**Biostatistics- BIOM 8110**

Final Take Home Exam

Submitted to

Dr. ERNO LINDNER

Professor, Biomedical Engineering

Prepared by

Md Juber Rahman

U 00617285

EECE, University of Memphis

Problem 1:

In a recent paper James et. al. compared two theoretical models to calculate the current in electrochemical experiments which can be used for the determination of diffusion coefficients. They results provided vs. data points where (peak current) is linearly increasing with (the square root of the scan rate). The slopes of the fitted lines as well as the intercepts of the fitted lines can be used for the determination of diffusion coefficients in electrochemical experiments.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (Peak currents) (A) | | |  | |
| Model -1 | Model - 2 | (V/s) | |
| 3.99E-11 | 3.89E-11 | 0.00316 | |
| 4.019E-11 | 3.987E-11 | 0.01 | |
| 4.449E-11 | 4.32E-11 | 0.0316 | |
| 4.75E-11 | 4.57E-11 | 0.0447 | |
| 4.99E-11 | 4.77E-11 | 0.0547 | |
| 5.59E-11 | 5.28E-11 | 0.0774 | |
| 6.199E-11 | 5.83E-11 | 0.1 | |
| 7.36E-11 | 6.92E-11 | 0.141 | |
| 8.26E-11 | 7.80E-11 | 0.173 | |
| 1.036E-10 | 9.830E-11 | 0.245 | |
| 1.245E-10 | 1.1900E-10 | 0.316 | |
| 1.632E-10 | 1.576E-10 | 0.447 | |

The authors were interested whether the slopes and intercepts of the lines fitted to the data points, generated by the two models, can be considered the same. In other words, the authors were interested whether the differences in the slopes and intercepts are statistically significant.

* Formulate the null hypothesis reflecting the questions of the authors.
* Answer these questions with p values reflecting your statistical analysis.

**Answer to problem 1:**

****

**Fig.1: Two models showing relationship of square root of scan rate and current**

At first, we test whether the slopes for two regression models are equal, i.e. we test the following null and alternative hypotheses:

H0: *β*1*= β*2

HA: *β*1*≠ β*2

**Assumptions for *t* test:**

1. The data points are independent of each other.
2. For any given value of the independent variable, the possible values of the dependent variable are distributed normally.
3. The mean of the population of the dependent variable at a given value of the independent variable increases ( or decreases ) linearly as the independent variable increases.
4. That data are homoscedastic (have the same standard deviation in different groups).

Degrees of freedom, v = n1+n2- 4= 12+12- 4= 20

With v=20 and ɑ=0.05 , tcritical = ±2.09 (two-tailed)

From MS-Excel(attached),

b1= 2.82081E-10

b2= 2.69239E-10

Now, tcalculated:

****

Where, 

Using Minitab software,

tcalculated= -1.92

P-value= 0.069

Since, | tcalculated|<|tcritical| and p>.05 we do not have enough evidence to reject the null hypothesis that two slopes are equal, i.e. the difference between the slopes is not statistically significant (P=0.069).

Now, we test whether the intercepts for two regression models are equal, i.e. we test the following null and alternative hypotheses:

H0: *a1=a2*

HA: *a1≠a2*

Degrees of freedom, v = n1+n2- 4= 12+12- 4= 20

With v=20 and ɑ=0.05 , tcritical = ±2.09 (two-tailed)

From MS-Excel,

a1= 3.51907E-11

a2= 3.38139E-11

Now, tcalculated:



Where, 

Using Minitab software,

tcalculated = -1.09

P-value= 0.290

**Since, | tcalculated|<|tcritical| and p>.05 we do not have enough evidence to reject the null hypothesis that two intercepts are equal, i.e. the difference between the intercepts is not statistically significant (P=0.290).**

**Problem 2:**

The following data represent the systolic blood pressure readings on 30 females preselected by age, covering the ages between 40 and 85.

|  |  |  |
| --- | --- | --- |
| Patient # | Age (years) | Systolic Blood pressure  mmHg |
| 1 | 52 | 114 |
| 2 | 58 | 120 |
| 3 | 76 | 125 |
| 4 | 74 | 132 |
| 5 | 64 | 115 |
| 6 | 72 | 122 |
| 7 | 82 | 140 |
| 8 | 66 | 112 |
| 9 | 78 | 122 |
| 10 | 81 | 130 |
| 11 | 68 | 117 |
| 12 | 56 | 112 |
| 13 | 88 | 138 |
| 14 | 72 | 121 |
| 15 | 68 | 122 |
| 16 | 62 | 119 |
| 17 | 77 | 135 |
| 18 | 90 | 130 |
| 19 | 72 | 115 |
| 20 | 74 | 124 |
| 21 | 74 | 134 |
| 22 | 63 | 114 |
| 23 | 86 | 125 |
| 24 | 92 | 135 |
| 25 | 80 | 124 |
| 26 | 76 | 124 |
| 27 | 79 | 128 |
| 28 | 82 | 122 |
| 29 | 78 | 123 |
| 30 | 74 | 123 |

