Solution

1.

A bunch of people review a movie from 1-star to 5-stars as shown in the table to the right. The mean and median score is calculated. Two more people then add their scores, one grading it as a 1-star and the other as a 5-star. Both of these statements are correct.

I. The shape of the distribution does not change.
II. Both the mean and median do not change.

Stars	1	2	3	4	5
People	5	4	8	2	1

I is correct as the both distributions will be skewed right.
II is incorrect. The median (middle score) of the original data is 3 and adding one to the 1-star and 5-star data doesn't change that.
But the mean of the original data is $\frac{5+8+24+8+5}{20} = 2.5.$
Adding 2 pieces of data whose average is 3 will increase the mean.
False

Solution

2.

Gloria and Howard are two senior golfers who are married to each other.
The table below shows how they did in a recent round of golf as well the
average and standard deviation for other married senior golfers in their gender.

Both statements below are correct.

I. Gloria scored better (lower) than Howard relative to their gender categories.

II. A sample of married senior women scores as well as their husband's scores are collected with the idea of showing that married senior men score better than their wife (raw score). To test that theory, a 2-sample *t*-test is used.

	Score	$\bar{x}$	s
Gloria	86	89.5	2.2
Howard	81	84.3	2.7

$z_G = \dfrac{86-89.5}{2.2} = -1.6 \qquad z_H = \dfrac{81-84.3}{2.7} = -1.2$

I. Since Gloria's standardized score is lower than Howard's, she scored better.

II. Since there is a relationship between the man and woman, each couple's scores is subtracted making it a matched pairs one-sample test.

False

Solution

3. During the COVID-19 outbreak, shoppers at a market were allowed a maximum of 3 cleaning/disinfecting products during a store visit. The number of such products brought by shoppers during a day are summarized in the table below. Someone purchasing 3 products was within 1 standard deviation of the mean.

products bought	0	1	2	3
percent buying	15%	10%	35%	40%

$$\bar{x} = 0.15(0) + 0.10(1) + 0.35(2) + 0.40(3) = 2.0$$

$$s = \sqrt{0.15(0-2)^2 + 0.1(1-2)^2 + 0.35(2-2)^2 + 0.4(3-2)^2}$$

$$s = \sqrt{0.60 + 0.1 + 0 + 0.4} = \sqrt{1.1} = 1.05$$

$2 + 1.05 = 3.05$ so 3 is within 1 sd of the mean

True

Solution

4.

The 5-number summary of a data set is 7, 20, 21, 28, and 33.
It can be determined that the mean of the data is less than 21 and there is at least one outlier.

The IQR is Q3 – Q1 = $28-20=8$.

$28+1.5(8)=40$ so no outliers on the high end

$20-1.5(8)=8$ so at least one outlier on the low end

This means that the data is skewed left and the mean will be less than the median.

True

Solution

5.

A tropical storm is expected to drop a lot of rain on an island in the Bahamas. There are many rain gauges on the island but obviously not everywhere. The 95% confidence interval for the amount of rain the island will get is $[5.5 \text{ to } 9.5]$ inches. This can be interpreted as being 95% sure that the average amount of rain throughout the island is between 5.5 and 9.5 inches. The following interpretations are both correct as well.

I. The probability that the average amount of rain throughout the island is between 5.5 to 9.5 inches is 95%.

II. If the procedure of checking rain gauges were done repeatedly, 95% of the resulting confidence intervals would contain the true population mean.

I is false. The average amount of rain on the island is either between 5.5 and 9.5 inches or not.

II is true. A confidence interval interpretation is about procedure. You take a sample of gauges and compute the confidence interval. Do it with other samples. 95% of them will catch the true mean. So you are reasonable sure that the given CI contains the true average amount of rain throughout the island.

False

Solution

6.

A survey was conducted in a state. People who play the lottery regularly were asked whether they had ever won a jackpot of over \$500 (winners). A scatterplot is shown with each dot representing a city within the state, examining the city's population in thousands versus the number of winners. The equation of the least square regression line is

$\widehat{\text{Winners}} = 0.0154 \cdot \text{People} + 13, \; r^2 = 0.88$

The following statements are both correct:

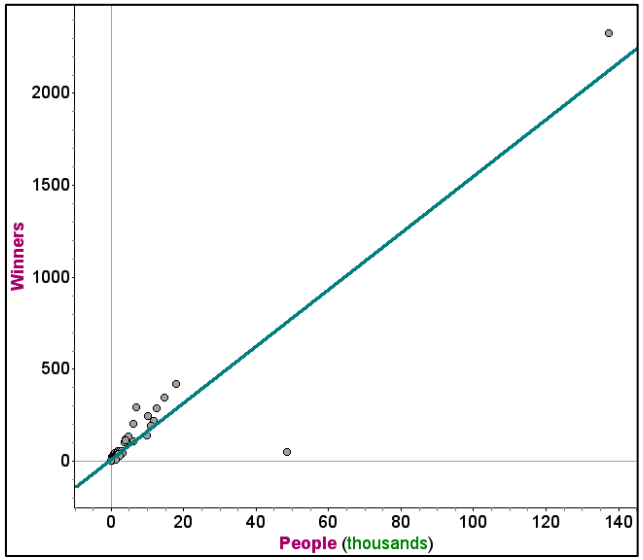
I. As a good rule of thumb, for every 1000 people playing the lottery in the state, about 15 of them win over \$500.

II. There is one outlier.

I. With $r^2 = 0.88, r = 0.94$ which is a strong positive association. The slope of the line is 0.0154 which is about 15 out of 1000.

II. The point at about $(50000,100)$ is an outlier as it corresponds to 2 people out of 1000 as winners. The point at the top right is not an outlier as it is close to the regression line.

True



Solution

7. On June 1, a used car dealer increases the price of every vehicle on his lot by 5% and then takes $100 off the new price. If the mean, median, and standard deviation of the car prices were taken on May 31 and then again on June 1, all will have increased by 5% and then lowered by $100.

