according to the specifics stated in problem 10, and according to the general conventions we have adopted for the course, what is the height of the bar with base that is the mode?

    A) 4 B) 5 C) 6 D) 7 E) 8
    F) 9 G) 10 H) 11 I) 12 J) 13

Answer: D) 7

Solution

The counts for the 10 bins listed in the preceding problem are 1, 3, 2, 1, 4, 7, 3, 0, 1, 2. The largest of these numbers is the answer to the question.

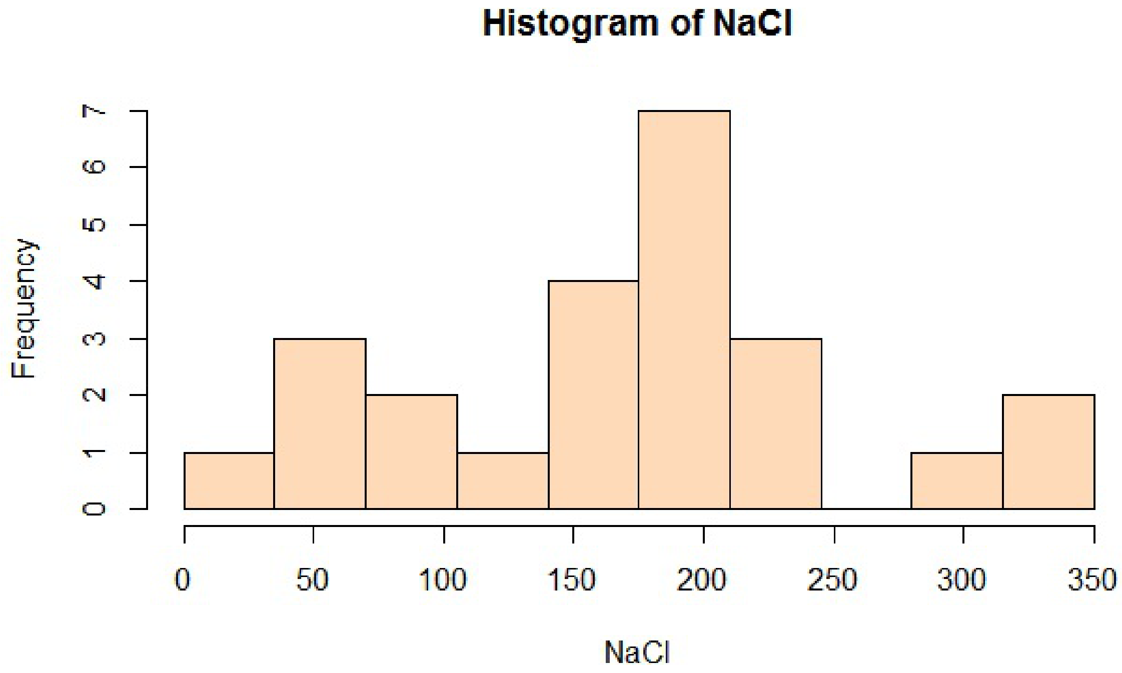
12. The distribution of the sodium observations in problem 10 is

    A) unimodal and skewed left B) unimodal and symmetric
    C) unimodal and skewed right D) bimodal and skewed left
    E) bimodal and symmetric F) bimodal and skewed right
    G) multimodal and symmetric H) multimodal and asymmetric
    I) uniform J) none of the preceding shapes

Answer: H) multimodal and asymmetric

Solution

The bin counts given in the preceding problem suffice to lead us to an answer, but one picture is worth n words for any value of n, so here is the picture generated by `hist(NaCl, breaks = nodes, right = FALSE, col = "peachpuff")` where `NaCl` is the vector of data, `nodes` is the vector of class limits, and `"peachpuff"` is the lovely color you are looking at, if your display is not monochrome. (Note: the terms *unimodal*, *bimodal*, and *multimodal* refer to *mode* in its least strict sense: that of a local maximum. There are three modes when that definition is used.)



