**PART 1:**

**1.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **20703.83** | **586.5061** | **35.30028** | **2.8E-264** | **19554.22** | **21853.43** | **19554.22** | **21853.43** |
| **X Variable 1** | **-469.785** | **129.9824** | **-3.61422** | **0.000302** | **-724.562** | **-215.008** | **-724.562** | **-215.008** |

sales = 20703.83 – 469.785 \* (price)

1. Therefore, the estimated intercept is 20703.83 and the estimated slope is -469.785
2. For the null hypothesis β1 = 0, the P-value = 0.0003 < 0.05. Therefore, we reject the null hypothesis and conclude that β1 is not equal to 0

**2.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **17631.21** | **716.4206** | **24.61014** | **1.2E-131** | **16226.96** | **19035.46** | **16226.96** | **19035.46** |
| **X Variable 1** | **5114.472** | **1084.992** | **4.713833** | **2.45E-06** | **2987.788** | **7241.155** | **2987.788** | **7241.155** |

sales = 17631.21 + 5114.472 \* (region)

Interpretation:

* The estimated intercept is 17631.21. This implies that if the region is CN, the expected sales would be 17631.21
* The estimated slope coefficient is 5114.471. This means that if the region is US, the expected sales would be (17631.21+5114.472) = 22745.682 for the app

**3.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **16541.26** | **788.1403** | **20.98771** | **1.12E-96** | **14996.43** | **18086.09** | **14996.43** | **18086.09** |
| **X Variable 1** | **4964.451** | **1085.646** | **4.572809** | **4.84E-06** | **2836.486** | **7092.415** | **2836.486** | **7092.415** |
| **X Variable 2** | **3910.127** | **1179.941** | **3.313833** | **0.000922** | **1597.335** | **6222.919** | **1597.335** | **6222.919** |

sales = 16541.26 + 4964.451 \* (region) + 3910.127 \* (ads)

Interpretation:

* The estimated intercept is 16541.26. If the app is present in CN and has no in app ads then the expected sales are 16541.26.
* The estimated coefficient for region is 4964.541. This implies that if the app is launched in US and has no in app ads then the expected sales is (16541.26+4964.541) = 21505.801
* The estimated coefficient for in app ads is 3910.127. This implies that if the app is launched in US and has in app ads then the expected sales is (16541.26+4964.541+3910.127) = 25415.93

**4.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **8.970884** | **0.009973** | **899.5355** | **0** | **8.951336** | **8.990431** | **8.951336** | **8.990431** |
| **X Variable 1** | **-0.06169** | **0.009962** | **-6.19279** | **6.03E-10** | **-0.08122** | **-0.04217** | **-0.08122** | **-0.04217** |

Log(sales) = 8.970884 – 0.06169 \* log(price)

Interpretation:

* The estimated coefficient for log(price) is -0.06169. This implies that if the price of an app is increased by 1% then the estimated sales reduces by 0.06%.
* The estimated intercept is 8.97. This means that for price = 1, the estimated sales of the app would be e8.97.

**5.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **8.729101** | **0.052291** | **166.9331** | **0** | **8.626606** | **8.831596** | **8.626606** | **8.831596** |
| **log\_price** | **-0.08567** | **0.010898** | **-7.86059** | **4.03E-15** | **-0.10703** | **-0.06431** | **-0.10703** | **-0.06431** |
| **log\_filesize** | **0.036788** | **0.005847** | **6.292231** | **3.2E-10** | **0.025328** | **0.048248** | **0.025328** | **0.048248** |
| **log\_screenshot** | **-0.02136** | **0.02067** | **-1.0334** | **0.301429** | **-0.06188** | **0.019155** | **-0.06188** | **0.019155** |
| **log\_rating** | **0.106074** | **0.024847** | **4.269101** | **1.97E-05** | **0.057372** | **0.154777** | **0.057372** | **0.154777** |

Log(sales) = 8.729101 – 0.08567 \* log(price) + 0.036788 \* log(filesize) – 0.02136 \* log(screenshot) + 0.106074 \* log(rating)

Interpretation:

Controlling for all other variables, a 1% increase in:

* price leads to a reduction of 0.08% in sales
* filesize leads to an increase of 0.03% in sales
* number of screenshots leads to a reduction of 0.02% in sales
* rating leads to an increase of 0.1% in sales
* If all the independent variables have a value of 1, then the expected sales are exp(8.729)

**6.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***Coefficients*** | ***Standard Error*** | ***t Stat*** | ***P-value*** | ***Lower 95%*** | ***Upper 95%*** | ***Lower 95.0%*** | ***Upper 95.0%*** |
| **Intercept** | **8.751645** | **0.052631** | **166.2836** | **0** | **8.648484** | **8.854807** | **8.648484** | **8.854807** |
| **log\_price** | **-0.10288** | **0.011856** | **-8.67784** | **4.36E-18** | **-0.12612** | **-0.07964** | **-0.12612** | **-0.07964** |
| **log\_filesize** | **0.030285** | **0.006106** | **4.959982** | **7.11E-07** | **0.018317** | **0.042253** | **0.018317** | **0.042253** |
| **log\_screenshot** | **-0.01959** | **0.020669** | **-0.94758** | **0.343353** | **-0.0601** | **0.020928** | **-0.0601** | **0.020928** |
| **log\_rating** | **0.100605** | **0.024883** | **4.043121** | **5.3E-05** | **0.051832** | **0.149378** | **0.051832** | **0.149378** |
| **log\_price\*purchases** | **0.059231** | **0.016089** | **3.681514** | **0.000233** | **0.027696** | **0.090767** | **0.027696** | **0.090767** |

