W271-Lab1 Spring 2016

Charles Kekeh

Thursday, January 14, 2016

Question 1

ML = 36
Stat = 28
Awesome = 18
ML ∩ Stat = 22
ML ∩ Awesome = 12
Stat ∩ Awesome = 9
ML ∪ Stat ∪ Awesome = 48

1)

$$ML \cup Stat = ML + Stat - ML \cap Stat = 42$$

$$Awesome.Other = (ML \cup Stat \cup Awesome) - (ML \cup Stat) = 6$$

$$Awesome = Awesome.Other \cup (Stat \cap Awesome) \cup (ML \cap Awesome) - (ML \cap Stat \cap Awesome)$$

$$18 = 6 + 9 + 12 - (ML \cap Stat \cap Awesome)$$

$$ML \cap Stat \cap Awesome = 9$$

$$\mathbf{Pr(ML \cap Stat \cap Awesome) = \frac{ML \cap Stat \cap Awesome}{ML \cup Stat \cup Awesome} = \frac{9}{48}}$$

2)

$$Pr(Awesome|ML) = \frac{Pr(Awesome \cap ML)}{Pr(ML)}$$

$$Pr(Awesome|ML) = \frac{12}{36} = \frac{1}{3}$$

$$\mathbf{Pr(!Awesome|ML) = 1 - Pr(Awesome|ML) = 1 - \frac{12}{36} = 1 - \frac{1}{3} = \frac{2}{3}}$$

3)

$$Pr(ML \cup Stat|Awesome) = \frac{Pr((ML \cup Stat) \cap Awesome)}{Pr(Awesome)}$$

$$Pr(ML \cup Stat|Awesome) = \frac{Pr(ML \cup Awesome) + Pr(Stat \cup Awesome) - Pr(Stat \cap ML \cap Awesome)}{Pr(Awesome)}$$

$$\mathbf{Pr(ML \cup Stat|Awesome) = \frac{12 + 9 - 9}{48} = \frac{12}{48} = \frac{1}{4}}$$

Question 2

$$Pr(A) = p \leq \frac{1}{2}, Pr(B) = q, where \frac{1}{4} < q < \frac{1}{2}$$

1)

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap\ B)$$

Pr(A ∪ B) is maximized when Pr(A ∩ B) is minimized and P(A) and P(B) are maximized. In this case that would be:

$$min(Pr(A \cap B)) = 0, (A\ and\ B\ are\ independent)$$

$$max(Pr(A)) = 1/2$$

$$max(Pr(B)) = 1/2 - \epsilon, where\ \epsilon\ approaches\ 0$$

$$\mathbf{max(Pr(A \cup B)) = 1 - \epsilon, where\ \epsilon\ approaches\ 0}$$

Alternately, Pr(A ∪ B) is minimized when A and B are completely overlapping, A = B. In this case

$$min(Pr(A \cup B) = max(min(Pr(A), Pr(B))) = 1/4 + \epsilon, where\ \epsilon\ approaches\ 0$$

$$\mathbf{min(Pr(A \cup B)) = 1/4 + \epsilon, where\ \epsilon\ approaches\ 0}$$

2)

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$$

Pr(A|B) is maximized when A is completely contained in B. In this case, Pr(A|B) = 1.

$$\mathbf{max(Pr(A|B)) = 1}$$

Pr(A|B) is minimized when Pr(A ∩ B) = 0 (A and B are independent). When Pr(A ∩ B) = 0, then Pr(A ∪ B) = 0.

$$\mathbf{min(Pr(A|B)) = 0}$$

Question 3

1)
Given that the server's lifespan is a randon uniform distribution over the range [0,k], the probability of every additional year of operation is independent of the time elapsed and is equal to

$$\mathbf{Pr(1yearofoperation)} = \mathbf{\frac{1}{k}}$$

2)
We know that:

$$E(g(x)) = \int_{x=0}^{k} g(x) f_x(x) dx$$

Considering g to be our refund function over time t:

$$E(g(t)) = \int_{t=0}^{1} \frac{\theta}{k} dt + \int_{t=1}^{k/2} \frac{A(k-t)^{\frac{1}{2}}}{k} dt + \int_{t=k/2}^{\frac{3k}{4}} \frac{\theta}{10k} dt$$

$$E(g(t)) = \frac{\theta}{k}[t]_0^1 + [\frac{-2}{3}(k-t)^{\frac{3}{2}}]_1^{k/2} + [\frac{\theta}{10k}t]_{\frac{k}{2}}^{\frac{3k}{4}}$$

$$\mathbf{E(g(t))} = \frac{\theta}{\mathbf{k}} + \frac{\theta}{\mathbf{40}} + \mathbf{\frac{2}{3}(k-1)^{\frac{3}{2}}} - \mathbf{\frac{k}{2}}^{\mathbf{\frac{3}{2}}}$$

