

HH/HLST 2300: Statistical Methods in Health Studies
Winter Term Assignment 4

Assigned: Friday March 5, 2021; Due 5PM Friday March 19, 2021

Submit 1 file for Assignment 4: PDF

PDF document name: LASTNAME_FIRSTNAME_WTAssignment4

Submit via eClass

Note1: WT Assignment 4 is worth a total of 50 marks. Therefore, assignments submitted late were deducted 2.5 marks per day ($50 * 0.05 = 2.5$)

Note2: If you did not use the proper assignment naming convention, 4 marks were deducted.

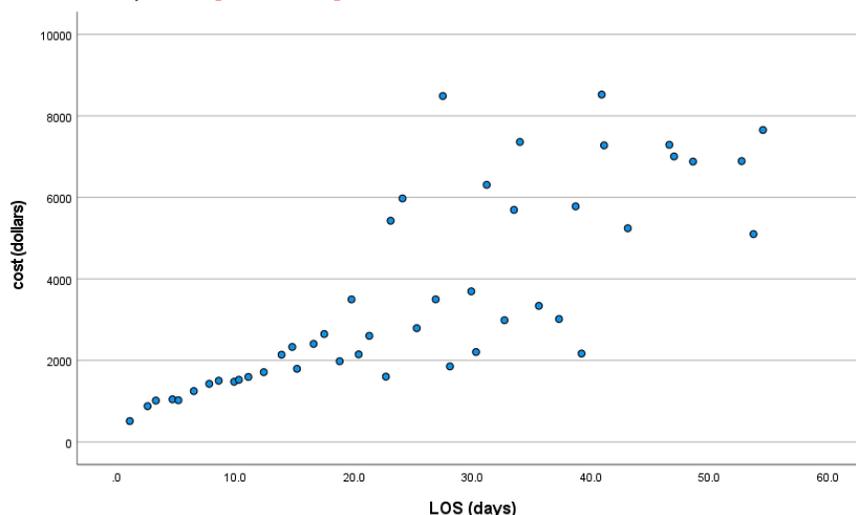
As noted in my eClass announcement on Nov 5, 2020, you will be deducted marks if you submit a file other than a pdf file and if that file is named incorrectly. **For WT Assignment 4, the deduction is 4 marks for incorrect file name. The submission file type has been set up in eClass such that the only accepted file type is PDF.**

Other reminders that you should take care to ensure while completing your assignment:

- Questions involving a data file must be answered using SPSS
- HLST 2300 rounding rules apply unless otherwise stated
- Screenshots of any hand-written work and SPSS must be of high resolution and be pasted upright (not sideways) so that they can be easily read and graded
- Answers to questions must directly follow the question asked – do not change the order of the questions
- If you fail to include the SPSS output instructed of you, you will receive zero for any subsequent questions that rely on that output

1. Excel file: 2300 WT Assignment4 (worksheet: simple_reg) consists of a random sample of 47 inpatient admissions at Hospital ABC. We are provided with hospital length of stay (LOS) in days (range 1.1-54.5 days) and cost of care in dollars. The Chief Medical Officer wants to determine how well hospital LOS can predict cost of care. Conduct the regression analysis to answer the following questions:

An assumption of linear regression is that the relationship between LOS and cost of care shows a linear pattern. To verify this, we will graph a scatterplot with LOS (predictor) on the x-axis and cost (outcome) on the y-axis: **[2 MARKS]**



Some of the typical errors found in Question 1 (assumption for simple linear regression):

- Did not produce a scatterplot of cost and LOS confirming linear trend. [-1 MARK]
- Variables in x- and y-axes incorrect (the predictor variable LOS belongs in the x-axis and the outcome variable cost belongs on the y-axis). [-1 MARK]

a) Is this model useful in predicting cost of care? Copy and paste the relevant SPSS output table(s) when reporting results. **[4 MARKS]**

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	168361806.7	1	168361806.7	78.550	.000 ^b
	Residual	96451265.26	45	2143361.450		
	Total	264813071.9	46			

a. Dependent Variable: cost (dollars)

b. Predictors: (Constant), LOS (days)

The ANOVA shows us that our regression model with the predictor variable LOS significantly predicts cost of care, $F(1, 45) = 78.55$, $p < .001$.

