Questions 3 and 4: ACT scores, part 2

In this 3-part question, you will convert raw ACT scores to Z-scores and answer some questions about them.

Convert `act_scores` to Z-scores. Recall from Data Visualization (the second course in this series) that to standardize values (convert values into Z-scores, that is, values distributed with a mean of 0 and standard deviation of 1), you must subtract the mean and then divide by the standard deviation. Use the mean and standard deviation of `act_scores`, not the original values used to generate random test scores.

Question 3a

1.0/1.0 point (graded)

What is the probability of a Z-score greater than 2 (2 standard deviations above the mean)?

0.0233 ✔ **Answer:** 0.0233

0.0233

Explanation

The probability can be calculated using the following code:

```
z_scores <- (act_scores - mean(act_scores))/sd(act_scores)
mean(z_scores > 2)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 3b

1/1 point (graded)

What ACT score value corresponds to 2 standard deviations above the mean (Z = 2)?

32.1906

✔ **Answer:** 32.2

32.1906

Explanation

The score value can be calculated using the following code:

```
2*sd(act_scores) + mean(act_scores)
```

Submit You have used 2 of 10 attempts

Answers are displayed within the problem

Question 3c

1/1 point (graded)

A Z-score of 2 corresponds roughly to the 97.5th percentile.

Use `qnorm()` to determine the 97.5th percentile of normally distributed data with the mean and standard deviation observed in `act_scores`.

What is the 97.5th percentile of `act_scores`?

31.96338

✔ **Answer:** 32.0

31.96338

Explanation

The 97.5th percentile can be calculated using the following code:

```
qnorm(.975, mean(act_scores), sd(act_scores))
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

In this 4-part question, you will write a function to create a CDF for ACT scores.

Write a function that takes a value and produces the probability of an ACT score less than or equal to that value (the CDF). Apply this function to the range 1 to 36.

Question 4a

1/1 point (graded)

What is the minimum integer score such that the probability of that score or lower is at least .95?

Your answer should be an integer 1-36.

31

✔ **Answer:** 31

31

Explanation

The minimum score can be calculated using the following code:

```
cdf <- sapply(1:36, function (x){
  mean(act_scores <= x)
})
min(which(cdf >= .95))
```

Submit

You have used 1 of 10 attempts

Answers are displayed within the problem

Question 4b

1/1 point (graded)

Use `qnorm()` to determine the expected 95th percentile, the value for which the probability of receiving that score or lower is 0.95, given a mean score of 20.9 and standard deviation of 5.7.

What is the expected 95th percentile of ACT scores?

30.27567

✔ **Answer:** 30.3

30.27567

Explanation

The expected 95th percentile can be calculated using the following code:

```
qnorm(.95, 20.9, 5.7)
```

Submit You have used 1 of 10 attempts

Answers are displayed within the problem

Question 4c

1/1 point (graded)

As discussed in the Data Visualization course, we can use `quantile()` to determine sample quantiles from the data.

Make a vector containing the quantiles for `p <- seq(0.01, 0.99, 0.01)`, the 1st through 99th percentiles of the `act_scores` data. Save these as `sample_quantiles`.

In what percentile is a score of 26?

Your answer should be an integer (i.e. 60), not a percent or fraction. Note that a score between the 98th and 99th percentile should be considered the 98th percentile, for example, and that quantile numbers are used as names for the vector `sample_quantiles`.

82 **Answer:** 82

82

Explanation

The percentile for a score of 26 can be calculated using the following code:

```
p <- seq(0.01, 0.99, 0.01)
sample_quantiles <- quantile(act_scores, p)
names(sample_quantiles[max(which(sample_quantiles < 26))])
```

