***Solution Section* 4.1 − Inferences about Two Portions**

***Exercise***

A Student surveyed her friends and found that among 20 males, 4 smoke and among 30 female, 6 smoke. Give two reasons why these results should not be used for a hypothesis test of the claim that the proportions of male smokers and female smokers are equal.

***Solution***

There are two requirements for using the methods of this section, and each of them is violated.

1. The samples should be 2 sample random samples that are independent. These samples are convenience samples, not simple random samples. These samples are likely not independent. Since she surveyed her friends, she may well have males and females that are dating each other (or least that associate with each other) – and people tend to associate with those that have similar behaviors.
2. The number of successes for each sample should be at least 5, and the number of failures for each sample be at least 5. This is not true for the males, for which *x* = 4.

Using to estimate *p* and  to estimate *q*.

These inequalities state that the number of successes must be greater than 5, and the number of failures must be greater than 5.

***Exercise***

In clinical trials of the drug Zocor, some subjects were treated with Zocor and other were given a placebo. The 95% confidence interval estimate of the difference between the proportions of subjects who experienced headaches is . Write a statement interpreting that confidence interval.

***Solution***

We have 95% confidence that the limits of −0.0518 and 0.01094 contain the true difference between the population proportions of subjects who experience headaches. Repeating the trials many times would result in confidence limits that would include the true difference between the population proportions 95% of the time. Since the interval includes the value 0, there is no significant difference between the two population proportions.

***Exercise***

Among 8834 malfunctioning pacemakers, in 15.8% the malfunctions were due to batteries. Find the number of successes *x*.

***Solution***





***Exercise***

Among 129 subjects who took Chantix as an aid to stop smoking, 12.4% experienced nausea. Find the number of successes *x*.

***Solution***





***Exercise***

Among 610 adults selected randomly from among the residents of one town, 26.1% said that they have favor stronger gun-control laws. Find the number of successes *x*.

***Solution***





***Exercise***

A computer manufacturer randomly selects 2,410 of its computers for quality assurance and finds that 3.13% of these computer are found defective. Find the number of successes *x*.

***Solution***





***Exercise***

Assume that you plan to use a significance level of  to test the claim that . Use the given sample sizes and number of successes to find the pooled estimate 

1. 
2. 

***Solution***

1.  



1.  



***Exercise***

The numbers of online applications from simple random samples of college applications for 2003 and for the current year are given below.

|  |  |  |
| --- | --- | --- |
|  | 2003 | Current Year |
| Number of application in sample | 36 | 27 |
| Number of online applications in sample | 13 | 14 |

Assume that you plan to use a significance level of  to test the claim that  . Find

1. The pooled estimate 
2. The *x* test statistic
3. The critical *z* values
4. The *P*−value

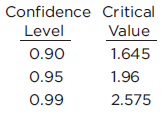
Assume 95% confidence interval

1. The margin of error E
2. The 95% confidence interval.

***Solution***



1. 
2. 





1. For , the critical values are



1.  



1. 





1. 





***Exercise***

Chantix is a drug used as an aid to stop smoking. The numbers of subjects experiencing insomnia for each of two treatment groups in a clinical trial of the drug Chantix are given below:

|  |  |  |
| --- | --- | --- |
|  | Chantix Treatment | Placebo |
| Number in group | 129 | 805 |
| Number experiencing insomnia | 19 | 13 |

Assume that you plan to use a significance level of  to test the claim that  . Find

1. The pooled estimate 
2. The *x* test statistic
3. The critical *z* values
4. The *P*−value

Assume 95% confidence interval

1. The margin of error E
2. The 95% confidence interval.

***Solution***



1. 
2. 





1. For , the critical values are











1. 



1. 





1. 





***Exercise***

In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Use a 0.05 significance level to test the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

***Solution***



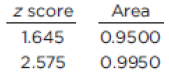




Original Claim: 



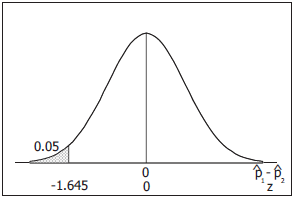
 

, the critical value





|  |  |
| --- | --- |
|  |  |



|  |  |
| --- | --- |
|  |  |

***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

***Exercise***

In a 1993 survey of 560 college students, 171 said that they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

***Solution***

|  |  |
| --- | --- |
|  |  |

Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and that the proportion of college students using illegal drugs in 1993 was less than it is now.

***Exercise***

A simple random sample of front-seat occupants involved in car crashes is obtained. Among 2823 occupants not wearing seat belts, 31 were killed. Among 7765 occupants wearing seat belts, 16 were killed. Construct a 90% confidence interval estimate of the difference between the fatality rates for those not wearing seat belts and those wearing seat belts. What does the result suggest about the effectiveness of seat belts?

***Solution***

|  |  |
| --- | --- |
|  |  |

Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and that seat belts are effective because the proportion of non-users who killed is greater than the proportion of users who are killed.

***Exercise***

A Pew Research Center poll asked randomly selected subjects if they agreed with the statement that “It is morally wrong fir married people to have an affair” Among the 386 women surveyed, 347 agrees with the statement. Among the 359 men surveyed, 305 agreed with the statement.

1. Use a 0.05 significance level to test the claim that the percentage of women who agree is difference from the percentage of men who agree. Does there appear to be a difference in the way women and men feel about this issue?
2. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.05 significance level. What conclusion does the confidence interval suggest?

***Solution***

1.  





*Original Claim*: 





, 









|  |  |
| --- | --- |
|  |  |

|  |  |
| --- | --- |
|  |  |

***Conclusion***:

Reject ; there is sufficient evidence to conclude that  (in fact, that . There is sufficient evidence to support the claim that the percentage of women who agree is different from the percentage of men who agree. Yes; there does appear to be a difference in the way that women and men feel about the issue.

1. 











Since the confidence interval does not include the value 0, it suggests that the two population proportions are not equal and the percentage of women who agree is different from the percentage of men who agree. Since the interval includes only positive values, conclude that the percentage of women who agree is greater than the percentage of men who agree.

***Exercise***

Tax returns include an option of designating $3 for presidential election campaigns, and it does not cost the taxpayer anything to make that designation. In a simple random sample of 250 tax returns from 1976, 27.6% of the returns designated the $3 for the campaign. In a simple random sample of 300 recent tax returns, 7.3% of the returns designated the $3 for the campaign. Use a 0.05 significance level to test the claim that the percentage of returns designating the $3 for the campaign was greater in 1973 than it is now.

***Solution***







*Original Claim*: 

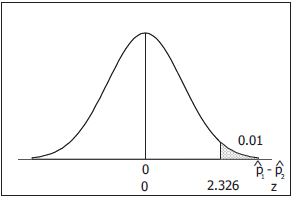


















|  |  |
| --- | --- |
|  |  |

***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the percentage of returns designated funds for campaigns was greater on 1976 than it is now.

***Exercise***

In an experiment, 16% of 734 subjects treated with Viagra experienced headaches. In the same experiment, 4% of 725 subjects given a placebo experienced headaches.

1. Use a 0.01 significance level to test the claim that the proportion of headaches is greater for those treated with Viagra. Do headaches appear to be a concern for those who take Viagra?
2. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.01 significance level. What conclusion does the confidence interval suggest?

***Solution***

1.  





*Original Claim*: 



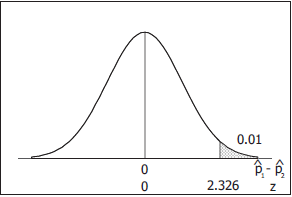












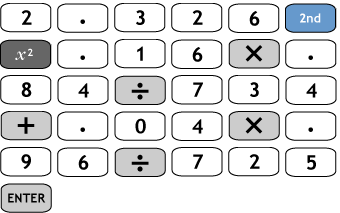




|  |  |
| --- | --- |
|  |  |

***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the proportion of persons experiencing headaches is greater for those treated with Viagra. Yes; headaches do appear to be a concern for those who take Viagra.

1. 









Since the confidence interval does not include the value 0, there is a significant difference the two proportions. Since the confidence interval includes only positive values, the proportion of persons experiencing headaches is greater for those treated with Viagra.

***Exercise***

Two different simple random samples are drawn from two different populations. The first sample consists of 20 people with 10 having a common attribute. The second sample consists of 2000 people with 1404 of them having the same common attribute. Compare the results from a hypothesis test of  (with a 0.05 significance level) and a 95% confidence interval estimate of .

***Solution***





*Original Claim*: 





, 









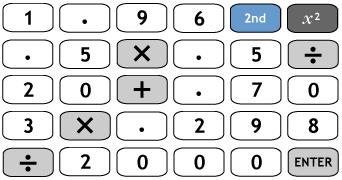
|  |  |  |
| --- | --- | --- |
|  | |  |
|  |  | |

***Conclusion***:

Reject ; there is sufficient evidence to conclude that  and conclude that  (in fact, that ).

The confidence interval is:











Since the confidence interval includes the value 0, could have the same values and one should not reject the claim that .

The test of hypothesis and the confidence interval lead to different conclusions. In this instance, they are not equivalent.

***Exercise***

A report on the nightly news broadcast stated that 11 out of 142 households with pet dogs were burglarized and 21 out of 217 without pet dogs were burglarized. Find the *z* test statistic for the hypothesis test.

Assume that you plan to use a significance level of  to test the claim that .

***Solution***











***Exercise***

Assume that the samples are independent and that they have been randomly selected. Construct a 90% confidence interval for the difference between population proportions



***Solution***















***Exercise***

The sample size needed to estimate the difference between two population proportions ti within a margin of error *E* with a confidence level of 1 − *α* can be found as follows:

.

In this expression, replace  and  by  (assuming both samples have the same size) and replace each of  by 0.5 (because their values are not known). Then solve for n.

Use this approach to find the size pf each sample of you want to estimate the difference between the proportions of men and women who plan to vote in the next presidential election. Assume that you want 99% confidence that your error is no more than 0.05.

***Solution***













***Solution Section* 4.2 − Inferences About Two Means: Independent Samples**

***Exercise***

If the pulse rates of men and women shown in the data below

Women:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 76 | 72 | 88 | 60 | 72 | 68 | 80 | 64 | 68 | 68 | 80 | 76 | 68 | 72 | 96 | 72 | 68 | 72 | 64 | 80 |
| 64 | 80 | 76 | 76 | 76 | 80 | 104 | 88 | 60 | 76 | 72 | 72 | 88 | 80 | 60 | 72 | 88 | 88 | 124 | 64 |

Men:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 68 | 64 | 88 | 72 | 64 | 72 | 60 | 88 | 76 | 60 | 96 | 72 | 56 | 64 | 60 | 64 | 84 | 76 | 84 | 88 |
| 72 | 56 | 68 | 64 | 60 | 68 | 60 | 60 | 56 | 84 | 72 | 84 | 88 | 56 | 64 | 56 | 56 | 60 | 64 | 72 |

These data are used to construct 95% confidence interval for the difference between the two population means, the result is , where pulse rates of men correspond to population 1 and pulse rates of women correspond to population 2. Express the confidence interval with pulse rates of women being population 1 and pulse rates of men being population 2.

***Solution***

Reversing the designation of which sample is considered group 1 and which sample is considered group 2 changes the sign of the point estimate and the signs of the endpoints of the interval estimate. The confidence interval using the new designation is 

***Exercise***

Assume that you want to use a 0.01 significance level to test the claim that the mean pulse rate of men is less than the mean pulse rate of women. What confidence level should be used if you want to test that claim using a confidence interval?

***Solution***

A one-tailed test of hypothesis at the 0.01 level of significance corresponds to a two-sided confidence interval at the 2(0.01) = 0.02 level of significance –i.e., to an interval with a confidence level of 98%

***Exercise***

To test the effectiveness of Lipitor, cholesterol levels are measured in 250 subjects before and after Lipitor treatments. Determine whether this sample is independent or dependent.

***Solution***

Dependent, since cholesterol levels and determined by many factors that the Lipitor treatment cannot change. Treatments to lower cholesterol typically reduce everyone’s levels by a certain amount, by persons who were high compared to the others before the treatment, for example, will likely still be high compared to the others after the treatment.

***Exercise***

On each of 40 different days, you measured the voltage supplied to your home and you also measured the voltage produced by the gasoline-powered generator. One sample consists of the voltages in the house and the second sample consists of the voltages produced by the generator. Determine whether this sample is independent or dependent.

***Solution***

Independent, since there is no relationship between the voltage supplied to the house by the power company and the voltage generated by a completely separate gasoline-powered generator.

