Statistics 427: Sample Final Exam

Instructions: The following sample exam was given several quarters ago in Stat 427. The same topics were covered in the class that year. This sample exam is meant to be representative of the types of questions that might be asked on your final exam. Some of the material examined here might not show up on your final exam, and your final exam might have material that is not covered in the sample exam. In addition to working through these problems, be sure to review the lecture notes and the homework assignments.

Question 1

A store stocks a number of different personal computers. These can be classified as laptops or desktops and as 512Mb or 1Gb. Let $A$ be the event that the next computer sold is a laptop. Let $B$ be the event that the next computer sold is 1 Gb. It is known that 80% of recent sales have been 1Gb machines. It is also known that 70% of recent sales have been laptops. Of the laptop sales, 60% were 1Gb.

(a) What is the probability that the next computer sold will be a 1Gb laptop?

$$P(A \cap B) = P(B|A) P(A) = (0.60)(0.70) = 0.42$$

(b) What is the probability that the next 1Gb computer sold will be a laptop?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.42}{0.80} = 0.525$$

(c) Suppose that sales can be regarded as independent. Let $X$ be the number of 1Gb computers sold among the next $n = 10$ computers sold. Is it reasonable to assume that $X$ has a Binomial distribution? Why or why not?

1. $n = 10$ is fixed
2. $S = 1 \text{ Gb}$, $F = \text{ not } 1 \text{ Gb}$ on each trial
3. Sales are independent
4. $p = P(1 \text{ Gb}) = 0.8$ if we always have 1Gb computers in stock.

$\Rightarrow X$ has a binomial distribution.
(d) Assume that $X$ in part (c) does, in fact, have a binomial distribution (no matter what you answered in part (c)). What is the probability that 7 of the next 10 sales will be for 1Gb computers?

$$P(X = 7) = \binom{10}{7} (0.8)^7 (1-0.8)^3 = 120 (0.8)^7 (0.2)^3$$

$$= 0.2013$$

$$\text{So} \quad P(X = 7) = P(X \leq 7) - P(X \leq 6) = 0.322 - 0.121 = 0.2013 \text{ from table.}$$

(e) Let $Y$ be the number of 1Gb computers sold among the next $n = 100$ computers sold and suppose that sales are still independent. Find $P(60 < Y < 75)$. Show your work very carefully.

We can use normal approximation to binomial.

$$P(60 < Y < 75)$$

$$\mu_Y = np = 80 \quad \sigma_Y = np(1-p) = 16$$

Let $T \sim N(80,16)$. Then:

$$P(60 < Y < 75) \approx P(61 \leq Y \leq 74)$$

Normal approx. with continuity correction

$$\approx P(60.5 < T < 74.5)$$

$$= P(T < 74.5) - P(T < 60.5)$$

$$= P\left(Z < \frac{74.5-80}{4}\right) - P\left(Z < \frac{60.5-80}{4}\right) , \quad Z \sim N(0,1)$$

$$= \Phi(-1.375) - \Phi(-4.875)$$

$$\approx 0.085 - 0 = 0.085 \text{ (approximately)}$$
Question 2

Two components of a minicomputer have the following joint distribution for their useful lifetimes:

\[ f(x, y) = \begin{cases} 
3e^{-x-3y} & \text{for } x > 0, y > 0, \\
0 & \text{otherwise.}
\end{cases} \]

The marginal pdf of \( Y \) can be shown to be (you do not have to derive this):

\[ f_Y(y) = \begin{cases} 
3e^{-3y} & \text{for } y > 0, \\
0 & \text{otherwise.}
\end{cases} \]

(a) Find the marginal pdf of \( X \).

\[
\begin{align*}
\int_x f(x, y) \, dy &= \int_0^\infty 3e^{-x-3y} \, dy \\
&= e^{-x} \left[ -e^{-3y} \right]_0^\infty \\
&= e^{-x} \left[ 0 - 0 \right] \\
&= e^{-x} \quad \text{for } x > 0 \\
&= 0 \quad \text{o.w.}
\end{align*}
\]

(b) Are \( X \) and \( Y \) independent? Why or why not?

\[
f_X(x) f_Y(y) = e^{-x} \cdot (3e^{-3y}) = 3e^{-x-3y} = f(x, y)
\]

So \( X \) and \( Y \) are independent.

