ⓘ **This quiz has been regraded; your score was affected.**

2020 Quiz 1

Due No due date **Points** 22 **Questions** 21
Available Jul 9 at 13:00 - Jul 9 at 15:00 about 2 hours
Time Limit None

Instructions

This quiz is open from 1pm AEST to 3pm AEST.

You may refer to your lecture notes in answering the quiz.

Answers must be your own work.

Plagiarism declaration

By submitting work for this quiz I hereby declare that I understand the University's policy on [academic integrity (https://academicintegrity.unimelb.edu.au/)](https://academicintegrity.unimelb.edu.au/) and that the work submitted is original and solely my work, and that I have not been assisted by any other person (collusion) apart from where the submitted work is for a designated collaborative task, in which case the individual contributions are indicated. I also declare that I have not used any sources without proper acknowledgment (plagiarism). Where the submitted work is a computer program or code, I further declare that any copied code is declared in comments identifying the source at the start of the program or in a header file, that comments inline identify the start and end of the copied code, and that any modifications to code sources elsewhere are commented upon as to the nature of the modification.

This quiz was locked Jul 9 at 15:00.

Attempt History

	Attempt	Time	Score	Regraded
LATEST	Attempt 1	112 minutes	18 out of 22	19 out of 22

Correct answers are no longer available.

Score for this quiz: **19** out of 22

Submitted Jul 9 at 14:52

This attempt took 112 minutes.

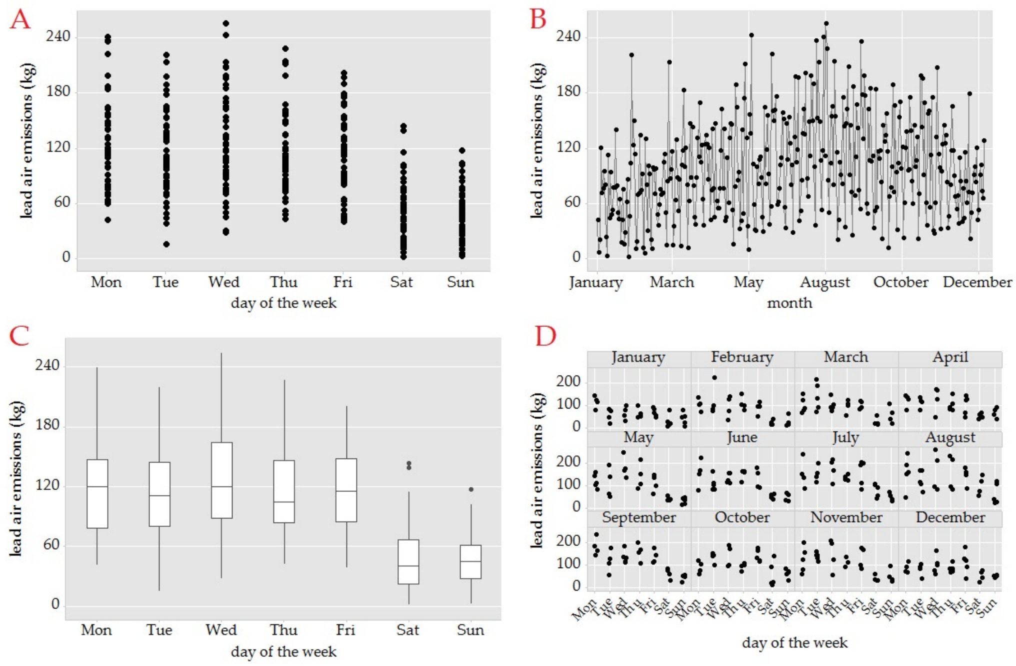
Question 1

1 / 1 pts

Your boss gives you the data in lead_smelter.mwx, and asks you to produce a graph to help examine the question of which day of the week has the "best" emissions.

The emissions referred to here consist of pollution, which is desirable to avoid. This means that low levels of emissions are "best".

You explore the data and produce four different graphs, shown below.



Which one of the graphs above is best suited to this purpose? (The graphs are labelled A to D.)