3)
We know that Var(X) = E[X^2] -μ^2 Thus Var(g(x)) = E[(g(x)^2] - [$E[g(X)]^2$] We previously computed E(g(x)). We now compute E[(g(x)^2]

$$E[(g(x)^2] = \int_0^1 \frac{\theta^2}{k} dt + \int_1^{k/2} \frac{A^2(k-t)}{k} dt + \int_{\frac{k}{2}}^{\frac{3k}{2}} \frac{Q^2}{100k} dt$$

$$\mathbf{E[(g(x)^2]} = \frac{\theta^{\mathbf{2}}}{\mathbf{k}} - \frac{\theta^{\mathbf{2}}}{\mathbf{400}} - \mathbf{\frac{A^2(3k^2-8k+4)}{8k}}$$

We substract $E(g(x))^2$ as previously computed to obtain the variance

Megan's Question 2 and 3 answers
2. Compute the expected payout from the contract, E(x).

$$E(g(x)) = \int_{x=0}^{\infty} g(x) f_x(x) dx$$

Considering g to be our refund function over time t:

$$E(g(t)) = \int_{t=0}^{1} \frac{\theta}{k} dt + \int_{t=1}^{k/2} \frac{2(k-t)^{\frac{1}{2}}}{k} dt + \int_{t=k/2}^{\frac{3k}{4}} \frac{\theta}{10k} dt + \int_{t=\frac{3k}{4}}^{\infty} 0 dt$$

$$E(g(t)) = \frac{\theta}{k}[t]_0^1 + \frac{-4}{3k}(k-t)^{\frac{3}{2}}[t]_1^{k/2} + \frac{\theta}{10k}[t]_{\frac{k}{2}}^{\frac{3k}{4}} + 0$$

$$\mathbf{E(g(t))} = \frac{\theta}{\mathbf{k}} + \mathbf{\frac{4}{3k}(k-1)^{\frac{3}{2}}} - \mathbf{\frac{4}{3k}(\frac{k}{2})^{\frac{3}{2}}} + \frac{\theta}{\mathbf{40}}$$

3. Compute the variance of the payout from the contract.
   We know that Var(X) = E[X^2] -μ^2 Thus Var(g(x)) = E[(g(x)^2] - [$E[g(X)]^2$] We previously computed E(g(x)). We now compute E[(g(x)^2]

$$E[(g(x)^2] = \int_{t=0}^{1} \frac{\theta^2}{k^2} dt + \int_{t=1}^{k/2} \frac{4(k-t)}{k^2} dt + \int_{t=\frac{k}{2}}^{\frac{3k}{4}} \frac{\theta^2}{100k^2} dt + \int_{t=\frac{3k}{4}}^{\infty} 0dt$$

$$E(g(t)) = \frac{\theta^2}{k^2}[t]_0^1 + \frac{-2t(2k-t)}{k^2}[t]]_1^{k/2} + \frac{\theta^2}{400k}[t]_{\frac{k}{2}}^{\frac{3k}{4}} + 0$$

$$\mathbf{E[(g(x)^2] = \frac{\theta^2}{k^2} + \frac{3}{2} + \frac{4k-2}{k^2} + \frac{\theta^2}{400k}}$$

We substract $E(g(x))^2$ as previously computed to obtain the variance

Question 4

f(x,y) = 2e^{-x}e^{-2y} for 0 < x < ∞, 0 < y < ∞, 0 otherwise

1)

$$Pr(x > a, y < b) = \int_{y=0}^{b} \int_{x=a}^{\infty} 2e^{-x}e^{-2y}$$

$$Pr(x > a, y < b) = 2\int_{y=0}^{b} e^{-2y} \int_{x=a}^{\infty} 2e^{-x}$$

$$Pr(x > a, y < b) = 2\int_{y=0}^{b} e^{-2y}(1 - [e^{-x}]_0^a)$$

$$Pr(x > a, y < b) = 2e^{-a}\int_{y=0}^{b} e^{-2y}$$

$$Pr(x > a, y < b) = 2e^{-a}[-\frac{1}{2}e^{-2y}]_0^b$$

$$Pr(x > a, y < b) = e^{-a}(1 - e^{-2b})$$

2)

$$Pr(x < y) = \int_{y=0}^{infty} \int_{x=0}^{y} 2e^{-x}e^{-2y} dxdy$$

$$Pr(x < y) = 2\int_{y=0}^{\infty} e^{-2y}dy \int_{x=0}^{y} e^{-x}dx$$

$$Pr(x < y) = 2\int_{y=0}^{\infty} e^{-2y}dy[-e^{-x}]_0^y$$

$$Pr(x < y) = 2\int_{y=0}^{\infty} e^{-2y}(1 - e^{-y})dy$$

$$Pr(x < y) = 2\int_{y=0}^{\infty} e^{-2y} - e^{-3y}dy$$

$$Pr(x < y) = 2[\frac{1}{6}e^{-3y}(2 - 3e^{y})]_0^{\infty}$$
$$Pr(x < y) = \frac{1}{3}$$