13. The average $\overline{\text{NaCl}}$ of the sodium observations in problem 10 is 164.5833. If we did not have a list of the 24 observations but had instead only the histogram specified in problem 10, then we could estimate $\overline{\text{NaCl}}$ by assuming that, for each bin, the average of the observations that fell in that bin was equal to the class mark of that bin. Then we could easily sum the observations in each bin, sum these bin totals, and divide the last sum by 24 to obtain an estimate of $\overline{\text{NaCl}}$. Using this method of estimation, what would be the bin total for the fifth bin, counting from the left?
A) 610 B) 614 C) 618 D) 622 E) 626
F) 630 G) 634 H) 638 I) 642 J) 646

Answer: F) 630

Solution

Going back to the solution of Problem 10, we see that the fifth bin when counting from the left is $[140, 175)$. Its class mark is $(140 + 175)/2$, or 157.5. Going back to the solution of Problem 11, we see that 4 observations fall into the fifth bin from the left. So the contribution from bin 5 to the total is 4×157.5, or 630. (Not that it was asked, but the actual contribution from bin 5 is $140+140+150+160$, or 590.)

14. Suppose that X is a numerical variable with N data values $x_1, x_2, \ldots, x_N$. Let $\overline{X}$ be the mean of X and set $Y = X - \overline{X}$. This means that the data values $y_1, y_2, \ldots, y_N$ of Y are given by the formula $y_j = x_j - \overline{X}$ for $1 \leq j \leq N$. Which of the 8 statistical measures among the answer choices *must* be the same for X and Y? (At least one of the listed measures s the same for X and Y, but no more than three of the measures are the same. Read all answer choices. If only one of the statistical measures is a correct answer, then choose the appropriate letter from (A) to (H). Otherwise, answer with either (I) or (J).)

A) mean	B) median
C) mode	D) lower quartile
E) upper quartile	F) IQR
G) the variance	H) the standard deviation
I) Exactly two of the cited measures	J) Exactly three of the cited measures

Answer: J) Exactly three of the cited measures

Solution

Notice that the word "must" has been emphasized in the statement of the problem. It may happen that for a special value of the mean of X the given statistic has the same value for X and Y. Clearly, if the mean of X is 0 then X and Y are one and the same, hence every statistic has the same value for X and Y. But knowing only what you are given, which means that you do not know the mean of X is 0, you cannot say that every statistic *must* have the same value for X and Y. From what we have been told, we cannot dispute that the mean of X can have any value. But the mean of Y must be 0. In particular, X and Y need not have the same mean. The observation or observations of X that yield the median of X transform to the observation or observations of Y that yield the median of Y. But, in general, the transformation from X to Y results in a different location for each observation, so X and Y need not have the same median. Any value that gives a mode for X results in a transformed value of Y that gives a mode for Y. But the transformed value of Y is, in general, not equal to the value in X that generated it. Thus, the mode is out. So are the lower and upper quartiles, for the same reason the median is eliminated. However the lower and upper quartiles of Y are obtained by shifting the lower and upper quartiles of X by the same amount $\overline{X}$. When the difference $Q_3 - Q_1$ is calculated, the shifts cancel in the subtraction, so IQR must be the same for X and Y. Similarly, when we calculate a deviation $y_j - \overline{Y}$, we have $y_j - \overline{Y} = \left(x_j - \overline{X}\right) - 0$, or $x_j - \overline{X}$. So X and Y have the same deviations from their means. Hence, X and Y have the same standard deviations and variances. Count 'em: IQR, standard deviation, variance. That makes three.

But these

15. Let X denote the distribution 1, 3, 4, 6, 7, 11, 12, 13, $x_{[9]}$, $x_{[10]}$, $x_{[11]}$ (given in nondecreasing order). If the IQR of X is 10, then what is x_9?

