Ian MacDonald
ESE 542
HW #4

1. We fit n data points to a line by minimizing RSS (least squares), x_0 is a new point on the line, and $\hat{u}_0$ is its estimate, given by

$$\hat{u}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

a) $$Var(\hat{u}_0) = Var(\hat{\beta}_0 + \hat{\beta}_1 x_0)$$

$$= Var(\hat{\beta}_0) + x_0^2 Var(\hat{\beta}_1) + 2x_0 Cov(\hat{\beta}_0, \hat{\beta}_1)$$

Given line is fit using least squares method,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \quad , \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{Cov(x,y)}{Var(x)}$$

$$Var(\hat{\beta}_0) = Var(\bar{y} - \hat{\beta}_1\bar{x})$$

$$= Var(\bar{y}) + \bar{x}^2 Var(\hat{\beta}_1) - Cov(\bar{y}, \hat{\beta}_1)$$

$$= Var\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) + \bar{x}^2 Var(\hat{\beta}_1) - 0$$

$$= \frac{\sigma^2}{n} + \bar{x}^2 Var(\hat{\beta}_1)$$

To find $Var(\hat{\beta}_1)$: $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})$

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

$$= \sum_{i=1}^{n}(x_i-\bar{x})y_i - \cancel{\bar{y}\sum_{i=1}^{n}x_i} + \cancel{\bar{y}K\frac{\sum_{i=1}^{n}x_i}{n}}$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})y_i$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})y_i}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$= \sum_{i=1}^{n} k_i y_i \ , \ k_i = \frac{(x_i-\bar{x})}{(x_i-\bar{x})^2}$$

$$Var(\hat{\beta}_1) = Var\left(\sum_{i=1}^{n} k_i y_i\right)$$

$$= \sum_{i=1}^{n} k_i^2 \, Var(y_i)$$

$$= \sigma^2 \sum_{i=1}^{n} \frac{\cancel{(x_i-\bar{x})^2}}{(x_i-\bar{x})^{\cancel{4}2}}$$

$$= \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \bar{x}^2 \, Var(\hat{\beta}_1)$$

$$= \frac{\sigma^2}{n} + \bar{x}^2\left(\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = Cov(\bar{y} - \hat{\beta}_1\bar{x}, \hat{\beta}_1)$$

$$= Cov(\bar{y}, \hat{\beta}_1) - Cov(\bar{x}\hat{\beta}_1, \hat{\beta}_1)$$

$$= 0 - \bar{x} \, Cov(\hat{\beta}_1, \hat{\beta}_1)$$

$$= \frac{-\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$Var(\hat{\mu}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right) + x_0^2\left(\frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$+ 2x_0\left(\frac{-\bar{x}\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

$$Var(\hat{u}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2 \bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} + \frac{x_0^2 \sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} - \frac{2 x_0 \bar{x} \sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

