Midterm 1 - Solution Set

Matthew Strathearn

March 12, 2015

Question 1

(a)

Let $f(x, \mu, \sigma)$ be the normal distribution PDF. Therefore;

$$f(x, \mu, \sigma) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma}) \tag{1}$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma}\phi(\frac{v-\mu}{\sigma})dv \tag{2}$$

$$F(x) = \Phi(\frac{v-\mu}{\sigma})\Big|_{-\infty}^{x} \tag{3}$$

$$F(x) = \Phi(\frac{x-\mu}{\sigma}) \tag{4}$$

In other words, if $x \sim N(\mu, \sigma^2)$, then $\frac{x-\mu}{\sigma} \sim N(0, 1)$. We want to find the probability that the students grade falls between 40 and 64 when $\mu = 70$ and $\sigma^2 = 10$.

$$\begin{aligned} P(40 \leq x \leq 64) &= F(64) - F(40) \\ &= \Phi(\frac{64-70}{\sqrt{10}}) - \Phi(\frac{40-70}{\sqrt{10}}) \\ &\approx \Phi(-1.90) - \Phi(-9.49) \\ &\approx .0287 - 0 \\ &\approx .0287 \end{aligned} \tag{5}$$

(b)

The probability of a student scoring higher than 64 is as follows;

$$\begin{aligned}
P(x > 64) &= P(x \leq \infty) - P(x \leq 64) \\
&= 1 - P(x \leq 64) \\
&= 1 - \Phi(\frac{64-70}{\sqrt{10}}) \\
&\approx 1 - \Phi(-1.90) \\
&\approx 1 - 0.0287 \\
&\approx 0.9713
\end{aligned} \tag{6}$$

Question 2

(a)

Assuming the normality of ϵ and that the regressor p is fixed among repeated samples;

$$\hat{\beta} \sim \mathcal{N}(\beta, Var(\hat{\beta})) \tag{7}$$

Where;

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{8}$$

As σ^2 is not known, thus has to be estimated with s^2, it then follows that

$$t = \frac{\hat{\beta} - \beta}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \sim t_{n-2}. \tag{9}$$

The 95% confidence level is as follows;

$$\begin{aligned}
P(-|t_{2.5\%,30-2}| \leq t \leq |t_{2.5\%,30-2}|) &= 95\% \\
P(-2.048 \leq t \leq 2.048) &= 95\% \\
P(-2.048 \leq \frac{\hat{\beta} - \beta}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \leq 2.048) &= 95\%
\end{aligned} \tag{10}$$

Solving for the 95% confidence interval;

$$\begin{aligned}
P\left\{\hat{\beta} - 2.048\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \leq \beta \leq \hat{\beta} + 2.048\sqrt{\frac{s^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right\} &= 95\% \\
P\left\{-9.835284 - 2.048 * 1.375 \leq \beta \leq -9.835284 + 2.048 * 1.375\right\} &\approx 95\% \\
P\left\{-12.58 \leq \beta \leq -7.11\right\} &\approx 95\%
\end{aligned} \tag{11}$$

Therefore $\beta \in [-12.58, -7.11]$. Also note that the confidence interval does not include zero and therefore the coefficient is statistically different from zero. There is a 95% chance that this confidence interval contains the true β.

(b)

$H_0 : \beta = 0$ is the null hypothesis while $H_a : \beta \neq 0$ is the alternative. From earlier, the t-statistic can be calculated as such;

$$\begin{aligned} t &= \frac{\hat{\beta} - \beta}{s/\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \\ t &= \frac{-9.835284 - 0}{1.375388} \\ t &= -7.150914 \end{aligned} \tag{12}$$

At the 5% significance level $t_{2.5\%,28} = \pm 1.071$. Since $|t| > |t_{2.5\%,28}|$ we reject the null hypothesis in favour of the alternative, at the 5% level.

(c)

All that changes is the t-critical value. $t_{5\%,28} = 1.071$, since $t < t_{5\%,28}$ we fail to reject the null hypothesis at the 5% level. To determine the P-Value we calculate the probability of observing a t-statistic at least as extreme as -7.150914.

$$P(t \geq -7.150914|\beta = 0) = 1 - P(t \leq -7.150914|\beta = 0) = 1 - 0 = 1 \tag{13}$$

Since $p > .05$, at the 5% level, we fail to reject the null hypothesis that $\beta = 0$.