Alternatively, since our regression model only has 1 predictor, we can look at either the t statistic (and its associated p value) or the 95% confidence interval for β in the coefficients table to answer the question of whether our model is useful:

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	406.813	421.741		.965	.340	-442.617	1256.242
	LOS (days)	127.293	14.363	.797	8.863	.000	98.366	156.221

a. Dependent Variable: cost (dollars)

The coefficients table shows us that the predictor variable LOS significantly predicts cost of care, $t = 8.86$ (in my slides I kept 3 decimal places to show that $t^2 = F$, so if reported $t = 8.863$, this is fine), $p < .001$.

The coefficients table also shows us that in the population there is an increase of as little as 98.37 to as much as 156.22 dollars in cost of care for each day increase in LOS. Since the 95% CI does not contain 0 (ie the true value of β is not 0), we can conclude that LOS significantly predicts cost of care.

Some of the typical errors found in Question 1a:

- Did not copy and paste the ANOVA table (if reported F) or the Coefficients table (if reported t or CI), or values in either table are different than those shown. [-2 MARKS]
- Did not report that model was statistically significant. [-1 MARK]
- Did not report correct F-value or t-value. [-0.5 MARKS]
- Did not report correct p-value if reported either F or t. [-0.5 MARKS]
- Did not state that the 95% CI does not contain 0 if reported the 95% CI. [-1 MARK]

b) What proportion of the variation in cost of care can be accounted for by the variation in LOS? Report results fit to the population rather than the sample. Copy and paste the relevant SPSS output table(s). **[3 MARKS]**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.797 ^a	.636	.628	1464.022

a. Predictors: (Constant), LOS (days)

If we generalize to the population, 62.8% of the variation in cost of care can be explained by the variation in LOS.

Some of the typical errors found in Question 1b:

- Did not copy and paste the Model Summary table or values in the table are different than those shown. [-2 MARKS]
- Did not report R square fit to population data (adjusted R square) or incorrectly described R square in relation to variables other than cost and LOS. [-1 MARK]

c) Interpret the coefficient for hospital LOS. Copy and paste the relevant output table(s) when reporting results. **[5 MARKS]**

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	406.813	421.741		.965	.340	-442.617	1256.242
	LOS (days)	127.293	14.363	.797	8.863	.000	98.366	156.221

a. Dependent Variable: cost (dollars)

LOS is a statistically significant predictor of cost of care ($p < .001$) with each additional day associated with an increase in \$127.293 dollars in cost of care.

Some of the typical errors found in Question 1c:

- Did not copy and paste the Coefficients table or values in the table are different than those shown. [-2 MARKS]

- Note: If reported results without producing the corresponding table where these results came from, no marks are earned.
- Did not state that coefficient (127.293 or 127.29) represents the increase in costs of care (outcome variable) with a one-unit increase in LOS (predictor variable). [-1 MARK]
- Did not report that LOS was a statistically significant predictor of costs of care. [-1 MARK]
- Did not report correct p-value. [-1 MARK]

d) State the regression equation and use it to predict the cost of care for a patient who spent 7 days in hospital. Show your work. Copy and paste the relevant SPSS output table(s) when reporting results. **[7 MARKS]**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	406.813	421.741		.965	.340	-442.617	1256.242
	LOS (days)	127.293	14.363	.797	8.863	.000	98.366	156.221

a. Dependent Variable: cost (dollars)

Regression Equation: Predicted cost = 406.813 + 127.293(LOS)

Therefore, the predicted cost of care for a patient who spent 7 days in hospital is:

$$\begin{aligned} \text{Predicted cost} &= 406.813 + 127.293(7) \\ &= 1297.864 \\ &= 1297.86 \text{ (rounded to 2 decimal places)} \end{aligned}$$

Some of the typical errors found in Question 1d:

- Did not copy and paste the Coefficients table or values in the table are different than those shown. [-2 MARKS]
 - Did not state that the outcome variable is predicted cost of care or \hat{y} . [-1 MARK]
 - Note: If reported results without producing the corresponding table where these results came from, no marks are earned.
 - Incorrect b coefficient. [-1 MARK]
 - Incorrect a value. [-1 MARK]
 - Plugged in incorrect value for LOS. [-1 MARK]
 - Incorrect calculation of predicted cost of care. [-1 MARK]
2. Excel file: 2300 WT Assignment4 (worksheet: multiple_reg) consists of a random sample of 657 patients that were admitted to hospital. To predict resource intensity weight (RIW), data has been collected on the following predictors: sex (1 = female, 2 = male), age in years, visit disposition (1 = died, 2 = discharged home, 3 = discharged home with supports, 4 = transferred to long-term care) and hospital length of stay (LOS) in days. Create dummy variables for the categorical predictors with female and discharged home being the reference categories for their respective predictor variables. Conduct the regression analysis to answer the following questions:

- a) Is this model useful in predicting RIW? Copy and paste the relevant SPSS output table(s) when reporting results. **[5 MARKS]**

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1840.645	6	306.774	143.730	.000 ^b
	Residual	1387.350	650	2.134		
	Total	3227.996	656			

a. Dependent Variable: RIW

b. Predictors: (Constant), transferLTC_dummy, Age (years), male_dummy, died_dummy, LOS (days), dchomesupports_dummy

The ANOVA shows us that the predictor variables together significantly predict RIW, $F(6, 650) = 143.73$, $p < .001$.

Some of the typical errors found in Question 2a:

- Did not copy and paste the ANOVA table or values are different than those shown. [-2 MARKS]
- Did not report correct F-value including degrees of freedom. [-1 MARK]
- Did not report correct p-value. [-1 MARK]
- Did not report that model was statistically significant. [-1 MARK]

- b) How much variation in RIW can be explained by our predictor variables? Report results fit to the population rather than the sample. Copy and paste the relevant output table(s) when reporting results. **[3 MARKS]**

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.755 ^a	.570	.566	1.4609534

a. Predictors: (Constant), transferLTC_dummy, Age (years), male_dummy, died_dummy, LOS (days), dchomesupports_dummy

If we generalize to the population, 56.6% of the variation in RIW can be explained by our predictor variables.

Some of the typical errors found in Question 2b:

- Did not copy and paste the Model Summary table. [-1 MARK]
- Values in the Model Summary table are different than those shown. [-1 MARK]
- Did not report correct R square fit to population data (adjusted R square) or incorrectly described R square in relation to an outcome variable other than RIW. [-1 MARK]. Note accepted adj R square as a decimal 0.566 or 0.57.

- c) Interpret the coefficients. Copy and paste the relevant output table(s) when reporting results. **[17 MARKS]**

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.822	.783		1.049	.295	-.716	2.359
	Age (years)	-.007	.009	-.021	-.810	.418	-.025	.010
	LOS (days)	.158	.006	.690	25.121	.000	.146	.171
	male_dummy	.112	.116	.025	.969	.333	-.115	.340
	died_dummy	1.783	.268	.197	6.660	.000	1.258	2.309
	dchomesupports_dummy	-.194	.147	-.044	-1.314	.189	-.483	.096
	transferLTC_dummy	-.473	.192	-.082	-2.468	.014	-.850	-.097

a. Dependent Variable: RIW

Age is not a significant predictor of RIW ($p = .418$) with each additional year of age associated with a decrease in RIW of .007 units (or a change in RIW of -.007 units), holding LOS, sex and discharge disposition constant.

Compared to the reference category of females, males have an increase in RIW by .112 units, holding age, LOS and discharge disposition constant. However, the difference is not statistically significant ($p = .333$).

Compared to the reference category of patients discharged home, patients discharged home with supports had a decrease in RIW of .194 units (or a change in RIW of -.194 units), holding age, LOS and sex constant. However, the difference is not statistically significant ($p = .189$).

LOS is a significant predictor of RIW ($p < .001$) with each additional day of LOS associated with an increase in RIW by .158 units, holding age, sex and discharge disposition constant.

Compared to the reference category of patients discharged home, patients who died had an increase in RIW of 1.783 units, holding age, LOS and sex constant. The difference is statistically significant ($p < .001$).

Compared to the reference category of patients discharged home, patients transferred to long-term care had a decrease in RIW of .473 units (or a change in RIW of -.473 units), holding age, LOS and sex constant. The difference is statistically significant ($p = .014$).