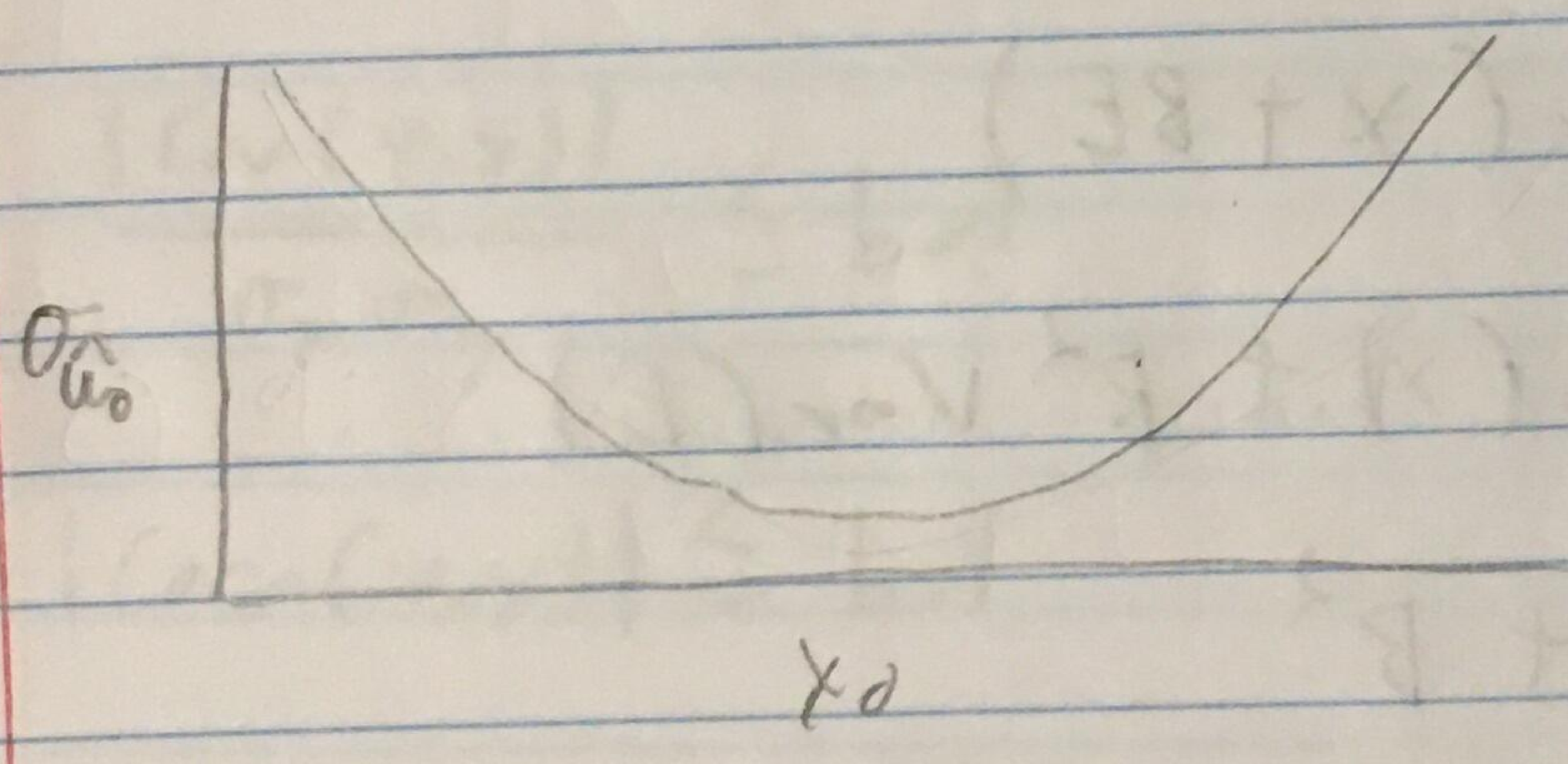
$$\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} + \bar{x}^2 + x_0^2 - 2x_0\bar{x}\right)$$

$$\frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\left(\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} + (x_0 - \bar{x})^2\right)$$

$$\frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

b) $Var(\hat{u}_0) = \frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

$$\sigma_{\hat{u}_0} = \sqrt{\frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$



$\sigma_{\hat{u}_0}$ gets larger the farther x_0 is away from $\bar{x}$, and is at it's minimum when $x_0 = \bar{x}$

c) 95% Confidence Interval for $\hat{u}_0$

$$\left(\hat{u}_0 - 1.96\sqrt{\frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}},\ \hat{u}_0 + 1.96\sqrt{\frac{\sigma^2}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right)$$

2. $X \sim N(0,1)$, $E \sim N(0,1)$, X and E independent, and $Y = X + BE$

Show that $r_{XY} = \dfrac{Cov(X,Y)}{\sigma_X \sigma_Y} = \dfrac{1}{\sqrt{B^2+1}}$

$$Cov(X,Y) = Cov(X+BE, X)$$
$$= Cov(X,X) + Cov(X,BE)$$
$$= Var(X) + 0$$
$$= 1$$

$Var(X) = 1$, $\sigma_X = \sqrt{1} = 1$

$$Var(Y) = Var(X+BE)$$
$$= Var(X) + B^2 Var(E)$$
$$= 1 + B^2$$

$\sigma_Y = \sqrt{B^2+1}$

$Cov(X,Y) = 1$, $\sigma_X = 1$, $\sigma_Y = \sqrt{B^2+1}$

$$r_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{1}{\sqrt{B^2+1}}$$

3. Suppose there are n data points on x, y (x and y both 1 dimensional). We fit lines $y = a + bx$ and $x = c + dy$

Show that $bd \leq 1$, and explain when $bd = 1$ and what it means

$$Var(x) = Var(c + dy)$$
$$= d^2 Var(y)$$

$$Var(y) = Var(a + bx)$$
$$= b^2 Var(x)$$

$$\sqrt{|Cov(x,y)|^2} \leq \sqrt{Var(x)\, Var(y)\, b^2 d^2}$$

$$|Cov(x,y)| \leq \sigma_x \sigma_y bd$$

$$\frac{|Cov(x,y)|}{\sigma_x \sigma_y} \leq bd$$

$$|Corr(x,y)| \leq bd$$

Since $|Corr(x,y)| \leq 1$, $bd \leq 1$

When $bd = 1$, then it means that lines x and y are perfectly correlated