JustQ1

2024-11-13

Q1

Suppose that six observations of the yield (Y) of a chemical process were taken at each of four temperature levels (X) for running the process, but you are only given information on the sample means and standard deviations for the observed yields at each temperature. The summary data are

Temperature (°C)	Sample Mean	Sample Variance	Sample Size
150	66	1.15	6
200	81	1.00	6
250	89	1.35	6
300	92	0.90	6

General Equations

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$Var(X) = \sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Week 10, Slide 10

$$b_0 = \bar{Y} - b_1\bar{x}$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})Y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

```
temperature <- c(150, 200, 250, 300)
sampleMean <- c(66, 81, 89, 92)
sampleVariance <- c(1.15, 1.00, 1.35, 0.90)

temperatureMean <- mean(temperature)
tempVar <- var(temperature)
responseMean <- mean(sampleMean)

num <- sum(6 * (temperature - temperatureMean)*(sampleMean - responseMean))
denom <- sum(6 * (temperature - temperatureMean)^2)
```

```r
b1 <- num/denom
b0 <- responseMean - (b1*temperatureMean)

b1
```

```
## [1] 0.172
```

```r
b0
```

```
## [1] 43.3
```

Week 10 Slide 15

$$Var(b_0) = \sigma^2 * \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right)$$

$$Var(b_1) = \sigma^2 * \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Week 10 Slide 26

$$\hat{\sigma}^2 = MS_{error} = SS_{error}/(n-2)$$

See below (ANOVA table calculations for reasoning)

$$SS_{error} = \sum_{j=1}^{4} (n_j - 1)\hat{\sigma}_j^2$$

```r
n <- 24
n_i <- 6

temperature <- c(150, 200, 250, 300)
sampleMean <- c(66, 81, 89, 92)
sampleVariance <- c(1.15, 1.00, 1.35, 0.90)

temperatureMean <- mean(temperature)
tempVar <- var(temperature)
responseMean <- mean(sampleMean)

hatY <- b0 + b1*temperature
hatYRep <- rep(hatY, each = 6)
sampleMeanRep <- rep(sampleMean, each = 6)
SSLack <- sum((sampleMeanRep - hatYRep)^2)
SSLack
```

```
## [1] 217.2
```

```r
nRep <- rep(6, 4)
pooledVariance <- sum((nRep - 1) * sampleVariance) / sum(nRep - 1)
pooledVariance
```

```
## [1] 1.1
```

```r
SSPure <- sum((n_i - 1) * sampleVariance)
SSPure
```

```
## [1] 22
```

```r
SSPE <- SSPure
SSLOF <- SSLack
SSE <- SSLOF + SSPE
SS_error <- SSE
MSE <- SS_error / 22
```

```r
temperature <- c(150, 200, 250, 300)
sampleMean <- c(66, 81, 89, 92)
sampleVariance <- c(1.15, 1.00, 1.35, 0.90)

temperatureMean <- mean(temperature)
tempVar <- var(temperature) * length(temperature)
tempRep <- rep(temperature, each = 6)

Varb1 <- MSE / sum((tempRep - temperatureMean)^2)
Varb0 <- MSE * ((1/n) + (temperatureMean^2 /sum((tempRep - temperatureMean)^2)))

SEb1 <- sqrt(Varb1)
SEb0 <- sqrt(Varb0)

SEb0
```

```
## [1] 2.791437
```

```r
SEb1
```

```
## [1] 0.01204034
```

Week 10 Slide 23

$$SS_{model} = b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2$$

```r
SSModel <- b1^2 * sum(6 * (temperature - temperatureMean)^2)
SSModel
```

```
## [1] 2218.8
```