The mean and median will be changed as stated. But the standard deviation will only be increased by 5% as the lowering of the price by $100 doesn't change the spread.

False

Solution

8.

Diane pays \$1 to play a game. She rolls two 10-sided dice, each containing the digits 0 – 9. If both come up a 7, she wins \$10. If one of them is a 7, she wins \$5. Otherwise she wins nothing. Diane can be expected to break even by playing this game.

There are $10(10) = 100$ possibilities.

The only way she can win \$10 is to roll 77.

To get one 7, there are 18 combinations: 07, 17, 27, 37, 47, 57, 67, 87, 97, 70,71, 72, 73, 74, 75, 76, 78, 79.

$\mu_x = 10\left(\frac{1}{100}\right) + 5\left(\frac{18}{100}\right) + 0\left(\frac{81}{100}\right) = 1$. She can be expected to win \$1 every time she plays but the cost of the game is \$1 so she breaks even.

True

Solution

9.

A car wash establishment has determined over time that the average amount of time that it takes a car to go through the process of driving up, paying, washing, and drying is 12 minutes, 15 seconds with a standard deviation of 2 minutes, 5 seconds. Because of a \$5 off promotion, they believe that on average, it will take a car at least 5 minutes longer to go through the car wash. Using a random sample of cars, the appropriate hypothesis test is used with a result of $p = 0.024$.

The hypothesis test is a one-sample z-test, the hypotheses are $\begin{cases} H_0 : \mu = 17.5 \\ H_a : \mu > 17.5 \end{cases}$ and the conclusion is that there is evidence that it takes at least 5 minutes longer to get through the car wash with the promotion.

This is a z-test as the population standard deviation is known. The parameter of interest is μ, the average amount of time to get through the car wash, which is believed to be greater than $12.25 + 5 = 17.25$. Since $p \le 0.05$, we reject H_0 and accept H_a.

True (Note that this falls short of saying the reason for the longer amount of time is the promotion itself. There could be other reasons why it takes longer such as fewer workers.)

Solution

10. A restaurant has on the menu a cold soup served in the husk of a melon. The price of the soup is related to the size of the husk which is offered in a variety of sizes. When examining the data points in the form $(x: \text{diameter of the melon}, y: \text{price of the soup})$, a statistician performed a transformation $(x, y) \Rightarrow (\log x, \log y)$. Statistics software yields the linear equation $y = a + bx$ where $a = -0.871$, $b = 2.54$, and $r^2 = 0.944$. The price P of the soup with diameter d is approximately given by $P = \dfrac{2.54^d}{10^{0.871}}$.

The $y = a + bx$ regression performed on $(\log x, \log y)$ is the product of a power regression on (x, y) with $y = a \cdot x^b$ so $P = 10^{-0.871} \cdot d^{2.54}$ or $P = \dfrac{d^{2.54}}{10^{0.871}}$.

False

Solution

11.

An Internet Service Provider gets frequent calls for support. The company wants to make a claim about the average wait on hold waiting to talk to a representative. It takes a sample of n callers, determines the average length of time on hold, and then creates a confidence interval. Both of these statements are correct:

I. Using $n = 100$ with 90% confidence will narrow the confidence interval from the one created with $n = 50$ and 95% confidence.
II. Whether the waiting times on hold are normally distributed does not affect the validity of the calculations in I.

The greater the confidence, the wider the interval. If we want a narrower, interval, we lower the confidence. A larger sample will make $\frac{s}{\sqrt{n}}$ smaller further narrowing the interval. The Central-Limit Theorem states that $n = 40$ is sufficient for these calculations, even if the population is not normally distributed.
True

Solution

12. A survey was taken as to the TV regular viewing habits of men who watch indoor sports and outdoor sports as shown in the table below.

	Football	Baseball	Golf
Basketball	40	32	14
Hockey	16	20	17
Indoor soccer	2	4	5

The percentage of people who watch basketball or football is about the same $(\text{within } 1\%)$ of the percentage of football watchers who also watch basketball.

There are 150 men total in the survey. The number of men who watch basketball or football is $40+16+2+32+14=104$.
The percentage is $104/150=69.3\%$.
There are $40+16+2=58$ football watchers and 40 also watch basketball. The percentage is $40/58=69.0\%$.
True

Solution

13.

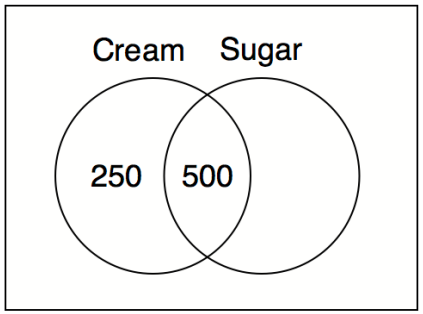
The managers of a restaurant believe that drinking coffee with cream and drinking it with sugar are independent events. In examining the drinking habits of 1200 coffee drinking customers in a week, the number of people using each can be partially summarized by the Venn diagram below. If 300 coffee drinkers use sugar only, the managers would be correct in their theory.

To be independent, $P(\text{cream}) \cdot P(\text{sugar}) = P(\text{both})$

$$\left(\frac{750}{1200}\right)\left(\frac{300+500}{1200}\right) = \left(\frac{5}{8}\right)\left(\frac{2}{3}\right) = \frac{5}{12} = \frac{500}{1200}$$

or independence: $P(\text{sugar} \mid \text{cream}) = P(\text{sugar}) \Rightarrow \frac{500}{750} = \frac{800}{1200}$

True

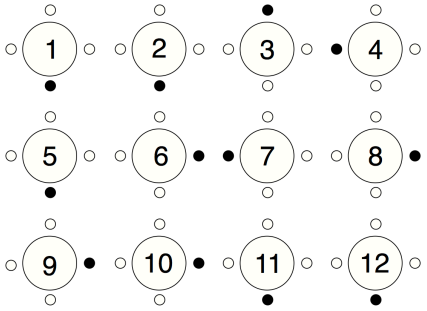


Solution

14. False (it could possibly be a simple random sample but the chance of choosing every table exactly once is remote.)

At a teacher meeting, there are 4 teachers seated at 12 tables. A sampling of teacher opinion about the meeting is taken with the teachers chosen by the sample indicated in black. This is most likely an example of multistage sampling and it eliminates bias in choosing the teachers.