- ○ B
- ○ D
- ○ A
- ◉ C

We are interested in how emissions vary according to day of the week, as is shown on the graphs on the left. The boxplot is preferable as it shows features of the distributions on each day that help with comparing the days. The dots on the dotplot are overlaid so we cannot make the comparison as readily.

Question 2

1 / 1 pts

Your boss examines the graphs above and asks you "Which day is the best? That is, which day has the lowest total emissions?"

Based on your examination of the graphs above, which one of the following statements is correct?

- One of the weekend days (Saturday or Sunday) will be best as emissions are clearly lower overall on the weekend days than on any of the weekdays (Monday to Friday).

- One of the weekdays will be best because emissions are generally higher on the weekdays than on weekends.

Sunday is best because it has the smallest maximum.

Saturday is best because it has the smallest median value.

Saturday is best because it has the smallest minimum value.

Based on the graphs, one of the weekend days will have the lowest total emissions. You need to consider the distribution of emissions on each day rather than just particular values such as the minimum or maximum. The lowest total emissions corresponds to the day with the lowest average emissions, and the lowest median does not necessarily correspond to the lowest average.

Question 3

1 / 1 pts

Now use Minitab to obtain some summary statistics to describe daily emissions by day of the week.

Examine the summary statistics. Again, assume we wish to minimize total emissions. Based on the summary statistics, on what day of the week does the plant generally operate at its "best"?

- ○ Friday
- ○ Thursday
- ○ Saturday
- ○ Wednesday
- ◉ Sunday
- ○ Tuesday
- ○ Monday

Based on the summary statistics, Sunday has the lowest mean emissions. Hence the total emissions on Sunday will be lowest. Remember: low emissions are preferable.

Question 4

1 / 1 pts

Your boss asks you to prepare a summary of the data on emissions to address his question. Based on your examination of the properties of the data, which of the following would you choose as the most informative set of information to present?

- ○ A graph of the data and a table of sample sizes and interquartile ranges
- ◉ A graph of the data, and a table of means and standard deviations
- ○ A table of means, standard deviations and variances
- ○ A table of sample sizes and means, and a list of the outliers
- ○ A table of means and medians, and a list of the outliers

Any summary of data should include a graph, unless there is a good reason not to. A table of means and standard deviations gives important information about location and spread. Sample sizes are useful information, but secondary to information about location.

Question 5

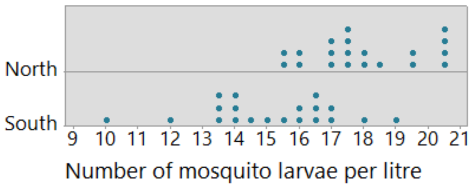
1 / 1 pts

You are attending a presentation where a researcher is presenting results from a survey of two different swamps. At each swamp, she took 20 observations of the density of mosquito larvae (in water samples), in units of larvae per litre. She presents summary statistics in the following table:

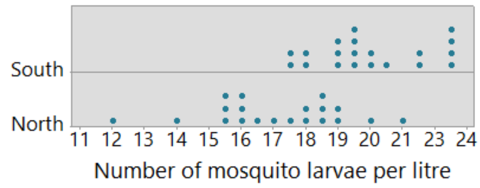
Swamp	**Number**	**Mean**	**Standard deviation**
South	20	18.0	1.7
North	20	15.1	2.1

She also presents a graph of the data. Assuming the summary statistics are correct, which one is the correct graph of the data in the set below?