3)

$$Pr(X < a) = \int_{x=0}^{a}\int_{y=0}^{\infty} 2e^{-x}e^{-2y}dxdy$$
$$Pr(X < a) = 2\int_{x=0}^{a} e^{-x}dx \int_{y=0}^{\infty} e^{-2y}dy$$
$$Pr(X < a) = 2\int_{x=0}^{a} e^{-x}dx[-\frac{1}{2}e^{-2y}]_0^{\infty}$$
$$Pr(X < a) = \int_{x=0}^{a} e^{-x}dx$$
$$Pr(X < a) = 1 - e^{-a}$$

Question 5

X random variable, x a real number.
Y = a + b (X - x^2)

1)

$$E(Y) = a + bE[(x - x^2)]$$
$$E(Y) = a + bE[X^2 - 2Xx - x^2]$$
$$E(Y) = a + b[E[X^2 - 2xE[X] + x^2]]$$

E(Y) is minimized when $\frac{d}{dx}$E(Y) = 0

$$\frac{d}{dx}E(Y) = -2bE(X) + 2bx$$
$$\frac{d}{dx}E(Y) = 0 \Rightarrow \mathbf{x} = \mathbf{E}(\mathbf{X})$$

2)

$$Whenx = E(X) : E(Y) = a + b[E[X^2] - 2(E[X])^2 + (E[X])^2]$$
$$E(Y) = a + b[E[X^2] - (E[X])^2]$$
$$\mathbf{E}(\mathbf{Y}) = \mathbf{a} + \mathbf{bVar}[\mathbf{X}]$$

3)
Y = ax + b(X - x^2)

$$E(Y) = ax + bE[(x - x^2)]$$
$$E(Y) = ax + bE[X^2 - 2Xx - x^2]$$
$$E(Y) = ax + b[E[X^2 - 2xE[X] + x^2]]$$

E(Y) is minimized when $\frac{d}{dx}$E(Y) = 0

$$\frac{d}{dx}E(Y) = a - 2bE(X) + 2bx$$
$$\frac{d}{dx}E(Y) = 0 \Rightarrow \mathbf{x} = \mathbf{E}(\mathbf{X}) - \frac{\mathbf{a}}{\mathbf{2b}}$$

Question 6

X, Y independent continuous variables, uniform over [0..1]
Z = X + Y

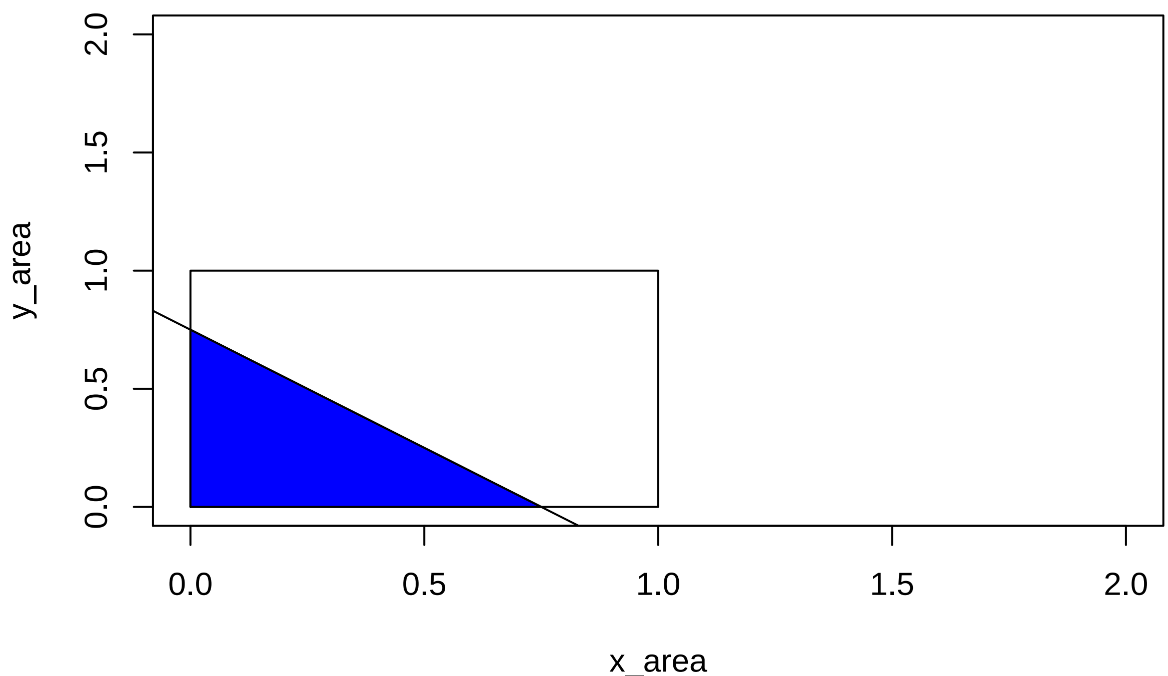
1)

```r
x_area = c(0:2)
y_area = c(0:2)

plot(x_area, y_area, type = "n")

xx = c(0, 1, 1, 0)
yy = c(0, 0, 1, 1)
polygon(xx, yy, density = 0, border = "black")

abline(.75, -1)
xz = c (0, .75, 0)
yz = c(0, 0, .75)
polygon(xz, yz, col = "blue", border = "black")
```