Multistage would choose several tables and ask several or all the teachers from those tables. For instance, it might choose tables 4, 7, 8 and 11 and randomly choose 3 teachers from each. This is most likely a stratified sample with one person chosen randomly from each table. It doesn't eliminate bias because once a teacher from a table is chosen, then no other teacher from that table can be chosen.



Solution

A convenience store has a soft-drink machine that with the press of a button is suppose to fill a cup with 12 ounces of the soda of choice. There have been complaints that it does not completely fill the cup. The store manager is uncertain whether the machine is actually broke. The appropriate hypotheses are $\begin{cases} H_0 : \mu = 12 \\ H_a : \mu < 12 \end{cases}$ and the cost of making a Type II error is the time and money it takes attempting to fix a machine that is not broke.

The hypotheses are correct. If the complaint was that the cup was not properly filling the cup, then it could be over or under-filling and $H_a : \mu \neq 12$. A Type I error is rejecting the null hypothesis when it is true. A Type I error decides that the machine is under filling the cup correctly when in fact it working OK. The cost of making a Type I error is the time and money it takes attempting to fix a machine that is not broke. A Type II accepts a false null hypothesis. It decides that the machine is filling the cups properly when it isn't. The error described is a type I error.

15. False

Solution

16. The chance that a randomly selected person will provide a suitable matching kidney for a transplant is 20%. The expected number of failures before a matching kidney is found is 4 and the chances that a matching kidney will be found within the first 3 donors is close to 50%.

This is a geometric probability question as we are waiting for a success.

The expected number of trials until a success is $\frac{1}{0.20} = 5$ so 4 failures are expected before success. For success within the first 3 donors, the probability is $0.2 + 0.8(0.2) + 0.8^2(0.2) = 0.488 \approx 50\%$.

True

Solution

17.

A summer camp has a policy that during the last activity period of the day starting at 4:00 PM, campers may be involved in activities or choose to read quietly. It was felt that too many were not active so the director insituted a "Pepsi break" at 4 PM. After institution of that policy, the average daily difference in percentage of campers who change from reading to active was 14% and the appropriate hypothesis test found this to be statistically significant. The following statement is correct:

A two-proportion *z*-test found that the Pepsi break was responsible for more campers being active than reading during the last period of the day.

The choice of test is correct. On a sample of days, you calculate the percentage of campers who are active and then do so after the Pepsi break policty is initiated. However, the statement is wrong when saying that it was the Pepsi break that was responsible for the change in percentage. There could have been other factors such as weather or illness.

False

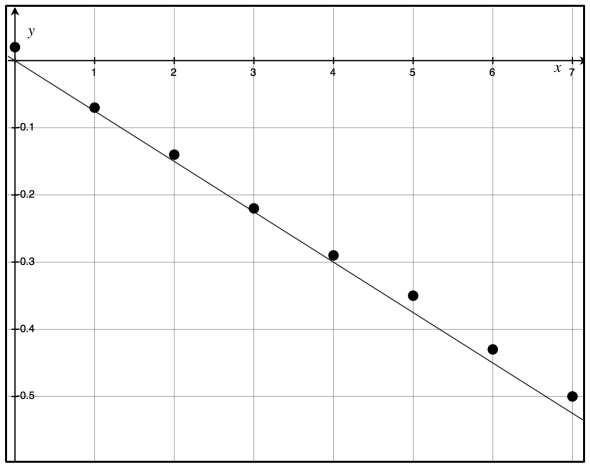
Solution

18.

Given 2 data sets x and y, the graph of x vs. $\log y$ is shown in the graph below. The equation of the least-squares regression line for this graph is $y = -0.071x$. The regression equation linking the original data sets x and y could be $y = -0.85^x$.

Since the graph of x vs. $\log y$ is quite linear, then the graph of x vs. y must be exponential. The equation is $y = 10^{-0.071x}$ which is approximately $y = 0.85^x$.

False



Solution

19.

In a simple random sample from a large condo, 64 of 91 said they were happy with their Comcast Digital Cable service while 38 of 62 said they were happy with their AT&T Fiber Optic service. The following two statements are correct:

I. A 95% confidence interval estimate for the difference between Comcast and AT&T user in this condo who are happy with their service is 0.09 ± 0.153.

II. Based on the CI, a greater percentage of people are happy with Comcast than those who are happy with AT&T.

$$n_C = 91, p_C = \frac{64}{91} = 0.703 \qquad n_A = 62, p_A = \frac{38}{62} = 0.613$$

$$\sigma = \sqrt{\frac{(0.703)(0.297)}{91} + \frac{(0.613)(0.387)}{62}} = 0.078$$

$$0.703 - 0.613 \pm 1.96(0.078) = 0.09 \pm 0.153$$

Doing on the calculator as a two-proportion z-interval: $[-0.06, 0.24]$

Since the 95% CI contains negative numbers, it is possible that a greater percent of people are happy with AT&T than Comcast.

False

Solution

On a commercial jet, all luggage is placed in a metal container weighing 250 lbs. The container is filled with a set number of large and small suitcases whose average weight and standard deviation is given in the table below.

Suitcase	Number in container	Average weight (lbs)	Standard Deviation (lbs)
Large	30	25	7
Small	50	15	4

The average weight of the filled containers is 1750 lbs.
The standard deviation of the filled containers is 410 lbs.

$$\mu = 30(25) + 50(15) + 250 = 1750 \text{ lbs}$$

$$\sigma = \sqrt{30(7^2) + 50(4^2)} = \sqrt{2270} \approx 47.6 \text{lbs}$$

20. False (To calculate the standard deviation, you must find the variance of the large and small suitcases, add them, then take the square root. Since the container has no variance, it is not included in the calculation.)

