IEOR 165 Midterm 1 (Lim, Spring 2012)

1.

	average	sd	$r=0.5$
income	90000	45000	
IQ	100	15	

(a),(c) $Y = IQ$; $X =$ Income

$\Rightarrow \hat{Y} = b_0 + b_1 X$

$b_1 = r\frac{S_y}{S_x} = 0.5 \times \frac{15}{45000} = \boxed{\frac{1}{6000}}$

$b_0 = \bar{Y} - b_1\bar{X} = 85$

$\hat{Y} = 85 + \frac{1}{6000}(120000) = \boxed{105}$

(b) $Y =$ Income $X = IQ$

$b_1 = r \cdot \frac{S_y}{S_x} = \frac{1}{2} \cdot \frac{45000}{15} = 1500$

$b_0 = \bar{Y} - b_1\bar{X} = -60000$

$\hat{Y} = -60000 + 1500(110) = \boxed{105000}$

2. (a) position correlation, linearly associated $\rightarrow$ ok to use a linear model

(b) a few outliers on the plots, can't remove outliers directly

(c) $r^2 = R^2$ $b_1 = r \cdot \frac{S_y}{S_x}$

$\Rightarrow r = 0.6846$, positive, weakly correlated

(d) $R^2 = 0.4687$ There are 46.87% of variability in Math score that can be explained in the model.

(e) Math score $= 209.55417 + 0.67507 \times$ Verbal score

for every increase in the Verbal score, we can expect to see 0.67507 increase in the Math score

(f) Verbal $= 650$

$\widehat{Math}|Verbal=650 = 648.3497$

(g) sd of Math score for Verbal $= 650, 700$

$Y_i = b_0 + b_1 X_i + \varepsilon_i$

$\hat{Y} = E[Y|X]$ $Y|X =$ constant $+ \varepsilon_i$

$Var(Y|X) = Var(\varepsilon_i) \Rightarrow sd(Y|X) = sd(\varepsilon_i)$

From output $= 71.75$ or $s_y\sqrt{1-r^2} \times \sqrt{\frac{n-1}{n-2}}$

(h) $Y|X=650 \sim N(648.3497, 71.75)$

$Pr(\hat{Y}|X=650 > 700) = Pr\left(\frac{\hat{Y}_{|X=650} - 648.3497}{71.75} > \frac{700 - 648.3497}{71.75}\right) = 0.2357$

3. SRS : 1000 people, 543 Democrats

(a) $\hat{p} = \frac{543}{1000}$ $\qquad \sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{1000}} = 0.01575$

(b) 95% CI $= \hat{p} \pm 1.96(0.01575)$

$= (0.5121, 0.5738)$

Assumption of CLT :

① SRS

② $1000 \ll$ size of population

③ $np \cong n\hat{p} \geq 10$, $n(1-p) \cong n(1-\hat{p}) \geq 10$

4. Given scatter plots (x_i, y_i), $(x_i, \log y_i)$, $(x_i, \sqrt{y_i})$

(a) $(x_i, y_i) \Rightarrow$ nonlinearity

$(x_i, \sqrt{y_i}) \Rightarrow$ variance increases as x increases

$\log y_i = 0.932 + 0.12276\,x$

$y = e^{0.932 + 0.12276x}$

(b) $\hat{y}_{|x=10} = e^{0.932 + 0.12276(10)} = 8.667$

(c) $\Pr(Y_{|x=10} > 15)$

$= \Pr(\log Y_{|x=10} > \log 15)$

$= \Pr\left(\frac{\log Y_{|x=10} - 8.667}{0.6682} > \frac{\log 15 - 8.667}{0.6682}\right)$

$= \Pr(Z > 0.8207) = 0.206$

$\log y = 0.932 + 0.12276x$

$\hat{y}_{|x=10} = E[Y \mid x=10] = 8.667$

$SE(\hat{y}_{|x=10}) = 0.6682$

$\log y_i \sim N(8.667, 0.6682)$