1. Plot the data along with the fitted regression line, the 95 % confidence interval of the line of means and the 95 % confidence interval for the observations.
2. Provide the slope, the standard error of the slope, the intercept, and the standard error of the intercept with adequate number of significant figures.
3. Provide the slope and the intercept values of the regression line with their 95 % confidence interval.
4. Provide the correlation coefficient and the coefficient of the determination.
5. Test the hypothesis that the slope = 0 (Let α = 0.05). What do you conclude? What is the p value?
6. Test the hypothesis that the correlation coefficient = 0 (Let α = 0.05). What do you conclude? What is the p value?
7. Based on the regression line provide the estimated systolic blood pressures in mmHg with its confidence interval for patients 70 and 88 years old.
8. Based on the regression line provide the 95% confidence interval of systolic blood pressures that one could measure for patients 70 and 88 years old
9. Assuming that you measure systolic blood pressure of **125 mmHg** for a female give your best estimate for her age. Provide also the range of your estimate and the minimum and maximum age values considering the confidence interval of the observations (systolic blood pressures).

Based on your answer in #9 comment whether one can use the systolic blood pressure to estimate the age of a female.

**Answer to Problem 2.**

First, we plot the dependent variable,

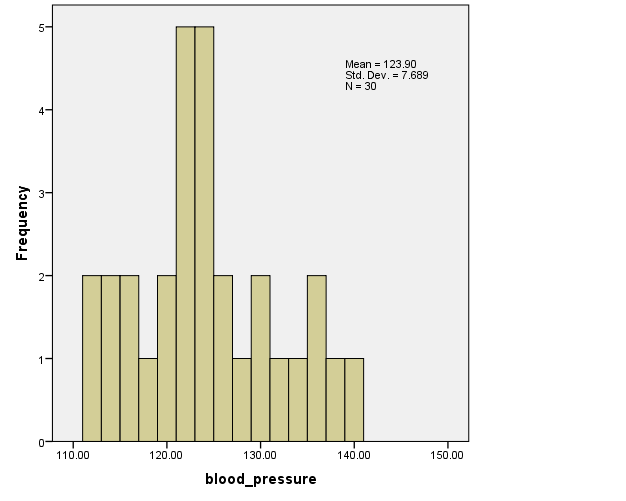
FF

Fig.2 Histogram of Blood Pressure (Dependent Variable)

Histogram shows approximate normal distribution; let us now look at the correlation between the two variables

Table I. Correlation of blood pressure with age

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Correlations** | | | | **Decision** |
|  | | age | blood\_pressure |
| age | Pearson Correlation | 1 | .749 | Strong correlation |
| Sig. (2-tailed) |  | .000 |
| N | 30 | 30 |

Using sigma plot software, we get,

The p-value for normality test is 0.116 and the data passed normality test

Also, the data passed constant variance test with p= 0.579

**So, we can use linear regression for this data with power 0.999 (ɑ=0.05)**

1.

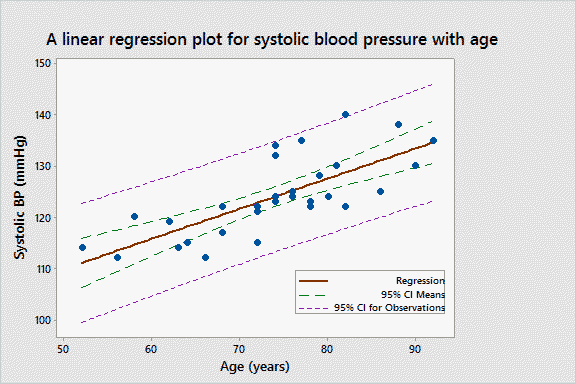


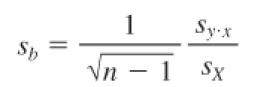
Fig 3. Linear regression plot with 95% confidence interval of the lines of means and 95% confidence intervals of the lines of observations for systolic blood pressure of females with age.

