Scientific Methods for Health Sciences: Fundamentals (HS550): Fall 2014 <u>http://www.socr.umich.edu/people/dinov/2014/Fall/HS550/</u> Homework 3¹ Solutions

Problem 1

Do a two-sample test between the MMSE scores of the two groups of patients defined by Group0 and Group 1.

$$t_{0} = \frac{\widehat{\mu_{3}} - \widehat{\mu_{4}}}{SE(\widehat{\mu_{3}} - \widehat{\mu_{4}})} = \frac{\widehat{\mu} - \widehat{\mu_{4}}}{\sqrt{\frac{s_{0}^{2}}{n_{0}} + \frac{s_{1}^{2}}{n_{1}}}} = \frac{27.48069 - 26.6625}{\sqrt{\frac{2.539484^{2}}{48} + \frac{2.609656^{2}}{240}}} = 3.4558$$
$$df = \frac{\left(\frac{s_{0}^{2}}{n_{0}} + \frac{s_{1}^{2}}{n_{1}}\right)^{2}}{\left(\frac{s_{0}^{2}}{n_{0}}\right)^{2}} = 470.9973$$
$$\frac{\left(\frac{s_{0}^{2}}{n_{0}}\right)^{2}}{\frac{s_{0}^{2}}{n_{0}} - 1} + \frac{\left(\frac{s_{1}^{2}}{n_{1}}\right)^{2}}{n_{1} - 1}}$$

The test statistic is 3.4558 and generates a p-value of 0.0005983. We have enough evidence to reject the null hypothesis of H_0 : $\mu_0 = \mu_1$ at 5% level of significance, and claim that there are significant difference between the MMSE scores of the two groups of patients of group 0 and group 1.

T-test result (using R-script below):

Welch Two Sample t-test

```
data: g0$MMSCORE and g1$MMSCORE
t = 3.4558, df = 470.997, p-value = 0.0005983
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.3529543 1.2834191
sample estimates:
mean of x mean of y
27.48069 26.66250
```

RCODE:

```
# problem 1
# Save the data (http://wiki.socr.umich.edu/index.php/SOCR_Data_AD_BiomedBigMetadata) in a
local file /data_folder/Homework3.csv or C:\data_folder\Homework3.csv
biom <- read.csv('C:\\data_folder\\Homework3.csv'')
summary(biom)
attach(biom)
g0 <- subset(biom,GDTOTAL==0)
mu0 <- mean(g0$MMSCORE)
n0 <- dim(g0)[1]
s0 <- sd(g0$MMSCORE)
g1 <- subset(biom,GDTOTAL==1)
mu1 <- mean(g1$MMSCORE)
n1 <- dim(g1)[1]</pre>
```

¹ http://www.socr.umich.edu/people/dinov/2014/Fall/HS550/HWs.html http://www.socr.umich.edu/people/dinov/2014/Fall/HS550/

s1 <- sd(g1\$MMSCORE)
$df <- (s0^{2}/n0+s1^{2}/n1)^{2}/((s0^{2}/n0)^{2}/(n0-1)+(s1^{2}/n1)^{2}/(n1-1))$
se <- sqrt(s0^2/n0+s1^2/n1)
t <- (mu0-mu1)/se ## 3.455795
p <- 2*pt(-abs(t),df=df) ## 0.0005983385
Or use the t.test function
t.test(g0\$MMSCORE,g1\$MMSCORE)
weich iwo Sample t-test
data: g0\$MMSCORE and g1\$MMSCORE t = 3.4558, df = 470.997, p-value = 0.0005983
Using SOCR Two independent sample t-test (pooled)
http://socr.ucla.edu/htmls/SOCR Analyses.html
Step 1: Input the data of MMSCORES for Group 0 and Group 1, the data can be generated in R:
SOCR data
write.csv(g0\$MMSCORE)
write.csv(g1\$MMSCORE)