***Exercise***

In a randomized controlled trial conducted with children suffering from viral croup, 46 children were treated with low humidity while 46 other children were treated with high humidity. Researchers used the Westley Croup Score to assess the results after one hour. The low humidity group had a mean score of 0.98 with standard deviation of 1.22 while the high humidity group had a mean score of 1.09 with standard deviation of 1.11.

1. Use a 0.05 significance level to test the claim that the two groups are from populations with the same mean. What does the result suggest about the common treatment of humidity?

*Assume that the two samples are independent simple random samples selected from normally distributed populations.*

1. Assume that , how are the results affected by this additional assumption?

***Solution***

1. *Original Claim*: 



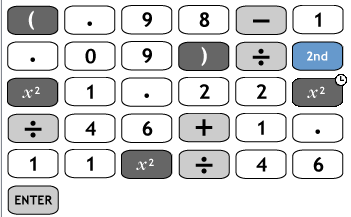




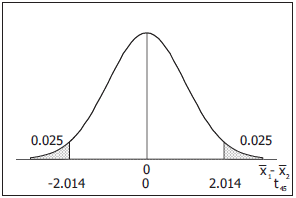


Critical value: 















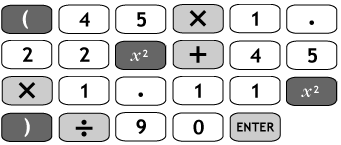
***Conclusion***:

Do not reject ; there is not sufficient evidence to reject the claim that . There is not sufficient evidence to reject the claim that the two groups are from populations with the same mean. The results suggest that increasing the humidity does not have a significant effect on the treatment of croup.

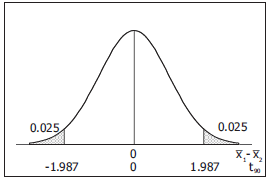
1. 









*Original Claim*: 







Critical value: 











***Conclusion***:

Do not reject ; there is not sufficient evidence to reject the claim that . There is not sufficient evidence to reject the claim that the two groups are from populations with the same mean. The results suggest that increasing the humidity does not have a significant effect on the treatment of croup.

When , the calculated *t* statistic does not change at all. The only difference the assumption of equal standard deviations makes in this instance is to change the df from 45 to 90 and the P-value from 0.6532 to 0.6521. The conclusion is unaffected.

***Exercise***

The mean tar content of a simple random sample of 25 unfiltered king size cigarettes is 21.1 mg, with a standard deviation of 3.2 mg. The mean tar content of a simple random sample of 25 filtered 100 mm cigarettes is 13.2 mg, with a standard deviation of 3.7 mg.

*Assume that the two samples are independent simple random samples selected from normally distributed populations in part a and b.*

1. Construct a 90% confidence interval estimate of the difference between the mean tar content of unfiltered king size cigarettes and the mean tar content of filtered 100 mm cigarettes. Does the result suggest that 100 mm filtered cigarettes have less tar than unfiltered king size cigarettes?
2. Use a 0.05 significance level to test the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. What does the result suggest about the effectiveness of cigarette filters?
3. Assume that , how are the results affected by this additional assumption?

***Solution***

1. Let the unfiltered cigarettes be group 1.



Critical value: 





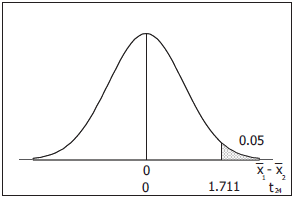






Yes; since the confidence interval includes only positive values, the results suggest that the filtered cigarettes have less tar than the unfiltered ones.

1. *Original Claim*: 





Critical value: 











***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. The results suggest that filters are effective in reducing the tar content cigarettes.

1. 



















Yes; since the confidence interval includes only positive values, the results suggest that the filtered cigarettes have less tar than the unfiltered ones.

When  the value of is unchanged. The only difference the assumption of equal standard deviations makes in this instance is to change the *df* from 24 to 48 and the from 1.711 to 1.676. This makes the interval slightly narrower, but the conclusion is unaffected.

***Exercise***

The heights are measured for the simple random sample of supermodels Crawford, Bundchen, Pestova, Christenson, Hume, Moss, Campbell, Schiffer, and Taylor. They have a mean of 70.0 in. and a standard deviation of 1.5 in. 40 women who are not supermodels, listed below and they have heights with means of 63.2 in. and a standard deviation of 2.7 in.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 64.3 | 66.4 | 62.3 | 62.3 | 59.6 | 63.6 | 59.8 | 63.3 | 67.9 | 61.4 | 66.7 | 64.8 | 63.1 | 66.7 | 66.8 |
| 64.7 | 65.1 | 61.9 | 64.3 | 63.4 | 60.7 | 63.4 | 62.6 | 60.6 | 63.5 | 58.6 | 60.2 | 67.6 | 63.4 | 64.1 |
| 62.7 | 61.3 | 58.2 | 63.2 | 60.5 | 65.0 | 61.8 | 68.0 | 67.0 | 57.0 |  |  |  |  |  |

1. Use a 0.01 significance level to test the claim that the mean height of supermodels is greater than the mean height of women who are not supermodels
2. Construct a 98% confidence interval level for the difference between the mean height of supermodels and the mean height of women who are not supermodels. What does the result suggest about those two means?

***Solution***

1. Let the supermodels be group 1. For which 

*Original Claim*: 



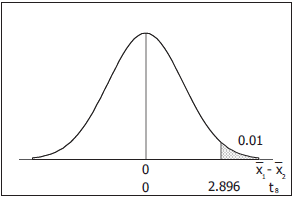




Critical value: 















***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the mean height of supermodels is greater than the mean height of women who are not supermodels.

1. Let the supermodels be group 1. For which .









Since the confidence interval includes only positive values, the results suggest that the mean height of supermodels is greater than the mean height of women who are not supermodels.

***Exercise***

Many studies have been conducted to test the effects of marijuana use on mental abilities. In one such study, groups of light and heavy users of marijuana in college were tested for memory recall, with the results given below. Use a 0.01 significance level to test the claim that the population of heavy marijuana users has a lower mean than the light users. Should marijuana use be of concern to college students?





***Solution***

*Original Claim*: 



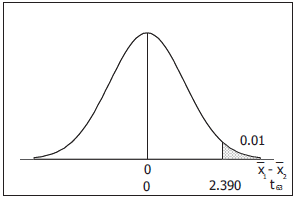




Critical value: 













***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that heavy marijuana users have a lower mean number of recalled items than do light users.

Yes; marijuana use should be of concern to college students – and an even more valuable study might one comparing light users to those who do not use marijuana at all.

***Exercise***

The trend of thinner Miss America winners has generated charges that the contest encourages unhealthy diet habits among young women. Listed below are body mass indexes (BMI) for Miss America winners from two different time periods. Consider the listed values to be simple random samples selected from larger populations.

1. Use a 0.05 significance level to test the claim that recent winners have a lower mean BMI than winners from the 1920s and 1930s.
2. Construct a 90% Confidence interval for the difference between the mean BMI of recent winners and the mean BMI of winners from the 1920s and 1930s.



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| BMI (from recent winners): | 19.5 | 20.3 | 19.6 | 20.2 | 17.8 | 17.9 | 19.1 | 18.8 | 17.6 | 16.8 |
| BMI (from 1920s and 1930s): | 20.4 | 21.9 | 22.1 | 22.3 | 20.3 | 18.8 | 18.9 | 19.4 | 18.4 | 19.1 |

***Solution***

|  |  |
| --- | --- |
| ***Group*** 1: recent (*n* = 10) |  |
| ***Group*** 2: 1920,1930 (*n* = 10) |  |

1. *Original Claim*: 



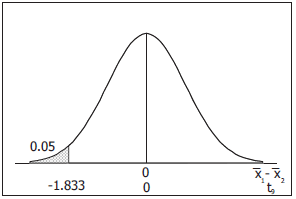




Critical value: 















***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that recent winners have a lower mean BMI than winners from the 1920s and 1930s.

1. .







***Exercise***

Listed below are amounts of strontium-90 (in millibecquerels or mBq per gram of calcium) in a simple random sample of baby teeth obtained from Pennsylvania residents and New York residents born after 1979.

1. Use a 0.05 significance level to test the claim that the mean amount of strontium-90 from Pennsylvania residents is greater than the mean amount from New York residents.
2. Construct a 90% Confidence interval for the difference between the mean amount of strontium-90 from Pennsylvania residents and the mean amount from New York residents.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Pennsylvania: | 155 | 142 | 149 | 130 | 151 | 163 | 151 | 142 | 156 | 133 | 138 | 161 |
| New York: | 133 | 140 | 142 | 131 | 134 | 129 | 128 | 140 | 140 | 140 | 137 | 143 |

***Solution***

|  |  |
| --- | --- |
| ***Group*** 1: PA (*n* = 12) |  |
| ***Group*** 2: NY (*n* = 12) |  |

1. *Original Claim*: 



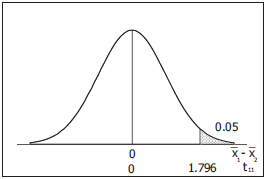
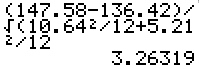




Critical value: 





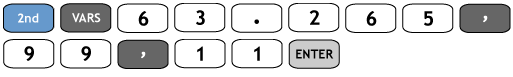










***Conclusion***:

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the mean amount of Strontium-90 from Pennsylvania residents is greater than the mean amount from N.Y. residents.

1. .







***Exercise***

Listed below are the word counts for male and female psychology students.

1. Use a 0.05 significance level to test the claim that male and female psychology students speak the same mean number of words in a day.
2. Construct a 95% Confidence interval estimate of the difference between the mean number of words spoken in a day by male and female psychology students. Do the confidence interval limits include 0, and what does that suggest about the two means?

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Male*** | 21143 | 17791 | 36571 | 6724 | 15430 | 11552 | 11748 | 12169 | 15581 | 23858 | 5269 |
|  | 12384 | 11576 | 17707 | 15229 | 18160 | 22482 | 18626 | 1118 | 5319 |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Female*** | 6705 | 21613 | 11935 | 15790 | 17865 | 13035 | 24834 | 7747 | 3852 | 11648 | 25862 |
|  | 17183 | 11010 | 11156 | 11351 | 25693 | 13383 | 19992 | 14926 | 14128 | 10345 | 13516 |
|  | 12831 | 9671 | 17011 | 28575 | 23557 | 13656 | 8231 | 10601 | 8124 |  |  |

*Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.*

***Solution***

|  |  |
| --- | --- |
| ***Group*** 1: Males (*n* = 20) | ***Group*** 2: Females (*n* = 31) |

1. *Original Claim*:  ***words/day***



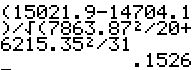
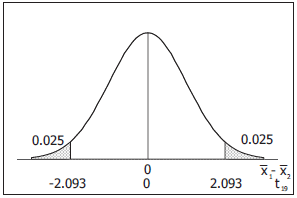




Critical value: 















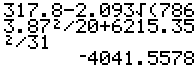
***Conclusion***:

Do not reject ; there is not sufficient evidence to reject the claim . There is not sufficient evidence to reject the claim that the male and female students speak the same mean number of words per day.

1. .







Yes; since the confidence interval includes zero, there does not appear to be significant difference between the mean number of words spoken by the male and female students.

***Exercise***

Refer to the tables below and test the claim that they contain the same amount of cola, the mean weight of cola cans of regular Coke is the same as the mean weight of cola in cans of Diet Coke. If there is a difference in the mean weights, identify the most likely explanation for that difference.