Submit You have used 10 of 10 attempts

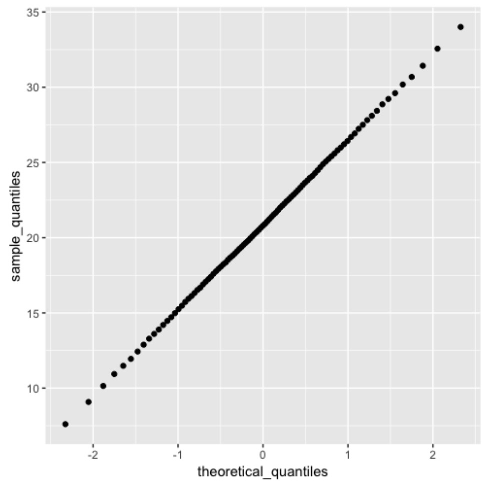
Answers are displayed within the problem

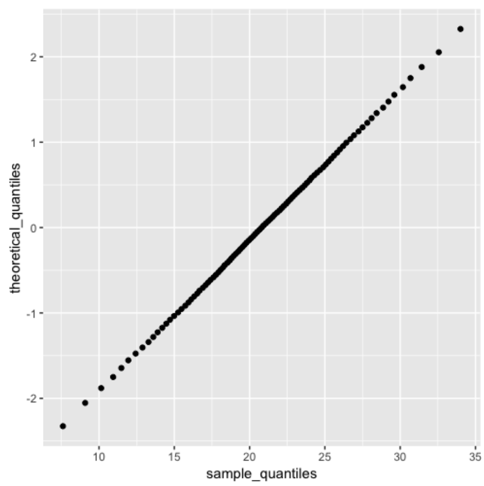
Question 4d

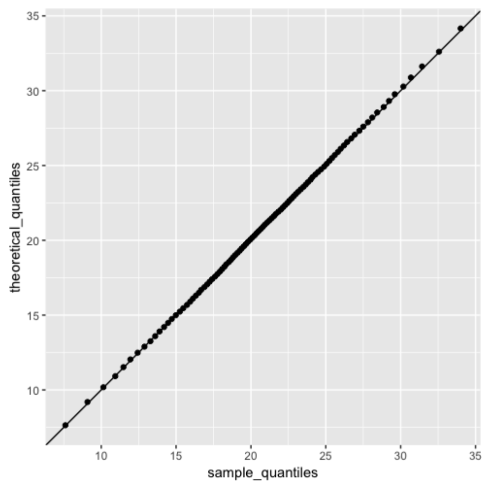
1/1 point (graded)

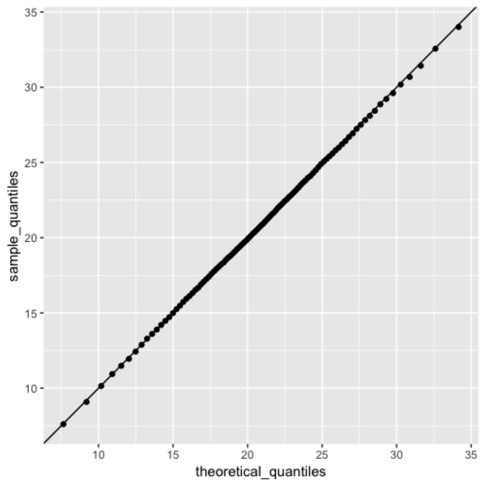
Make a corresponding set of theoretical quantiles using `qnorm()` over the interval `p <- seq(0.01, 0.99, 0.01)` with mean 20.9 and standard deviation 5.7. Save these as `theoretical_quantiles`. Make a QQ-plot graphing `sample_quantiles` on the y-axis versus `theoretical_quantiles` on the x-axis.

Which of the following graphs is correct?









Explanation

The fourth graph is correct and can be generated using the following code:

```
p <- seq(0.01, 0.99, 0.01)
sample_quantiles <- quantile(act_scores, p)
theoretical_quantiles <- qnorm(p, 20.9, 5.7)
qplot(theoretical_quantiles, sample_quantiles) + geom_abline()
```

The first graph uses standard normal theoretical quantiles, the result of omitting the mean and standard deviation in `qnorm`. The second graph uses standard normal theoretical quantiles and inverts the x and y axes. The third graph inverts the x and y axes.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem