# Problem Set 2 - SOLUTIONS

Statistics - NYU, Summer 2016
(Econ UA - 18.001)

**[1]** a) 227, 224, 682 $\times(1+r)^{14} = 313,877,061 \implies r = 0.233$ or $r = 2.3\%$.

b) $313,877,061(1.009)(1+r)^7 = 341,387,000 \implies r = 0.108 = 1.08\%$.

**[2]** a) Denote Education and Family Income variables with $e$ and $f$, respectively. Then sample means are:

$$\bar{e} = \frac{\sum_{i=1}^{10} e_i}{10} = \frac{18+14+\ldots+12+18}{10} = 15.1$$
$$\bar{f} = \frac{\sum_{i=1}^{10} f_i}{10} = \frac{110+80+\ldots+60+70}{10} = 88.5$$

And sample variances are:

$$var(e) = s_e^2 = \frac{\sum_{i=1}^{10}(e_i - \bar{e})^2}{10-1} = \frac{(18-15.1)^2 + \ldots + (12-15.1)^2 + (18-15.1)^2}{9} = 5.66$$
$$var(f) = s_f^2 = \frac{\sum_{i=1}^{10}(f_i - \bar{f})^2}{10-1} = \frac{(110-88.5)^2 + \ldots + (60-88.5)^2 + (70-88.5)^2}{9} = 866.94$$

b) $CV(e) = \left(\frac{\sqrt{5.66}}{15.1}\right) \cdot 100 = 15.75$ and $CV(f) = \left(\frac{\sqrt{866.94}}{88.5}\right) \cdot 100 = 33.27$
Therefore, family income varies more around its mean value.

c) Covariance between education and family income:

$$Cov(e,f) = \frac{\sum_{i=1}^{10}(e_i - \bar{e})(f_i - \bar{f})}{10-1}$$
$$= \frac{(18-15.1)(110-88.5) + \ldots + (12-15.1)(60-88.5) + (18-15.1)(70-88.5)}{9}$$
$$= 36.15$$

And the correlation coefficient:

$$r = \frac{Cov(e,f)}{s_e s_f} = \frac{36.15}{\sqrt{5.66}\sqrt{866.94}} = 0.57$$

d) Since covariance is greater than zero, we can say that there is a positive relation between the two variables, and correlation coefficient says strength of the relation is reasonably high.

e) But these findings does NOT imply any causal relationship between $e$ and $f$. A reasonable direction would be from family income to education, i.e. those with high family incomes are more educated. But it's also possible to think as follows: those families with high incomes are already have higher education levels in compare to families with lower incomes.

**[3]** a) $\bar{X} = 39.8$, $\bar{Y} = 38.1$

b) $s_x^2 = 263.511$, $s_y^2 = 110.767$, and $s_{xy} = 100.578$.

c) $r = \frac{Cov(x,y)}{s_x s_y} = 0.589$

d) The sample correlation coefficient is generally a more useful measure, as it provides both the direction and the strength of the relationship.

**[4]** a) The completed table is as follows:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $X_{new}$ | 56 | 37 | 13 | 66 | 31 | 61 | 31 | 53 | 42 | 38 |
| $Y_{new}$ | 77 | 55 | 61 | 113 | 75 | 99 | 85 | 87 | 45 | 95 |

b) $\bar{X}_{new} = 42.8 = \bar{X} + 3$, $\bar{Y}_{new} = 79.2 = 2 * \bar{Y} + 3$

c) $s_{x_{new}} = 16.233 = s_x$, $s_{y_{new}} = 21.049 = 2 * s_y$

d) $s_{x_{new}y_{new}} = 201.155 = 2 * s_{xy}$, $r_{x_{new}y_{new}} = 0.588 = r_{xy}$

**[5]** a)

$$
\begin{aligned}
\bar{Y} &= \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) \\
&= a\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} b \\
&= a\bar{X} + \frac{1}{n}nb \\
&= a\bar{X} + b
\end{aligned}
$$

b)

$$\begin{aligned} s_Y^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{Y})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + b - a\bar{X} - b)^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(ax_i - a\bar{X})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n} a^2(x_i - \bar{X})^2 \\ &= a^2 s_X^2 \end{aligned}$$

**[6]** a) Denote Experience and Weekly Sales variables with $e$ and $s$ respectively. Then sample means are:

$$\begin{aligned} \bar{e} &= \frac{\sum_{i=1}^{8} e_i}{8} = \frac{2+4+\ldots+6+2}{8} = 3.875 \\ \bar{s} &= \frac{\sum_{i=1}^{8} s_i}{8} = \frac{5+10+\ldots+20+4}{8} = 10.75 \end{aligned}$$