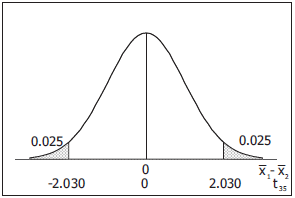
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Coke*** | 0.8192 | 0.815 | 0.8163 | 0.8211 | 0.8181 | 0.8247 | 0.8062 | 0.8128 | 0.8172 | 0.811 |
| 0.8251 | 0.8264 | 0.7901 | 0.8244 | 0.8073 | 0.8079 | 0.8044 | 0.817 | 0.8161 | 0.8194 |
| 0.8189 | 0.8194 | 0.8176 | 0.8284 | 0.8165 | 0.8143 | 0.8229 | 0.815 | 0.8152 | 0.8244 |
| 0.8207 | 0.8152 | 0.8126 | 0.8295 | 0.8161 | 0.8192 |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Diet*** | 0.7773 | 0.7758 | 0.7896 | 0.7868 | 0.7844 | 0.7861 | 0.7806 | 0.783 | 0.7852 | 0.7879 |
| 0.7881 | 0.7826 | 0.7923 | 0.7852 | 0.7872 | 0.7813 | 0.7885 | 0.776 | 0.7822 | 0.7874 |
| 0.7822 | 0.7839 | 0.7802 | 0.7892 | 0.7874 | 0.7907 | 0.7771 | 0.787 | 0.7833 | 0.7822 |
| 0.7837 | 0.791 | 0.7879 | 0.7923 | 0.7859 | 0.7811 |  |  |  |  |

*Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.*

***Solution***

|  |  |
| --- | --- |
| ***Group*** 1: Regular Coke (*n* = 36) | ***Group*** 2: Diet Coke (*n* = 36) |

*Original Claim*:  ***lbs***



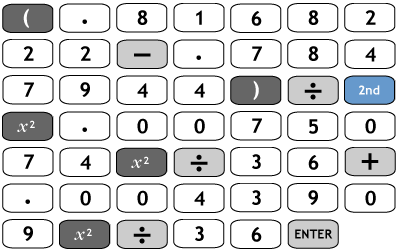


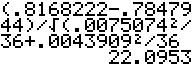


Critical value: 













***Conclusion***:

Reject ; there is sufficient evidence to reject the claim that  and conclude that  ( in fact, that ). There is sufficient evidence to reject the claim that the mean weight of cola in cans of regular Coke is the same as the mean weight of cola in cans of Diet Coke. The regular Coke may weigh more because it contains sugar.

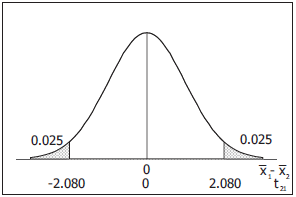
***Exercise***

An Experiment was conducted to test the effects of alcohol. Researchers measured the breath alcohol levels for a treatment group of people who drank ethanol and another group given a placebo. The results are given in the accompanying table. Use a 0.05 significance level to test the claim that the two sample groups come from populations with the same mean.

|  |  |  |  |
| --- | --- | --- | --- |
| Treatment Group: |  |  |  |
| Placebo Group: |  |  |  |

***Solution***

*Original Claim*:  ***lbs***





Critical value: 











***Conclusion***:

Reject ; there is sufficient evidence to reject the claim that  and conclude that  ( in fact, that ). There is sufficient evidence to reject the claim that the two sample groups come from populations with the same mean.

The fact that there was no variation in the second sample did not affect the calculations or present any special problems. Since there is no variation in , it is really equivalent to the constant value zero – and the test is mathematically equivalent to the one-sample test  for which 

***Exercise***

A researcher was interested in comparing the GPAs of students at two different colleges. Independent simple populations. Do samples of 8 students from college *A* and 13 students from college *B* yielding the following GPAs.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **College *A*** | 3.7 | 3.2 | 3.0 | 2.5 | 2.7 | 3.6 | 2.8 | 3.4 |  |  |  |  |  |
| **College *B*** | 3.8 | 3.2 | 3.0 | 3.9 | 3.8 | 2.5 | 3.9 | 2.8 | 4.0 | 3.6 | 2.6 | 4.0 | 3.6 |

Construct a 95% confidence interval for . The difference between the mean GPA of college *A* students and the mean GPA of college *B* students.



***Solution***



Critical value: 









***Exercise***

Assume that the two samples are independent simple random samples selected from normal distributed populations. Do not assume that the population standard deviations are equal.

A researcher was interested in comparing the heights of women in two different countries. Independent simple random samples of 9 women from country ***A*** and 9 women from ***B*** yielded to the following heights (in inches).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Country A*** | 64.1 | 66.4 | 61.7 | 62.0 | 67.3 | 64.9 | 64.7 | 68.0 | 63.6 |
| ***Country B*** | 65.3 | 60.2 | 61.7 | 65.8 | 61.0 | 64.6 | 60.0 | 65.4 | 59.0 |

Construct a 90% confidence interval for  the difference between the mean height of women in country ***A*** and the mean height of women in country ***B***. Round to two decimal places.

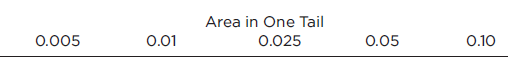


***Solution***















***Solution Section* 4.3 − Inferences from Dependent Samples**

***Exercise***

Listed below are the time intervals (in minutes) before and after eruptions of the Old Faithful geyser. Find the values of  and . In general, what does  represent?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Time interval before eruption* | 98 | 92 | 95 | 87 | 96 |
| *Time interval after eruption* | 92 | 95 | 92 | 100 | 90 |

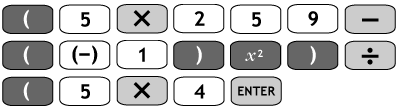
***Solution***

The difference values are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 98 | 92 | 95 | 87 | 96 |
|  | 92 | 95 | 92 | 100 | 90 |
| ***Difference = d*** | 6 | −3 | 3 | −13 | 6 |
|  | 36 | 9 | 9 | 169 | 36 |













In general,  represents the true mean of the differences from the population of matched pairs (which is mathematically equivalent to the true of the difference between the means of the two populations).

***Exercise***

Listed below are measured fuel consumption amount (in miles/gal) from a sample of cars.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *City fuel consumption* | 18 | 22 | 21 | 21 |
| *Highway fuel consumption* | 26 | 31 | 29 | 29 |

Assume that wou want to use a 0.05 significance level to test the claim that the paired sample data come from a population for which the mean differecne is . Find

1. 
2. 
3. The *t* test statistic
4. The critical values.

***Solution***

The difference values are:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 18 | 22 | 21 | 21 |
|  | 26 | 31 | 29 | 29 |
| ***Difference = d*** | -8 | -9 | -8 | -8 |
|  | 64 | 81 | 64 | 64 |



1. 
2. 



1. 
2. With , the critical values are 

***Exercise***

Listed below are predicted high temperatures that were forecast different days.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Predicted high temperatures forecast* 3 *days ahead* | 79 | 86 | 79 | 83 | 80 |
| *Predicted high temperatures forecast* 5 *days ahead* | 80 | 80 | 79 | 80 | 79 |

Assume that wou want to use a 0.05 significance level to test the claim that the paired sample data come from a population for which the mean differecne is . Find

1. 
2. 
3. The *t* test statistic
4. The critical values.

***Solution***

The difference values are:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 79 | 83 | 79 | 83 | 80 |
|  | 80 | 80 | 79 | 80 | 79 |
| ***Difference = d*** | −1 | 6 | 0 | 3 | 1 |
|  | 1 | 36 | 0 | 9 | 1 |



1. 
2. 



1. 
2. With , the critical values are 

***Exercise***

Listed below are body mass indices (BMI). The BMI of each student was measured in September and April of the freshman year.

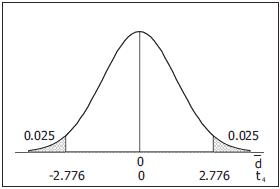
1. Use a 0.05 significance level to tet the claim that the mean change in BMI for all students is equal to 0. Does BMI appear to change during freshman year?
2. Construct a 95% confidence interval estimate of the change in BMI during freshman year. Does the confidence interval include 0, and what does that suggest about BMI during freshman year?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *April BMI* | 20.15 | 19.24 | 20.77 | 23.85 | 21.32 |
| *September BMI* | 20.68 | 19.48 | 19.59 | 24.57 | 20.96 |

***Solution***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 20.15 | 19.24 | 20.77 | 23.85 | 21.32 |
|  | 20.68 | 19.48 | 19.59 | 24.57 | 20.96 |
| ***Difference = d*** | −0.53 | −0.24 | 1.18 | −0.72 | 0.36 |
|  | .2809 | .0576 | 1.3924 | .5184 | .1296 |



1. Original claim: 



















***Conclusion:***

Do not reject ; there is not sufficient evidence to reject the claim . There is not sufficient evidence to reject the claim that the mean change in BMI for all students is equal to 0.

No; BMI does not appear to change during the freshman year.

1. 







Yes; the confidence interval includes 0, which suggests that the mean of the differences could be 0 and that there is no change in BMI during the freshman year

***Exercise***

Listed below are body temperature (in °F) of subjects measured at 8:00 AM and at 12:00 AM. Construct a 95% confidence interval estimate of the difference between the 8:00 AM temperatures and the 12:00 AM temperatures. Is body temperature basically the same at both times?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 8:00 AM | 97.0 | 96.2 | 97.6 | 96.4 | 97.8 | 99.2 |
| 12:00 AM | 98.0 | 98.6 | 98.8 | 98.0 | 98.6 | 97.6 |

***Solution***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 97.0 | 96.2 | 97.6 | 96.4 | 97.8 | 99.2 |
|  | 98.0 | 98.6 | 98.8 | 98.0 | 98.6 | 97.6 |
| ***Difference = d*** | −1.0 | −2.4 | −1.2 | −1.6 | −0.8 | 1.6 |
|  | 1 | 5.76 | 1.44 | 2.56 | .64 | 2.56 |

















Yes; since the confidence intervals includes 0, body temperature is basically the same at both times.

***Exercise***

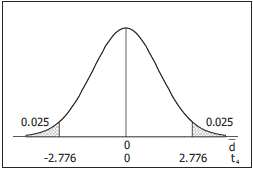
Listed below are systolic blood pressure measuremetns (mm Hg) taking rom the right and left arms of the same woman. Use a 0.05 significance level to test for a difference the measurements from the two arms. What do you conclude?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Right arm | 102 | 101 | 94 | 79 | 79 |
| Left arm | 175 | 169 | 182 | 146 | 144 |

***Solution***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 102 | 101 | 94 | 79 | 79 |
|  | 175 | 169 | 182 | 146 | 144 |
| ***Difference = d*** | −73 | −68 | −88 | −67 | −65 |
|  | 5329 | 4624 | 7744 | 4489 | 4225 |



















***Conclusion:***

Reject ; there is sufficient evidence to conclude that  (in fact, that ). There is sufficient evidence to support the claim that there is a difference in measurements between the two arms. The statistical conclusion is that the right arm. Since the right and left arms should yield the same measurements, the practical conclusion is that a mistake has been made.

The most obvious explanation is that diastolic (and not the systolic) values were mistakenly recorded for the right arm. Further investigation is definitely in order.

***Exercise***

As part of the National Health and Nutrition Examination Survey, the Department of Health and Human Services obtained self-reported heights and measured heights for males ages 12 − 16. All measurement are in inches. Listed below are sample results

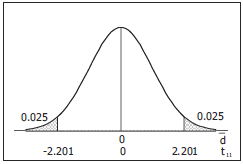
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Reported height* | 68 | 71 | 63 | 70 | 71 | 60 | 65 | 64 | 54 | 63 | 66 | 72 |
| *Measured height* | 67.9 | 69.9 | 64.9 | 68.3 | 70.3 | 60.6 | 64.5 | 67.0 | 55.6 | 74.2 | 65.0 | 70.8 |

1. Is there sufficient evidence to support the claim that there is a difference between self-reported heights and measured heights of males? Use a 0.05 significance level.
2. Construct a 95% confidence interval estimate of the man difference between reported heights and measured heights. Interpret the resulting confidence interval, and comment on the implications of whether the confidence interval limits contain 0.

***Solution***

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 68 | 71 | 63 | 70 | 71 | 60 | 65 | 64 | 54 | 63 | 66 | 72 |
|  | 67.9 | 69.9 | 64.9 | 68.3 | 70.3 | 60.6 | 64.5 | 67.0 | 55.6 | 74.2 | 65.0 | 70.8 |
| ***Difference = d*** | 0.1 | 1.1 | −1.9 | 1.7 | 0.7 | −0.6 | 0.5 | −3.0 | −1.6 | −11.2 | 1.0 | 1.2 |
|  | .01 | 1.21 | 3.61 | 2.89 | .49 | .36 | .25 | 9 | 2.56 | 125.44 | 1 | 1.44 |









Original claim: 



















***Conclusion:***

Do not reject ; there is not sufficient evidence to reject the claim . There is not sufficient evidence to support the claim that there is a difference between self-reported heights and measured height of such males.

1. 







Since the confidence interval contains 0, there is no significant difference between the reported and measured heights.