Some of the typical errors found in Question 2c:

- Did not copy and paste the Coefficients table. [-1 MARK]
- Values in the Coefficients table are different than those shown. [-1 MARK]
- Note: If reported results without producing the corresponding table where these results came from, no marks are earned.
- Did not conclude that age is not a statistically significant predictor of RIW. [-1 MARK]
- Did not conclude that males are not statistically significantly different than females in terms of RIW, or that sex is not a statistically significant predictor of RIW. [-1 MARK]

- Did not conclude that patients discharged home with supports are not statistically significantly different than patients discharged home in terms of RIW. [-1 MARK]
 - For LOS, did not state that for one unit increase in LOS, RIW will increase by 0.158 (or 0.16 if rounded) units. [-1 MARK]
 - For LOS, did not state that result adjusts for (or takes into account, or holds constant) the other predictors. [-1 MARK]
 - For LOS, did not report that result was statistically significant. [-1 MARK]
 - For LOS, did not report correct p-value. [-1 MARK]
 - For patients that died, did not state that RIW is increased by 1.783 (or 1.78 if rounded) compared to reference category discharge home. [-1 MARK]
 - For patients that died, did not state that result adjusts for (or takes into account, or holds constant) the other predictors [-1 MARK]
 - For patients that died, did not report that result was statistically significant. [-1 MARK]
 - For patients that died, did not report correct p-value. [-1 MARK]
 - For patients transferred to LTC, did not state that RIW is decreased by .473 (or .47 if rounded), or changed by -.473 or -.47 units, compared to reference category discharge home. [-1 MARK]
 - For patients transferred to LTC, did not state that result adjusts for (or takes into account, or holds constant) the other predictors [-1 MARK]
 - For patients transferred to LTC, did not report that result was statistically significant. [-1 MARK]
 - For patients transferred to LTC, did not report correct p-value. [-1 MARK]
- d) State the regression equation and use it to predict the RIW of male patient aged 75 who was discharged home and had an LOS of 5 days. Show your work. Copy and paste the relevant output table(s) when reporting results. **[4 MARKS]**

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.822	.783		1.049	.295	-.716	2.359
	Age (years)	-.007	.009	-.021	-.810	.418	-.025	.010
	LOS (days)	.158	.006	.690	25.121	.000	.146	.171
	male_dummy	.112	.116	.025	.969	.333	-.115	.340
	died_dummy	1.783	.268	.197	6.660	.000	1.258	2.309
	dchomesupports_dummy	-.194	.147	-.044	-1.314	.189	-.483	.096
	transferLTC_dummy	-.473	.192	-.082	-2.468	.014	-.850	-.097

a. Dependent Variable: RIW

Regression Equation:

Predicted RIW

$$= .822 - .007(\text{age}) + .158(\text{LOS}) + .112(\text{male_dummy}) + 1.783(\text{died_dummy}) - .194(\text{dchomesupports_dummy}) - .473(\text{transferLTC_dummy})$$

Therefore, the predicted RIW of a male patient aged 75 who was discharged home and had an LOS of 5 days is:

$$\begin{aligned}\text{Predicted RIW} &= .822 - .007(75) + .158(5) + .112(1) + 1.783(0) - .194(0) - .473(0) \\ &= .822 - .525 + .79 + .112 \\ &= 1.199 \\ &= 1.20 \text{ (rounded to 2 decimal places)}\end{aligned}$$

Some of the typical errors found in Question 2d:

- Did not copy and paste the Coefficients table or values in the table are different than those shown. [-1 MARK]
- Note: If reported results without producing the corresponding table where these results came from, no marks are earned.
- Did not state that the outcome variable is predicted RIW or \hat{y} . [-1 MARK]
- Incorrect right-hand side of equation. Note names of predictor variables may be slightly different or if used x_1, x_2, \dots, x_6 this is also fine. [-1 MARK]
- Incorrect calculated final answer. [-1 MARK]

School of Health Policy and Management

Assignment Attachment Form

Student Name:

Student Number:

Course Code:

Assignment Title:

Due Date:

Tutorial Leader (if applicable):

Please check each box after reading, to acknowledge agreement with each statement.

- I have read and understand the Senate Policy on Academic Honesty found on website at the following York Secretariat website on Academic Honesty.
- I have read and understood the assignment submission described in the course outline (syllabus)
- I have read and understood the criteria used for assessment in this assignment
- I have read and understood and followed the referencing guidelines required for assignments submitted at York University
- This assignment is entirely my own work, except where I have given documented references to work of others
- This assignment or substantial parts of it has not previously been submitted for assessment in any formal course of study, unless acknowledged in the assignment and previously agreed to by my Tutorial Leader and Course Director
- I understand that this assignment may undergo electronic detection for plagiarism and a copy of the assignment may be retained on the database and used to make comparisons with other assignments in the future

Signature: _____ Date: _____