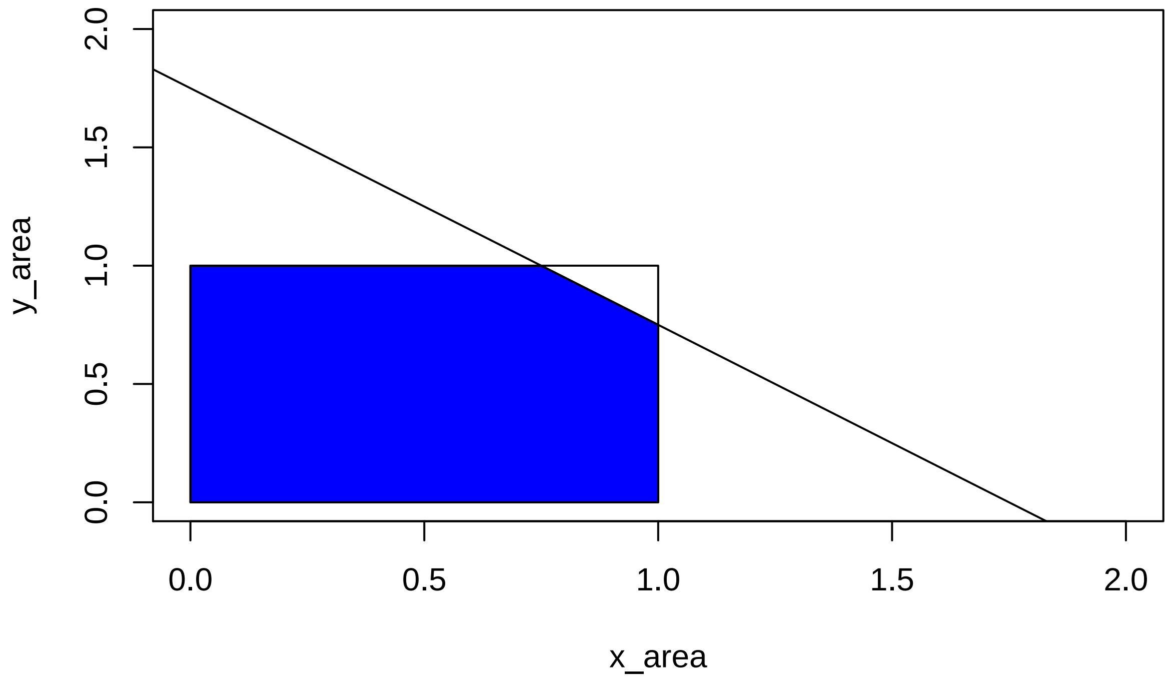



```r
plot(x_area, y_area, type = "n")

xx = c(0, 1, 1, 0)
yy = c(0, 0, 1, 1)
```

```r
polygon(xx, yy, density = 0, border = "black")

abline(1.75, -1)
xz = c (0, 1, 1, .75, 0, 0)
yz = c(0, 0, .75, 1, 1, 0)
polygon(xz, yz, col = "blue", border = "black")
```