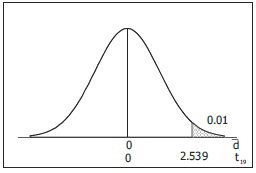
***Exercise***

Listed below are combined city – highway fuel consumption ratings (in miles/gal) for different cars measured under both the old rating system and a new rating system introducing in 2008. The new ratings were implemented in response to complaints that the old ratings were too high. Use a 0.01 significance level to test the claim the old ratings are higher than the new ratings.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Old rating* | 16 | 18 | 27 | 17 | 33 | 28 | 33 | 18 | 24 | 19 | 18 | 27 | 22 | 18 | 20 | 29 | 19 | 27 | 20 | 21 |
| *New rating* | 15 | 16 | 24 | 15 | 29 | 25 | 29 | 16 | 22 | 17 | 16 | 24 | 20 | 16 | 18 | 26 | 17 | 25 | 18 | 19 |

***Solution***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 16 | 18 | 27 | 17 | 33 | 28 | 33 | 18 | 24 | 19 | 18 | 27 | 22 | 18 | 20 | 29 | 19 | 27 | 20 | 21 |
|  | 15 | 16 | 24 | 15 | 29 | 25 | 29 | 16 | 22 | 17 | 16 | 24 | 20 | 16 | 18 | 26 | 17 | 25 | 18 | 19 |
| ***Diff = d*** | 1 | 2 | 3 | 2 | 4 | 3 | 4 | 2 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 3 | 2 | 2 | 2 | 2 |
|  | 1 | 4 | 9 | 4 | 16 | 9 | 16 | 4 | 4 | 4 | 4 | 9 | 4 | 4 | 4 | 9 | 4 | 4 | 4 | 4 |







Original claim: 



















***Conclusion:***

Reject ; there is sufficient evidence to conclude that . There is sufficient evidence to support the claim that the old ratings are higher than the new ratings.

***Exercise***

Listed below are 2 tables. Construct a 95% confidence interval estimate of the mean of the differences between weights of discarded paper and weights of discarded plastic. Which seems to weigh more: discarded paper or discarded plastic?

*Paper*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.41 | 7.57 | 9.55 | 8.82 | 8.72 | 6.96 | 6.83 | 11.42 | 16.08 | 6.38 | 13.05 | 11.36 | 15.09 |
| 2.80 | 6.44 | 5.86 | 11.08 | 12.43 | 6.05 | 13.61 | 6.98 | 14.33 | 13.31 | 3.27 | 6.67 | 17.65 |
| 12.73 | 9.83 | 16.39 | 6.33 | 9.19 | 9.41 | 9.45 | 12.32 | 20.12 | 7.72 | 6.16 | 7.98 | 9.64 |
| 8.08 | 10.99 | 13.11 | 3.26 | 1.65 | 10.00 | 8.96 | 9.46 | 5.88 | 8.26 | 12.45 | 10.58 | 5.87 |
| 8.78 | 11.03 | 12.29 | 20.58 | 12.56 | 9.92 | 3.45 | 9.09 | 3.69 | 2.61 |  |  |  |

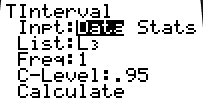
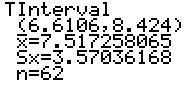
*Plastic*

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0.27 | 1.41 | 2.19 | 2.83 | 2.19 | 1.81 | 0.85 | 3.05 | 3.42 | 2.10 | 2.93 | 2.44 | 2.17 |
| 1.41 | 2.00 | 0.93 | 2.97 | 2.04 | 0.65 | 2.13 | 0.63 | 1.53 | 4.69 | 0.15 | 1.45 | 2.68 |
| 3.53 | 1.49 | 2.31 | 0.92 | 0.89 | 0.80 | 0.72 | 2.66 | 4.37 | 0.92 | 1.40 | 1.45 | 1.68 |
| 1.53 | 1.44 | 1.44 | 1.36 | 0.38 | 1.74 | 2.35 | 2.30 | 1.14 | 2.88 | 2.13 | 5.28 | 1.48 |
| 3.36 | 2.83 | 2.87 | 2.96 | 1.61 | 1.58 | 1.15 | 1.28 | 0.58 | 0.74 |  |  |  |

***Solution***

Using Ti-84, store paper into List 1 and plastic in List 2

To create list 3: **[2nd]** **1** (L1) − **[2nd]** **2** (L2) [***STO→***] **[2nd]** **3** (L3)



Since the confidence interval includes only positive values, there discarded paper appears to weigh more than the discarded plastic.

***Exercise***

Suppose you wish to test the claim that , the mean value of the differences d for a population of paired data, is different from 0. Given a sample of  and a significance level of  what criterion would be used for rejecting the null hypothesis?

***Solution***

Given:   and 





To reject null hypothesis if test statistic is:  or 

***Exercise***

Assume that he paired data came from a population that is normally distributed. Using a 0.05 significance level, find , , the *t* test statistic, and the critical values to test the claim that 

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 14 | 8 | 4 | 14 | 3 | 12 | 4 | 13 |
| ***y*** | 15 | 8 | 7 | 13 | 5 | 11 | 6 | 15 |

***Solution***

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 14 | 8 | 4 | 14 | 3 | 12 | 4 | 13 |  |
|  | 15 | 8 | 7 | 13 | 5 | 11 | 6 | 15 |  |
| ***Difference = d*** | −1 | 0 | −3 | 1 | −2 | 1 | −2 | −2 |  |
|  | 0 | 1 | 4 | 4 | 1 | 4 | 1 | 1 | 16 |















***Exercise***

Assume that he paired data came from a population that is normally distributed. Using a 0.05 significance level, find , , the *t* test statistic, and the critical values to test the claim that 

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 12 | 5 | 1 | 20 | 3 | 16 | 12 | 8 |
| ***y*** | 7 | 10 | 5 | 15 | 7 | 14 | 10 | 13 |

***Solution***

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 12 | 5 | 1 | 20 | 3 | 16 | 12 | 8 |
|  | 7 | 10 | 5 | 15 | 7 | 14 | 10 | 13 |
| ***Difference = d*** | 5 | −5 | −4 | 5 | −4 | 2 | 2 | −5 |
|  | 25 | 25 | 16 | 25 | 16 | 4 | 4 | 25 |















***Solution*** ***Section* 4.4 − Correlation**

***Exercise***

For each of several randomly selected years, the total number of points scored in the Super Bowl football game and the total number of new cars sold in The U.S. are recorded. For this sample of paired data

1. What does *r* represent?
2. What does *ρ* represent?
3. With our doing any research or calculations, estimate the value of *r*.

***Solution***

1. *r* = the correlation in the sample. In this context, *r* is the linear correlation coefficient computed using the chosen paired (points in Super Bowl, number of new cars sold) values for the randomly selected years in the sample.
2. *ρ*  = the correlation in the population. In this context, *ρ* is the linear correlation coefficient computed using the paired (points in Super Bowl, number of new cars sold) values for every year there has been a Super Bowl.
3. Since there is no relationship between the number of points scored in a Super Bowl and the number of new cars sold that year, the estimated value of *r* is 0.

***Exercise***

The heights (in inches) of a sample of eight mother/daughter pairs of subjects measured. Using Excel with the paired mother/daughter heights, the linear correlation coefficient is found to be 0.693. Is there sufficient evidence to support the claim that there is a linear correlation between the heights of mothers and the heights of their daughters? Explain.

***Solution***

From the table for . Therefore *r* = 0.693 indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the heights of mothers and the heights of their daughters,

***Exercise***

The heights and weights of a sample of 9 supermodels were measured. Using a TI calculator, the linear correlation coefficient is found to be 0.360. Is there sufficient evidence to support the claim that there is a linear correlation between the heights and weights of supermodels? Explain.

***Solution***

From the table for . Therefore *r* = 0.360 does not indicate a significant linear correlation. No; there is not sufficient evidence to support the claim that there is a linear correlation between the heights and weights of supermodels.

***Exercise***

Given the table below

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 10 | 8 | 13 | 9 | 11 | 14 | 6 | 4 | 12 | 7 | 5 |
| ***y*** | 9.14 | 8.14 | 8.74 | 8.77 | 9.26 | 8.10 | 6.13 | 3.10 | 9.13 | 7.26 | 4.74 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.
3. Identify the feature of the data that would be missed if part (*b*) was completed without constructing the scatterplot.

***Solution***

1. Excel produces the following

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 10 | 9.14 | 91.4 | 100 | 83.5396 |
| 8 | 8.14 | 65.12 | 64 | 66.2596 |
| 13 | 8.74 | 113.62 | 169 | 76.3876 |
| 9 | 8.77 | 78.93 | 81 | 76.9129 |
| 11 | 9.26 | 101.86 | 121 | 85.7476 |
| 14 | 8.1 | 113.4 | 196 | 65.6100 |
| 6 | 6.13 | 36.78 | 36 | 37.5769 |
| 4 | 3.1 | 12.4 | 16 | 9.6100 |
| 12 | 9.13 | 109.56 | 144 | 83.3569 |
| 7 | 7.26 | 50.82 | 49 | 52.7076 |
| 5 | 4.74 | 23.7 | 25 | 22.4676 |
| 99 | 82.51 | 797.59 | 1001 | 660.176 |

1. 





From table *A*-5; 

Therefore *r* = 0.816 indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the 2 variables.

1. The scatterplot indicates that the relationship between the variables is quadratic, not linear.

***Exercise***

Given the table below

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 10 | 8 | 13 | 9 | 11 | 14 | 6 | 4 | 12 | 7 | 5 | |
| ***y*** | 7.46 | 6.77 | 12.74 | 7.11 | 7.81 | 8.84 | 6.08 | 5.39 | 8.15 | 6.42 | 5.73 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.
3. Identify the feature of the data that would be missed if part (*b*) was completed without constructing the scatterplot.

***Solution***

1. Excel produces the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 10 | 7.46 | 74.60 | 100 | 55.6516 |
| 8 | 6.77 | 54.16 | 64 | 45.8329 |
| 13 | 12.74 | 165.62 | 169 | 162.3076 |
| 9 | 7.11 | 63.99 | 81 | 50.5521 |
| 11 | 7.81 | 85.91 | 121 | 60.9961 |
| 14 | 8.84 | 123.76 | 196 | 78.1456 |
| 6 | 6.08 | 36.48 | 36 | 36.9664 |
| 4 | 5.39 | 21.56 | 16 | 29.0521 |
| 12 | 8.15 | 97.80 | 144 | 66.4225 |
| 7 | 6.42 | 44.94 | 49 | 41.2164 |
| 5 | 5.73 | 28.65 | 25 | 32.8329 |
| 99 | 82.50 | 797.47 | 1001 | 659.9762 |

1. 





From table *A*-5; 

Therefore *r* = 0.816 indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the 2 variables.

1. The scatterplot indicates that the relationship between the variables is essentially a perfect straight line except for one point, which is likely an error or an oulier.

***Exercise***

The paired values of the Consumer Price Index (CPI) and the cost of a slice of pizza are shown below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***CPI*** | 30.2 | 48.3 | 112.3 | 162.2 | 191.9 | 197.8 |
| ***Cost of Pizza*** | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and find the critical values if *r*, using α = 0.05.
3. Determine whether there is sufficient evidence to support the claim of a linear correlation between the CPI and the cost of a slice of pizza?

***Solution***

1. Excel produces the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 30.2 | 0.15 | 4.53 | 912.04 | 0.0225 |
| 48.3 | 0.35 | 16.905 | 2332.89 | 0.1225 |
| 112.3 | 1.00 | 112.3 | 12611.29 | 1.00 |
| 162.2 | 1.25 | 202.75 | 26308.84 | 1.5625 |
| 191.9 | 1.75 | 335.825 | 36825.61 | 3.0625 |
| 197.8 | 2.00 | 395.60 | 39124.84 | 4.00 |
| 742.7 | 6.50 | 1067.91 | 118115.5 | 9.77 |

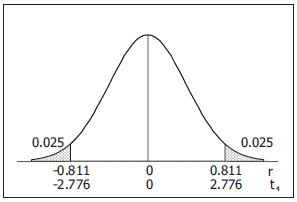
1. 





1. 





Critical value: 









***Conclusion***:

Reject ; there is sufficient evidence to conclude that  (in fact, that ).

Yes; there is sufficient evidence to support the claim of a linear correlation between the CPI and the cost of a slice of pizza.

***Exercise***

Listed below are systolic blood pressure measurements (in mm HG) obtained from the same woman.

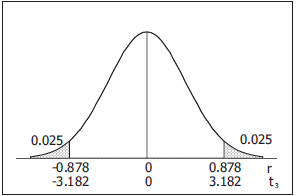
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Right Arm*** | 102 | 101 | 94 | 79 | 79 |
| ***Left Arm*** | 175 | 169 | 182 | 146 | 144 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and find the critical values if *r*, using α = 0.05.
3. Determine whether there is sufficient evidence to support the claim of a linear correlation between the right and left arm systolic blood pressure measurements?