OCR Analyses	CALCUL	ATE EXA	MPLE 1	EXAMPLE 2	EXAMPLE	3 EXAN	MPLE 4 E	XAMPLE 5	5 CLEAR R	ANDOM EXA	AMPLE			
Two Independent Sample T Test (Pooled)							1	DATA	MAPPING	RESULT				
ABOUT HELP SNAPSHOT	MMG0	MMG1	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C
	29	28												+
COPY PASTE FILE OPEN	25	20				_						_		+
	28	21	_					_				_		+
Result RoundOff: 0.001 0.00001 Al	29	29			-			-				-	_	+
	20	20	_	-						-	-	_		+
	29	2.9								-	8	-		+
	29	29	-	-	-			_	10			-		+
	30	27	-		-		-			-		-	-	+
	26	30				-	-				8	-		+
	30	27										-		+
	30	24					-							+
	27	22		- C			1							T
	28	30					1							T
	28	29										-		T
	30	21				1								
	30	30										1		
	24	25				1	1		1					
	30	24				1	1		1					
	29	30	_									_		+
	25	29				_	j.			_				+
	20	29	-			-	- U			_		-		+
	30	22	-	-		-	-			-		-	_	+
	61	50												

Step 2: Mapping

CALCULATE EXAMPLE 1 EXAMPLE 2 EXAMPLE 3 EXAMPLE 4 EXAMPLE 5 CLEAR RANDOM EXAMPLE]
DATA MAPPING RESULT	
C3 VARIALBE 1	MMG0
C4 ADD	
CS REMOVE	
C7	
C8	
C11	
C12	
C15	
C16	
VARIABLE 2	MMG1
ADD	
KEMOVE	

Step 3: Calculate Step 4: Check t-test result:

ALCULATE	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4	EXAMPLE 5 CLEAR RANDOM EXAMPLE
					DATA MAPPING RESULT
			Result of Two Variable 1 = M Sample Size = Sample Mean Sample Variar Sample SD = 2 Variable 2 = M Sample Size = Sample Mean Sample Variar Sample D = 2	Independent : MG0 233 = 27.481 cce = 6.449 2.539 MG1 240 = 26.662 cce = 6.810 0.610	Sample T-Test:
			Degrees of Fro Pooled Sample Cooled Sample Statistics (P Dne-Sided P-V Wo-Sided P-V	eedom = 471 e Variance = (e SD = 2.575 ooled) = -3.45 'alue (Pooled) 'alue (Pooled)	6.632 54 () = .000 () = .001

The result is very close to the one we got in R and the conclusion is also consistent, we reject the null hypothesis of no significant difference in the MMSCORE score in group 0 and group 1 at 5% level of significance and claim that the MMSCORE scores in group 0 and group 1 differ significantly.

Problem 2

Do a test on the proportions of patients with {GDTOTAL >0} in two groups Group3 (CDGLOBAL=1) vs. Group4 (CDGLOBAL=0). Null hypothesis $P_3 = P_4$. Test statistics (t-test) for proportions in several groups without continuity correction (see Yates' continuity correction for details):

$$Z_{0} = \frac{\widehat{p_{3}} - \widehat{p_{4}}}{SE(\widehat{p_{3}} - \widehat{p_{4}})} = \frac{\widehat{p_{3}} - \widehat{p_{4}}}{\sqrt{\frac{\widehat{p_{3}}(1 - \widehat{p_{3}})}{n_{3}} + \frac{\widehat{p_{4}}(1 - \widehat{p_{4}})}{n_{4}}}}{\frac{39}{48} - \frac{471}{695}}$$
$$= \frac{\frac{\sqrt{0.8125 * (1 - 0.8125)}}{\sqrt{\frac{0.8125 * (1 - 0.8125)}{48} + \frac{0.6776978 * (1 - 0.6776978)}{695}}}}{Z \sim N(0,1)} = 2.282454$$

p value = 0.02246257, we have enough evidence to reject the null hypothesis of equal proportion at 5% level of significance and claim that there are significant difference between the two proportions, that is the proportion of patients with GDTOTAL >0 in group3 where CDGLOBAL=1 differs significantly from that proportion in group 4 where CDGLOBAL=0.



p value can also be observed from this, $value = P(Z > Z_{0.025}) + P(Z < Z_{0.975}) = 0.11231 + 0.11231 \approx 0.22462.$

```
RCODE:
```

```
## problem 2
g3 <- subset(biom,CDGLOBAL==1)
n3 <- dim(g3)[1]
p3 <- sum(as.numeric(g3$GDTOTAL>0))/n3
g4 <- subset(biom,CDGLOBAL==0)
n4 <- dim(g4)[1]
p4 <- sum(as.numeric(g4$GDTOTAL>0))/n4
n4 <- dim(g4)[1]
se2 <- sqrt(p3*(1-p3)/n3+p4*(1-p4)/n4)
z2 <- (p3-p4)/se2
p2 <- 2*(1-pnorm(z2,0,1))</pre>
```