2)
From the areas above, we derive that:
For 0 <= z <= 1:

$$\mathbf{Pr(Z < z) = \frac{z^2}{2}}$$

For 1 < z <= 2:

$$\mathbf{Pr(Z < z) = 1 - \frac{(2-z)^2}{2}}$$

Hence:
For 0 <= z <= 1:

$$\mathbf{f(z) = \frac{d}{dz}\frac{z^2}{2} = z}$$

For 1 < z <= 2:

$$\mathbf{f(z) = \frac{d}{dz} 1 - \frac{(2-z)^2}{2} = 2 - z}$$

Question 7

1)

Event Class	Sum of Dices	Events in Class	Pr(Sum of Dices)
House wins	2	(1,1)	$\frac{1}{36}$
	3	(1,2)(2,1)	$\frac{1}{18}$
	12	(6,6)	$\frac{1}{36}$
You win	7	(3,4)(4,3)(5,2)(2,5)(1,6)(6,1)	$\frac{1}{6}$
	11	(5,6)(6,5)	$\frac{1}{18}$
X	4	(2,2)(3,1)(1,3)	$\frac{1}{12}$
	5	(2,3)(3,2)(1,4)(4,1)	$\frac{1}{9}$
	6	(3,3)(4,2)(2,4)(5,1)(1,5)	$\frac{5}{36}$
	8	(4,4)(6,2)(2,6)(3,5)(5,3)	$\frac{5}{36}$
	9	(3,6)(6,3)(5,4)(4,5)	$\frac{1}{9}$
	10	(5,5)(6,4)(4,6)	$\frac{1}{12}$

We can now define

$$E(Y_{Playerwins}) = Pr(Playerwinsinone) * 1 +$$
$$\sum_{n=0}^{\infty}(Pr(4))^2(1-(Pr(4)+Pr(7)))^n(n+2)+$$
$$\sum_{n=0}^{\infty}(Pr(5))^2(1-(Pr(5)+Pr(7)))^n(n+2)+$$
$$\sum_{n=0}^{\infty}(Pr(6))^2(1-(Pr(6)+Pr(7)))^n(n+2)+$$
$$\sum_{n=0}^{\infty}(Pr(8))^2(1-(Pr(8)+Pr(7)))^n(n+2)+$$
$$\sum_{n=0}^{\infty}(Pr(9))^2(1-(Pr(9)+Pr(7)))^n(n+2)+$$
$$\sum_{n=0}^{\infty}(Pr(10))^2(1-(Pr(10)+Pr(7)))^n(n+2)$$

$$E(Y_{Housewins}) = Pr(Housewinsinone) * 1 + \sum_{n=0}^{\infty} Pr(X)Pr(7)(1-Pr(7))^n(n+2)$$

We know:

$$\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots$$
$$\frac{1}{(1-x)^2} - 1 = 2x + 3x^2 + 4x^3 + \cdots$$
$$\frac{1}{x(1-x)^2} - \frac{1}{x} = 2 + 3x + 4x^2 + 5x^3 \cdots$$
$$\frac{1-(1-x)^2}{x(1-x)^2} = 2 + 3x + 4x^2 + 5x^3 \cdots$$

Thus:

$$
\begin{aligned}
E(Y_{PlayerWins}) = Pr(Playerwin)*1 + \Pr(4)^2 \frac{1-(Pr(4)+Pr(7))^2}{(Pr(4)+Pr(7))^2(1-(Pr(4)+Pr(7)))} + \\
\Pr(5)^2 \frac{1-(Pr(5)+Pr(7))^2}{(Pr(5)+Pr(7))^2(1-(Pr(5)+Pr(7)))} + \\
\Pr(6)^2 \frac{1-(Pr(6)+Pr(7))^2}{(Pr(6)+Pr(7))^2(1-(Pr(6)+Pr(7)))} + \\
\Pr(8)^2 \frac{1-(Pr(8)+Pr(7))^2}{(Pr(8)+Pr(7))^2(1-(Pr(8)+Pr(7)))} + \\
\Pr(9)^2 \frac{1-(Pr(9)+Pr(7))^2}{(Pr(9)+Pr(7))^2(1-(Pr(9)+Pr(7)))} + \\
\Pr(10)^2 \frac{1-(Pr(10)+Pr(7))^2}{(Pr(10)+Pr(7))^2(1-(Pr(10)+Pr(7)))} +
\end{aligned}
$$