***Solution***

1. Excel produces the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 102 | 175 | 17850 | 10404 | 30625 |
| 101 | 169 | 17069 | 10201 | 28561 |
| 94 | 182 | 17108 | 8836 | 33124 |
| 79 | 146 | 11534 | 6241 | 21316 |
| 79 | 144 | 11376 | 6241 | 20736 |
| 455 | 816 | 74937 | 41923 | 134362 |

1. 





1. 





Critical value: 





***Conclusion***:

Do not reject ; there is no sufficient evidence to conclude that .

No; there is not sufficient evidence to support the claim of a linear correlation between the right and left arm systolic blood pressure measurements.

***Exercise***

Listed below are costs (in dollars) of air fares for different airlines from NY to San Francisco. The costs are based on tickets purchased 30 days in advance and one day in advance.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **30 *Days*** | 244 | 260 | 264 | 264 | 278 | 318 | 280 |
| ***One Day*** | 456 | 614 | 567 | 943 | 628 | 1088 | 536 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and find the critical values if *r*, using α = 0.05.
3. Determine whether there is sufficient evidence to support the claim of a linear correlation between costs of tickets purchased 30 days in advance and those purchased one day in advance?

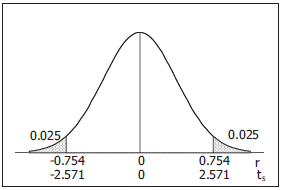
***Solution***

1. Excel produces the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 244 | 456 | 111264 | 59536 | 207936 |
| 260 | 614 | 159640 | 67600 | 376996 |
| 264 | 567 | 149688 | 69696 | 889249 |
| 264 | 943 | 248952 | 69696 | 889249 |
| 278 | 628 | 174584 | 77284 | 394384 |
| 318 | 1088 | 345984 | 101124 | 1183744 |
| 280 | 536 | 150080 | 78400 | 287296 |
| 1908 | 4832 | 1340192 | 523336 | 3661094 |

1. 



1. 





Critical value: 







***Conclusion***:

Do not reject ; there is no sufficient evidence to conclude that .

No; there is not sufficient evidence to support the claim of a linear correlation between the costs of tickets purchased 30 days in advance and those purchased one day in advance.

***Exercise***

Listed below are repair costs (in dollars) for cars crashed at 6 mi/h in full-front crash tests and the same cars crashed at 6 mi/f in full-rear crash tests.

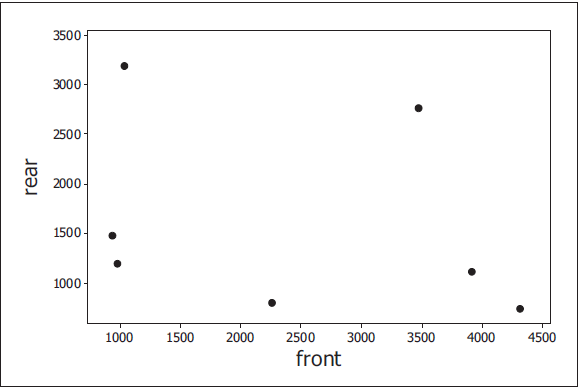
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Front** | 936 | 978 | 2252 | 1032 | 3911 | 4312 | 3469 |
| ***Rear*** | 1480 | 1202 | 802 | 3191 | 1122 | 739 | 2767 |

1. Construct a scatterplot
2. Find the value of linear correlation coefficient *r* and find the critical values if *r*, using α = 0.05.
3. Determine whether there is sufficient evidence to support the claim of a linear correlation between costs from full-front crashes and full-rear crashes?

***Solution***

1. Excel produces the following

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 936 | 1480 | 1385280 | 876096 | 2190400 |
| 978 | 1202 | 1175556 | 956484 | 1444804 |
| 2252 | 802 | 1806104 | 5071504 | 643204 |
| 1032 | 3191 | 3293112 | 1065024 | 10182481 |
| 3911 | 1122 | 4388142 | 15295921 | 1258884 |
| 4312 | 739 | 3186568 | 18593344 | 546121 |
| 3469 | 2767 | 9598723 | 12033961 | 7656289 |
| 16890 | 11303 | 24833485 | 53892334 | 23922183 |



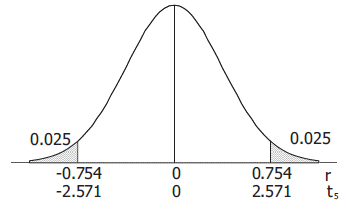
1. 





1. 





Critical value: 







***Conclusion***:

Do not reject ; there is no sufficient evidence to conclude that .

No; there is not sufficient evidence to support the claim of a linear correlation between the costs from full-front crashes and full-rear crashes.

***Solution*** ***Section* 4.5 − Regression**

***Exercise***

A physician measured the weights and cholesterol levels of a random sample of men. The regression equation is , where *x* represents weight (in pounds). What does the symbol  represent? What does the predictor variable represent? What does the response variable represent?

***Solution***

The symbol  represents the predicted cholesterol level. The predictor variable x represents weight. The response variable represents cholesterol level.

***Exercise***

In what sense is the regression line the straight line that “best” fits the points in a scatterplot?

***Solution***

The regression line is the best fit for the points of a scatterplot in the sense that it minimizes the sum of the squared differences between the observed *y* values and the *y* values predicted by the regression line.

***Exercise***

In a study, the total weight (in pounds) of garbage discarded in one week and the household size were recorded for 62 households. The linear correlation coefficient is  and the regression equation, where x represents the total weight of discarded garbage. The mean of the 62 garbage weights is 27.4 lb. and the 62 households have a mean size of 3.71 people. What is the best predicted number of people in a household that discards 50 lb. of garbage?

***Solution***

For *n* = 62, the critical value = .

Since *r* = 0.759 > 0.254, use the regression line for prediction.

\



***Exercise***

A sample of 8 mother/daughter pairs of subjects was obtained, and their heights (in inches) were measured. The linear correlation coefficient is 0.693 and the regression equation , where *x* represents the height of the mother. The mean height of the mothers is 63.1 in. and the mean height of the daughters is 63.3 in. Find the best predicted height of a daughter given that the mother has a height of 60 in.

***Solution***

For *n* = 8, the critical value = ±0.707.

Since *r* = 0.693 < 0.707, use the regression line for prediction. 



***Exercise***

A sample of 40 women is obtained, and their heights (in inches) and pulse rates (in beats per minute) are measured. The linear correlation coefficient is 0.202 and the equation of the regression line is , where *x* represents height. The mean of the 40 heights is 63.2 in. and the mean of the 40 pulse rates is 76.3 beats per minute. Find the best predicted pulse rate of a woman who is 70 in. tall.

***Solution***

For *n* = 40, the critical value = ±0.312.

Since *r* = 0.202 < 0.312, use the regression line for prediction. 



***Exercise***

Heights (in inches) and weights (in pounds) are obtained from a random sample of 9 supermodels. The linear correlation coefficient is 0.360 and the equation of the regression line is , where *x* represents height. The mean of the 9 heights is 69.3 in. and the mean of the 9 weights is 117 lb. Find the best predicted weight of a supermodel with a height of 72 in.?

***Solution***

For *n* = 9, the critical value = ±0.666.

Since *r* = 0.360 < 0.666, use the regression line for prediction. 



***Exercise***

Find the equation of the regression line for the given data below

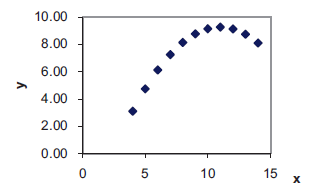
|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 10 | 8 | 13 | 9 | 11 | 14 | 6 | 4 | 12 | 7 | 5 |
| ***y*** | 9.14 | 8.14 | 8.74 | 8.77 | 9.26 | 8.10 | 6.13 | 3.10 | 9.13 | 7.26 | 4.74 |

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 10 | 9.14 | 91.40 | 100 | 83.5396 |
| 8 | 8.14 | 65.12 | 64 | 66.2596 |
| 13 | 8.74 | 113.62 | 169 | 76.3876 |
| 9 | 8.77 | 78.93 | 81 | 76.9129 |
| 11 | 9.26 | 101.86 | 121 | 85.7476 |
| 14 | 8.10 | 113.40 | 196 | 65.61 |
| 6 | 6.13 | 36.78 | 36 | 37.5769 |
| 4 | 3.10 | 12.40 | 16 | 9.61 |
| 12 | 9.13 | 109.56 | 144 | 83.3569 |
| 7 | 7.26 | 50.82 | 49 | 52.7076 |
| 5 | 4.74 | 23.70 | 25 | 22.4676 |
| 99 | 82.51 | 797.59 | 1001 | 660.1763 |



















The scatterplot indicates that the relationship between the variables is quadratic, not linear.

***Exercise***

Find the equation of the regression line for the given data below

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | 10 | 8 | 13 | 9 | 11 | 14 | 6 | 4 | 12 | 7 | 5 |
| ***y*** | 7.46 | 6.77 | 12.74 | 7.11 | 7.81 | 8.84 | 6.08 | 5.39 | 8.15 | 6.42 | 5.73 |

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

***Solution***

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***x*** | ***y*** | ***xy*** |  |  | | 10 | 7.46 | 74.60 | 100 | 55.6516 | | 8 | 6.77 | 54.16 | 64 | 45.8329 | | 13 | 112.74 | 165.62 | 169 | 162.3076 | | 9 | 7.11 | 63.99 | 81 | 50.5521 | | 11 | 7.81 | 85.91 | 121 | 60.9961 | | 14 | 8.84 | 123.76 | 196 | 78.1456 | | 6 | 6.08 | 36.48 | 36 | 36.9664 | | 4 | 5.39 | 21.56 | 16 | 29.0521 | | 12 | 8.15 | 97.80 | 144 | 66.4225 | | 7 | 6.42 | 44.94 | 49 | 41.2164 | | 5 | 5.73 | 28.65 | 25 | 32.8329 | | 99 | 82.50 | 797.47 | 1001 | 659.9762 | |



















The scatterplot indicates that the relationship between the variables is essentially a perfect straight line except for one point, which is likely an error or an outlier.

***Exercise***

Find the equation of the regression line for the given data below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***CPI*** | 30.2 | 48.3 | 112.3 | 162.2 | 191.9 | 197.8 |
| ***Cost of Pizza*** | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |

Let the first variable be the predictor (*x*) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

***Solution***

Excel produces the following

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 30.2 | 0.15 | 4.53 | 912.04 | 0.0225 |
| 48.3 | 0.35 | 16.905 | 2332.89 | 0.1225 |
| 112.3 | 1.00 | 112.3 | 12611.29 | 1.00 |
| 162.2 | 1.25 | 202.75 | 26308.84 | 1.5625 |
| 191.9 | 1.75 | 335.825 | 36825.61 | 3.0625 |
| 197.8 | 2.00 | 395.60 | 39124.84 | 4.00 |
| 742.7 | 6.50 | 1067.91 | 118115.5 | 9.77 |























***Exercise***

Find the equation of the regression line for the given data below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***CPI*** | 30.2 | 48.3 | 112.3 | 162.2 | 191.9 | 197.8 |
| ***Subway fare*** | 0.15 | 0.35 | 1.00 | 1.35 | 1.5 | 2.00 |

Let the first variable be the predictor (*x*) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

***Solution***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |  |
| 30.2 | 0.15 | 4.53 | 912.04 | 0.0225 |
| 48.3 | 0.35 | 16.905 | 2332.89 | 0.1225 |
| 112.3 | 1.00 | 112.3 | 12611.29 | 1.00 |
| 162.2 | 1.35 | 218.97 | 26308.84 | 1.8225 |
| 191.9 | 1.50 | 287.85 | 36825.61 | 2.25 |
| 197.8 | 2.00 | 395.60 | 39124.84 | 4.00 |
| 742.7 | 6.35 | 1036.155 | 118115.51 | 9.2175 |























***Exercise***

Listed below are systolic blood pressure measurements (in mm HG) obtained from the same woman.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Right Arm*** | 102 | 101 | 94 | 79 | 79 |
| ***Left Arm*** | 175 | 169 | 182 | 146 | 144 |

Find the best predicted systolic blood pressure in the left arm given that the systolic blood pressure in the right arm is 100 mm Hg.