2. Slope,

[](../../Erno/Classes/Biostatistics/Lectures/Uploaded/# 17-chapter 8/Table-8-1.xls)

Slope = **0.588** [Using MS Excel & SPSS software]

The standard error of the slope



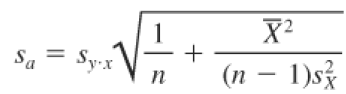
The standard error of the slope = **0.098** [Using MS Excel & SPSS software]

Intercept,

[](../../Erno/Classes/Biostatistics/Lectures/Uploaded/# 17-chapter 8/Table-8-1.xls)

Intercept = **80.503** [Using SPSS software]

The standard error of the intercept,



The standard error of the intercept= **7.314** [Using MS Excel & SPSS software]

3. Confidence interval of the slope=



Using SPSS software,

95% confidence interval for slope **(0.387, 0.789)**

Confidence interval of the intercept=



Using SPSS software,

95% confidence interval for intercept **(65.521, 95.484)**

4. Using SPSS software,

Correlation coefficient = **0.749**

Coefficient of the determination= **0.561**

5. Test the hypothesis that the slope = 0

H0 : slope =0

HA : slope≠ 0

Assumptions for t test:

1. The mean of the population of the dependent variable at a given value of the independent variable increases ( or decreases ) linearly as the independent variable increases.
2. For any given value of the independent variable, the possible values of the dependent variable are distributed normally.
3. The standard deviation of population of the dependent variable about its mean at any given value of the independent variable is the same for all values of the independent variable.

Tcritical= 2.05 with degrees of freedom =30-2=28

Using SPSS software

Tcalculated= 5.984

Tcalculated > Tcritical

p- Value≈ 0.00

**so, we have enough evidence to reject the null hypothesis (P=0.0)**

6. Test the hypothesis that the correlation coefficient = 0

H0 : r =0

HA : r ≠ 0

Assumptions for t test:

1. The mean of the population of the dependent variable at a given value of the independent variable increases ( or decreases ) linearly as the independent variable increases.
2. For any given value of the independent variable, the possible values of the dependent variable are distributed normally.
3. The standard deviation of population of the dependent variable about its mean at any given value of the independent variable is the same for all values of the independent variable.

Tcritical= 2.05 with degrees of freedom =30-2=28

Using SPSS,

T calculated= 11.007

Tcalculated> Tcritical

p-value ≈ 0

**We have, enough evidence to reject the null hypothesis (P=0)**

7. The regression line is, Systolic BP= 80.5 + 0.588 \* Age

So, for age= 70, Estimated Systolic BP= 80.5 + 0.588 \* 70 = **121.7**

95% confidence interval = tɑSŷ, tɑSŷ )

= (121.66-1.96x 5.18, 121.66+1.96x 5.18)

= (**110.84, 132.49**) mm of Hg [using Minitab software]

For age =88, Estimated Systolic BP= 80.5 + 0.588 \* 88 = **132.3**

95% confidence interval = tɑSŷ, tɑSŷ )

= (132.25-1.96x 5.75, 132.25+1.96x 5.75)

= (**121.08, 143.41**) mm of Hg [using Minitab software]

8.

95% confidence interval of systolic blood pressures that one could measure

For patients of 70 years old= **(119.6, 123.8)** mm of Hg [using Minitab software]

For patients of 88 years old= **(128.8, 135.7)** mm of Hg [using Minitab software]

9.

For 125 mmHg, estimate of age= (blood\_pressure-80.5) / 0.588 = 75.68 ≈ **76**

For 125 mmHg,

Confidence interval of observations = **(114.2, 135.8)** mm Hg [using Minitab]

Estimate of minimum age= 57

Estimate of maximum age= 94

Range= 94-57 = **37**

From the step 9 we see that though one can use the systolic blood pressure to estimate the age of a female the range is comparatively big.

**Problem 3**

A new approach to prenatal care is proposed for pregnant women living in rural community. The new program involves in home visits during the pregnancy in addition to the usual scheduled office visits. A pilot randomized trial with 15 pregnant women is designed to evaluate whether women who participate in the program deliver healthier babies than women receiving the usual care. The outcome is the APGAR score measured 5 minutes after birth. The APGAR scores range from 0 to 10, where a score of 7 or larger is considered normal, 4-6 is low, 0-3 is critically low. The data are shown below:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Usual care | 8 | 7 | 6 | 2 | 5 | 8 | 7 | 3 |
| New program | 9 | 8 | 7 | 8 | 10 | 9 | 6 |  |

Is there statistical evidence of a difference in the APGAR scores in women receiving the new care versus the usual care?