Week 10 Slide 25

$$SS_{model} = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y}_i)^2$$

$$SS_{error} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

Week 11 Slide 40

$$SS_{error} = SS_{pureerror} + SS_{lack-of-fit}$$

$$SS_{lack-of-fit} = \sum_{i} \sum_{j} (\bar{Y}_{i.} - \hat{Y}_i)^2$$

$$\hat{Y}_i = b_0 + b_1 x_i$$

```
n <- 24
temperature <- c(150, 200, 250, 300)
sampleMean <- c(66, 81, 89, 92)
sampleVariance <- c(1.15, 1.00, 1.35, 0.90)

temperatureMean <- mean(temperature)
tempVar <- var(temperature)
responseMean <- mean(sampleMean)

varB0 <- 1/n + (temperatureMean^2 / sum(6 * (temperature - temperatureMean)^2) )
SEb0 <- sqrt(varB0)
varB1 <- 1 / sum(6 * (temperature - temperatureMean)^2)
SEB1 <- sqrt(varB1)

hatY <- b0 + b1*temperature
hatYRep <- rep(hatY, each = 6)
sampleMeanRep <- rep(sampleMean, each = 6)
SSLack <- sum((sampleMeanRep - hatYRep)^2)
SSLack
```

```
## [1] 217.2
```

$$SS_{pureerror} = \sum_{i} \sum_{j} (Y_{ij} - \bar{Y}_i) = \sum_{i=1}^{4} \sum_{j=1}^{6} (Y_{ij} - \bar{Y}_{i.})^2$$

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2$$

$$SS_{pureerror} = \sum_{i=1}^{4} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i.})^2 = \sum_{i=1}^{4} (n_i - 1) s_i^2$$

```r
n <- 24
n_i <- 6

temperature <- c(150, 200, 250, 300)
sampleMean <- c(66, 81, 89, 92)
sampleVariance <- c(1.15, 1.00, 1.35, 0.90)

temperatureMean <- mean(temperature)
tempVar <- var(temperature)
responseMean <- mean(sampleMean)

SSPure <- sum((n_i - 1) * sampleVariance)
SSPure
```

```
## [1] 22
```

Giving the following table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Regression on X	1	2218.8	2218.8
Residuals	22	239.2	10.87273
- Lack-of-fit	2	217.2	108.60000
- Pure error	20	22	1.1
Total	23	2458	

```r
temperature <- c(150, 200, 250, 300)
sample_mean <- c(66, 81, 89, 92)
sample_variances <- c(1.15, 1.00, 1.35, 0.90)
n_i <- 6

SSPE <- SSPure
SSLOF <- SSLack
SSR <- SSModel

temperatureMean <- mean(temperature)
responseMean <- mean(sample_mean)

df_regression <- 1
df_residual <- length(temperature) * n_i - 2
df_lack_of_fit <- length(temperature) - 2
df_pure_error <- df_residual - df_lack_of_fit
df_total <- length(temperature) * n_i - 1

SSE <- SSLOF + SSPE
SST <- SSR + SSE

MSR <- SSR / df_regression
MSE <- SSE / df_residual
MSLOF <- SSLOF / df_lack_of_fit
MSPE <- SSPE / df_pure_error

anova_table <- data.frame(
```

```r
  "Source of Variation" = c("Regression on X", "Residuals", "- Lack-of-fit", "- Pure error", "Total"),
  "Degrees of Freedom" = c(df_regression, df_residual, df_lack_of_fit, df_pure_error, df_total),
  "Sum of Squares" = c(SSR, SSE, SSLOF, SSPE, SST),
  "Mean Square" = c(MSR, MSE, MSLOF, MSPE, NA)
)

# Display the table
print(anova_table)
```

```
##   Source.of.Variation Degrees.of.Freedom Sum.of.Squares Mean.Square
## 1     Regression on X                  1         2218.8  2218.80000
## 2           Residuals                 22          239.2    10.87273
## 3       - Lack-of-fit                  2          217.2   108.60000
## 4        - Pure error                 20           22.0     1.10000
## 5               Total                 23         2458.0          NA
```

```r
MSE <- sum(n_i * (n_i - 1) * sampleVariance) / (n-2)
MSE
```

```
## [1] 6
```