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 102 | 175 | 17850 | 10404 | 30625 |
| 101 | 169 | 17069 | 10201 | 28561 |
| 94 | 182 | 17108 | 8836 | 33124 |
| 79 | 146 | 11534 | 6241 | 21316 |
| 79 | 144 | 11376 | 6241 | 20736 |
| 455 | 816 | 74937 | 41923 | 134362 |



















 ***No significant correlation***

***Exercise***

Find the best predicted height of runner-up Goldwater, given that the height of the winning presidential candidate is 75 in. Is the predicted height of Goldwater close to his actual height of 72 in.?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Winner*** | 69.5 | 73 | 73 | 74 | 74.5 | 74.5 | 71 | 71 |
| ***Runner-Up*** | 72 | 69.5 | 70 | 68 | 74 | 74 | 73 | 76 |

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 69.5 | 72 | 5004 | 4830.25 | 5184 |
| 73 | 69.5 | 5073.5 | 5329 | 4830.25 |
| 73 | 70 | 5110 | 5329 | 4900 |
| 74 | 68 | 5032 | 5476 | 4624 |
| 74.5 | 74 | 5513 | 5550.25 | 5476 |
| 74.5 | 74 | 5513 | 5550.25 | 5476 |
| 71 | 76 | 5183 | 5041 | 5329 |
| 71 | 76 | 5396 | 5041 | 5776 |
| 580.5 | 576.5 | 41824.5 | 42146.75 | 41595.25 |



















 ***No significant correlation***

***Exercise***

Find the best predicted amount of revenue (in millions of dollars), given that the amount has a size 87 thousand *ft*2. How does the result compare to the actual revenue of $65.1 million?

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Size*** | 160 | 227 | 140 | 144 | 161 | 147 | 141 |
| ***Revenue*** | 189 | 157 | 140 | 127 | 123 | 106 | 101 |

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 160 | 189 | 30240 | 25600 | 35721 |
| 227 | 157 | 35639 | 51529 | 24649 |
| 140 | 140 | 19600 | 19600 | 19600 |
| 144 | 127 | 18288 | 20736 | 16129 |
| 161 | 123 | 19803 | 25921 | 15129 |
| 147 | 106 | 15582 | 21609 | 11236 |
| 141 | 101 | 14241 | 19881 | 10201 |
| 1120 | 943 | 153393 | 184876 | 132665 |



















 ***No significant correlation***

The predicted value is far from the actual value. Since there is no significant correlation, the mean is used for all predictions – but the *x* = 87 thousand ft2 is well outside the range of x values used to construct the predictive regression equation.

***Exercise***

Find the best predicted new mileage rating of a jeep given that old rating is 19 mi/gal. Is the predicted value close to the actual value of 17 mi/gal?

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Old*** | 16 | 27 | 17 | 33 | 28 | 24 | 18 | 22 | 20 | 29 | 21 |
| ***New*** | 15 | 24 | 15 | 29 | 25 | 22 | 16 | 20 | 18 | 26 | 19 |

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 16 | 15 | 240 | 256 | 225 |
| 27 | 24 | 648 | 729 | 576 |
| 17 | 16 | 272 | 289 | 256 |
| 33 | 29 | 957 | 1089 | 841 |
| 28 | 25 | 700 | 784 | 625 |
| 24 | 22 | 528 | 576 | 484 |
| 18 | 16 | 288 | 324 | 256 |
| 22 | 20 | 440 | 484 | 400 |
| 20 | 18 | 360 | 400 | 324 |
| 29 | 26 | 754 | 841 | 676 |
| 21 | 19 | 399 | 441 | 361 |
| 255 | 230 | 5586 | 6213 | 5024 |





















Yes; the predicted value is close to the actual value of 17 mpg.

***Exercise***

Find the best predicted temperature for a recent year in which the concentration (in parts per million) of CO2 is 370.9. Is the predicted temperature close to the actual temperature of 14.5° C??

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO2** | 314 | 317 | 320 | 326 | 331 | 339 | 346 | 354 | 361 | 369 |
| ***Temperature*** | 13.9 | 14.0 | 13.9 | 14.1 | 14.0 | 14.3 | 14.1 | 14.5 | 14.5 | 14.4 |

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 314 | 13.9 | 4364.6 | 985696 | 193.21 |
| 317 | 14 | 4438 | 100489 | 196 |
| 320 | 13.9 | 4448 | 102400 | 193.21 |
| 326 | 14.1 | 4596.6 | 106276 | 198.81 |
| 331 | 14 | 4634 | 109561 | 196 |
| 339 | 14.3 | 4847.7 | 114921 | 204.49 |
| 346 | 14.1 | 4878.6 | 119716 | 198.81 |
| 354 | 14.5 | 5133 | 125316 | 210.25 |
| 361 | 14.5 | 5234.5 | 130321 | 210.25 |
| 369 | 14.4 | 5313.6 | 136161 | 207.36 |
| 3377 | 141.7 | 47888.6 | 1143757 | 2008.39 |





















Yes; the predicted temperature is equal to the actual temperature of 14.5 °C..

***Exercise***

Find the best predicted IQ score of someone with a brain size of 1275 

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Brain Size*** | 965 | 1029 | 1030 | 1285 | 1049 | 1077 | 1037 | 1068 | 1176 | 1105 |
| ***IQ*** | 90 | 85 | 86 | 102 | 103 | 97 | 124 | 125 | 102 | 114 |

***Solution***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***y*** | ***xy*** |  |  |
| 965 | 90 | 86850 | 931225 | 8100 |
| 1029 | 85 | 87465 | 1058841 | 7225 |
| 1030 | 86 | 88580 | 1060900 | 7396 |
| 1285 | 102 | 131070 | 1651225 | 10404 |
| 1049 | 103 | 108047 | 1100401 | 10609 |
| 1077 | 97 | 104469 | 1159929 | 9409 |
| 1037 | 124 | 128588 | 1075369 | 15376 |
| 1068 | 125 | 133500 | 1140624 | 15625 |
| 1176 | 102 | 119952 | 1382976 | 10404 |
| 1105 | 114 | 125970 | 1221025 | 12996 |
| 10821 | 1028 | 1114491 | 11782515 | 107544 |





















*No significant correlation*

***Exercise***

Listed below are the word counts for men and women.

***Male***

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 27531 | 15684 | 5638 | 27997 | 25433 | 8077 | 21319 | 17572 | 26429 | 21966 | 11680 | 10818 |
| 12650 | 21683 | 19153 | 1411 | 20242 | 10117 | 20206 | 16874 | 16135 | 20734 | 7771 | 6792 |
| 26194 | 10671 | 13462 | 12474 | 13560 | 18876 | 13825 | 9274 | 20547 | 17190 | 10578 | 14821 |
| 15477 | 10483 | 19377 | 11767 | 13793 | 5908 | 18821 | 14069 | 16072 | 16414 | 19017 | 37649 |
| 17427 | 46978 | 25835 | 10302 | 15686 | 10072 | 6885 | 20848 |  |  |  |  |

***Female***

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 20737 | 24625 | 5198 | 18712 | 12002 | 15702 | 11661 | 19624 | 13397 | 18776 | 15863 | 12549 |
| 17014 | 23511 | 6017 | 18338 | 23020 | 18602 | 16518 | 13770 | 29940 | 8419 | 17791 | 5596 |
| 11467 | 18372 | 13657 | 21420 | 21261 | 12964 | 33789 | 8709 | 10508 | 11909 | 29730 | 20981 |
| 16937 | 19049 | 20224 | 15872 | 18717 | 12685 | 17646 | 16255 | 28838 | 38154 | 25510 | 34869 |
| 24480 | 31553 | 18667 | 7059 | 25168 | 16143 | 14730 | 28117 |  |  |  |  |

Find the best predicted word count of a woman given that her male partner speaks 6,000 words in a day.

***Solution***

Using Excel spread sheet - ***Regression***

|  |  |
| --- | --- |
|  | *Coefficients* |
| Intercept | 13438.884 |
| X Variable 1 | 0.302 |







***Exercise***

According the least-squares property, the regression line minimizes the sum of the squares of the residuals. Listed below are the paired data consisting of the first 6 pulse and the first systolic blood pressures of males.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Pulse*** (***x***) | 68 | 64 | 88 | 72 | 64 | 110 |
| ***Systolic*** (***y***) | 125 | 107 | 126 | 110 | 72 | 107 |

1. Find the equation of the regression line.
2. Identify the residuals, and find the sum of squares of the residuals.
3. Show that the equation  results in a larger sum of squares of residuals.

***Solution***

***x*** = pulse rate

***y*** = systolic blood pressures

1. Using Excel spread sheet ***– Data Analysis*** - ***Regression***

|  |  |
| --- | --- |
|  | *Coefficients* |
| Intercept | 71.678 |
| X Variable 1 | 0.5956 |

The equation of the regression line: 

1. = residuals for the regression line

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 68 | 125 | 112.208 | 12.792 | 163.635 |
| 64 | 107 | 109.824 | -2.824 | 7.975 |
| 88 | 126 | 124.128 | 1.872 | 3.504 |
| 72 | 110 | 114.592 | -4.592 | 21.086 |
| 64 | 110 | 109.824 | 0.176 | 0.031 |
| 72 | 107 | 114.592 | -7.592 | 57.638 |
| 428 | 685 | 684.997 | 0.003 | 253.866 |

The table indicates that the sum of the squares of the residuals is 253.866

1. = residuals for the regression line where 

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| 68 | 125 | 104.000 | 21.000 | 441.000 |
| 64 | 107 | 102.000 | 5.000 | 25.000 |
| 88 | 126 | 114.000 | 12.000 | 144.000 |
| 72 | 110 | 106.000 | 4.000 | 16.000 |
| 64 | 110 | 102.000 | 8.000 | 64.000 |
| 72 | 107 | 106.000 | 1.000 | 1.000 |
| 428 | 685 | 634.0 | 51.0 | 691.0 |

The table indicates that the sum of the squares of the residuals is 691, which is greater the the 253.866 of the least squares regression equation.

***Solution*** ***Section* 4.6 − Variation and Prediction Intervals**

***Exercise***

A height of 70 in. is used to find the predicted weight is 180 lb. In your own words, describe a prediction interval in this situation.

***Solution***

A prediction interval is an interval estimate for a predicted value. In this situation it will be a range of weights centered at the prediction’s point estimate of 180 lbs.

***Exercise***

A height of 70 in. is used to find the predicted weight is 180 lb. What is the major advantage of using a prediction interval instead of the predicted weight of 180 lb.? Why is the terminology of prediction interval used instead of confidence interval?

***Solution***

By providing a range of values instead of a single point, a prediction interval gives an indication of the accuracy of the prediction. A confidence interval is an internal estimate of a parameter – i.e., of a conceptually fixed, although unknown, value. A prediction interval is an interval estimate of a random variable – i.e., of a value from a distribution of values.

***Exercise***

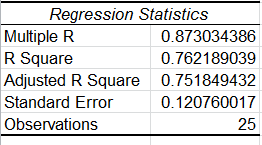
Use the value of the linear correlation  to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 16 | 13 | 16 | 9 | 14 | 13 | 12 | 14 | 14 | 13 | 13 | 16 | 13 | 13 | 18 |
| 9 | 19 | 2 | 13 | 14 | 14 | 15 | 16 | 6 | 8 |  |  |  |  |  |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1.1 | 0.8 | 1 | 0.9 | 0.8 | 0.8 | 0.8 | 0.8 | 0.9 | 0.8 | 0.8 | 1.2 | 0.8 | 0.8 | 1.3 |
| 0.7 | 1.4 | 0.2 | 0.8 | 1 | 0.8 | 0.8 | 1.2 | 0.6 | 0.7 |  |  |  |  |  |

***Solution***

The coefficient of determination is 

The portion of the total variation in *y* explained by the regression is 

***Exercise***

Use the value of the linear correlation  to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 41 | 20 | 116 | 70 | 75 | 52 | 120 | 65 | 6.5 | 60 | 125 | 20 | 5 | 150 |
| 4.5 | 7 | 100 | 30 | 225 | 70 | 80 | 40 | 70 | 50 | 74 | 200 | 113 | 68 |
| 72 | 160 | 68 | 29 | 132 | 40 |  |  |  |  |  |  |  |  |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 117 | 5 | 103 | 66 | 121 | 116 | 101 | 100 | 55 | 104 | 213 | 34 | 12 | 290 |
| 47 | 10 | 111 | 100 | 322 | 19 | 117 | 48 | 228 | 47 | 17 | 373 | 380 | 118 |
| 114 | 120 | 101 | 120 | 234 | 209 |  |  |  |  |  |  |  |  |

***Solution***

The coefficient of determination is 

The portion of the total variation in *y* explained by the regression is 

***Exercise***

Use the value of the linear correlation  to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables



***Solution***

The coefficient of determination is 

The portion of the total variation in *y* explained by the regression is 

***Exercise***

Use the value of the linear correlation  to find the coefficient of determination and the percentage of the total variation that can be explained by the linear relationship between the 2 variables



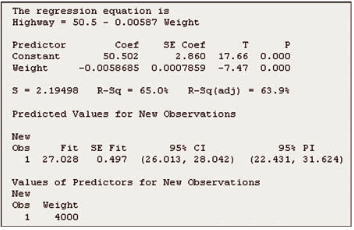
***Solution***

The coefficient of determination is 

The portion of the total variation in *y* explained by the regression is 

***Exercise***

Refer to the display obtained by using the paired data consisting of weights (in *lb*.) of 32 cars and their highway fuel consumption amounts (in *mi/gal*). A car weight of 4000 *lb*. to be used for predicting the highway fuel consumption amount



1. What percentage of the total variation in highway fuel consumption can be explained by the linear correlation between weight and highway fuel consumption?
2. If a car weighs 4000 *lb*., what is the single value that is the best predicted amount of highway fuel consumption? (Assume that there is a linear correlation between weight and highway fuel consumption.)