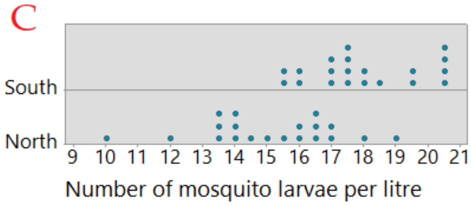
A



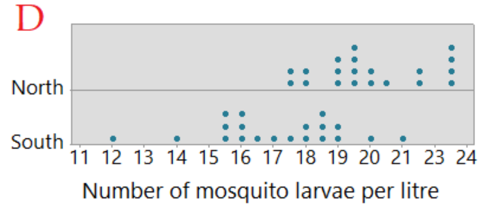
B



C



D



- ○ A
- ○ B
- ◉ C
- ○ D

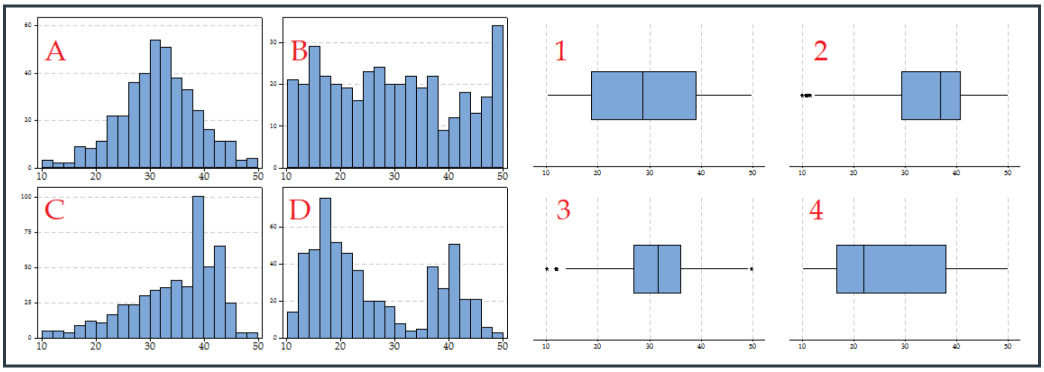
The correct answer is the bottom left; make sure the labels for North and South are the correct way around.

Question 6

1 / 1 pts

Consider the histograms and boxplots below. They represent four different data sets. There is a

matching histogram and boxplot for each data set.
Match the histogram with the corresponding boxplot.



Histogram A Boxplot 3

Histogram B Boxplot 1

Histogram C Boxplot 2

Histogram D Boxplot 4

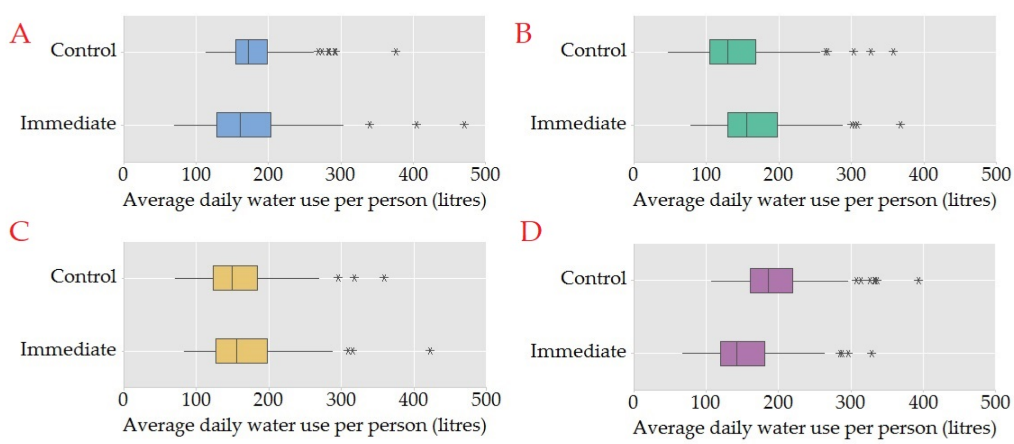
Histogram A - Boxplot 3
Histogram B - Boxplot 1
Histogram C - Boxplot 2
Histogram D - Boxplot 4

Question 7

1 / 1 pts

Consider the following study designed to assess the effectiveness of a household intervention for reducing water use. The intervention involves auditing the house, making changes to the house and training for household members. A random sample of households are approached to participate in the study; households agreeing to participate are randomly assigned to get the intervention immediately, or in a year (control with delayed intervention). The intervention takes around one month to complete. The water use is measured in both groups at the start of the study (baseline), before the intervention starts. There are four pairs of boxplots below; the boxplots within a pair are the same colour. Each compares the average daily water use per person (litres) by two groups --- the immediate intervention and the control (delayed intervention).

Which corresponds best to a plausible pattern for the average daily water use at baseline?