Question 3

(a)

Assumption 1: ϵ_i, $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.), and $\epsilon_i \sim N(0, \sigma^2)$.
Assumption 2: ϵ_i and x_i are independent for every $i \in [1, n]$.
Assumption 3: x_i, $i = 1, \ldots, n$, are fixed regressors.

(b)

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{14}$$

Using the sample moments;

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i) &= \frac{1}{n}\sum_{i=1}^{n}(y_i) - \hat{\alpha} - \hat{\beta}\frac{1}{n}\sum_{i=1}^{n}(x_i) = 0 \\ \frac{1}{n}\sum_{i=1}^{n}(x_i\epsilon_i) &= \frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - \hat{\alpha}\frac{1}{n}\sum_{i=1}^{n}(x_i) - \hat{\beta}\frac{1}{n}\sum_{i=1}^{n}(x_i^2) = 0 \end{aligned} \tag{15}$$

Solving the two equations yields;

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \tag{16}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_i\epsilon_i) = \frac{1}{n}\sum_{i=1}^{n}(x_iy_i) - (\bar{y} - \hat{\beta}\bar{x})\frac{1}{n}\sum_{i=1}^{n}(x_i) - \hat{\beta}\frac{1}{n}\sum_{i=1}^{n}(x_i^2) = 0 \tag{17}$$

$$\frac{1}{n}\sum_{i=1}^{n}(x_iy_i) - \bar{y}\frac{1}{n}\sum_{i=1}^{n}(x_i) = \hat{\beta}\frac{1}{n}\sum_{i=1}^{n}(x_i^2) - \hat{\beta}\bar{x}\frac{1}{n}\sum_{i=1}^{n}(x_i) \tag{18}$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{19}$$

(c)

$$y_i - \bar{y} = \beta(x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon}) \tag{20}$$

Substituting 23 into 22;

$$\begin{aligned}
\hat{\beta} &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\beta(x_i - \bar{x}) + (\epsilon_i - \bar{\epsilon}))}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\hat{\beta} &= \beta + \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\
\mathbb{E}[\hat{\beta}] &= \beta + \frac{\sum_{i=1}^{n}(x_i - \bar{x})\mathbb{E}[\epsilon_i - \bar{\epsilon}]}{\sum_{i=1}^{n}(x_i - \bar{x})^2}
\end{aligned} \tag{21}$$

Therefore $\mathbb{E}[\hat{\beta}] = \beta$ since $\mathbb{E}[\epsilon_i - \bar{\epsilon}] = 0$. This demonstrates the unbiasedness of $\hat{\beta}$.

$$\hat{\alpha} = \alpha + \bar{x}(\beta - \hat{\beta}) + \bar{\epsilon} \tag{22}$$

$$\mathbb{E}[\hat{\alpha}] = \alpha + x(\beta - \mathbb{E}[\hat{\beta}]) + \mathbb{E}[\bar{\epsilon}] \tag{23}$$

This implies that $\mathbb{E}[\hat{\alpha}] = \alpha$ since $\mathbb{E}[\hat{\beta}] = \beta$ and $\mathbb{E}[\bar{\epsilon}] = 0$.

(d)

Notice that

$$\hat{\beta} - \beta = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{24}$$

$E[\hat{\beta}] = \beta$, and

$$\mathrm{Var}[\hat{\beta} - \beta] = \frac{1}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\mathrm{Var}[\epsilon_i - \bar{\epsilon}] \tag{25}$$

$$\mathrm{Var}[\hat{\beta} - \beta] = \frac{1}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma^2 \tag{26}$$

$$\mathrm{Var}[\hat{\beta} - \beta] = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{27}$$

Because ϵ_i are i.i.d. $N(0, \sigma^2)$ and x_i, $i = 1, \ldots, n$, are fixed, as a summation of normal random variables is also a normal random variable, it then follows that

$$\hat{\beta} - \beta \sim \mathcal{N}\left(0, \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right). \tag{28}$$