Using SOCR: Chi-Square Test Contingency Table: http://socr.ucla.edu/htmls/SOCR Analyses.html Step 1: Input the contingency table

		Mas
SOCR Analyses	CALCULATE CLEAR	
Chi-Square Test Contingency Table	INPUT RESULT	
ABOUT HELP SNA	Select Significance Level: 0.05 ‡	
COPY PASTE FILE	Select Number of Rows: 2 \$	
Result RoundOff: 💿 0.001 🗌 0.00	Select Number of Columns: 2 🛟	
	GDTOTAL > 0 GDTOTAL = 0	0
	Group 3 39 9	
	Group 4 471 224	
A		

Step 2: Calculate Step 3: Check the result:

CALCULATE CLEAR					
			INPUT	RESULT	
	Results of C Number of R Number of C	hi-Square Test fo lows = 2 columns = 2	r Independent o	r Homogeneity	
		GDTOTAL > 0	GDTOTAL = 0	Row Total	
	Group 3	39.0 (32.948)	9.0 (15.052)	48	
	Group 4	471.0 (477.05	2)	224.0 (217.948)	695
	Col Total	510	233	743	
	Degrees of F	Freedom = 1			
	Pearson Chi	-Square Statistic:	s = 3.790		
	P-Value = 0.0	05154803681843	456		

The chi-square test has test statistics = 3.790, which is slightly smaller compared to the threshold of X^2 =3.84. We don't have enough evidence to reject the null hypothesis of no significant difference in the proportions of GDTOTAL > 0 between group 3 and group 4 at 5% level of significance. The result is slightly different compared to the result concluded in R.

Problem 3

For the MCI-to-AD Converters (DX_Conversion) variable, the summary of the dataset suggests that there 1 missing value, 735 records with DX_Conversion=0 (No conversion or Reversion to NL/MCI), 7 records with DX_Conversion=1, that is has Conversion to NL/MCI, and 1 record with DX_Conversion =2, that is has Reversion to NL/MCI. To compare their performances in MMSCORE scores, I choose to compare two groups of Group with no conversion or Reversion to NL/MCI and group with only Conversion as well as http://www.socr.umich.edu/people/dinov/2014/Fall/HS550/ 5

group with no conversion or reversion and group with either Conversion or Reversion and the t test result on H_0 : $\mu_0 = \mu_1$ and another t test H_0 : $\mu_0 = \mu_{12}$ are as following:

a. No vs. Conversion only

Welch Two Sample t-test

data: DX0\$MMSCORE and DX1\$MMSCORE t = 0.5758, df = 6.278, p-value = 0.5848 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -1.211888 1.968351 sample estimates: mean of x mean of y 26.80680 26.42857

Conclusion: there aren't any significant difference in the MMSCORE for group with no Conversion or Reversion to NL/MCI at 5% level of significance.

b. No vs. Conversion or Reversion:

Welch Two Sample t-test

data: DX0\$MMSCORE and DX12\$MMSCORE t = -0.0941, df = 7.265, p-value = 0.9276 alternative hypothesis: true difference in means is not equal to 0 95 percent confidence interval: -1.769420 1.633026 sample estimates: mean of x mean of y 26.8068 26.8750

Conclusion: there aren't any significant difference in MMSCORE score between group with no Conversion or Reversion and group with either Conversion or Reversion to NL/MCI at 5% level of significance.

Hence, the fact whether patients have Conversion or Reversion to NL/MCI didn't have any significant influence on the MMSCORE scores. They aren't significantly associated.

RCODE:

problem 3
summary(DX_Conversion)
##. 0 1 2
1 735 7 1
data3 <- subset(biom,DX_Conversion!='.')
summary(data3\$DX_Conversion)
DX0 <- subset(data3,data3\$DX_Conversion==0)
DX1 <- subset(data3,data3\$DX_Conversion==1)
DX12 <- subset(data3,data3\$DX_Conversion!=0)
t.test(DX0\$MMSCORE,DX1\$MMSCORE)
t.test(DX0\$MMSCORE,DX12\$MMSCORE)</pre>