A) 13.5	B) 14	C) 14.5	D) 15	E) 15.5
F) 16	G) 16.5	H) 17	I) 17.5	J) 18

Answer: H) 17

Solution

The size of the distribution is 11, which is odd. Hence, there is a middle value, $x_{[6]}$, that is the median: $Q_2 = x_{[6]} = 11$. The lower quartile Q_1 is the median of 1, 3, 4, 6, 7, 11, which is the average of 4 and 6, namnely 5. The upper quartile Q_3 is the median of 11, 12, 13, $x_{[9]}$, $x_{[10]}$, $x_{[11]}$, which is the average of 13 and $x_{[9]}$. On the other hand, $Q_3 = Q_1 + \text{IQR} = 5 + 10 = 15$. It follows that $\frac{1}{2}\left(13 + x_{[9]}\right) = 15$, or $x_{[9]} = 2 \times 15 - 13 = 17$.

16. This problem and the next one pertain to a distribution of size 6 for which five of the deviations from the mean are -8, -3, 0, 1, and 1. What is the sixth deviation from the mean?

A) 1	B) 2	C) 3	D) 4	E) 5
F) 6	G) 7	H) 8	I) 9	J) 10

Answer: I) 9

Solution

The sum of all deviations from the mean is 0. Always. The sum of the given deviations from the mean is -9. So the missing deviation from the mean must be +9.

17.) What is the standard deviation of the distribution described in problem 16?

A) 3.662	B) 4.143	C) 4.624	D) 5.105	E) 5.586
F) 6.067	G) 6.548	H) 7.029	I) 7.510	J) 7.991

Answer: E) 5.586

Solution

We calculate

$$\text{sd} = \sqrt{\frac{1}{6-1}\left((-8)^2 + (-3)^2 + (0)^2 + (1)^2 + (1)^2 + (9)^2\right)} = 5.586.$$

18. The mean of the first 50 observations of a distribution X is 20 and the mean of the remaining 30 observations of X is 12. What is the mean of X?

A) 16.25	B) 16.5	C) 16.75	D) 17	E) 17.25
F) 17.5	G) 17.75	H) 18	I) 18.25	J) 18.5

Answer: D) 17

Solution

Because

$$\frac{x_1 + x_2 + \cdots + x_{50}}{50} = 20 \qquad \text{and} \qquad \frac{x_{51} + x_{52} + \cdots + x_{80}}{30} = 12,$$

we see that $x_1 + x_2 + \cdots + x_{50} = 50 \times 20 = 1000$ and $x_{51} + x_{52} + \cdots + x_{80} = 30 \times 12 = 360$. The mean of X is given by

$$\overline{X} = \frac{x_1 + x_2 + \cdots + x_{80}}{80} = \frac{(x_1 + x_2 + \cdots + x_{50}) + (x_{51} + x_{52} + \cdots + x_{80})}{80} = \frac{1000 + 360}{80} = 17.$$

19. This problem and the next pertain to the horizontal Tukey boxplot for the data set 17, 22, 27, 28, 29, 30, 31, 33, 34, 34. At what number is the fence on the right drawn?

A) 37.5	B) 38	C) 38.5	D) 39	E) 39.5
F) 40	G) 40.5	H) 41	I) 41.5	J) 42

Answer: J) 42

Solution

The median of 17, 22, 27, 28, 29, 30, 31, 33, 34, 34 is (29 + 30)/2, or 29.5. The lower quartile, Q_1, is the middle value, 27, of the smallest five observations. The upper quartile, Q_3, is the middle value, 33, of the largest five observations. The interquartile range is given by $\text{IQR} = Q_3 - Q_1 = 33 - 27 = 6$. The right fence is drawn at $Q_3 + 1.5 \times \text{IQR}$, or $33 + 1.5 \times 6$, or 42.