Solution

21.

Angelo is a very controversial radio talk-show host.
It is felt that people either really like him or can't stand him.
To test this, he does an online poll. He asks his listeners to
visit the radio station's website and participate in the poll and
rate Angelo in terms of how much they like him on a scale to
1 to 5. The hypothesis test that should be used is a two-sided *t*-test
but the results will most likely be flawed because of undercoverage.

Since it is believed that Angelo doesn't have many listeners who are ambivalent about him, the test would be two-sided: $\begin{cases} H_0 : \mu = 2.5 \\ H_a : \mu \neq 2.5 \end{cases}$

A *t*-test is appropriate because the population standard deviation
is unknown. Since people who can't stand Angelo are most likely
not listening, the poll will most likely be biased towards people who
like him a lot. This is voluntary response which usually results
in undercoverage.

True

Solution

22.

A large study is done with parents and children as to whether they were forced to eat everything on their plate at dinner. Parents are asked 2 questions:

1. Did your parents make you eat everything on your plate?
2. Do you make your kids eat everything on your plate?

Children are asked 2 questions:

1. Do your parents make you eat everything on your plate?
2. Will you make your kids eat everything on your plate?

The results are summarized in the table to the right.

		Parents: Yes	Parents: No
Kids	Yes	22%	7%
	No	11%	60%

Both statements are true:

I. If a parent is chosen who was forced to eat everything, it is more likely that her kids will do the same.

II. If a child is chosen who isn't forced to eat everything, it is more likely that her parents didn't do so as well.

$$P\left(\text{Kids yes}\middle|\text{Parents yes}\right)=\frac{22\%}{22\%+11\%}\approx 67\%$$

$$P\left(\text{Parents no}\middle|\text{Kids no}\right)=\frac{60\%}{11\%+60\%}\approx 85\%$$

True

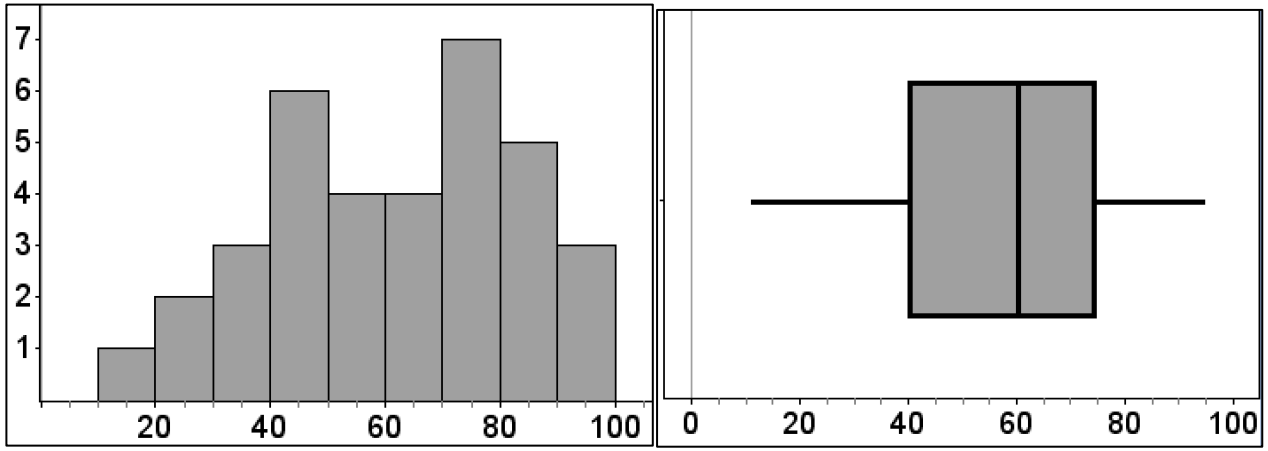
Solution

23.

The histogram and boxplot below could represent the same data set.

Counting the heights of the histogram bars, there are 35 pieces of data so the median is in the bar where the 18th piece of data is which is between 50 and 60. The median in the boxplot is 60. The data in the histogram tails to the left so that the data is skewed left which is borne out by the boxplot. The hisogram puts the minimum between 10 and 20 and the maximum between 90 and 100 which is consistent with the boxplot. There is nothing in either that says that they do not come from the same data.

True.



Solution

24.

A car company claims that its new electric vehicle can travel 200 miles on an electric charge. A consumer's research company believes this figure is too high and does a study. It makes 30 trips with the same car and finds that the average number of miles that the car can travel is on a charge is 193 with a standard deviation of 13.9. Their conclusion of the research is:

The test statistic value of -2.758 with degrees of freedom 29 is sufficient evidence to reject the car company's claim and that this electric vehicle gets less than the advertised 200 miles on a charge.

Based on this description, the conclusion is valid.

If $\mu =$ the mileage that this type of car gets on a single charge,

$\begin{cases} H_0 : \mu = 200 \\ H_a : \mu < 200 \end{cases}$ This is a one-sample t-test with $\bar{x} = 193, n = 30, s = 13.9$

$n = 30$ meets the conditions of a t-test. The degrees of freedom is $n - 1 = 14$ and from the calculator, $t = -2.758$ and $p = 0.005$. We reject H_0 and accept H_a. However, the fact that the same car was used over and over introduces bias and invalidates the test.

False

Solution

John pays \$1 to play a game of chance. He spins both spinners below, with the numbers on each spinner having the same probability of coming out. If he gets 2 odd numbers, he wins \$4. If he gets 2 even numbers, he wins nothing. Otherwise, he wins \$1.50. John can expect to make money playing this game if he plays it many times.

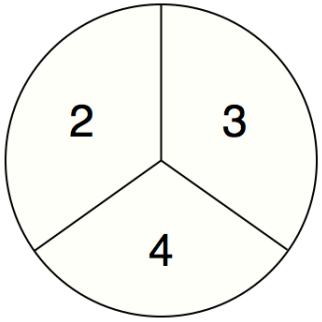
25.