Using SOCR: Two Independent Sample T-test Pooled: <u>http://socr.ucla.edu/htmls/SOCR_Analyses.html</u> Step 1: Input data:

					and the second			1.1	110	5
SOCR Analyses	CALCULATE	EXAMPLE 1 EXAMPLE 2	EXAMPL	E 3 EXA	AMPLE 4	EXAMPL	E5 C	LEAR	RANDOM EX.	AMPLE
Two Independent Sample T Test (Pooled)						DATA	MAP	PING	RESULT)
ABOUT HELP SNAPSHOT	DX_CONVERS	ION=0 DX_Conversion=1	C3	C4	C5	C6	C7	C8	C9	C10
COPY PASTE FILE OPEN	20	24								
	29	26								
Result RoundOff: • 0.001 0.00001 All	28	28						_		
	29	25								
	29									
	30									
	30									
	25									

Step 2: Mapping

	Shanna a shi ka shi Na shi ka shi		<u> </u>					HIG'S .		
×	CALCULATE	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4	EXAMPLE 5	CLEAR	RANDOM EXAMPLE]	
					(DATA N	IAPPING	RESULT		
	C3 C4 C5 C6 C7 C8 C9 C10 C11				VARIA ADD REMO	ALBE 1				DX_CONVERSION=0
	C12 C13 C14 C15 C16									
د					VARIA	ABLE 2				DX_Conversion=1
					REMO	VE				

Step 3: Calculate

Step 4: Check result:



Similar for comparison of MMSCORE score between group with DX_Conversion =0 and group with DX_Conversion \neq 0 (which has one more point with DX_Conversion=2 and MMSCORE = 30 compared to the group with DX_Conversion=1): Result:

CALCULATE	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4	EXAMPLE 5	CLEAR	RANDOM EXAMPLE
					DATA	APPING	RESULT
		F	lesult of Two	Independent S	Sample T-Test	:	
		V	ariable 1 = D) ample Size =	X_CONVERSI 735	ON=0		
		5	ample Mean ample Varian	= 26.807 ice = 7.107			
		s v	ariable 2 = D	 X Conversion	=1		
		S	ample Size = ample Mean	8 = 26.875			
		5	ample Varian ample SD = 2	ice = 4.125 2.031			
		C P	egrees of Fre	edom = 741 Variance = 7	.079		
		F	ooled Sample -Statistics (P	e SD = 2.661 ooled) = .072			
			wo-Sided P-V	alue (Pooled) alue (Pooled)	= .471 = .943		

The result is consistent with the conclusion from R, we reject the null hypothesis at 5% level of significance and claim that whether patients have Conversion or Reversion to NL/MCI didn't have any significant influence on the MMSCORE scores. They aren't significantly associated.

Problem 4

Do a Chi-square test on the standard deviation of MMSCORE with the null hypothesis of H_0 : $\sigma_0^2 = 2.5^2$. The test statistic ~ $X_{df=n-1}^2$

$$X_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(744-1)*2.657866^2}{2.5^2} = 839.7985.$$

$$df = n - 1 = 743$$

The corresponding p value is 0.007617524, so we reject the null hypothesis at 5% level of significance and claim that the standard deviation of MMSCORE scores are significantly different from 2.5.

To check on the p-value of the chi-square test: http://socr.ucla.edu/htmls/SOCR Distributions.html



By selecting the degree of freedom of 743 and roughly a test score at around 839.7985 gives a p.value at around 0.0076175 (*p. value* = $P(X_{df=n-1}^2 > X_o^2) = 0.0076175$), which suggest that we have enough evidence to reject the null hypothesis of $\sigma_0^2 = 2.5^2$ at 5% level of significance and claim that the standard deviation of the MMSCORE score is significantly different from 2.5.

RCODE:

MM.std <- sd(MMSCORE) n <- length(MMSCORE) chi.test <- (n-1)*MM.std^2/2.5^2 p.value <- pchisq(chi.test,df=n-1,lower.tail=F)

Problem 5

Correlation between systolic and diastolic blood pressure within group 3 and group 4 are 0.4052743 and 0.439872 respectively.

Using Fisher's transformation to test for comparing the two correlations using Normal distribution on null hypothesis $H_0: r_1 = r_2$, transform the two correlations into $r_{11} = \frac{1}{2} \ln || \frac{1+r_1}{1-r_1} ||$ and $r_{22} = \frac{1}{2} \ln || \frac{1+r_2}{1-r_2} ||$, the test statistic follows a standard normal distribution N(0,1):

$$Z_0 = \frac{r_{11} - r_{22}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = \frac{0.9441442/2 - 0.8598874/2}{\sqrt{\frac{1}{48 - 3} + \frac{1}{695 - 3}}} = -\frac{0.5476845}{2} = -0.2738423.$$

Since $|Z_0| < 1.96$, we don't have enough evidence to reject the null hypothesis of $r_1 = r_2$ at 5% level of significance. The conclusion is that the correlation between systolic and diastolic blood pressure didn't differ significantly, in fact they are very similar to each other.