- D
- C

A

B

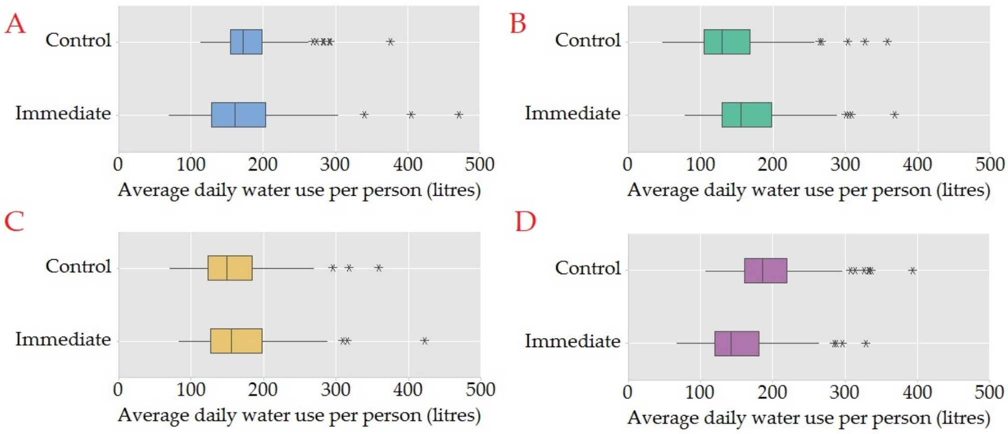
The two right hand pairs show a difference in mean between the two groups. When the groups are randomised, this is not expected at baseline, at the time of randomisation. In the top left hand pair, the means are similar, but the intervention group has a larger variance. This is also unexpected at baseline. The bottom left hand pair have similar shapes, and hence similar means and variances. This is what is expected at baseline, at the time of randomisation.

correct

Question 8

0 / 1 pts

Consider the water use study again. There are four pairs of boxplots below. Each compares the average daily water use (litres) per person at four months after agreement to participate by two groups --- the immediate intervention and the control (delayed intervention).



Which pair that shows the strongest evidence of success of the household intervention at four months, relative to the background variation within the groups?

- B
- D
- A
- C

If the intervention is effective, water consumption at four months will be lower, on average, in the "immediate intervention" group than in the "delayed (control)" group. In the bottom left pair the means are about the same, and in the top right pair the intervention mean is higher, so neither of these gives evidence of an effective intervention. In the top left pair the intervention group has a slightly lower mean, perhaps, but a larger variance. The bottom right pair shows the strongest evidence for the intervention being effective: the mean in the intervention group is clearly lower, and the variance in the two groups is about the same.