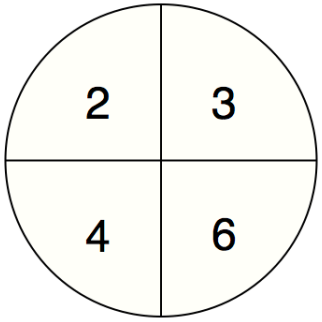
$$P(OO)=\frac{1}{3}\left(\frac{1}{4}\right)=\frac{1}{12} \quad P(EE)=\frac{2}{3}\left(\frac{3}{4}\right)=\frac{1}{2} \quad \text{So } P(\text{OE or EO})=1-\frac{1}{12}-\frac{1}{2}=\frac{5}{12}$$

$$\mu=4\left(\frac{1}{12}\right)+0\left(\frac{1}{2}\right)+1.5\left(\frac{5}{12}\right)=\frac{11.5}{12}=0.958$$

Since the game costs \$1.00, he can expect to lose $1-0.958=4.2$ cents on each play

False





Solution

26.

You get on a train which costs \$10 a ride. There is no conductor checking tickets but there are plain-clothed officers assigned to the train to check on random passengers. If any are found without a ticket, they are arrested and fined heavily. The chances of being chosen to have a ticket check is 2%. If the cost of the fine is \$500, it is not worth it to ride for free and gamble that you will not be checked.

Your point of view: X how much you gain by gambling

X	10	$-t$
$p(X)$	0.98	0.02

$$\mu_X = 10(0.98) - t(0.02) = 0 \Rightarrow 0.02t = 9.8 \Rightarrow t = 490$$

If the fine is more than \$490, it is not worth it to gamble.

True (many cities have this type of arrangement and the cost of the fine is so high that the gamble is not worth it.)

Solution

27.

In a pet shop with many fish tanks, at feeding time, the average amount of time it takes for a tank to receive food is 18 minutes, 30 seconds with a standard deviation of 6 minutes, 45 seconds. Because of the layout of the store with many tanks clustered together and others far apart as well as different fish having different feeding requirements, the distribution of these times is approximately normal. The probability that a tank chosen at random containing 5 fish will be fed within the first 15 minutes of feeding time is about 30%.

The number of fish in the tank is irrelevant. This is simply a normal distribution problem:

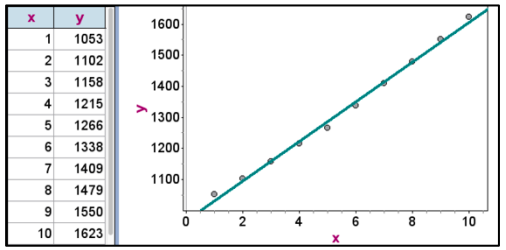
$\text{normalcdf}\left(0,15,18.5,6.75\right) \approx 29.9\% \approx 30\%.$

True

Solution

28.

A set of data points (x, y) appears below along with the graph of the Least-Squares Regression Line which is $\hat{y} = 63.9x + 967, r^2 = 1.00$. The predicted value of y when $x = 20$ is 2245.



x	y
1	1053
2	1102
3	1158
4	1215
5	1266
6	1338
7	1409
8	1479
9	1550
10	1623

It appears that a line is a good fit, especially with $r = 1$.
But there appears to be a pattern to the residuals with the first few above the curve, then the next few below the curve and then the next few above the curve again. To check this out, we check to see if the y values are increasing by a common value.

$1002 - 1053 = 51 \quad 1158 - 1102 = 56 \quad 1215 - 1158 = 57 \ldots 1623 - 1550 = 73$

It appears that the differences are increasing.

$$\frac{1102}{1053} = 1.046 \quad \frac{1158}{1102} = 1.051 \quad \frac{1210}{1158} = 1.045$$

It appears that the ratios are quite similar so the data is exponential.
Trying to use the linear equation to determine the value of y when $x = 20$ would significantly under-approximate the real value.
False

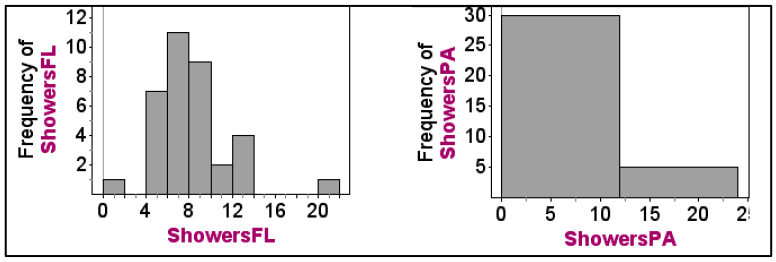
Solution

29.

35 randomly selected people were polled both in Florida and Pennsylvania as to the number of showers they take weekly. The results are shown in the two histograms below. This evidence shows that the Florida data has a greater standard deviation than the Pennsylvania data.

It is impossible to tell without seeing the data. Comparisons of this nature are only possible if the bin width of the histograms are the same and even then, only in obvious cases can standard deviations be compared. In fact, the data in the two histograms are the same.

False



Solution

30.

