**Chapter 2**

We assume that the normal error regression model is applicable. This model is:

$$Y\_{i}=β\_{0}+β\_{1}X\_{i}+ε\_{i}$$

where:

$β\_{0}$ and $β\_{1}$, are parameters

$X\_{i}$ are known constants

$ε\_{i}$ are independent $N (0, σ^{2})$

$$E\left(Y\_{i}\right)=β\_{0}+β\_{1}X\_{i}$$

**Sampling Distribution of** $\hat{β\_{1}}$

$$\hat{β\_{1}}=b\_{1}=\sum\_{i=1}^{n}\frac{\left(X\_{i}-\overbar{X}\right)\left(Y\_{i}-\overbar{Y}\right)}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}$$

$$E\left(\hat{β\_{1}}\right)=β\_{1}$$

$$σ^{2}\left(\hat{β\_{1}}\right)=\frac{σ^{2}}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}$$

$$s^{2}\left(\hat{β\_{1}}\right)=\frac{MSE}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}$$

$$\frac{b\_{1}-β\_{1}}{s\left(b\_{1}\right)}\~t\_{\left(n-2\right)}$$

**Confidence Interval for** $β\_{1}$

$$P\left[b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)\leq β\_{1}\leq b\_{1}+t\_{\left(1-α/2,n-2\right)}s\left(b\_{1}\right)\right]=1-α$$

C.I $\left(1-α\right)\%$ for $β\_{1}$

$$b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)\leq β\_{1}\leq b\_{1}+t\_{\left(1-α/2,n-2\right)}s\left(b\_{1}\right)$$

**Tests Concerning** $β\_{1}$

|  |
| --- |
| 1. Hypothesis  |
| $$H\_{0}:β\_{1}=β\_{10} $$$$H\_{1}:β\_{1}\ne β\_{10}$$ | $$H\_{0}:β\_{1}=β\_{10} $$$$H\_{1}:β\_{1}>β\_{10}$$ | $$H\_{0}:β\_{1}=β\_{10} $$$$H\_{1}:β\_{1}<β\_{10}$$ |
| 2. Test statistic |
| $$T\_{0}=\frac{b\_{1}-β\_{10}}{s\left(b\_{1}\right)}$$ |
| 3. Decision: Reject $H\_{0}$ if |
| $$\left|T\_{0}\right|>t\_{\left(1-\frac{α}{2},n-2\right)}$$ | $$T\_{0}>t\_{\left(1-α,n-2\right)}$$ | $$T\_{0}<t\_{\left(α,n-2\right)}$$ |
| P-value: Reject $H\_{0}$ if $p-value<α$ |
| p-value=$2P\left(t\_{\left(n-2\right)}>\left|T\_{0}\right|\right)$ | $$p-value=P\left(t\_{\left(n-2\right)}>T\_{0}\right)$$ | $$p-value=P\left(t\_{\left(n-2\right)}<T\_{0}\right)$$ |

**Page 114**

**Q2.4**. Refer to **Grade point average** Problem 1.19.

**a. Obtain a 99 percent confidence interval for** $β\_{1}$**. Interpret your confidence interval. Does it include zero? Why might the director of admissions be interested in whether the confidence interval includes zero?**

Solution:

By using Minitab:

$$Stat\rightarrow Regression\rightarrow Regression\rightarrow Fit Regression Mode$$





**Regression Analysis: Yi versus Xi**

Analysis of Variance

Source DF Seq SS Contribution Adj SS Adj MS F-Value P-Value

Regression 1 3.588 7.26% 3.588 3.5878 9.24 0.003

 Xi 1 3.588 7.26% 3.588 3.5878 9.24 0.003

Error n-2=118 45.818 92.74% SSE=45.818 MSE=0.3883

Lack-of-Fit 19 6.486 13.13% 6.486 0.3414 0.86 0.632

 Pure Error 99 39.332 79.61% 39.332 0.3973

Total 119 49.405 100.00%

Model Summary

 S R-sq R-sq(adj) PRESS R-sq(pred)

0.623125 7.26% 6.48% 47.6103 3.63%

Coefficients

Term Coef SE Coef 99% CI T-Value P-Value VIF

Constant 2.114 0.321 ( 1.274, 2.954) 6.59 0.000

Xi 0.0388 0.0128 **(0.0054, 0.0723)** 3.04 **0.003** 1.00

Regression Equation

Yi = 2.114 + 0.0388 Xi

99% C.I for $β\_{1}$: $b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)\leq β\_{1}\leq b\_{1}+t\_{\left(1-α/2,n-2\right)}s\left(b\_{1}\right)$

$$0.0054\leq β\_{1}\leq 0.0723$$

**Interpret your confidence interval. Does it include zero? No**

**Why might the director of admissions be interested in whether the confidence interval includes zero?**

If the C.I of $β\_{1}$ include zero, then $β\_{1}$ can tack zero and $β\_{1}=0$

**b. Test, using the test statistic t\*, whether or not a linear association exists between student's ACT score (X) and GPA at the end of the freshman year (Y). Use a level of significance of 0.01 State the alternatives, decision rule, and conclusion.**

$$α=0.01$$

1. Hypothesis

$$H\_{0}:β\_{1}=0 $$

$$H\_{1}:β\_{1}\ne 0$$

2. Test statistic

$$T\_{0}=\frac{b\_{1}-β\_{10}}{s\left(b\_{1}\right)}=\frac{b\_{1}}{s\left(b\_{1}\right)}=\frac{0.0388}{0.0128}=3.04$$