1. Set up the null and alternative hypothesis.
2. Select the appropriate statistics
3. Set up the decision rule (the critical value of the test statistics)
4. Compute the test statistics
5. Provide the conclusion

**Answer to problem. 3**

From the given data, we get the following descriptive statistics:

Table II. Descriptive statistics for APGAR score of usual care and new program

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Approach | Sample Size | Mean | Median | Standard Deviation |
| Usual Care | 8 | 5.75 | 6.50 | 2.25 |
| New Program | 7 | 8.14 | 8.0 | 1.34 |

From the descriptive statistics, we see that in the usual Care group mean and median value differs from each other.

To have a better understanding the data has been visualized by Boxplot, Histogram, Q-Q plot below:

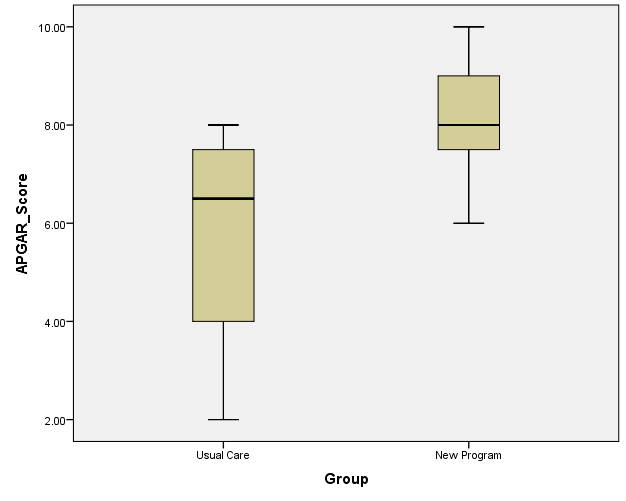


Fig. 4: Boxplot of APGAR score for usual care and New program

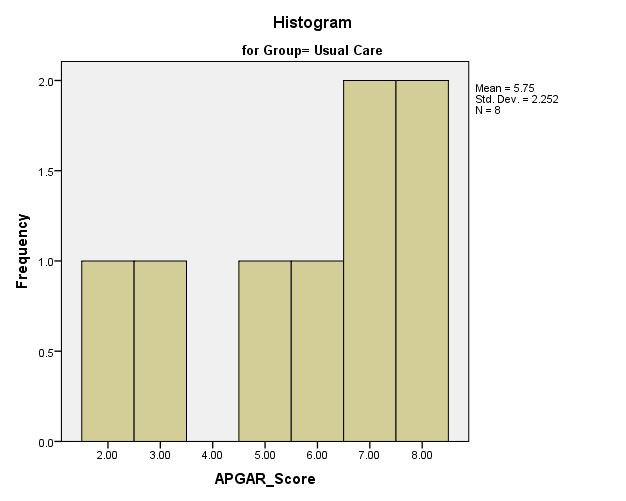


Fig. 5: Histogram of APGAR score for usual care

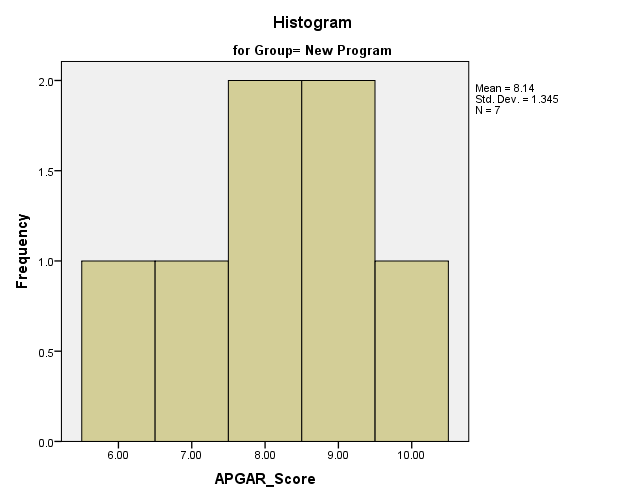


Fig. 6: Histogram of APGAR score for New Program

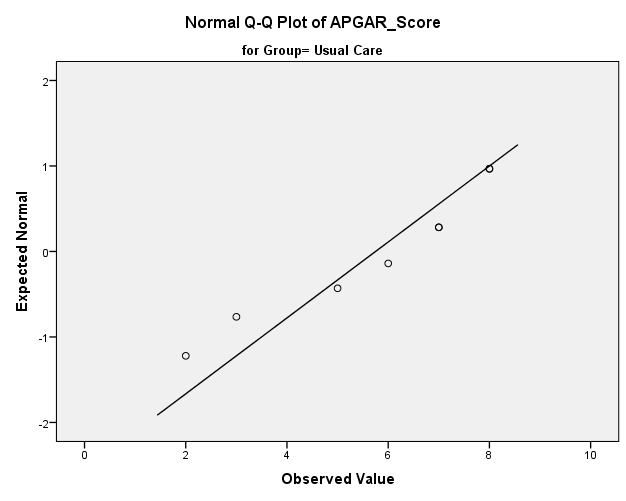


Fig. 7: Q-Q plot of APGAR score for Usual Care

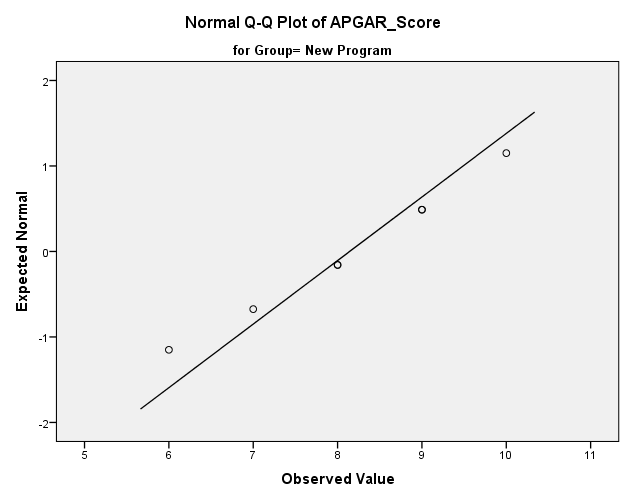


Fig. 8: Q-Q plot of APGAR score for New Program

**The boxplot, Histogram and Q-Q plot all shows that data in the usual care group is left skewed but data in the new approach is symmetric.**

**Normality Test:**

Table III. Result of Normality test

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tests of Normality** | | | | | | | | Decision |
|  | Group | Kolmogorov-Smirnov | | | Shapiro-Wilk | | |
| Statistic | df | Sig. | Statistic | df | Sig. |  |
| APGAR\_Score | Usual Care | .211 | 8 | .200 | .889 | 8 | .230 | Passed |
| New Program | .172 | 7 | .200 | .967 | 7 | .873 | Passed |

df- degrees of freedom

The Normality test was done using SPSS software

**Equal Variance Test:**

Using SPSS software we get:

Table IV. Result of Equal Variance test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Levene's Test for Equality of Variances | | Decision |