In a large apartment complex, 40% of the residents follow the recycle program rather than throwing everything in the trash. A letter is sent to all residents warning them that if they don't recycle, they could be fined. A month later, the building managers observe a sample of residents to see if the warning letter had any effect. Both statements are true:

I. This is an observational study.
II. The appropriate hypothesis test is a 1-proportion z-test.

It certainly is a 1-proportion z-test with $H_0 : \rho = 0.4$ and $H_a : \rho > 0.4$
But it is an experiment, not an observational study. The letter is the treatment and the goal is to find whether it had an effect on resident's behavior. (note that it would be more effective to send some residents the letter and others not).
False

Solution

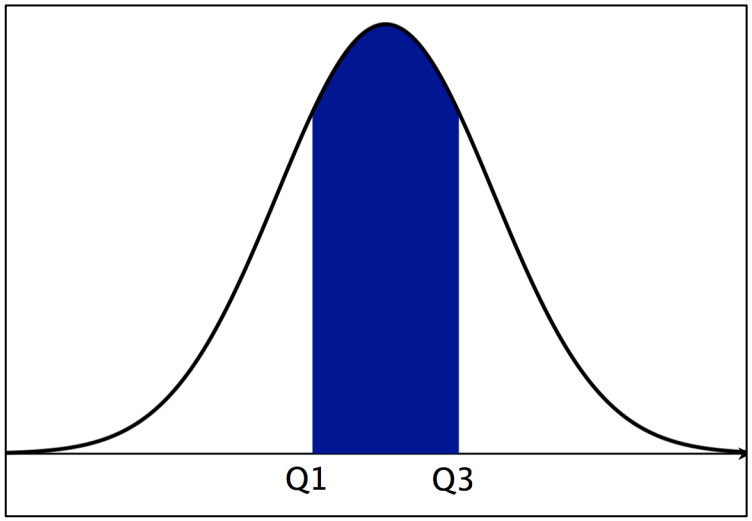
31.

The distribution of times it takes to takes to shop in a supermarket is approximately normal with a mean of 33 minutes and a standard deviation of 6 minute, 30 seconds. The shaded region in the figure below represents the percentage of people who take between 26 minutes, 30 seconds and 39 minutes, 30 seconds to shop in this market.

Q1 represents the 25th percentile. To find it, use invnorm $(0.25,33,6.5) = 28.6$

Q3 represents the 75th percentile. To find it, use invnorm $(0.75,33,6.5) = 37.4$

False



Solution

32.

Researchers in a drug company are doing human trials on a vaccine to combat COVID-19. In deciding their goals, they conclude that they want to use a hypothesis test for the vaccine's effectiveness with a high degree of statistical power. This means that if their formula is not effective, they want to be sure as possible before they promote it as a vaccine.

The hypotheses are: $\begin{cases} H_0 : \text{the vaccine is not effective for COVID-19} \\ H_a : \text{the vaccine is effective for COVID-19} \end{cases}$

A type I error is concluding the vaccine is effective when it isn't.

A type II error is concluding the vaccine is not effective when it is.

Power = 1 – the probability of a type II error

If power is high, the probability of a type II error is low.

That means that if they really do have an effective vaccine, they need to be as confident as possible that it works.

False

Solution

33.

Prime time for watching TV is between 8 PM and 11 PM.
A survey was done among TV viewers and the 95% confidence interval for the mean number of hours that viewers watch TV during prime time is $[1.87, 2.26]$. It is possible to use this information to perform a one-sample t-test to conclude that the average number of hours that a viewer watches TV may be 2.25 hours.

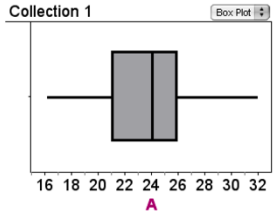
Since the 95% CI was found, a hypothesis test can be performed under the same conditions. Since the α level is 0.05 and the CI is 95%, the conditions are the same. The test looks like this: $\begin{cases} H_0 : \mu = 2.25 \\ H_a : \mu \neq 2.25 \end{cases}$

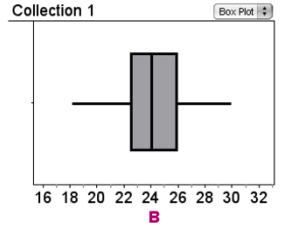
That means that all the values in the CI are plausible values for μ.
So we would not reject the null hypothesis and conclude that there is no evidence that the average number of hours a view watches TV is not 2.25. That does doesn't mean it is though. The key word is "may."
True

Solution

34.

I play a game 60 times using strategy A that has probability of winning of 40%. I then switch to strategy B that has a 20% higher probability of winning than A but I play it 20 fewer times. If I do this for a large number of days the boxplots representing the number of times I win each day are appropriately labeled A and B.





The game results will have a binomial distribution.

$\mu=np, \sigma=\sqrt{np(1-p)}$

$\mu_A = 60(0.4) = 24 \quad \mu_B = 40(0.6) = 24$

The center of distribution appears to be the same.

$\sigma_A = \sqrt{60(0.4)(0.6)} \quad \sigma_B = \sqrt{40(0.6)(0.4)}$

Since $\sigma_A > \sigma_B$, the spread in boxplot A should be greater than the spread in boxplot B. It appears to be.

True

Solution

35.

A COVOD-19 study shows that people now wash their hands an average of 6.8 times a day with a standard deviation of 2.2. Assuming a normal distribution, the probability that a group of 25 random people will wash their hands on average between 6 and 8 times a day will be at least 10% higher than that of a group of 10 random people.

Since the distribution is normal, we can apply the central limit theorem without having to worry about sample size: Using the calculator:

$\text{normalcdf}\left(6,8,6.8,2.2/\sqrt{10}\right)=0.833$ $\text{normalcdf}\left(6,8,6.8,2.2/\sqrt{25}\right)=0.962$

$0.962-0.833=0.129\geq 0.10$

True

Solution

36. Ralph bowls in two leagues, the Thunderbirds and the Raccoons. He averages a score of 150 in each league. In the Thunderbirds, his average is at the 40th percentile while in the Raccoons, his average is at the 70th percentile.
If a bowler is chosen at random from each league, the probability that one or both will have an average equal to or higher than Ralph's is 90%.

If Ralph is at the 40th percentile in the Thunderbirds, the probability that a random bowler will have a higher average is 60%. If Ralph is at the 70th percentile in the Raccoons, the probability that a random bowler will have a higher average is 30%. $60\% + 30\% = 90\%$ but we must subtract the possibility that they both have scores above 150 which is $(0.6)(0.3) = 18\%$. So the probability that one or both will have an average equal to or higher than Ralph is $90\% - 18\% = 72\%$.
False

Solution

37.