(c) Find \( P(-1 < Y < 2) \).

\[
P(-1 < Y < 2) = P(0 < Y < 2) = \int_0^2 f_Y(y) \, dy \\
= \int_0^2 3e^{-3y} \, dy = -e^{-3y} \bigg|_0^2 = 1 - e^{-6} = 0.9975
\]

(d) Find \( P(X = 2.1, Y = 1.7) \).

\[
P(X = 2.1, Y = 1.7) = 0 \quad \text{because } X \text{ and } Y \text{ are continuous RVs.}
\]
(e) What is the cdf of $Y$? Use this to find $P(Y < 1.5)$.

$$F_Y(y) = \int_0^y 3e^{-3t} \, dt = -e^{-3t} \bigg|_0^y = \left\{ \begin{array}{ll} 1-e^{-3y} & y > 0 \\ 0 & \text{o.w.} \end{array} \right.$$ 

So $P(Y < 1.5) = F_Y(1.5) = 1-e^{-3(1.5)} = 0.9888$

Question 3

Customers entering a particular computer store can be regarded as randomly selected customers. Define two random variables as follows. On a randomly selected weekday:

$X =$ number of customers buying a computer

$Y =$ number of customers buying a printer

The joint probability distribution of $X$ and $Y$ is known to be:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.15</td>
<td>0.09</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>0.13</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>0.09</td>
<td>0.11</td>
<td>0.07</td>
</tr>
</tbody>
</table>

(a) The store gets a bonus from the manufacturer if $x^2y$ on a given weekday is bigger than 12. Calculate $E(X^2Y)$ for this store.

$$E(X^2Y) = 0 + (1)^2(1)(0.13) + (1)^2(2)(0.21) + (1)^2(3)(0.02) + (2)^2(1)(0.09) + (2)^2(2)(0.11) + (2)^2(3)(0.07)$$

$$= 2.69$$
In the following parts, you may need some of the following information:

\[ E(X) = 0.94, \quad E(Y) = 1.45, \quad \text{Var}(X) = 0.6164, \quad \text{Var}(Y) = 0.8475, \quad E(XY) = 1.65. \]

(b) Using the pmf of \( X \), verify that \( \text{Var}(X) = 0.6164 \) and then find the standard deviation of \( X \).

\[
\text{Var}(X) = E(X^2) - E(X)^2
\]

\[
E(X^2) = 1^2 \cdot 0.38 + 2^2 \cdot (0.28) = 1.5
\]

\[
\text{Var}(X) = 1.5 - (0.94)^2 = 0.6164
\]

\[
\text{SD}(X) = \sqrt{0.6164} = 0.7851
\]

(c) Calculate the covariance of \( X \) and \( Y \).

\[
\text{Cov}(X,Y) = E(XY) - E(X)E(Y)
\]

\[
= 1.65 - (0.94)(1.45)
\]

\[
= 0.287
\]

(d) Calculate the correlation between \( X \) and \( Y \).

\[
\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{0.287}{\sqrt{0.6164 \times 0.8475}} = 0.3871
\]

(e) The correlation in part (d) is positive. Why is the following statement false?

"A decrease in the number of computers bought causes a decrease in the number of printers bought." (A brief answer is fine).

Correlation gives no information about causation.
Question 4

The time to failure (measured in hundreds of hours) for a particular mechanical component is known to have a Gamma distribution with mean 20 (hundred hours) and variance 200 (hundred hours squared).

