Name: OUTLINE SOLUTIONS

University of Chicago
Graduate School of Business

Business 41000: Business Statistics
Solution Key

Special Notes:

1. This is a closed-book exam. You may use an 8×11 piece of paper for the formulas.
2. Throughout this paper, $N(\mu, \sigma^2)$ will denote a normal distribution with mean μ and variance σ^2.
3. This is a 1 hr 45 min exam.

Honor Code: By signing my name below, I pledge my honor that I have not violated the Booth Honor Code during this examination.

Signature:

Problem A. True or False: Please Explain your answers in detail. Partial credit will be given (50 points)

1. For independent random variables X and Y, we have $var(X-Y) = var(X) - var(Y)$.

   False. $Var(X-Y) = Var(X) + Var(Y)$

2. Suppose that you toss a coin 5 times. Then there are 10 ways of getting 3 heads.

   True. We see this by computing:

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

3. Suppose that you have a one in a hundred chance of hitting the jackpot on a slot machine. If you play the machine 100 times then you are certain to win.

   False. The expected waiting time until you hit the jackpot is 100 times, but since each outcome of the game is independent of all previous outcomes, there is no guarantee about the number of plays until a jackpot it hit.

4. In a Markov process, tomorrow's value is independent of today's.

False. Tomorrow's value is only dependent on today's by the Markov property.

5. A chip manufacturer needs to add the right amount of chemicals to make the chips resistant to heat. On average the population of chips needs to be able to withstand heat of 300 degrees. Suppose you have a random sample of 30 chips with a mean of 305 and a standard deviation of 8. Then you can reject the null hypothesis $H_0 : \mu = 300$ versus $H_a : \mu > 300$ at the 5% level.

True. Since the sample size is $n \geq 30$ the t test is effectively the same as the z test. Therefore, we find the following z value:

$$z = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{305 - 300}{8/\sqrt{30}} = 3.4233$$

Here we use a one-sided test:

$$z_\alpha = 1.65$$

Since $z > z_\alpha$, we can reject H_0 in favor of H_a at the 5% level.

6. There is a 95% probability that a normal random variable lies between $\mu \pm \sigma$.

False. There is a 95% probability that it lies between $\mu \pm 1.96\sigma$.

7. Bayes' rule states that $p(A|B) = p(B|A)$.

False. Recall Bayes' rule:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

8. A recent CNN poll found that 49% of 10000 voters said they would vote for Obama versus Romney if that was the election in November 2012. A 95% confidence interval for the true proportion of voters that would vote for Obama is then 0.49 ± 0.03.

False. The confidence interval for the proportion is given by:

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$$

This gives:

$$0.49 \pm 1.96\sqrt{(0.49)(0.51)/10000} = 0.49 \pm 0.009798$$

9. The soccer team Manchester United scores on average two goals per game. Given that the distribution of goals is Poisson, the chance that they score two or less goals is 87%

False. We have that :

$$E[g] = 2 \quad \Rightarrow \quad g \sim Pois(\lambda = 2)$$

So, we find the chance that they score two or fewer goals is:

$$P(g \leq 2) = 0.6767$$

In `R: ppois(2,2)=0.67667`

10. The Central Limit Theorem guarantees that the distribution of $\bar{X}$ is constant.

False. For some set of independent random variables, the mean of a sufficiently large number of them (each with finite mean and variance) will have a distribution which is Normal.

Problem B. (20 points)

A cable television company has 10000 subscribers in a suburban community. The company offers two premium channels, HBO and Showtime. Suppose 2750 subscribers receive HBO and 2050 receive Showtime and 6200 do not receive any premium channel

1. What is the probability that a randomly selected subscriber receives both HBO and Showtime.

$$\frac{2750 + 2050 + 6200 - 10000}{10000} = \frac{1000}{10000} = 0.10$$

2. What is the probability that a randomly selected subscriber receives HBO but not Showtime.

$$\frac{2750 - 1000}{10000} = \frac{1750}{10000} = 0.175$$

You now obtain a new dataset, categorized by gender, on the proportions of people who watch HBO and Showtime given below

Cable	Female	Male
HBO	0.14	0.48
Showtime	0.17	0.21

1. Conditional on being female, what's the probability you receive HBO?

$$P(HBO|F) = \frac{0.14}{0.14 + 0.17} = 0.4516$$

2. Conditional on being female, what's the probability you receive Showtime?

$$P(Show|F) = \frac{0.17}{0.14 + 0.17} = 0.5484$$

Problem C. (20 points)
The following table shows the descriptive statistics from 1000 days of returns on IBM and Exxon's stock prices.

	N	Mean	StDev	SE Mean
IBM	1000	0.0009	0.0157	0.00049
Exxon	1000	0.0018	0.0224	0.00071

Here is the covariance table

	IBM	Exxon
IBM	0.000247	
Exxon	0.000068	0.00050

1. What is the variance of IBM returns?

$$\sigma^2_{IBM} = 0.000247$$

2. What is the correlation between IBM and Exxon's returns?

$$\rho = \frac{cov(IBM, Exxon)}{\sigma_{IBM}\sigma_{Exxon}} = \frac{0.000068}{\sqrt{0.000247}\sqrt{0.00050}} = 0.1935$$

3. Consider a portfolio that invests 50% in IBM and 50% in Exxon. What are the mean and variance of the portfolio? Do you prefer this portfolio to just investing in IBM on its own?

$$\mu_p = w_I\mu_I + w_E\mu_E = (0.5)(0.0009) + (0.5)(0.0018) = 0.00135$$

$$\sigma^2_p = w^2_I\sigma^2_I + w^2_E\sigma^2_E + 2w_Iw_Ecov(I, E)$$
$$= (0.5)^2(0.000247) + (0.5)^2(0.00050) + 2(0.5)(0.5)(0.000068) = 0.000221$$

Yes, because the portfolio has a higher mean return but with a lower variance.

Explain your answers clearly.

Problem D. (20 points)

The quality of Nvidia's graphic chips have the probability that a randomly chosen chip being defective is only 0.1%. You have invented a new technology for testing whether a given chip is defective or not. This test will always identify a defective chip as defective and only "falsely" identify a good chip as defective with probability 1%

This give the following probabilities (D represents "defective chip" and T represents "test result indicates defective chip"):

$$P(D) = 0.001$$

$$P(T|D) = 1$$

$$P(T|\bar{D}) = 0.01$$

1. What are the sensitivity and specificity of your testing device?

   Sensitivity: $P(T|D) = 1$

   Specificity: $P(\bar{T}|\bar{D}) = 1 - P(T|\bar{D}) = 1 - 0.01 = 0.99$

2. Given that the test identifies a defective chip, what's the posterior probability that it is actually defective?

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})P(\bar{D})}$$

$$= \frac{(1)(0.001)}{(1)(0.001) + (0.01)(1 - 0.001)} = 0.090992$$

3. What percentage of the chips will the new technology identify as being defective?

$$P(T) = P(T|D)P(D) + P(T|\bar{D})P(\bar{D})$$

$$= (1)(0.001) + (0.01)(1 - 0.001) = 0.01099$$

4. Should you advise Nvidia to go ahead and implement your testing device? Explain.

   No, only 9% of those chips indicated to be defective by the test will actually be defective. Essentially, we would be throwing away 91% of the chips indicated to be defective even though they are perfectly fine!