To check this with the normal distribution: http://socr.ucla.edu/htmls/SOCR_Distributions.html



Note: We reject the null hypothesis if the test statistic falls in the red region of (-1.96, 1.96). To calculate the p-value, we have



RCODE:

```
corr3 <- cor(g3$VSBPSYS,g3$VSBPDIA) # r1=0.4052743
corr4 <- cor(g4$VSBPSYS,g4$VSBPDIA) # r2=0.439872
r11 <- log((1+corr3)/(1-corr3),base=exp(1))
r22 <- log((1+corr4)/(1-corr4),base=exp(1))
z5 <- (r11-r22)/sqrt(1/(n3-3)+1/(n4-3))
p.value <- 2*(1-pnorm(z5,0,1,lower.tail=F))</pre>
```

Problem 6

Fit a simple linear regression of MMSCORE on VSTEMP and Weight_Kg and a brief summary of the model is given as below:

Call: Im(formula = MMSCORE ~ VSTEMP + Weight_Kg) Residuals: Min 1Q Median 3Q Max -8.7497 -1.7785 0.3089 2.1841 3.6242 Coefficients: Estimate Std. Error t value Pr(>/t/) (Intercept) 22.465619 2.548285 8.816 <2e-16 *** VSTEMP 0.093785 0.067553 1.388 0.1655 Weight_Kg 0.012355 0.006523 1.894 0.0586 ---Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.652 on 741 degrees of freedom Multiple R-squared: 0.006825, Adjusted R-squared: 0.004144 F-statistic: 2.546 on 2 and 741 DF, p-value: 0.07909

From the regression model result, we can tell that the model didn't fit very well, p value of the coefficient of VSTEMP is 0.1655, which is not significant at all while the p value of the coefficient of Weight_Kg is 0.0586, which is right above 5% boundary and didn't seem to be significant either. Given the test on the coefficient is a test of trivial slope at the variable H_0 : $\beta = 0$, and we fail to reject the null hypothesis for both cases. Hence, we can conclude there are trivial slope of the regression curve on VSTEMP and Weight_Kg at 5% level of significance.

RCODE:

```
model <- lm(MMSCORE~ VSTEMP+Weight_Kg)
summary(model)</pre>
```

http://www.socr.umich.edu/people/dinov/2014/Fall/HS550/

Using SOCR multiple regression analysis to fit simple linear regression of VSTEMP and Weight_Kg w.r.t. MMSCORE <u>http://www.socr.ucla.edu/htmls/ana/SimpleRegression Analysis.html</u> Step 1: input data

												and the second	
SOCR Analyses	× ×	CALCULAT	E EXAN	IPLE 1 EXA	MPLE 2	EXAMPL	E 3 EX	AMPLE 4	EXAMPL	.E 5 EXA	MPLE 6	CLEAR	
Multiple Regression Analysis \$								DA	ta M	APPING	RESUL	T GR	APH
ABOUT HELP SNAPSHOT		MMSCORE	VSTEMP	Weight_Kg	C4	C5	C6	C7	C8	C9	C10	C11	C12
		2.8	35.7	89									
COPY PASTE FILE OPEN		20	36.6	74									
		29	35.9	88									
		25	36.1	69.39972									
Result RoundOff: 💿 0.001 🕓 0.00001 🕓 All		28	35.6	60									
		24	35.6	68									
		29	36.1	87.08984									
		29	36.1	64.81843									
		21	35.9	69.76259									-
		29	35.7	61.23504									-
		30	36.5	71.66768									
		29	36	83.5									1
		30	36.1	94.61948									+

Step 2: Mapping



Step 3: Calculate

Step 4: Check result:



From the result, we can see that the p-value for estimate of coefficients of Weight_Kg and VSTEMP are 0.059 and 0.165 respectively, which suggest that we don't have enough evidence to reject the null hypothesis of coefficient equals zero for both cases. Hence, the conclusion is also consistent with the result from R that there are trivial slope of the regression curve on VSTEMP and Weight_Kg at 5% level of significance.