***Solution***

1. 
2. The given point estimate is 

***Exercise***

The paired values of the Consumer Price Index (CPI) and the cost of a slice of pizza are shown below

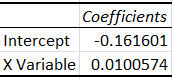
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***CPI*** | 30.2 | 48.3 | 112.3 | 162.2 | 191.9 | 197.8 |
| ***Cost of Pizza*** | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |

1. Find the explained variation
2. Find the unexplained variation
3. Find the total variation
4. Find the coefficient of determination
5. Find the standard error of estimate 
6. Find the predicted cost of a slice of pizza for the year 2001, when the CPI was 187.1.
7. Find a 95% prediction interval estimate of the cost of a slice of pizza when the CPI was 187.1

*In each case, there is sufficient evidence to support a claim of a linear correlation so that it is reasonable to use the regression equation when making predictions.*

***Solution***

The predicted values:





|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** |  |  |  |  |  |  |  |  |
| 30.2 | 0.15 | 0.142 | 1.083 | −0.940 | 0.886 | 0.008 | 0.000 | −0.930 | 0.871 |
| 48.3 | 0.35 | 0.324 | 1.083 | −0.760 | 0.576 | 0.023 | 0.001 | −0.730 | 0.538 |
| 112.3 | 1.00 | 0.968 | 1.083 | −0.120 | 0.013 | 0.032 | 0.001 | −0.080 | 0.007 |
| 162.2 | 1.25 | 1.470 | 1.083 | 0.386 | 0.149 | −0.220 | 0.048 | 0.167 | 0.028 |
| 191.9 | 1.75 | 1.768 | 1.083 | 0.685 | 0.469 | −0.018 | 0.000 | 0.667 | 0.444 |
| 197.8 | 2.00 | 1.828 | 1.083 | 0.744 | 0.554 | 0.172 | 0.030 | 0.917 | 0.840 |
| 742.7 | 6.50 | 6.50 | 6.50 | 0.0 | 2.648 | 0.0 | 0.08 | 0.0 | 2.728 |

1. The explained variation is 
2. The unexplained variation is 
3. The total variation is 
4. ******
5. ******



|  |  |  |
| --- | --- | --- |
| ***x*** | ***y*** |  |
| 30.2 | 0.15 | 912.04 |
| 48.3 | 0.35 | 2332.89 |
| 112.3 | 1.00 | 12611.29 |
| 162.2 | 1.25 | 26308.84 |
| 191.9 | 1.75 | 36825.61 |
| 197.8 | 2.00 | 39124.84 |
| 742.7 | 6.50 | 118115.5 |

1. 



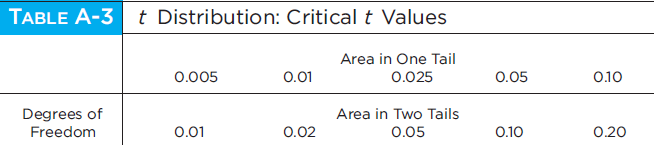


1. Preliminary calculations for *n* = 6



 (2-tails)



















***Exercise***

The paired values of the Consumer Price Index (CPI) and the subway fare are shown below

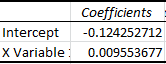
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***CPI*** | 30.2 | 48.3 | 112.3 | 162.2 | 191.9 | 197.8 |
| ***Subway fare*** | 0.15 | 0.35 | 1.00 | 1.35 | 1.5 | 2.00 |

1. Find the explained variation
2. Find the unexplained variation
3. Find the total variation
4. Find the coefficient of determination
5. Find the standard error of estimate 
6. Find the predicted cost of subway fare for the year 2001, when the CPI was 187.1.
7. Find a 95% prediction interval estimate of the cost of subway fare when the CPI was 187.1

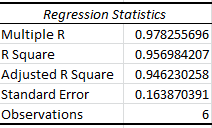
*In each case, there is sufficient evidence to support a claim of a linear correlation so that it is reasonable to use the regression equation when making predictions.*

***Solution***

The predicted values (from Excel):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** |  |  |  |  |  |  |  |  |
| 30.2 | 0.15 | 0.164 | 1.058 | −0.890 | 0.799 | −0.014 | 0.0 | −0.910 | 0.825 |
| 48.3 | 0.35 | 0.337 | 1.058 | −0.720 | 0.520 | 0.013 | 0.0 | −0.710 | 0.502 |
| 112.3 | 1.00 | 0.949 | 1.058 | −0.110 | 0.012 | 0.051 | 0.006 | 0.292 | 0.085 |
| 162.2 | 1.35 | 1.425 | 1.058 | 0.367 | 0.135 | −0.075 | 0.006 | 0.292 | 0.085 |
| 191.9 | 1.50 | 1.709 | 1.058 | 0.651 | 0.423 | −0.209 | 0.044 | 0.442 | 0.195 |
| 197.8 | 2.00 | 1.765 | 1.58 | 0.707 | 0.500 | 0.235 | 0.055 | 0.942 | 0.887 |
| 742.7 | 6.35 | 6.350 | 6.350 | 0.0 | 2.930 | 0.0 | 0.104 | 0.0 | 2.497 |

1. The explained variation is 
2. The unexplained variation is 
3. The total variation is 
4. ******
5. ******

|  |  |  |
| --- | --- | --- |
| ***x*** | ***y*** |  |
| 30.2 | 0.15 | 912.04 |
| 48.3 | 0.35 | 2332.89 |
| 112.3 | 1.00 | 12611.29 |
| 162.2 | 1.35 | 26308.84 |
| 191.9 | 1.50 | 36825.61 |
| 197.8 | 2.00 | 39124.84 |
| 742.7 | 6.35 | 118115.51 |



1. 

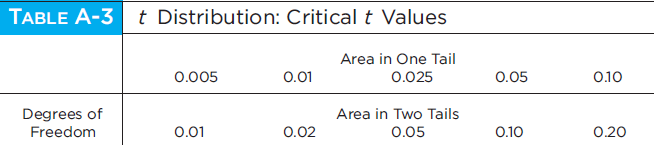


1. Preliminary calculations for *n* = 6



 (2-tails)



















***Exercise***

Find the best predicted temperature for a recent year in which the concentration (in parts per million) of CO2 and temperature (in °C) for different years

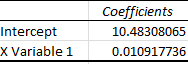
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CO2** | 314 | 317 | 320 | 326 | 331 | 339 | 346 | 354 | 361 | 369 |
| ***Temperature*** | 13.9 | 14.0 | 13.9 | 14.1 | 14.0 | 14.3 | 14.1 | 14.5 | 14.5 | 14.4 |

1. Find the explained variation
2. Find the unexplained variation
3. Find the total variation
4. Find the coefficient of determination
5. Find the standard error of estimate 
6. Find the predicted temperature (in °C) when CO2 concentration is 370.9 parts per million.
7. Find a 99% prediction interval estimate temperature (in °C) when CO2 concentration is 370.9 parts per million

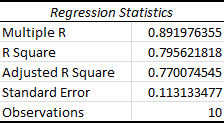
*In each case, there is sufficient evidence to support a claim of a linear correlation so that it is reasonable to use the regression equation when making predictions.*

***Solution***

The predicted values (from Excel):

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***x*** | ***y*** |  |  |  |  |  |  |  |  |
| 314 | 13.9 | 13.911 | 14.17 | −0.259 | 0.067 | −0.011 | 0.0 | −0.27 | 0.073 |
| 317 | 14 | 13.944 | 14.17 | −0.266 | 0.051 | 0.056 | 0.003 | −0.17 | 0.029 |
| 320 | 13.9 | 13.977 | 14.17 | −0.193 | 0.037 | −0.077 | 0.006 | −0.27 | 0.073 |
| 326 | 14.1 | 14.042 | 14.17 | −0.128 | 0.016 | 0.058 | 0.003 | −0.07 | 0.005 |
| 331 | 14 | 14.097 | 14.17 | −0.073 | 0.005 | −0.097 | 0.009 | −0.17 | 0.029 |
| 339 | 14.3 | 14.184 | 14.17 | 0.014 | 0.0 | 0.116 | 0.013 | 0.13 | 0.017 |
| 346 | 14.1 | 14.261 | 14.17 | 0.091 | 0.008 | −0.161 | 0.026 | −0.07 | 0.005 |
| 354 | 14.5 | 14.348 | 14.17 | 0.178 | 0.032 | 0.152 | 0.023 | 0.33 | 0.109 |
| 361 | 14.5 | 14.424 | 14.17 | 0.254 | 0.065 | 0.076 | 0.006 | 0.33 | 0.109 |
| 369 | 14.4 | 14.512 | 14.17 | 0.342 | 0.117 | −0.112 | 0.012 | 0.23 | 0.053 |
| 3377 | 141.7 | 141.7 | 141.70 | 0.0 | 0.399 | 0.0 | 0.102 | 0.0 | 0.501 |

1. The explained variation is 
2. The unexplained variation is 
3. The total variation is 
4. ******

|  |  |  |
| --- | --- | --- |
| ***x*** | ***y*** |  |
| 314 | 13.9 | 985696 |
| 317 | 14 | 100489 |
| 320 | 13.9 | 102400 |
| 326 | 14.1 | 106276 |
| 331 | 14 | 109561 |
| 339 | 14.3 | 114921 |
| 346 | 14.1 | 119716 |
| 354 | 14.5 | 125316 |
| 361 | 14.5 | 130321 |
| 369 | 14.4 | 136161 |
| 3377 | 141.7 | 1143757 |

1. ******



1. 

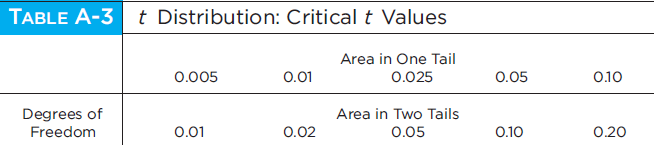


1. Preliminary calculations for *n* = 8



 (2-tails)



















***Exercise***

Find a prediction interval data listed below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cost of Pizza | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |
| Subway Fare | 0.15 | 0.35 | 1.00 | 1.35 | 1.50 | 2.00 |

Using: 

***Solution***

The predicted values (from Excel):



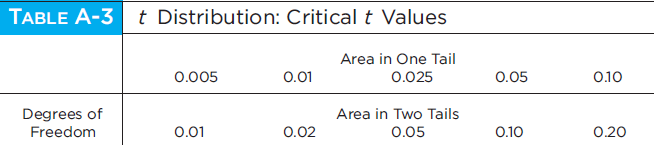














|  |  |
| --- | --- |
| *x* |  |
| 0.15 | 0.0225 |
| 0.35 | 0.1225 |
| 1 | 1 |
| 1.25 | 1.5625 |
| 1.75 | 3.0625 |
| 2 | 4 |
| 6.5 | 9.77 |













***Exercise***

Find a prediction interval data listed below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cost of Pizza | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |
| Subway Fare | 0.15 | 0.35 | 1.00 | 1.35 | 1.50 | 2.00 |

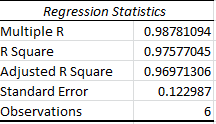
Using: 

***Solution***

The predicted values (from Excel):



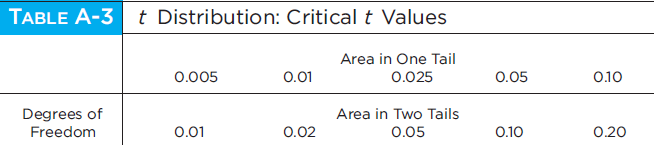














|  |  |
| --- | --- |
| *x* |  |
| 0.15 | 0.0225 |
| 0.35 | 0.1225 |
| 1 | 1 |
| 1.25 | 1.5625 |
| 1.75 | 3.0625 |
| 2 | 4 |
| 6.5 | 9.77 |