3. Decision: Reject $H\_{0}$ if $ \left|T\_{0}\right|>t\_{\left(1-\frac{α}{2},n-2\right)}$, $ 3.04>t\_{\left(0.995,118\right)}=$2.61814

Then reject $H\_{0}$

**c. What is the P-value of your test in part (b)? How does it support the conclusion reached in part (b)?**

p-value= 0.003<0.01, then we reject $H\_{0}$.

Q2.5. Refer to **Copier maintenance Problem** 1.20.

$$n=45, \sum\_{i=1}^{n=45}X\_{i}=230, \sum\_{i=1}^{45}Y\_{i}=3432, \sum\_{i=1}^{45}X\_{i}^{2}=1516, \sum\_{i=1}^{45}X\_{i}Y\_{i}=22660$$

$$SSE=3416.377$$

**a. Estimate the change in the mean service time when the number of copiers serviced increases by one. Use a 90 percent confidence interval. Interpret your confidence interval.**

90% C.I for $β\_{1}$: $b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)\leq β\_{1}\leq b\_{1}+t\_{\left(1-α/2,n-2\right)}s\left(b\_{1}\right)$

$$α=1-0.9=0.1$$

$$b\_{1}=\sum\_{i=1}^{n}\frac{\left(X\_{i}-\overbar{X}\right)\left(Y\_{i}-\overbar{Y}\right)}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}=\sum\_{i=1}^{n}\frac{\left(X\_{i}Y\_{i}-\overbar{X}Y\_{i}-X\_{i}\overbar{Y}+\overbar{X}\overbar{Y}\right)}{\sum\_{i=1}^{n}\left(X\_{i}^{2}-2\overbar{X}X\_{i}+\overbar{X}^{2}\right)}=\frac{\sum\_{i=1}^{n}X\_{i}Y\_{i}-n\overbar{X}\overbar{Y}}{\sum\_{i=1}^{n}X\_{i}^{2}-n\overbar{X}^{2}}=\frac{22660-45\*5.1111\*76.2667}{1516-45\*5.1111^{2}}=15.035$$

$$s^{2}\left(b\_{1}\right)=\frac{MSE}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}=\frac{3416.377/(45-2)}{1516-45\*5.1111^{2}}=0.23337$$

$$s\left(b\_{1}\right)=\sqrt{0.2337}=0.48308$$

$$t\_{\left(1-\frac{α}{2},n-2\right)}=t\_{\left(0.95,43\right)}=1.68107$$

$$b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)=15.035-1.68107\*0.48308=14.222$$

$$b\_{1}+t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)=15.035+1.68107\*0.48308=15.84709$$

$$14.222\leq β\_{1}\leq 15.847$$

**b. Conduct a t test to determine whether or not there is a linear association between X and Y here; control the** $α$ **a risk at 0.01. State the alternatives, decision rule, and conclusion. What is the P-value of your test?**

$$α=0.01$$

1. Hypothesis

$$H\_{0}:β\_{1}=0 $$

$$H\_{1}:β\_{1}\ne 0$$

2. Test statistic

$$T\_{0}=\frac{b\_{1}-β\_{10}}{s\left(b\_{1}\right)}=\frac{b\_{1}}{s\left(b\_{1}\right)}=\frac{15.035}{0.48308}=31.123$$

3. Decision: Reject $H\_{0}$ if $ \left|T\_{0}\right|>t\_{\left(1-\frac{α}{2},n-2\right)}$, $ 31.123>t\_{\left(0.995,43\right)}=2.695$

Then reject $H\_{0}$

p-value=$2P\left(t\_{\left(n-2\right)}>\left|T\_{0}\right|\right)=2\left(1-P\left(t\_{\left(43\right)}<31.123\right)\right)=2\left(1-1\right)$

$0.00<0.01$, then we reject $H\_{0}$.

**c. Are your results in parts (a) and (b) consistent? Explain.**

Yes, the C.I of $β\_{1}$ does not include zero, and we reject $H\_{0}$.