A supermarket sells on average of 54.5 bags of a premium coffee a day with a standard deviation of 10.6 bags. The company then changes the formula and places "new and improved" on the bag. They then take a random sample of 30 days over the next half-year and check coffee sales. Averaging 57.7 bags a day would give the minimum evidence at the 5% level that the new and improved coffee sells better.

Since the population standard deviation is given, this is a one-sided z-test.

It can be done by finding the z-score. $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{57.7 - 54.5}{10.6/\sqrt{30}} = 1.654.$

The corresponding p-value is approximately 0.05.

This can also be be found on the calculator: $\text{invNorm}(0.95, 0, 1) \approx 1.645.$

OR the entire thing can be done on the calculator as a z-test.

$\mu_0 = 54.5, \sigma = 10.6, \bar{x} = 57.7, n = 30$

$\begin{cases} H_0 : \mu = 54.5 \\ H_a : \mu > 54.5 \end{cases} \Rightarrow z = 1.654, p = 0.049$

True

Solution

38.

When given a choice at a barbecue, 60% of kids normally prefer hamburgers to hot dogs. If a random sample of 100 kids were taken, the probability that 55 or fewer will prefer hamburgers can be approximated by finding the probability

$$p\left(z < \frac{0.55-0.6}{\sqrt{\frac{(0.6)(0.4)}{100}}}\right)$$

This is a binomial distribution as kids prefer hamburgers or hot dogs.
There are 100 trials. We want 55 or fewer successes. The calculation is
$\text{binomialcdf}(100, 0.6, 55)$

But since that is not asked, the probability can be approximated using a normal distribution with mean 0.6 and st. dev. of a proportion is $\sqrt{\frac{p(1-p)}{n}}$.

The calculation is: $\text{normalcdf}\left(0, 0.55, 0.6, \sqrt{\frac{(0.6)(0.4)}{100}}\right)$

True

Solution

39.

Out of 1,000 children taken at random, 2% have a peanut allergy.
A blood test is used to try and confirm it, but it is not always accurate.
For those with the allergy, the test that confirms it is 90% accurate.
For those without the allergy, the test that confirms it is 95% accurate.
If you complete the table, you will find that the probability of
a false positive (choosing someone who tests positive for the peanut allergy but doesn't have it) is 73%.

	Tests Yes	Tests No	Total
Has peanut allergy			
No peanut allergy			
Total			1000

	Test Yes	Test No	Total
Peanut allergy	18	2	20
No peanut allergy	49	931	980
Total	67	933	1000

$$P(\text{No peanut allergy} \mid \text{Test Yes}) = \frac{49}{67} \approx 73\%$$

True

Solution

40.

You are a racing bicylist and like to go on long rides.
We are not sure whether to carry a tire-patching kit
and a bike pump on one of your rides. You make a judgement
based on the null hypothesis that you will not get a flat
tire. The ramifications of the following errors are:

Type I error: Getting a flat tire with no way to fix it.
Type II error: Needlessly carrying extra weight slowing you down.

They are reversed.
A Type I error is rejecting H_0 when H_0 is true. In this case, you do not get a flat tire. You reject this and carry the kit and pump, needlessly as it turns out. (However you did have piece of mind throughout the ride.)
A Type II error is rejecting the alternative hypothesis when the alternative is true. In this case, the altenative hypothesis is that you will get a flat tire. But you reject that and don't carry the kit. You get the flat tire and cannot fix it. (what you were worried about)
False

Solution

The first 4 integer powers of 3 $\{1, 3, 9, 27\}$ has mean $\mu = 10$ and standard deviation $\sigma = 10.25$. If I sample 3 numbers from this set with replacement, I create a sampling distribution with N possibilities, each with its own mean. The sampling distribution has mean μ_x and standard deviation σ_x. The values of the following variables are:

$N = 64, \mu_x = 10, \sigma_x = 3.42$

The size of the sampling distribution is $4^3 = 64$.

The Central Limit Theorem says the mean of the sampling distribution is the same as the mean of the population so $\mu_x = 10$. But the standard deviation of the sampling distribution is the standard devation of the population divided by the square root of the sample size: $\sigma_x = \dfrac{10.25}{\sqrt{3}} = 5.92$.

41. False

Solution

42.

Whether or not people stop or go through a yellow light is being surveyed.
On one day, 50 random drivers are chosen and whether or not they stop on yellow at a particular intersection is monitored. Both these statements are true:

I. Choosing the first 50 cars seen is as likely as choosing any other set of 50 cars in the population of cars that go through that intersection.
II. The percentage of drivers in the sample who go through the yellow light at that intersection will always be very similar if not equal to the percentage of drivers in the population of cars that go through the yellow light.

I is true. Choosing the first 50 cars is extremely unlikely but it is just as likely as choosing any other group of cars.
II is not necessarily true. We hope it is true because our hypothesis tests make judgements based on our belief that the percentage of the population will be the same as the percentage of the sample. But there is no guarantee.
False

Solution

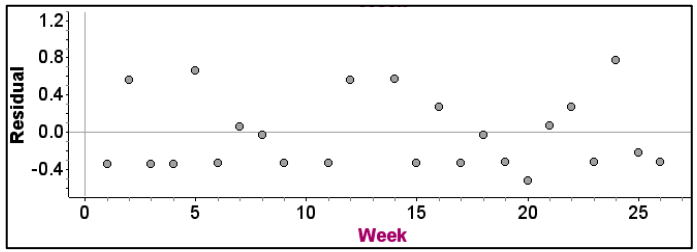
43.

As an experiment, Marci bought the exact same items at a market every week. If x represents the time in weeks and y represents the price of the items and linear regression was performed on x vs. y, the standard deviation of x is 7.6, the standard deviation of y is \$1.37 and $r = 0.89$. A residual plot is below. All of this information allows us to claim that there is fairly good evidence that the average price increase is about \$0.16 a week.

Using the formula $m = r\left(\frac{s_y}{s_x}\right), m = 0.89\left(\frac{1.37}{7.6}\right) = \$0.16$

The slope of the LSRL represents the change in y for one unit change in x. No pattern to the residual plot states that the best model is a linear one. The r value of 0.89 is fairly good evidence that the average price increase is about \$0.16 a week.