***Exercise***

Find a prediction interval data listed below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cost of Pizza | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |
| Subway Fare | 0.15 | 0.35 | 1.00 | 1.35 | 1.50 | 2.00 |

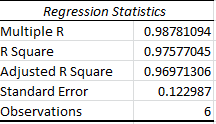
Using: 

***Solution***

The predicted values (from Excel):



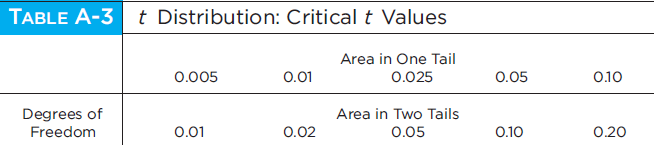


























***Exercise***

Find a prediction interval data listed below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cost of Pizza | 0.15 | 0.35 | 1.00 | 1.25 | 1.75 | 2.00 |
| Subway Fare | 0.15 | 0.35 | 1.00 | 1.35 | 1.50 | 2.00 |

Using: 

***Solution***

The predicted values (from Excel):

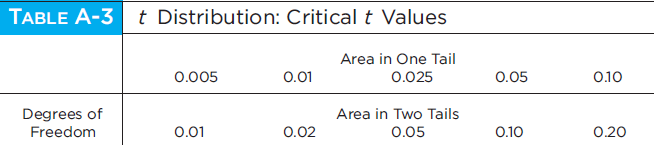
 

























***Solution Section* 4.7 − Goodness-of-Fit**

***Exercise***

A poll typically involves the selection of random digits to be used for telephone numbers. The New York Times states that “within each (telephone) exchange, random digits were added to form a complete telephone number, thus permitting access to listed and unlisted numbers. “When such digits are randomly generated, what is the distribution of those digits? Given such randomly generated digits, what is a test for “goodness-off-fit”?

***Solution***

When digits are randomly generated they should form a uniform distribution – i.e., a distribution in which each of the digits is equally likely. The test for goodness-to-fit is a test of the hypothesis that the sample data fit the uniform distribution.

***Exercise***

When generating random digits, we can test the generated digits for goodness-of-fit with the distribution in which all of the digits are equally likely. What does an exceptionally large value of the  test statistic suggest about the goodness-of-fit? What does an exceptionally small value of the  test statistic (such as 0.002) suggest about the goodness-of-fit?

***Solution***

The calculated  is a measure of the discrepancy between the hypothesis distribution and the sample data. An exceptionally large value of the  test statistic suggests a large discrepancy between the hypothesized distribution and the sample data – that there is not goodness-of-fit, and that the observed and expected frequencies are quite different. An exceptionally small of the  test statistic suggests an extremely good fir – that the observed and expected values are almost identical.

***Exercise***

You purchased a slot machine, and tested it by playing it 1197 times. There are 10 different categories of outcome, including no win, win jackpot, win with three bells, and so on. When testing the claim the observed outcomes agree with the expected frequencies, the author obtained a test statistic of . Use a 0.05 significance level to test the claim that the actual outcomes agree with the expected frequencies. Does the slot machine appear to be functioning as expected?

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

***Solution***

Original claim: the actual outcomes agree with the expected frequencies

 The actual outcomes agree with the expected frequencies

 At least one outcome is not as expected





*Calculations*:



*P*-value = 

***Conclusion***

Do not reject ; there is not sufficient evidence to reject the claim that the actual outcomes agree with the expected frequencies. There is no reason to say the slot machine is not functioning as expected.

***Exercise***

Do “*A*” students tend to sit in a particular part of the classroom? The author recorded the locations of the students who received grades *A*, with these results: 17 sat in the front, 9 sat in the middle, and 5 sat in the back of the classroom. When testing the assumption that the “*A*” students are distributed evenly throughout the room, the author obtained the test statistic of . If using a 0.05 significance level, is there sufficient evidence to support the claim that the “*A*” students are not evenly distributed throughout the classroom? If so, does that mean you can increase your likelihood of getting an *A* by sitting in the front of the room?

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

***Solution***

Original claim: “*A*” student are not evenly distributed throughout the classroom

 “*A*” students are evenly distributed throughout the classroom

 “*A*” students are not evenly distributed throughout the classroom









*Calculations*:



*P*-value = 

***Conclusion***

Reject ; there is sufficient evidence to support the claim “*A*” students are not evenly distributed throughout the classroom.

***Exercise***

Randomly selected nonfat occupational injuries and illnesses are categorized according to the day of the week that they first occurred, and the results are listed below. Use a 0.05 significance level to test the claim that such injuries and illness occur with equal frequency on the different days of the week

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Day*** | *Mon* | *Tues* | *Wed* | *Thurs* | *Fri* |
| ***Number*** | 23 | 23 | 21 | 21 | 19 |

***Solution***

Original Claim: The injuries and illnesses occur with equal frequencies on the different days.



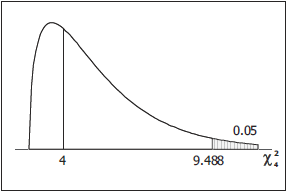








*Calculations*:

|  |  |  |  |
| --- | --- | --- | --- |
| *Day* | *O* | *E* |  |
| *M* | 23 | 21.4 | 0.1196 |
| *T* | 23 | 21.4 | 0.1196 |
| *W* | 21 | 21.4 | 0.0075 |
| *Th* | 21 | 21.4 | 0.075 |
| *F* | 19 | 21.4 | 0.2693 |
|  | 107 | 107 | 0.5234 |



*P*-value = 

***Conclusion***

Do not reject ; there is not sufficient evidence to reject the claim that  for each day. There is no sufficient evidence to reject the claim that the injuries and illnesses occur with equal frequencies on the different days of the week.

***Exercise***

Records of randomly selected births were obtained and categorized according to the day of the week that they occurred. Because babies are unfamiliar with our schedule of weekdays, a reasonable claim is that occur on the different days with equal frequency. Use a 0.01 significance level to test that claim. Can you provide an explanation for the result?

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Day*** | *Sun* | *Mon* | *Tues* | *Wed* | *Thurs* | *Fri* | *Sat* |
| ***Number of births*** | 77 | 110 | 124 | 122 | 120 | 123 | 97 |

***Solution***

Original Claim: births occur on the different days with equal frequency.











*Calculations*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Day* | *O* | *E* |  |  |
| *S* | 77 | 110.43 | 10.119 |
| *M* | 110 | 110.43 | 0.0017 |
| *T* | 124 | 110.43 | 1.6679 |
| *W* | 122 | 110.43 | 1.2125 |
| *Th* | 120 | 110.43 | 0.8296 |
| *F* | 123 | 110.43 | 1.4312 |
| *S* | 97 | 110.43 | 1.6330 |
|  | 773 | 773 | 16.8952 |



*P*-value = 

***Conclusion***

Reject ; there is sufficient evidence to support the claim that  for each day. There is sufficient evidence to reject the claim that births occur on the different days with equal frequency. Births that do not occur naturally (induced, Caesarean sections) are typically not scheduled for Saturday and Sunday, accounting for the smaller than expected numbers of births on those days.

***Exercise***

The table below lists the frequency of wins for different post positions in the Kentucky Derby horse race. A post position of 1 is closest to the inside rail, so that horse has the shortest distance to run. (Because the number of horses varies from year to year, only the first ten post positions are included.) Use a 0.05 significance level to test the claim that the likelihood of winning is the same for the different post positions. Based on the result, should bettor consider the post position of a horse racing in the Kentucky Derby?

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Post Position*** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| ***Wins*** | 19 | 14 | 11 | 14 | 14 | 7 | 8 | 11 | 5 | 11 |

***Solution***

Original Claim: The likelihood of winning is the same for all post positions.















*Calculations*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Position* | *O* | *E* |  |  |
| 1 | 19 | 11.4 | 5.0667 |
| 2 | 14 | 11.4 | 0.5930 |
| 3 | 11 | 11.4 | 0.0140 |
| 4 | 14 | 11.4 | 0.5930 |
| 5 | 14 | 11.4 | 0.5930 |
| 6 | 7 | 11.4 | 1.6982 |
| 7 | 8 | 11.4 | 1.0140 |
| 8 | 11 | 11.4 | 0.0140 |
| 9 | 5 | 11.4 | 3.5930 |
| 10 | 11 | 11.4 | 0.0140 |
|  | 114 | 114.0 | 13.193 |



*P*-value = 

***Conclusion***

Do not reject ; there is not sufficient evidence to reject the claim that  for each position. There is no sufficient evidence to reject the claim that the likelihood of winning is the same for all post positions. Based on these results, post position is not a significant consideration when betting on the Kentucky Derby.

***Exercise***

The table below lists the cases of violent crimes are randomly selected and categorized by month.

Use a 0.01 significance level to test the claim that the rate of violent crime is the same for each month. Can you explain the result?

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Month*** | Jan. | Feb. | Mar. | Apr. | May | June | July | Aug. | Sept. | Oct. | Nov. | Dec. |
| ***Number*** | 786 | 704 | 835 | 826 | 900 | 868 | 920 | 901 | 856 | 862 | 783 | 797 |

***Solution***

Original Claim: The occurrence of violent crime is the same for each month.



















*Calculations*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Month* | *O* | *E* |  |  |
| Jan | 786 | 836.5 | 3.0487 |
| Feb | 704 | 836.5 | 20.9877 |
| Mar | 835 | 836.5 | 0.0027 |
| Apr | 826 | 836.5 | 0.1318 |
| May | 900 | 836.5 | 4.8204 |
| Jun | 868 | 836.5 | 1.1862 |
| Jul | 920 | 836.5 | 8.3350 |
| Aug | 901 | 836.5 | 4.9734 |
| Sep | 856 | 836.5 | 0.4546 |
| Oct | 862 | 836.5 | 0.7773 |
| Nov | 783 | 836.5 | 3.4217 |
| Dec | 797 | 836.5 | 1.8652 |
|  | 10038 | 10038.0 | 50.0048 |



*P*-value = 

***Conclusion***

Reject ; there is sufficient evidence to support the claim that  for each month. There is sufficient evidence to reject the claim that the occurrence of violent crime is the same for each month. A major factor involved in this conclusion is the large contribution of the month of February to the calculated  statistic. The comparison of frequencies for each month is not fair because not all months have the same number of days.

***Exercise***

The table below lists the results of the Advanced Placement Biology class conducted genetics experiments with fruit flies. Use a 0.05 significance level to test the claim that the observed frequencies agree with the proportions that were expected according to principles of genetics

Conduct the hypothesis test and the test statistic, critical value and/or *P*-value, and state the conclusion.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Characteristic*** | ***Red eye /***  ***normal wing*** | ***Sepia eye /***  ***normal wing*** | ***Red eye /***  ***vestigial wing*** | ***Sepia eye /***  ***vestigial wing*** |
| ***Frequency*** | 59 | 15 | 2 | 4 |
| ***Expected proportion*** |  |  |  |  |

***Solution***

Original Claim: Observed frequencies fit the expected proportions.



 is not as claimed









*Calculations*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Day* | *O* | *E* |  |  |
| 1 | 59 |  | 4.3556 |
| 2 | 15 |  | 0.00 |
| 3 | 2 |  | 11.2667 |
| 4 | 4 |  | 0.200 |
|  | 80 | 80 | 15.8222 |



*P*-value = 

***Conclusion***

Reject ; there is sufficient evidence to reject the claim that the proportions are as claimed. There is sufficient evidence to reject the claim that observed frequencies fit the proportions that were expected according to the principles of genetics

***Exercise***

The table below lists the claims that its M&M plain candies are distributed with the following color percentages: 16% green, 20% orange, 14% yellow, 24% blue, 13% red, and 13% brown. Use a 0.05 significance level to test the claim that the color distribution is as claimed.

***Solution***

Original Claim: The color distribution is as stated



 is not as claimed









*Calculations*:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Day* | *O* | *E* |  |  |
| G | 19 |  | 0.5625 |
| O | 25 |  | 1.2500 |
| Y | 8 |  | 2.5714 |
| Bl | 27 |  | 0.3750 |
| R | 13 |  | 0.0 |
| Br | 8 |  | 1.9231 |
|  | 100 | 100 | 6.6820 |



*P*-value = 

***Conclusion***

Do not reject ; there is not sufficient evidence to reject the claim that the proportion are as stated. There is no sufficient evidence to reject the claim that the color distribution is as stated.