20. For the Tukey boxplot of the preceding problem, how long is the whisker on the left?

A) 5	B) 5.5	C) 6	D) 6.5	E) 7
F) 7.5	G) 8	H) 8.5	I) 9	J) 9.5

Answer: A) 5

Solution

The left fence is at $Q_1 - 1.5 \times \text{IQR}$, or $27 - 1.5 \times 6$, or 18. The whisker extends leftward from Q_1 to the smallest observation that is not smaller than 18. That observation is 22. The length of the left whisker is $27 - 22$, or 5.

21. One evening Alison, who has an interest in mood-altering drugs, and Cosima, a biology geek, each took an exam. They obtained identical class percentiles. Alison's grade in her pharmacology class was 22 on an exam with mean 17.7 and standard deviation 6.43. The mean and standard deviation on Cosima's evo-devo exam were 76.4 and 10.02 respectively. Assuming that both exam results could be modelled by a normal distribution, what was Cosima's exam grade?

A) 80	B) 81	C) 82	D) 83	E) 84
F) 85	G) 86	H) 87	I) 88	J) 89

Answer D) 83

Solution

Alison's z-score was $(22-17.7)/6.43$, or 0.6687403. Because Cosima's exam result x has the same z-score, it satisfies the equation $0.6687403 = (x - 76.4)/10.02$, or $x = 76.4 + (0.6687403)(10.02)$, or $x = 83.1$.

22. This problem and the next two pertain to the fuel efficiency of a large fleet of vehicles that is modelled by the normal distribution with mean 28 mpg and standard deviation 6 mpg. What percentage of vehicles in the fleet have a fuel efficiency greater than 20 mpg?

A) 86.5	B) 87.4	C) 88.3	D) 89.2	E) 90.0
F) 90.9	G) 91.8	H) 92.6	I) 93.5	J) 94.4

Answer: F) 90.9

Solution

The z-score of 20 is $(20 - 28)/6$, or -1.333. The fraction of the fleet with greater fuel efficiency is $1 - \Phi(-1.333)$, or $1 - \big(1 - \Phi(1.333)\big)$, or $\Phi(1.333)$. Using the given table, we calculate

$$\begin{aligned}
\Phi(1.333) &= \Phi\left(1.33 + \frac{3}{10}\,(0.01)\right) \\
&= \Phi\left(1.33 + \frac{3}{10}\,(1.34 - 1.33)\right) \\
&\approx \Phi(1.33) + \frac{3}{10}\,(\Phi(1.34) - \Phi(1.33)) \\
&= 0.9082 + \frac{3}{10}\,(0.9099 - 0.9082) \\
&= 0.90871.
\end{aligned}$$

In R, the command `pnorm( (20 - 28)/6, lower.tail = FALSE)*100` returns 90.87888.

23. What percentage of vehicles in the fleet have a fuel efficiency between 26 mpg and 32 mpg?

A) 36.061	B) 36.934	C) 37.807	D) 38.680	E) 39.553
F) 40.426	G) 41.299	H) 42.172	I) 43.045	J) 43.918

Answer: C) 37.807

Solution

The z-scores of 26 and 32 are $(26 - 28)/6$, or -0.3333, and $(32 - 28)/6$, or 0.6667. The answer we seek is $\Phi(0.6667) - \Phi(-0.3333)$. Let us calculate $\Phi(0.6667)$ first:

$$\begin{aligned}
\Phi(0.6667) &= \Phi\left(0.66 + \frac{67}{100}(0.01)\right) \\
&= \Phi\left(0.66 + \frac{67}{100}(0.67 - 0.66)\right) \\
&\approx \Phi(0.66) + \frac{67}{100}\left(\Phi(0.67) - \Phi(0.66)\right) \\
&= 0.7454 + \frac{67}{100}(0.7486 - 0.7454) \\
&= 0.747544.
\end{aligned}$$