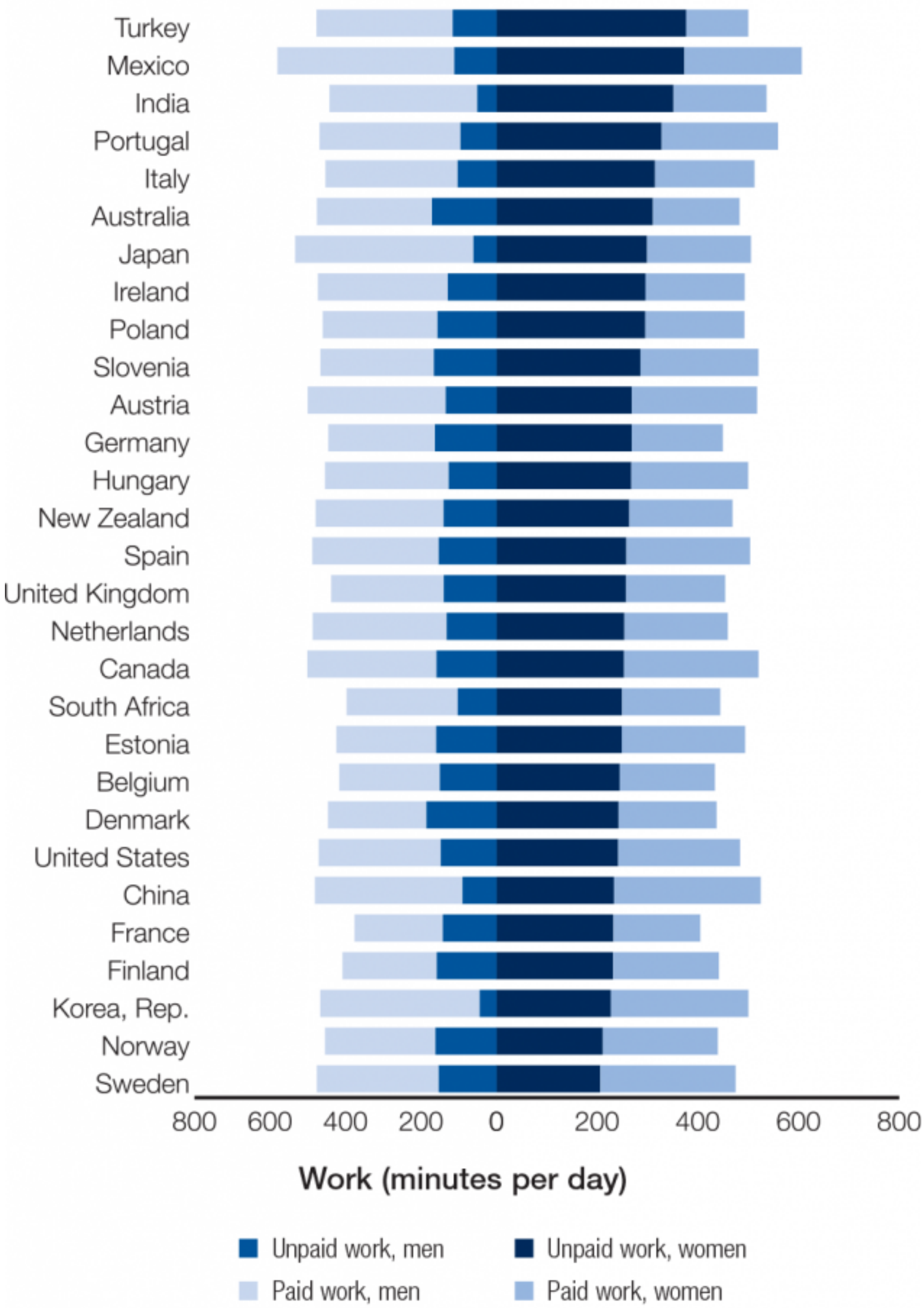
correct

Question 9

0 / 2 pts

This graph was published in a report by the World Economic Forum in 2016.
The graph shows the average number of minutes of paid and unpaid work per day done by men and women in 29 different countries. This includes weekdays and weekends.
The commentary on the graph claimed that it revealed “the strong gender gaps in distribution and the longer time spent by women on all forms of work, across most economies.”

Figure 12: Paid and unpaid work (minutes per day) for men and women, by country



Source: OECD Social Protection and Wellbeing Database.

Considering the principles of good graphs presented in this subject, which of the following are true? (Tick as many as apply)

The visual encoding for estimation of differences between males and females in the average time spent in paid work, for any country, is transparent.

The graph makes good use of a single measurement scale.

The consistency of the data with the claim that “women spend more time on all forms of work, across most economies” is difficult to evaluate in the graph.

The graph does not conform to the principle of using good alignment to a common scale.

The visual encoding on the graph is difficult as the men and women are aligned to different scales; the graph does not make good use of a single measurement scale.

Question 10

1 / 1 pts

Casey, a Venus chocolate bar lover, buys a Venus Bar every day in the hope of winning chocolate in a promotion promising "one in four wrappers is a winner".

Assume that the conditions of the binomial distribution apply: that the outcomes for Casey's purchases are independent and that the population of bars is effectively infinite.

What is the distribution of the number of winning wrappers in ten days?

- ○ Bi(0.25,10)
- ○ N(10,0.25)
- ○ N(1,4)
- ○ Bi(4,1)
- ○ Bi(1,4)
- ◉ Bi(10,0.25)

This context fits the structure of a binomial random variable. We specify the applicable binomial distribution in terms of the number of trials (10 in this case) and the probability of success (0.25 in this case).