Log(sales) = 8.751645 – 0.10288 \* log(price) + 0.030285 \* log(filesize) – 0.01959 \* log(screenshot) + 0.100605 \* log(rating) + 0.059231 \* log(price) \* purchase

Interpretation:

Controlling for all other variables, a 1% increase in:

* price leads to a reduction of 0.10% in sales
* filesize leads to an increase of 0.03% in sales
* number of screenshots leads to a reduction of 0.019% in sales
* rating leads to an increase of 0.1% in sales
* if all the independent variables have a value of 1, then the expected sales are exp(8.75)

Also, if an app offers in app purchases then controlling for all other variables, a 1% increase in price leads an increase of (-0.1+0.059) = 0.049% in sales

**PART 2:**

**Exercise 1:**

Null Hypothesis: The average viewing time is less than or equal to 12 minutes

Alternative Hypothesis: The average viewing time is more than 12 minutes

Mean (24 members) = 13.92 minutes

S.D. (24 members) = 5.34

T-stat (23,0.05) = 1.714

T = (13.92 – 12)/(5.34/sqrt(24)) = 1.761

Therefore, since T > T-stat we reject the null hypothesis and conclude that the average viewing time for the website is more than 12 minutes.

**Exercise 2:**

1. Null Hypothesis: The new website design is not more likely to lead to more sales

Alternative Hypothesis: The new website design is more likely to lead to more sale conversions

In order to determine the result, we will conduct a test of proportions:

N1 = 477, N2 = 93

P1 = 11.53%, P2 = 25.81%

s.e. = sqrt((P1(1-P1)/N1) + (P2(1-P2)/N2) = 0.04679

Z = P1-P2 / s.e. = -2.99 < -Z0.05

Therefore, we reject the null hypothesis and conclude that the new website leads to more sale conversions

1. Null Hypothesis: There is no statistical difference between the mean of time spent by customers on the new and old website

Alternative Hypothesis: There is a statistical difference between the mean of time spent by customers on the new and old website

In order to determine the result, we will conduct a test of proportions:

N1 = 477, N2 = 93

x1 = 6.47, x2 = 8.98

s1 = 3.26, s2 = 3.47

s.e. = sqrt((s12/N1) + (s22/N2) = 0.3826

t = x1-x2 / s.e. = -6.56 < -1.96

Therefore, we reject the null hypothesis and conclude that there is a statistical difference between the mean time spent by customers on the new and old website

1. Yes, the data does suggest that the assignment of customers to the new and old website was not completely random. Of all the customers assigned to the old website, only 26% were members whereas this number was 73% for customers assigned to the new website. However, we shall check this using a hypothesis test.

Checking for randomness:

P1 = proportion of members in the cohort assigned to old website

P2 = proportion of members in the cohort assigned to new website

Null hypothesis: p1 – p2 = 0: the proportions of members in both groups is similar

Alternative Hypothesis: p1 – p2 != 0: the proportions of members in both groups is not similar.

P1 = 0.27, P2 = 0.73

N1 = 477, N2 = 93

s.e. = sqrt((P1(1-P1)/N1) + (P2(1-P2)/N2) = 0.05

Z = P1-P2 / s.e. = -9.2 < -Z0.05

Therefore, we reject the null hypothesis and conclude that the proportions of members in the cohort assigned to the old and new website are different

If the customers were truly segmented randomly, we should check if the membership has any effect on sales conversion. Thus, we can now check whether membership has any effect on sale conversion using a hypothesis test.

P1 = proportion of non-members who converted to a sale

P2 = proportion of members who converted to a sale

Null hypothesis: p1 – p2 = 0: the proportions of non-members and members converting to a sale is similar

Alternative Hypothesis: p1 – p2 != 0: the proportions of non-members and members converting to a sale is not similar

P1 = 0.02, P2 = 0.35

N1 =374, N2 = 196

s.e. = sqrt((P1(1-P1)/N1) + (P2(1-P2)/N2) = 0.035

Z = P1-P2 / s.e. = -9.43 < -Z0.05

Therefore, we reject the null hypothesis and conclude that membership does have an effect on sales conversion. Due to this reason, the data does suggest that the answer for part a could vary if the selection of customer traffic was truly random.

1. Null Hypothesis: There is no significant increase in returns due to the new features

Alternative Hypothesis: There is a significant increase in returns due to the new features

N1 = 477, N2 = 93

P1 = 6.50%, P2 = 9.68%

s.e. = sqrt((P1(1-P1)/N1) + (P2(1-P2)/N2) = 0.0327

Z = P1-P2 / s.e. = -0.97 > Z0.05

Therefore, we fail to reject the null hypothesis and conclude that there is no significant increase in returns due to the addition of new features