**Problem Set 4 PSY 507, Fall 2016**

**Hyun Joon Park**

1. At 10 grocery stores, a market researcher records the number of cases of a product sold both before and after coupons were mailed to households in the area. The data below are the changes in the number of cases sold (positive means more cases were sold after the ad campaign, negative means fewer cases were sold, and zero means there was no change).

|  |  |
| --- | --- |
| Store | Change |
| A | 5 |
| B | 12 |
| C | -3 |
| D | 20 |
| E | -5 |
| F | 15 |
| G | 11 |
| H | 0 |
| I | 8 |
| J | 7 |

1. Calculate the mean and standard deviation of the change scores by hand (i.e., using excel).



1. Using the sum of squared error (SSE) as your aggregate measure of error,

compute by hand/excel the ERROR for a simple model

that predicts no change for each store (i.e., a change of 0).

That ERROR equals SSE(C), the sum of squared errors for the compact model.



1. Still using SSE, compute by hand/excel the ERROR for the simple model that uses the mean change score from the DATA. That ERROR equals SSE(A), the sum of squared errors for the augmented model.



1. Use SSE(C) and SSE(A) to compute PRE.

PRE = {SSE( C ) –SSE ( A )} / SSE ( C )

0.461394

1. Write out MODEL A, MODEL C, and the null hypothesis.



Model A:

Model C: 

Null hypothesis: 

1. Should MODEL C be rejected in favor of MODEL A? Be sure to report any relevant statistics that justify your decision.

For our model, n= 10 and we have only used one estimator in Model A. So, our degree of freedom would be 9 (n- PA).

With n-PA=9 and with alpha value 0.05 the cut-off point to reject model C is 0.362 (from the critical PRE value table).

Since our PRE value for our Model A and Model C is 0.461394 we can reject the Model C in favor of Model A.

1. Enter these data in your software package (e.g., SPSS or R) to verify the means and sums of squares you calculated.

Provide and annotate your output to explain what you see.

Everything has the same value as I have calculated by hands when I ran analysis on R.

This result shows us that we can safely reject MODEL( C ), because MODEL ( A ) reduces the error by significant amount.

> data<- c(5, 12, -3, 20, -5, 15, 11, 0, 8, 7)

> mean(data)

[1] 7

dataf <- data.frame(y= c( 5, 12, -3, 20, -5, 15, 11, 0, 8, 7), x= rep(0,10))

test <- lm(y~x, data= dataf)

summary(test)

Call:

lm(formula = y ~ x, data = dataf)

Residuals:

Min 1Q Median 3Q Max

-12.00 -5.75 0.50 4.75 13.00

Coefficients: (1 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) **7.000** 2.521 2.777 0.0215 \*

x NA NA NA NA

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: **7.972** on 9 degrees of freedom

> anova(test)

Analysis of Variance Table

Response: y

Df Sum Sq Mean Sq F value Pr(>F)

Residuals 9 **572** 63.556

1. Write a brief **lay** explanation of your findings. Think of this as a “5 o’clock news” summary; keep it simple and leave out any jargon.

MODEL ( A ) reduces the squared sum of error significantly than MODEL ( C) . So we can discard (reject) MODEL ( C ) and select MODEL ( A) to explain our phenomena of how sending out coupons influenced the sales of a product.