Question 11

1 / 1 pts

Find the probability that Casey gets no winning wrappers in ten days.

Give the answer rounded to three decimal places.

0.056

Let X = Number of winning wrappers in ten days.

The distribution of X is Bi(10,0.25), and we want to find Pr(X = 0).

From Minitab:

Probability Density Function Binomial with n = 10 and p = 0.25

x P(X = x)

0   0.0563135

This rounds to 0.056.

This may also be calculated directly: the chance of no winners in 10 days is $(0.75)^{10}$ = 0.056.

Question 12

1 / 1 pts

Casey gets no winning wrappers for the first nine days of the ten days.

What is the chance that he will get a winning wrapper on the tenth day?

0.25

We have assumed that Casey's purchases are independent. This means that the probability of winning on the tenth day is unrelated to the outcomes on previous days. The chance of winning is 0.25 on any day.

Question 13

1 / 1 pts

Casey buys a bar every day for four weeks. Find the probability that he buys at least five winning wrappers.

Give your answer to three decimal places.

0.865

In four weeks, there are 28 days, so Casey is carrying out 28 trials.

The probability of success is 0.25.

The probability of buying at least five winning wrappers is the probability of buying more than four winning wrappers: Pr $(X > 4)$.

Here is the relevant Minitab output:

Cumulative Distribution Function Binomial with $n = 28$ and $p = 0.25$

x	P(X <= x)
4	0.135387

This gives the probability of fewer than five winning wrappers, that is, 4 or fewer.

Hence $\Pr(X \geq 5) = \Pr(X > 4) = 1 - \Pr(X \leq 4) = 1 - 0.135387 = 0.865$.

The answer can also be obtained in Minitab via Graph > Probability Distribution Plots > View Probability > Binomial > Shaded Area.

Question 14

1 / 1 pts

Casey decides he wants to make the chance of at least one winning wrapper 0.95 or more. How many days of purchases are required to achieve this?

Trial and error in Minitab, or any other suitable method, may be used to work this out.

11

If we want the chance of at least one winning wrappers to be 0.95 or more, then the chance of no winning wrappers will be 0.05 or less.

The chance of no winning wrappers in n days is 0.75n. We want this probability to be less than 0.05. Solving $0.75^n \leq 0.05$ gives $n \geq 11$.

Alternatively, we may use a trial and error approach in Minitab.

We want Pr(X = 0) to be less than 0.05.

From question 2 we know that the chance of no winning wrappers in 10 days is 0.056, so Casey needs to enter for more than 10 days. That is, 10 days are not enough; so try 11 days:

Cumulative Distribution Function
Binomial with n = 11 and p = 0.25
x P(X ≤ x)

0 0.0422351

The minimum number of days for which Pr(X = 0) < 0.05 is 11.

Question 15

1 / 1 pts

Scores on a national tertiary entrance examination have a mean of 30 and a standard deviation of 5. The scores are not necessarily whole numbers, and are Normally distributed.

Find the probability of scoring two standard deviations or more above the mean. Give your answer to four decimal places. (Report the probability as a number between 0 and 1.)

0.0228

Let X be the score on the entrance examination. If the standard deviation is $\sigma = 5$, then $2 \times \sigma = 10$.

Hence we need to find $\Pr(X > 30 + 10) = \Pr(X > 40)$.

From Minitab, the answer rounds to 0.0228 (to four decimal places).

Question 16

1 / 1 pts

Scores on a national tertiary entrance examination have a mean of 30 and a standard deviation of 5. The scores are not necessarily whole numbers, and are Normally distributed.

Find the probability of obtaining a score lower than one standard deviation below the mean. Give your answer to four decimal places. (Report the probability as a number between 0 and 1.)

0.1587

Let X be the score on the entrance examination. The standard deviation is 5. Hence we need to find Pr(X < 30 - 5) = Pr(X < 25).

Question 17

Original Score: 0 / 1 pts Regraded Score: 1 / 1 pts

This question has been regraded.