**d. The manufacturer has suggested that the mean required time should not increase by more than 14 minutes for each additional copier that is serviced on a service call. Conduct a test to decide whether this standard is being satisfied by Tri-City. Control the risk of a Type I error at 0.05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?**

$$α=0.05$$

1. Hypothesis

$$H\_{0}:β\_{1}\leq 14 $$

$$H\_{1}:β\_{1}>14$$

2. Test statistic

$$T\_{0}=\frac{b\_{1}-β\_{10}}{s\left(b\_{1}\right)}=\frac{b\_{1}-14}{s\left(b\_{1}\right)}=\frac{15.035-14}{0.48308}=2.143$$

3. Decision: Reject $H\_{0}$ if $ T\_{0}>t\_{\left(1-α,n-2\right)}$, $ 2.143>t\_{\left(0.95,43\right)}=1.861$

Then reject $H\_{0}$

p-value=$P\left(t\_{\left(n-2\right)}>T\_{0}\right)=\left(1-P\left(t\_{\left(43\right)}<2.143\right)\right)=\left(1-0.981\right)=0.019<0.05$

, then we reject $H\_{0}$.

**Q2.6. Refer to Airfreight breakage Problem 1.21.**

$\overbar{X}=1, \overbar{Y}=14.2$, $\sum\_{i=1}^{n=10}\left(X\_{i}-\overbar{X}\right)\left(Y\_{i}-\overbar{Y}\right)=40$

$\sum\_{i=1}^{10}\left(X\_{i}-\overbar{X}\right)^{2}=10$, $MSE=2.2$

**a. Estimate** $β\_{1}$ **with a 95 percent confidence interval. Interpret your interval estimate.**

95% C.I for $β\_{1}$: $b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)\leq β\_{1}\leq b\_{1}+t\_{\left(1-α/2,n-2\right)}s\left(b\_{1}\right)$

$$α=1-0.95=0.05$$

$$b\_{1}=\hat{β\_{1}}=\sum\_{i=1}^{n=120}\frac{\left(X\_{i}-\overbar{X}\right)\left(Y\_{i}-\overbar{Y}\right)}{\sum\_{i=1}^{n=120}\left(X\_{i}-\overbar{X}\right)^{2}}=4$$

$$s^{2}\left(b\_{1}\right)=\frac{MSE}{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}=\frac{2.2}{10}=0.22$$

$$s\left(b\_{1}\right)=\sqrt{0.22}=0.469$$

$$t\_{\left(1-\frac{α}{2},n-2\right)}=t\_{\left(0.975,8\right)}=2.306$$

$$b\_{1}-t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)=4-2.306\*0.469=2.918$$

$$b\_{1}+t\_{\left(1-\frac{α}{2},n-2\right)}s\left(b\_{1}\right)=4+2.306\*0.469=5.081$$

$$2.918\leq β\_{1}\leq 5.081$$

**b. Conduct a t test to decide whether or not there is a linear association between number of times a carton is transferred (X) and number of broken ampules (Y). Use a level of significance of 0.05. State the alternatives, decision rule, and conclusion. What is the P-value of the test?**

$$α=0.05$$

1. Hypothesis

$$H\_{0}:β\_{1}=0 $$

$$H\_{1}:β\_{1}\ne 0$$

2. Test statistic

$$T\_{0}=\frac{b\_{1}-β\_{10}}{s\left(b\_{1}\right)}=\frac{b\_{1}}{s\left(b\_{1}\right)}=\frac{4}{0.469}=8.528$$

3. Decision: Reject $H\_{0}$ if $ \left|T\_{0}\right|>t\_{\left(1-\frac{α}{2},n-2\right)}$, $8.528>t\_{\left(0.975,8\right)}=2.308$

Then reject $H\_{0}$

p-value=$2P\left(t\_{\left(n-2\right)}>\left|T\_{0}\right|\right)=2\left(1-P\left(t\_{\left(8\right)}<8.528\right)\right)=2\left(1-0.9999\right)$

$0.0002<0.05$, then we reject $H\_{0}$.

Analysis of Variance

Source DF Seq SS Contribution Adj SS Adj MS F-Value P-Value

Regression 1 160.000 90.09% 160.000 160.000 72.73 0.000

 Xi 1 160.000 90.09% 160.000 160.000 72.73 0.000

Error 8 17.600 9.91% 17.600 2.200

 Lack-of-Fit 2 0.933 0.53% 0.933 0.467 0.17 0.849

 Pure Error 6 16.667 9.38% 16.667 2.778

Total 9 177.600 100.00%

Model Summary

 S R-sq R-sq(adj) PRESS R-sq(pred)

1.48324 90.09% 88.85% 25.8529 85.44%

Coefficients

Term Coef SE Coef 95% CI T-Value P-Value VIF

Constant 10.200 0.663 (8.670, 11.730) 15.38 0.000

Xi 4.000 0.469 (2.918, 5.082) 8.53 0.000 1.00

Regression Equation

Yi = 10.200 + 4.000 Xi

**H.W:**

**Q2.7 Refer to Plastic hardness Problem 1.22.**

**a. Estimate the change in the mean hardness when the elapsed time increases by one hour. Use a 99 percent confidence interval. Interpret your interval estimate.**

**b. The plastic manufacturer has stated that the mean hardness should increase by 2 Brinell units per hour. Conduct a two-sided test to decide whether this standard is being satisfied; use** $α=0.01$**. State the alternatives, decision rule, and conclusion. What is the P-value of the test?**