Answer Key – Midterm, Fall 2007

Problem 1.

a) Let event weight be X, and event EPA be Y.

$$\hat{\mu}_X = \frac{1800 + 2000 + 2500}{3} = 2100$$

$$\hat{\mu}_Y = \frac{30 + 25 + 23}{3} = 26$$

b) We may find the covariance and standard deviations of both events,

$\hat{\sigma}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)$ =1/2[(1800-2100)(30-26)+(2000-2100)(25-26)+(2500-2100)(23-26)] = -1150

$$\hat{\sigma}_x = sqrt(\hat{\sigma}_x^2) = sqrt[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu}_x)^2] = sqrt\{\frac{1}{2}[(1800-2100)^2 + (2000-2100)^2 + (2500-2100)^2]\}$$

$$= 100\sqrt{13} \approx 360.555$$

$$\hat{\sigma}_y = sqrt(\hat{\sigma}_y^2) = sqrt[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{\mu}_x)^2] = sqrt\{\frac{1}{2}[(30-26)^2 + (25-26)^2 + (23-26)^2]\}$$

$$= \sqrt{13} \approx 3.6056$$

Hence $\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = -0.8846$

c) Version A: the total weight without kids is 3*2050 = 6150;
the total weight with kids is 1800 + 2000 +2500 = 6300;
the difference of upper two is the weight of the kids, 6300 – 6150 = 150.

Version B: the total weight with empty tank is 3*2080 = 6240;
the total weight with the heaviest one having a full tank is 1800 + 2000 +2500 = 6300;
the difference of upper two is the gas weight, 6300 – 6240 = 60.

d) 30*0.45*1.61 = 21.735

Problem 2

a)

HHH HHT HTH THH HTT THT TTH TTT

b)

HHT HTH THH

Are the outcomes that are of interest. Each outcome is equally likely. Therefore, the probability is 3/8.

c) Let event A be getting two heads or less, and event B be getting more than one tail. Note that:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = 7/8 + 4/8 - 4/8 = 7/8$$

Problem 3

a. When choosing a random number from a continuous distribution, the chance of picking any given number is infinitesimally small. The chance of running *exactly* 3 miles is zero.

b.
Version A
The procedure is to convert 2 and 4 to z-scores and look them up in the standard normal tables.

$$z = \left(\frac{x - \mu_x}{\sigma_x}\right)$$

$$\left(\frac{2 - 2.5}{1}\right) = -0.5 \quad \left(\frac{4 - 2.5}{1}\right) = 1.5$$

Now we want to subtract Pr(z<-0.5) from Pr(z<1.5) to find the probability of landing between them.

P(z<-0.5) = 1 - Pr(z<0.5) = 1 – 0.6915 = 0.3085

P(z<1.5) = 0.9332

Pr(z<1) - Pr(z<-0.5) = 0.6247

Version B
The procedure is to convert 3 and 6 to z-scores and look them up in the standard normal tables.

$$z = \left(\frac{x - \mu_x}{\sigma_x}\right)$$

$$\left(\frac{3 - 4}{2}\right) = -0.5 \quad \left(\frac{6 - 4}{2}\right) = 1$$

Now we want to subtract Pr(z<-0.5) from Pr(z<1) to find the probability of landing between them.

P(z<-0.5) = 1 - Pr(z<0.5) = 1 – 0.6915 = 0.3085

P(z<1) = 0.8413

Pr(z<1) - Pr(z<-0.5) = .5328

c.
Version A
Here the operative word is OR which implies the union of two events,

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

A = runs between 1 and 2 miles per week B = studies between 4 and 5 hours per day

Since the student study hours per week are uniform there are equal chances of working each number of hours (1/8) so P(B) = 1/8

We need to calculate the z-scores of 1 and 2 which (using the above formula) are -1.5 and -.5. If we use the normal table and subtract as before, we get that:

Pr(A) = N(-.5) – N(-1.5) = N(1.5) – N(.5) =.9332 - .6915=.241

Since A and B are independent,

$P(A \cap B) = P(A) * P(B) = 1/8 * .241$

$P(A\ or\ B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3365$

Version B

Since the student study hours per week are uniform there are equal chances of working each number of hours (1/6) so P(B) = 1/6

We need to calculate the z-scores of 1 and 2 which (using the above formula) are -1.5 and -1. If we use the normal table and subtract as before, we get that:

Pr(A) = N(-1) – N(-1.5) = .1587 - .0669 = .0918

Since A and B are independent,

$P(A \cap B) = P(A) * P(B) = 1/6 * .0918 = 0.0153$

$P(A\ or\ B) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2432$

Problem 4

a. To compute the beta's use the formulas

$$\hat{\beta}_0 = \hat{\mu}_y - \hat{\beta}_1 \hat{\mu}_x \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \hat{\mu}_x)(y_i - \hat{\mu}_y)}{\sum_{i=1}^{n} (x_i - \hat{\mu}_x)^2}$$

First compute the means, then deviations from means, then the squares of means and covariation.

	Tips (y)	Originals (x)	x - μ$_x$	y - μ$_y$		(x - μ$_x$)^2	(x - μ$_x$)(y - μ$_y$)
	10	6	1	-5		1	-5
	5	9	4	-10		16	-40
	20	1	-4	5		16	-20
	25	4	-1	10		1	-10
μ =	15	5			Sum =	34	-75

$$\hat{\beta}_1 = \frac{-75}{34} = -2.21$$

$$\hat{\beta}_0 = 15 - (-2.21 * 5) = 26.03$$

b. Each additional original song seems to reduce tips by $2.21.

c. This would be if originals (x) = 0 which would be $\hat{\beta}_0$, $26.03

d. The formulas for R-squared are:

$$R^2 = 1 - \frac{SSR}{SST} \qquad SSR = \sum_{i=1}^{n} (\hat{\mu}_i)^2 \qquad SST = \sum_{i=1}^{n} (y_i - \hat{\mu}_y)^2$$

y_hat	u_hat	(u_hat)^2	(y - µy)^2
12.79412	-2.79412	7.807093	25
6.176471	-1.17647	1.384083	100
23.82353	-3.82353	14.61938	25
17.20588	7.794118	60.74827	100
	sum =	84.55882	250
		SSR	SST
	R^2 =	**0.661765**	

Extra Credit 1.

The amount of screaming per gig could be considered an omitted variable. Since screaming per gig will be correlated with originals and because it will be in the error term, we violate the assumption E(u|x) = 0. Since the screaming per gig has a negative effect on tips it will bias downward the coefficient on original songs making it more negative. Once we control for the amount of screaming, we may find that original songs don't reduce tips, they may even increase them.

Extra Credit 2

Causality