Scores on a medical test have a mean of 100 and a standard deviation of 20. The scores are not necessarily whole numbers, and are Normally distributed.

Is the probability of a score below 90 greater than the probability of a score above 115?

- Yes

No

You can obtain plots to find the probabilities in each tail. Use Graph > Probability Distribution Plot > View Probability. Choose a Normal Distribution and make the mean 100 and the standard deviation 20.

Click on the tab for the Shaded Area, and Define Shaded Area By the X-value. Use the left tail and the X value of 90 for the plot shown on the left below, and the right tail and the X value of 115 for the plot on the right.

Also: Since the Normal distribution is symmetric about its mean, this Normal distribution is symmetric about its mean = 100. The distance from 115 to 100 is greater than the distance from 90 to 100, so it follows that $Pr(X < 90) > Pr(X > 115)$.

Question 18

1 / 1 pts

The final questions refer to the following information:

A caterer orders pies for spectators at one day cricket matches at a large sporting venue according to whether or not Australia is playing.

On days when Australia is playing, the number of pies consumed is Normally distributed with a mean of 25,000 and a standard deviation of 2,500.

On days when Australia is not playing, the number of pies consumed is Normally distributed with a mean of 20,000 and a standard deviation of 2,500.

For an upcoming event, the caterer assumes that Australia is not playing.

Let X be the number of pies consumed when Australia is not playing. The caterer wants to have a chance of only 5% of failing to meet the crowd demand, and needs to work out the number of pies she should order (x) to achieve that.

Which of the following probability statements represents the caterer's problem?

- $\Pr(X \leq x) = 0.05$
- $\Pr(X \geq 0.05) = x$
- $\Pr(X > x) = 0.05$
- $\Pr(X < 0.05) = x$

If the number of pies, X, is in the top 5% of the distribution, above the number ordered (x), then the need will not be met. To fail to meet the need with only a 5% probability, the upper tail of the distribution should have an area of 0.05, so x should be chosen so that

$$\Pr(X > x) = 0.05.$$

Question 19

1 / 1 pts

How many pies should the caterer order?

Give the answer rounded to the nearest whole number.

24,112

You can obtain the answer with a Probability Distribution Plot that shows that the number of pies for which Pr(X > x) = 0.05 is 24,112. That is, Pr(X > 24,112) = 0.05
The result can also be obtained using: Calc > Probability Distribution > Normal. Remember you need to ask for the Inverse cumulative probability, and to specify the Input constant (the cumulative probability) to be 0.95.

Question 20

1 / 1 pts

The caterer orders x pies for the one day cricket match (where x is the answer to the question above).

On the actual day, the caterer realizes that she has made a mistake: Australia **is** playing.

Let X_A be the number of pies consumed when Australia is playing.

Given that the caterer has ordered assuming Australia is not playing, the caterer now needs to know the probability (p) that she will have enough pies.

Which of the following statements represents this probability correctly?

- $\Pr(X_A \geq p) = x$
- $\Pr(X_A \geq x) = p$
- $\Pr(X_A \leq p) = x$
- $\Pr(X_A \leq x) = p$

The number of pies ordered x=24,112 (obtained earlier). We consider this observation in the context of a Normal distribution with a mean 25,000 and standard deviation 2,500, because that is the relevant distribution when Australia is playing.

The caterer will have enough pies if the number of pies is less than or equal to 24,112 . As 24,112 is below the mean, we know that the probability of having enough will be less than 0.5.

We need to consider the lower tail of the distribution. That is, we need to find $\Pr(X_A \leq x) = p$.

Question 21

1 / 1 pts

Given that Australia is playing, and the caterer has ordered assuming that Australia is not playing, what is the probability that the caterer will run out of food?

Give your answer to four decimal places. (Report the probability as a number between 0 and 1.)

0.6388

If the demand exceeds the amount ordered by the caterer, she will run out of food. Hence we are interested in the upper tail of the distribution.

The area under the pdf of the Normal distribution (with a mean 25,000 and standard deviation 2,500) that is above 24,112 is 0.6388.
(If you used 24,112.1 for the number of pies, the probability is the same to four decimal places.)

Quiz Score: **19** out of 22