| F | Sig. |
| APGAR\_Score | Equal variances assumed | 2.381 | .147 | Passed |

1.

H0 : µusual = µNew ; The mean in the two groups are equal

HA : µusual ≠ µNew ; The mean in the two groups are not equal

2. Although the data in the usual care group is left skewed, data in both the group passed the Normality test and equal variance test. So, I have decided to do independent samples *t* test (two-tailed) for two groups.

Assumptions for t test:

1. Each sample is independent.
2. Each is randomly selected from the population studied.
3. The sampling distribution is normal.
4. Population variances are equal.

3. Tcritical:

Degrees of freedom = n1+n2-2= 8+7-2= 13

Significance level, ɑ =0.05

From the *t*  distribution we get tcritical  = ± 2.16 (two-tailed)

4.

tcalculated  = ;

s is the pooled variance, n1= gr. 1 sample size, n2= gr. 2 sample size

tcalculated = -2.448 [using SPSS software]

P-Value: 0.029 (two tailed)

**5. Conclusion:**

**Since, |tcalculated | > |tcritical | and p < 0.05 we have enough evidence to reject the null hypothesis and claim that, there is a statistically significant difference between the APGAR score of usual care and new program (P = 0.029).**

Since, small sample always pass normality test we can also perform non-parametric test considering that data is heavily skewed, so the Man-Whitney test has been performed below:

Table V. Ranks for Man-Whitney U test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ranks** | | | | |
|  | Group | N | Mean Rank | Sum of Ranks |
| APGAR\_score | Usual Care | 8 | 5.69 | 45.50 |
| New Program | 7 | 10.64 | 74.50 |
| Total | 15 |  |  |

Table VI. Results for Man-Whitney U test

|  |  |
| --- | --- |
| **Test Statistics** | |
|  | APGAR\_score |
| Mann-Whitney U | 9.500 |
| Wilcoxon W | 45.500 |
| Z | -2.172 |
| Sig. (2-tailed) | **.030** |

**Hence, Man-Whitney U test also reveals that the difference in the APGAR score for the usual care and new approach is statistically significant (P=.03).**