$$
E(Y_{HouseWins}) = Pr(Housewin)*1 + \Pr(X)Pr(7)\frac{1-(Pr(7)+Pr(X))^2}{(1-(Pr(7)+Pr(X)))(Pr(7)+Pr(X))^2}
$$

```r
pr.house.wins.in.one <- 4/36
pr.player.wins.in.one <- 8/36
pr.seven <- 6/36
pr.x.events <- c(3/36, 4/36, 5/36/ 5/36, 4/36, 3/36)
pr.x.plus.seven.events <- pr.x.events + pr.seven
exp.y.player.wins <- pr.player.wins.in.one +
    sum((pr.x.events)^2 * (1 - pr.x.plus.seven.events^2)/((1- pr.x.plus.seven.events)*pr.x.plus.seven.e

exp.y.house.wins <- pr.house.wins.in.one +
    pr.seven*sum(pr.x.events)*(1 - (pr.seven + sum(pr.x.events))^2)/
    ((1 - (pr.seven + sum(pr.x.events)))*(pr.seven + sum(pr.x.events))^2)

print(sprintf("E(Y_Player_Wins) = %f", exp.y.player.wins))
```

```
## [1] "E(Y_Player_Wins) = 0.908914"
```

```r
print(sprintf("E(Y_House_Wins) = %f", exp.y.house.wins))
```

```
## [1] "E(Y_House_Wins) = 0.437681"
```

2)

$$
E(Payoff) = 100*Pr(Y=1) + 80*Pr(Y=2) + 60*Pr(Y=3) + 40*Pr(Y=4) + 0*Pr(Y=5)
$$
$$
Pr(Y=1) = Pr(Playerwinsinone)
$$

Question 8

$\mathrm{E}(\mathrm{Y}_1) = \mathrm{E}(\mathrm{Y}_2) = \dots = \mathrm{E}(\mathrm{Y}_n) = \mu$
$\mathrm{Var}(\mathrm{Y}_1) = \mathrm{Var}(\mathrm{Y}_2) = \dots = \mathrm{Var}(\mathrm{Y}_n) = \sigma^2$

1)

$$W = \sum_{i=1}^{n} a_i Y_i$$

For W to be un unbiased estimator of μ:

$$E(W) = \mu$$

$$\Rightarrow E(\sum_{i=1}^{n} a_i Y_i) = \mu$$

$$\Rightarrow \sum_{i=1}^{n} a_i E(Y_i) = \mu$$

$$\Rightarrow \sum_{i=1}^{n} a_i \mu = \mu$$

$$\Rightarrow \mu \sum_{i=1}^{n} a_i = \mu$$

$$\Rightarrow \mathbf{\sum_{i=1}^{n} a_i = 1}$$

2)

$$Var(W) = Var(\sum_{i=1}^{n} a_i Y_i)$$

$$Var(W) = \sum_{i=1}^{n} a_i^2 Var(Y_i)$$

$$\mathbf{Var(W)} = \sigma^{\mathbf{2}} \mathbf{\sum_{i=1}^{n} a_i^2}$$

3)
We know that:

$$\frac{1}{n}(\sum_{i=1}^{n} a_i)^2 \leq \sum_{i=1}^{n} a_i^2$$

Multiply each side of the equation by σ^2 and rewrite the inequality by swapping which terms are on each side. It does not change the value of the inequality by multiplying each side by a positive number and σ^2 is a positive number.

$$\sigma^2 \sum_{i=1}^{n} a_i^2 \geq \sigma^2 \frac{1}{n}(\sum_{i=1}^{n} a_i)^2$$

Substitute the variance found in part 2 on the left-hand side:

$$Var(W) \geq \sigma^2 \frac{1}{n}(\sum_{i=1}^{n} a_i)^2$$

When W is unbiased, we know that:

$$\sum_{i=1}^{n} a_i = 1$$

Substitute the above value in to the right-hand side:

$$Var(W) \geq \frac{\sigma^2}{n}$$

We know that:

$$Var(\bar{Y}) = \frac{\sigma^2}{n}$$

Substitute this value in to the right-hand side of the equation and the proof is complete.

$$\mathbf{Var(W) \geq Var(\bar{Y})}$$

Question 9

$W_1 = (\frac{n-1}{n})\bar{Y}$

$W_2 = k\bar{Y}$

1)

$$bias(W_1) = E((\frac{n-1}{n})\bar{Y}) - \mu$$

$$bias(W_1) = \frac{n-1}{n}E(\bar{Y}) - \mu$$

$$bias(W_1) = \frac{n-1}{n}\mu - \mu$$

$$bias(W_1) = \mu(\frac{n-1}{n} - 1)$$

$$\mathbf{bias(W_1)} = \frac{-\mu}{\mathbf{n}}$$

Similarly:

$$bias(W_2) = E(k\bar{Y}) - \mu$$

$$bias(W_2) = kE(\bar{Y}) - \mu$$

$$bias(W_2) = k\mu - \mu$$

$$\mathbf{bias(W_2)} = \mu(\mathbf{k-1})$$

Which is a consistent estimator?
W_1 is a consistent estimator of μ because as n goes to ∞, the difference between W_1 and μ goes to 0.