(a) Find the parameters of the Gamma distribution.

\[
\text{Mean} \rightarrow \alpha \beta = 20 \quad \beta = 10 \quad \alpha = 2
\]

\[
\text{Variance} \rightarrow \alpha \beta^2 = 200
\]

(b) What is the probability that a randomly selected mechanical component will last more than 10 (hundred hours)?

\[
P(X > 10) = 1 - P(X \leq 10) = 1 - F(10; \alpha = 2, \beta = 10)
\]

\[
= 1 - P\left(\frac{10}{10}; \alpha = 2\right) = 1 - 0.264 = 0.736
\]

from table

(c) Suppose \( n = 60 \) such components are tested. Let \( \bar{X} \) be the sample mean time to failure for the 60 components. What are the mean and variance of \( \bar{X} \) (measured in the same units as above)?

\[
\mu_{\bar{X}} = \mu = 20
\]

\[
\sigma^2_{\bar{X}} = \frac{\sigma^2}{n} = \frac{200}{60} = 3.333
\]

(d) What is the probability that the sample mean time to failure for the 60 components is greater than 10 (hundred hours)? State your choice of distribution carefully.

\( n = 60 \) is fairly large, so we could use Central Limit Theorem.

\( \bar{X} \) has approx. \( N(20, 3.333) \) distribution

\[
P(\bar{X} > 10) \approx P\left(Z > \frac{10 - 20}{\sqrt{3.333}}\right) = 1 - P\left(Z < \frac{-10}{1.8257}\right)
\]

\[
= 1 - \Phi(-5.773)
\]

\[
\approx 1 - 0 = 1.
\]
The salaries (measured in thousands) of employees in a large company follow a lognormal distribution with parameters $\mu = 3$ and $\sigma^2 = 2.56$. Let $X$ be the salary of a randomly selected employee, so that $Y = \ln(X)$ so that $Y$ has a normal distribution. Let $X$ be a salary of $\$20,000$. Let $X = 3$ and variance $\sigma^2 = 2.56$.

(a) What is the mean of $X$?

\[
\mu_X = e^{\mu + \frac{\sigma^2}{2}} = e^{3 + \frac{2.56}{2}} = 72.2404
\]

re. $\$72,240.4$

(b) Find the probability that a randomly selected employee has salary less than $\$100,000$.

\[
P(X < 100) = P(\ln(X) < \ln(100)) = P(\ln(X) < 4.6052)
\]

\[
= P\left(z < \frac{4.6052 - 3}{\sqrt{2.56}}\right)
\]

\[
= \Phi(1.0033)
\]

\[
= 0.8413
\]

(c) Find the 90th percentile of the distribution of $Y$, and then the 90th percentile of the salary distribution.

Want $P(Y < c) = 0.90$

\[
P(z < \frac{c - 3}{\sqrt{2.56}}) = 0.90
\]

From normal table, $\frac{c - 3}{\sqrt{2.56}} = 1.28 \Rightarrow c = 5.048$

Now $X = e^Y$ So 90th percentile of the salary distribution is $e^{5.048} = 155,710.7$

or $\$155,710.70$
Question 6

Shortly after being put into service, some buses manufactured by a certain company developed cracks. Suppose a particular city has 30 of these buses and cracks are now visible in 9 of them.

(a) If a sample of 10 buses are selected for thorough inspection, what is the probability that exactly 6 will have visible cracks? [Note: the formula (with numbers) is fine. There is no need to calculate the final numerical answer.]

\[ P(6 \text{ have cracks}) = \frac{\binom{9}{6} \binom{21}{4}}{\binom{30}{10}} = 0.0167 \]

(b) If a sample of 10 buses are selected for thorough inspection, what is the probability that 5 or 6 will have visible cracks? [Note: the formula (with numbers) is fine. There is no need to calculate the final numerical answer.]

\[ P(5 \text{ or } 6 \text{ have cracks}) = \frac{\binom{9}{5} \binom{21}{5}}{\binom{30}{10}} + \frac{\binom{9}{6} \binom{21}{4}}{\binom{30}{10}} \]