Next, let us calculate $\Phi(0.3333)$:

$$\begin{aligned}
\Phi(0.3333) &= \Phi\left(0.33 + \frac{33}{100}(0.01)\right) \\
&= \Phi\left(0.33 + \frac{33}{100}(0.34 - 0.33)\right) \\
&\approx \Phi(0.33) + \frac{33}{100}\left(\Phi(0.34) - \Phi(0.33)\right) \\
&= 0.6293 + \frac{33}{100}(0.6331 - 0.6293) \\
&= 0.630554.
\end{aligned}$$

The requested percentage is the percentage corresponding to the proportion $\Phi(0.6667) - \Phi(-0.3333)$, or $\Phi(0.6667) - \big(1 - \Phi(0.3333)\big)$, or $0.747544 - \big(1 - 0.630554\big)$, or 0.378098. The answer is 37.81% In R, the command `( pnorm( (32 - 28)/6 ) - pnorm( (26 - 28)/6) )*100` returns 37.80661.

24. In mpg, what is the fuel efficiency of a vehicle that is more efficient than 70% of the vehicles in the fleet?

A) 30.630	B) 31.146	C) 31.662	D) 32.178	E) 32.694
F) 33.210	G) 33.726	H) 34.242	I) 34.758	J) 35.274

Answer: B) 31.146

Solution

First we find the z-score of such a vehicle by solving $\Phi(z) = 0.70$. From the table, we see that $\Phi(0.52) = 0.6985$ and $\Phi(0.53) = 0.7020$. From a practical point of view, the answer choices are spread far enough apart that we could just split the difference between these two z-scores and take z = 0.525 for our estimated z-score. That would lead to a raw score x that satisfies (x - 28)/6 = 0.525, or x = 31.15. Answer B rounds to this number. If we wanted to be more accurate, we would get a better approximation to the z-score by interpolating:

$$z = 0.52 + \frac{0.7000 - 0.6985}{0.7020 - 0.6985}(0.01) = 0.5242857.$$

This more precise z-score leads to (x - 28)/6 = 0.5242857, or x = 31.14571. In `R`, the appropriate call, `qnorm(0.70, mean = 28, sd = 6)` returns 31.1464 with a minimum of fuss.

25. For a standard normal distribution and any real number z, let $\Phi(z)$ denote the fraction of observations in the distribution that do not exceed z. Suppose that X is a large distribution that follows a normal model with mean 16 and standard deviation 1/2. Using $\Phi(z)$ for an appropriate value of z (or for appropriate values of z), what fraction of observations in X fall between 14 and 18?

A) $\Phi(1/2)$	B) $1 - \Phi(1/2)$	C) $2\,\Phi(1/2) - 1$	D) $\Phi(2)$	E) $1 - \Phi(2)$
F) $2\,\Phi(2) - 1$	G) $\Phi(4)$	H) $1 - \Phi(4)$	I) $2\,\Phi(4) - 1$	J) $\Phi(18) - \Phi(14)$

Answer: I) $2\Phi(4) - 1$

Solution

The z-scores of 14 and 18 are $(14-16)/(1/2)$, or -4 and $(18-16)/(1/2)$, or 4. The fraction of observations within 4 standard deviations of the mean is $2\,\Phi(4) - 1$.

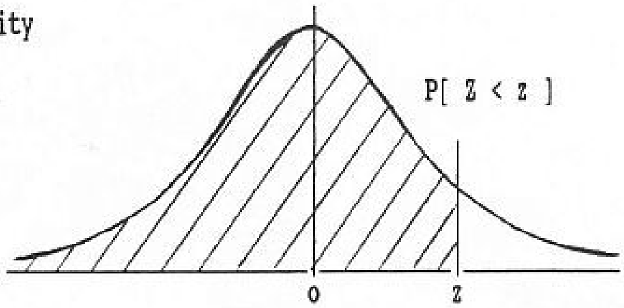
STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z

i.e.

$$P[\, Z < z \,] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7020	0.7054	0.7089	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000