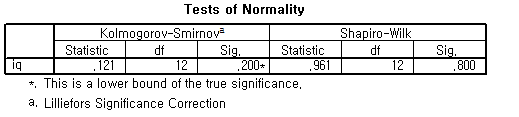
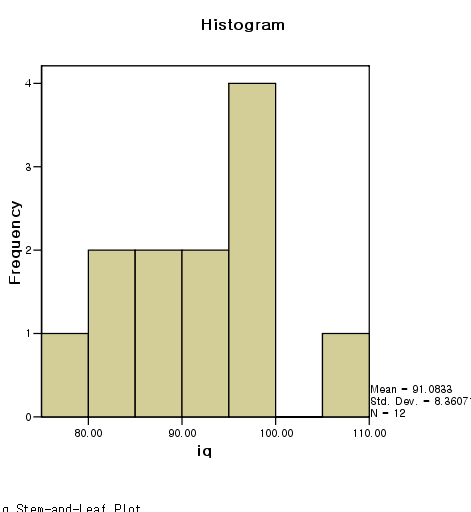
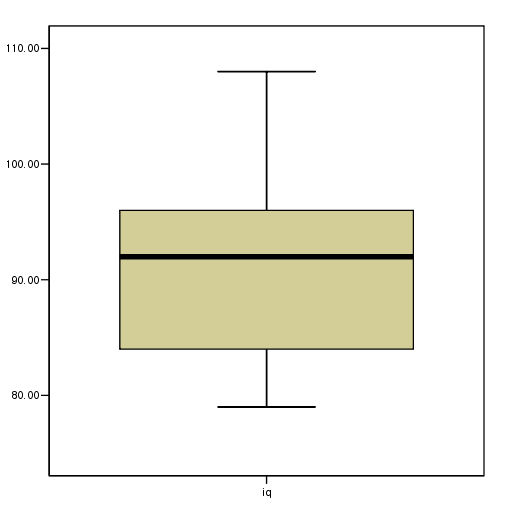
1. Open the ADD data set. This data set comes from a study of 386 children who had, and had not, exhibited symptoms of attention deficit disorder (ADD) during childhood.

For our purposes, we are going to look to see whether students who have repeated at least one grade (variable = REPEAT) have a mean IQ of 100,

the same as the general population. Before running the statistical test, you will need to select those cases that have repeated at least on grade

* 1. Do some exploratory data analysis (EDA) to get a sense of how the data look. Do the data look normally distributed? Do you think there is any reason to be concerned about the independence of errors assumption?

From the box-plot data seems normally distributed that 2nd and 3rd quartiles have similar size. However, the histogram and skewness seems to tell us that the data is skewed toward left. But, the test of normality tells us that it is normally distributed. But it is hard determine because there are very small sample size of n=12.



* 1. State the hypotheses in terms of the compact and augmented model (i.e., MODEL A and MODEL C).

Model A: 

Model C: 

Null hypothesis: 

* 1. Use SPSS or R to calculate the relevant sums of squares. To do this, you will need to calculate a new variable that subtracts 100 from the IQ variable.

df$iq\_new <- df$iq-100

df$zero <- rep(0,88)

df2 <- df[which(df$repeat.=='1'),]

analysis<- lm (df2$iq\_new~df2$zero, data= df2)

anova(analysis)

summary(analysis)

Analysis of Variance Table

Response: df2$iq\_new

Df Sum Sq Mean Sq F value Pr(>F)

Residuals 11 **768.92** 69.902

> summary(analysis)

* 1. Create an ANOVA summary table that summarizes your results.

**Tests of Between-Subjects Effects**

Dependent Variable: iq\_100

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
| Corrected Model | .000(a) | 0 | . | . | . |
| Intercept | 954.083 | 1 | 954.083 | 13.649 | .004 |
| Error | 768.917 | 11 | 69.902 |  |  |
| Total | 1723.000 | 12 |  |  |  |
| Corrected Total | 768.917 | 11 |  |  |  |

a R Squared = .000 (Adjusted R Squared = .000)

Analysis of Variance Table

Response: df2$iq\_new

Df Sum Sq Mean Sq F value Pr(>F)

Residuals 11 768.92 69.902

* 1. Use SPSS or R to conduct a One Sample T Test to test the hypothesis above (Set Test Value = 100). Verify that your results are consistent with those above.

One Sample t-test

data: df2$iq

t = -3.6945, df = 11, p-value = 0.003536

alternative hypothesis: true mean is not equal to 100

95 percent confidence interval:

85.77119 96.39548

sample estimates:

mean of x

91.08333

* 1. Write a one-sentence summary of your findings that would be suitable for a results section of an APA-style journal. This should include means, a test result, and a measure of effect size.

A one-sample *t*-test found significant difference between IQ of general population (M=100) and those children who had sign of ADD having to repeat at least one grade (M = 91.0833) , *t*(11) = -3.6945, *p* = .003536.