True



Solution

44.

At a catered lunch, 62% of people have coffee while the rest have lemonade. 70% of the coffee drinkers use an artificial sweetener while 40% of the lemonade drinkers use an artificial sweetener. If a person is chosen at random and uses an artificial sweetener, the chances that the person is a coffee drinker is greater than 3 chances out of 4.

Since coffee and lemonade are disjoint events, and there is no indication that having a drink and using an artifical sweetener are dependent,

$P(\text{coffee \& AS}) = (0.62)(0.7) = 0.434 \quad P(\text{lemonade \& AS}) = (0.38)(0.4) = 0.152$

$$P(\text{coffee} \mid \text{AS}) = \frac{0.434}{0.434 + 0.152} = \frac{0.434}{0.586} = 74.1\% \leq 75\%$$

OR

	AS	Total
Coffee	434	620
Lemonade	152	380
Total	586	1000

False

Solution

45.

Researchers are looking into the relationship between the weight of young boys and how often they wet the bed. The relationship is $B = -0.105W + 7.35,\ r^2 = 0.892,$ where W is average weight of the boy and B is the average number of times they wet the bed weekly.

The results of the research is that a strong negative linear association exists between weight and bed-wetting in young boys with a common response variable present.

Given the slope to be negative, the relationship is negative. That makes $r = -\sqrt{.892} = -0.944$ so the association is strong. But the weight doesn't cause boys to wet the bed more. The common response is age. Younger boys are lighter and will wet the bed more. Older boys are heavier and will wet the bed less. True

Solution

46.

The managers of a large company need to select 40 workers to be in a TV commercial. They randomly assign each worker a 3-digit number from 000 to 999. Men are given the numbers 000 - 459 while women are given the numbers 580 - 999. A table of random digits is used to choose 3-digit numbers. If the assignment of numbers to the workers is based on a probability sample, the number of women in the group selected to make the commercial is most likely to be 19.

Based on the numbers, there are $459-0+1=460$ men and $999-580+1=420$ women in the company. The probability of choosing a woman is $\frac{420}{880}$. So the expected number of women is $40\left(\frac{420}{880}\right)=19.09\approx 19$

True

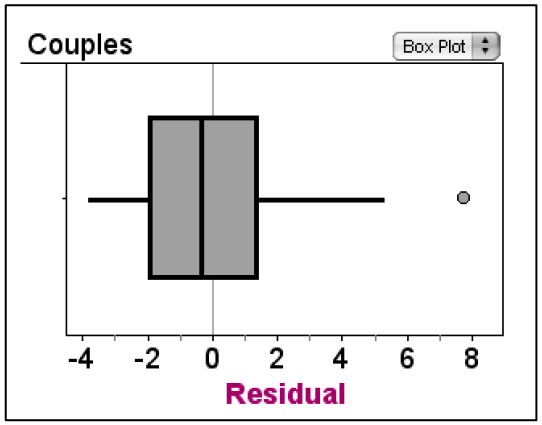
Solution

47.

A survey of 26 couples was made comparing the age of the husband and wife. The scatterplot below along with the LRSL is shown. If the residuals for this relationship are computed, a boxplot of their distributions cannot possibly be the one that is shown.

14 of the 26 points are clearly above the line meaning the majority of the residuals (observed data – predicted data) are positive. In this boplot, since the median is about –0.5, the majority of residuals are negative. Also the one outlier is about 8 meaning that there should be a point 8 units above the line.

True (boxplot is for the incorrect residual calculation as predicted – observed)



Solution

48.

Palm Beach County, FL has a population of 1.5 million people and on April 20, there were 2,150 people in that county who had COVID-19. We expect that most of them are quarantining so the chance of encountering one is close to zero. But with this virus, people can have it and not know it and be able to spread it to others. Assuming that the same number of people who had COVID-19 are asymptomatic (having COVID-19, but walking around not knowing it), the approximate probability of being in a movie theatre with 250 other people, with at least one of them having COVID-19 is about 30%.

The probability of someone having COVID-19 is $\frac{2150}{1500000} \approx 0.00143$.

So the probability of someone not having COVID-19 is $\left(1-\frac{2150}{1500000}\right)$.

Since with random people, having COVID-19 is mostly independent, the probability that all 250 will not have the virus is $\left(1-\frac{2150}{1500000}\right)^{250} \approx 0.699$

So the probability that at least one person has the virus is $1-0.699 \approx 30\%$.

True $\left(\begin{array}{l}\text{this explains why social distancing is so important. If 2,000 people were on} \\ \text{a beach, the probability that at least one person has the virus is about 94\%.}\end{array}\right)$

Solution

Those of us who own cats know that they occasionally throw up hairballs. A company that has a hairball medication on the market has developed a newer medication that will supposedly reduce hairballs. There is some concern, however, that whether a cat is a longhair or shorthair will change the effectiveness of the medication. If a randomized block exam is used, testing how the average number of hairballs a day changes based on medication, a two-sample t-test is appropriate with the blocking factor as whether the medication was the old or new one.

The test is correct but the blocking factor is not. Since we don't want the study to be biased on whether the cat is a shorthair or longhair, that is the blocking factor. The longhair cats randomly receive one of the medications and the average number of hairballs a day is measured for each med and since the population standard deviation is not known, a two-sample t-test is used. The same can be done on shorthairs.

49. False

Solution

50.

A movie theatre sells a large container of popcorn with an average of 54 ounces and a standard deviation of 1.6 ounces. The manager wishes to fill n cups with a margin of error of no more than 0.5 ounces with 95% confidence. He needs to fill at least 40 cups.

$$1.96\left(\frac{1.6}{\sqrt{n}}\right) < 0.5$$

$$\sqrt{n} > \frac{1.96(1.6)}{0.5} = 6.27$$

$$n > 39.34 \Rightarrow n = 40$$

True