And sample variances are:

$$\begin{aligned} var(e) = s_e^2 &= \frac{\sum_{i=1}^{8}(e_i - \bar{e})^2}{8-1} = \frac{(2-3.875)^2 + \ldots + (6-3.875)^2 + (2-3.875)^2}{7} = 2.696 \\ var(s) = s_s^2 &= \frac{\sum_{i=1}^{8}(s_i - \bar{s})^2}{8-1} = \frac{(110-88.5)^2 + \ldots + (60-88.5)^2 + (70-88.5)^2}{7} = 37.93 \end{aligned}$$

b) Covariance between experience and weekly sales:

$$\begin{aligned} Cov(e,s) &= \frac{\sum_{i=1}^{8}(e_i - \bar{e})(s_i - \bar{s})}{8-1} \\ &= \frac{(2-3.875)(5-10.75) + \ldots + (6-3.875)(20-10.75) + (2-3.875)(4-10.75)}{7} \\ &= 9.964 \end{aligned}$$

And the correlation coefficient:

$$r = \frac{Cov(e,s)}{s_e s_s} = \frac{9.964}{\sqrt{2.696}\sqrt{37.93}} = 0.985$$

**[7]** Please refer to Excel file.

**[8]** a) $(1+r)^5 = 1.25 \implies r = 0.0456$ or $r = 4.56\%$.

b) $(1+r)^5 = (1+0.05)^2(1+0.2)^3 \implies r = 0.1376$.

**[9]** a) The completed table is as follows:

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $Z$ | 205.25 | 134.5 | 66.25 | 257.75 | 129 | 234 | 135.25 | 202.5 | 143.25 | 162.5 |
| W | 410.5 | 269 | 132.5 | 515.5 | 258 | 468 | 270.5 | 405 | 286.5 | 325 |

b) By direct calculation, we obtain $Cov(Z, W) = 6598.0125$

Alternatively, note that $W = 2Z$. Therefore,

$$
\begin{aligned}
Cov(Z, W) &= Cov(Z, 2Z) \\
&= \frac{\sum_{i=1}^{10}(z_i - \bar{z})(2z_i - 2\bar{z})}{10-1} \\
&= \frac{\sum_{i=1}^{10}(z_i - \bar{z})(2)(z_i - \bar{z})}{10-1} \\
&= 2 \cdot \frac{\sum_{i=1}^{10}(z_i - \bar{z})(z_i - \bar{z})}{10-1} \\
&= 2 \cdot Var(Z) \\
&= 6598.0125
\end{aligned}
$$

Using this alternative approach, compute the covariance for the following cases:

i. $w_k = 3x_k + 1.25y_k + 7$ and $z_k = 6x_k + 1.25y_k$

ii. $w_k = 3x_k + 1.25y_k$ and $z_k = 18x_k + 7.5y_k + 4$

**[10]** a) Denote Test Score and Weekly Sales variables with $x$ and $y$ respectively. Then sample means are:

$$
\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{10} x_i}{10} = \frac{12 + 30 + \ldots + 19 + 27}{10} = 21.3 \\
\bar{y} &= \frac{\sum_{i=1}^{10} y_i}{10} = \frac{20 + 60 + \ldots + 32 + 57}{10} = 41.2
\end{aligned}
$$

And sample variances are:

$$var(x) = s_x^2 = \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{10-1} = \frac{(12-21.3)^2 + \ldots + (19-21.3)^2 + (27-21.3)^2}{9} = 42.01$$
$$var(y) = s_y^2 = \frac{\sum_{i=1}^{10}(y_i - \bar{y})^2}{10-1} = \frac{(20-41.2)^2 + \ldots + (32-41.2)^2 + (57-41.2)^2}{9} = 278.4$$

Covariance between test scores and weekly sales:

$$\begin{aligned} Cov(x,y) &= \frac{\sum_{i=1}^{10}(x_i - \bar{x})(y_i - \bar{y})}{10-1} \\ &= \frac{(2-3.875)(5-10.75) + \ldots + (6-3.875)(20-10.75) + (2-3.875)(4-10.75)}{9} \\ &= 106.93 \end{aligned}$$

b) And the correlation coefficient:

$$r = \frac{Cov(x,y)}{s_x s_y} = \frac{106.93}{\sqrt{42.01}\sqrt{278.4}} = 0.989$$