Biostat 200B HW1

1. $S_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n-1}}$ $S_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}$

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

$$\hat{\beta}_1 = r_{xy}\frac{S_y}{S_x} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cancel{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}}\cdot\frac{\cancel{\sqrt{\frac{1}{n-1}}}\cancel{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}}{\cancel{\sqrt{\frac{1}{n-1}}}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}$$

$$= \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \#$$

2. (a) $r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{92.41}{\sqrt{2379.93}\sqrt{49.405}} = \frac{92.41}{342.9} = 0.269 \#$

(b) ~~$\hat{\beta}_1 = r_{xy}\frac{S_y}{S_x} = 0.272\cdot\frac{\sqrt{\frac{1}{n-1}}\sqrt{49.405}}{\sqrt{\frac{1}{n-1}}\sqrt{2379.93}} = 0.272\cdot 0.144 = 0.039$~~

$\hat{\beta}_0 = \bar{Y} - b_1\bar{X} = 3.07 - 0.039\times 24.7 = 2.1067 \#$

(c) $\hat{\beta}_0 = 2.1067$ (score)

The satisfaction with health care provider will be 2.1067 when the age of patient is 0.

$\hat{\beta}_1 = 0.039$ (score/years)

The satisfaction with health care provider will increase by 0.039 as the age of patient goes up by 1 in general.

(d) $Y_i = \alpha_0 + \alpha_1(X_i - \bar{X}) + \varepsilon_i$

$\hat{Y}_i = \hat{\alpha}_0 + \hat{\alpha}_1(X_i-\bar{X}) = \bar{Y} + \hat{\beta}_1(X_i - \bar{X})$, where $\hat{\alpha}_0 = \bar{Y}$, $\hat{\alpha}_1 = \hat{\beta}_1$

$\therefore \hat{\alpha}_0 = \bar{Y} = 3.07$, $\hat{\alpha}_1 = \hat{\beta}_1 = 0.039$

(e) $\hat{\alpha}_0 = 3.07$ (score)

The estimated mean satisfaction with health care provider for patients aged 24.7 (the mean age of 120 young adult patients) is 3.07.

$\hat{\alpha}_1 = 0.039$ (score/years)

The satisfaction with health care provider will increase by 0.039 as the age of patients minus mean age of 120 young adult patients go up by 1 in general.

3. In the method of least squares, we consider the sum of squared deviations,

$$\sum_{i=1}^{n}(Y_i-\beta_0)^2 \text{, then } Q(\beta_0)=\sum_{i=1}^{n}(Y_i-\beta_0)^2$$

$$\frac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i-\beta_0) \xrightarrow{\text{let}} 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = n\beta_0$$

$$\Rightarrow \hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y} \text{ \#}$$

5. $$\cos\theta = \frac{|\text{adjacent}|}{|\text{hypotenuse}|} = \frac{|p|}{|u|} = \frac{\frac{u'v}{v'v}|v|}{|u|}$$

$$= \frac{\frac{u'v}{|v|^2}|v|}{|u|} = \frac{u'v}{|u||v|} \text{ \#}$$

4. (a)

```
proc reg data= hw1.senic;
  model risk = beds;
run;
```

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	3.72404	0.19517	19.08	<.0001
beds	1	0.00250	0.00061579	4.06	<.0001

The coefficient associated with **beds** is 0.0025, which means the infection risk will increase by 0.0025 as the average number of beds in hospital during study period goes up by 1 unit.

(b)

```
proc reg data= hw1.senic;
  model risk = svcs;
run;
```

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	2.78402	0.34882	7.98	<.0001
svcs	1	0.03640	0.00763	4.77	<.0001

The coefficient associated with **svcs** is 0.0364, which means the infection risk will increase by 0.0364 as available facilities and services goes up by 1 unit (percent).

(c)

```
proc reg data= hw1.senic;
  model nurses = age;
run;
```

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	311.06760	157.71376	1.97	0.0511
age	1	-2.58905	2.95251	-0.88	0.3824

The coefficient associated with **age** is -2.58905, which means the average number of full-time equivalent nurses during study period will decrease by 2.58905 as average age of patients goes up by 1 unit.

(d)

```
data hw1.senic1;
  set hw1.senic;
  med = msch;
  if med = 2 then med = 0;
run;

proc reg data= hw1.senic1;
  model nurses = med;
run;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	138.92708	11.54610	12.03	<.0001
med	1	228.13174	29.76803	7.66	<.0001

The coefficient associated with **med** is 228.13174, which means the average number of full-time equivalent nurses during study period will increase by 228.13174 as medical school affiliation goes up by 1 unit. That is to say, if the hospital is medical school affiliation, the average number of full-time equivalent nurses during study period is 228.13174 higher than those which are not medical school affiliation in general.