2)

$$Var(W_1) = Var((\frac{n-1}{n})\bar{Y})$$

$$Var(W_1) = \frac{(n-1)^2}{n^2}Var(\bar{Y})$$

$$Var(W_1) = \frac{(n-1)^2}{n^2}\frac{\sigma^2}{n}$$

$$\mathbf{Var(W_1)} = \frac{\mathbf{(n-1)^2}\sigma^{\mathbf{2}}}{\mathbf{n^3}}$$

Similarly:

$$Var(W_2) = Var(k\bar{Y})$$
$$Var(W_2) = k^2 Var(\bar{Y})$$
$$Var(W_2) = k^2 \frac{\sigma^2}{n}$$
$$\mathbf{Var(W_2) = k^2 \frac{\sigma^2}{n}}$$

Which estimator has lower variance?
The estimator that has lower variance depends on the values of n and k as follows:
For n = 1, k > 0, Var(W_1) is lower
For n > 1 and k = $\frac{n-1}{n}$, Var(W_1) = Var(W_2)
For n > 1 and k > $\frac{n-1}{n}$, Var(W_1) is lower
For n > 1 and k < $\frac{n-1}{n}$, Var(W_2) is lower

Question 10

$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}$(Y_{i} - $\bar{Y})^2$

1)

$$E(\bar{Y}) = E[\frac{\sum_{i=1}^{n} Y_i}{n}]$$

We know that E(X) is a linear function, thus:

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n} E[Y_i]$$

$$E(\bar{Y}) = \frac{1}{n}(n\mu)$$

$$\Rightarrow \mathbf{E(\bar{Y})} = \mu = \mathbf{E[Y_i], y = 1, \cdots, n}$$

2)

$$Var(\bar{Y}) = Var[\frac{\sum_{i=1}^{n} Y_i}{n}]]$$

We know that Var[$\sum_{i=1}^{n}$a$_i X_i$] = $\sum_{i=1}^{n} a_i^2$Var(X_i), thus:

$$Var(\bar{Y}) = \frac{1}{n^2}\sum_{i=1} n(Var[Y_i])$$

$$\Rightarrow \mathbf{Var(\bar{Y})} = \mathbf{\frac{1}{n^2} n}\sigma^{\mathbf{2}} = \mathbf{\frac{1}{n}Var[Y_i], y = 1, \cdots, n}$$

3)

$$\widehat{\sigma^2} = E[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2]$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} E[(Y_i - \bar{Y})^2]$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} E(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2)$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} E(Y_i^2) - 2E(Y_i)E(\bar{Y}) + \bar{Y}^2)$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} E(Y_i^2) - E(Y_i)^2 + E(\bar{Y}^2) - E(\bar{Y})^2$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} Var(Y_i) + Var(\bar{Y})$$

$$\widehat{\sigma^2} = \frac{1}{n}(n\sigma^2 + n\frac{\sigma^2}{n})$$

$$\widehat{\boldsymbol{\sigma^2}} = \frac{\mathbf{n+1}}{\mathbf{n}}\boldsymbol{\sigma^2}$$

4)
We have shown that E($\widehat{\sigma^2}$) != σ^2 and we conclude that $\widehat{\sigma^2}$ is a biased estimator for σ^2.

5)
We know that:

$$\widehat{\sigma^2} = \frac{n+1}{n}\sigma^2$$

is a biased estimator. Thus:

$$\frac{n}{n+1}\widehat{\sigma^2}$$

is an unbiased estimator of σ^2

Question 11

X, Y positive random variables. E(Y|X) = θ X

i)
We know that Z = $\frac{Y}{X}$
We first compute E(Z|X)

$$E(Z|X) = E(\frac{Y}{X}|X)$$

Using E(a(X) Y + b(X))|X) = a(X)E(Y|X) + bX, we derive:

$$E(Z|X) = \frac{1}{X}E(Y|X) = \frac{\theta X}{X} = \theta$$

Then, we knwo that E[E(Z|X)] = E(Z). Thus:

$$E(\theta) = E(Z)$$

Θ being a constant:

$$E(\theta) = \theta$$

and

$$\mathbf{E(Z)} = \theta$$

ii)
$W_1 = n^{-1}\sum_{i=1}^{n}\frac{Y_i}{X_i}(X_i, Y_i) : i = 1, 2, \cdots, n$ is the estimator.
We compute $E(W_1)$:

$$E(W_1) = n^{-1}E[\sum_{i=1}^{n}(\frac{Y_i}{X_i})]$$

$$E(W_1) = n^{-1}\sum_{i=1}^{n}E(Y_i/X_i)$$

$$\mathbf{E(W_1) = n^{-1}[nE(Z)] = E(Z)} = \theta$$

We conclude that W_1 is unbiased for θ

iii)
$W_2 = \frac{\bar{Y}}{\bar{X}}$

$$W_2 = \frac{n^{-1}\sum_{i=1}^{n}Y_i}{n^{-1}\sum_{i=1}^{n}X_i}$$

$$\Rightarrow W_2 = \frac{Y_1 + Y_2 + Y_3 + \cdots + Y_n}{X_1 + X_2 + X_3 + \cdots + X_n}$$

whereas:

$$W_1 = \frac{Y_1}{X_1} + \frac{Y_2}{X_2} + \frac{Y_3}{X_3} + \cdots \frac{Y_n}{X_n}$$

$W_2 and W_1$ are thus different estimators

$$E(W_2) = E[\frac{\bar{Y}}{\bar{X}}]$$

We know that

$$E(Y) = E(\bar{Y}) = E[E(Y|X)] = E[\theta X] = \theta E(X)$$

$$\Rightarrow E(W_2) = E[\frac{\theta E(X)}{E(X)}] = E(\theta)$$

$$\Rightarrow \mathbf{E(W_2)} = \theta$$

Question 12

i)
The null hypothesis is that $\mu = 0$
ii)
The alternative hypothesis hypothesis is that $\mu < 0$
iii)

$$\mu = 0, when\ the\ null\ hypothesis\ is\ true$$

$$n = 900$$

$$\bar{Y} = -32.8$$

$$s = 466.4$$

$$t = \frac{\bar{Y} - \mu}{\frac{s}{\sqrt{n}}} = -2.109777$$

$$p(z \leq t) = 0.0174$$

Thus, we reject the null hypothesis at the 5% significance level as $p(z \leq t) \leq 0.05$ We cannot reject the null hypothesis at the 1% significance level as $p(z \leq t) \geq 0.01$
iv)
The effect size is inferior to 10% of the variance of the State Liquor Consumption variable. That's an indication of a small practical effect. We also compute the correlation coefficient as

$$r = \sqrt{\frac{t^2}{t^2 + DF}}$$

$$r = \sqrt{\frac{-2.11^2}{-2.11^2 + 899}} = 0.07$$

The value of R also confirms the small practical effect despite the test being statistically significant because of the high sample size.
v)
What has been assumed is that the other determinates of liquor consumption have had no net effect over the two-year period that was analyzed.

Question 13

$Y_i = 1$ Shot made. $Y_i = 0$ Shot missed. $\theta = Pr(Makinga3ptshot)$ Bernouilli distribution. $\bar{Y} = \frac{FGM}{FGA}$ estimator of θ

i)

$$\theta = \frac{188}{429} = .4382284$$

ii)
Because Y has a Bernouilli distribution: $E(Y) = \theta$ Let $Y_i i \in 1, \cdots, n$ be an occurance of a free throw. We know that each Y_i is a Bernouilli variable. We can define:

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

Thus:

$$Var(\bar{Y}) = (\frac{1}{n})^2 Var(\sum_{i=1}^{n} Y_i)$$

$$Var(\bar{Y}) = (\frac{1}{n})^2 \sum_{i=1}^{n} Var(Y_i)$$

Thus:

$$Var(\bar{Y}) = (\frac{1}{n})^2 n\theta(1-\theta)$$

$$\Rightarrow Var(\bar{Y}) = \frac{\theta(1-\theta)}{n}$$

And:

$$sd(Y) = \sqrt{Var(\bar{Y})}$$

$$\Rightarrow sd(Y) = \sqrt{\frac{\theta(1-\theta)}{n}}$$

iii)

We know $se(\bar{\gamma}) = \sqrt{\frac{\bar{\gamma}(1-\bar{\gamma})}{n}}$ And $\frac{\bar{\gamma}-\theta}{se(\bar{Y})} \equiv Normal(0,1)$ We compute:

$$z_Y = \frac{\frac{188}{429} - .5}{se(\bar{Y})} = -2.588303$$

$$\Rightarrow p(z) = 0.0048$$

The p-value is significant at the 1% significance level and we reject the null hypothesis. iv)

Type I error is the probability of a false positive or the probability that the null hypothesis was rejected but it should not have been. It's the probability of saying there is a result when there is not one. v)

The probablity of the type error is the significance level and is 1%. vi)

Type II error is the probability of a false negative, or the probablity of failing to reject the null hypothesis when it should be rejected. vii)

viii)

The power of the test is the probabilty of rejecting the null hypothesis when its is actually false. ix)