Revision Paper Suggested Answer

<u>Question 1</u> (25 marks)

a. 19-21

Normal distribution is bell-shaped and symmetric, with mean equals to the mode, whereas the mode is the value which most frequently occurred. So, the range including the mode is most likely to happen.

b. P(X > 30) = P (Z > $\frac{30-20}{5}$)
= P (Z > 2.0)
= 1 - P (Z $\leq$ 2.0)
= 1 - 0.9772
= 0.0228

c. Let X = number of free pizza

X follows binomial (n=5, pi=0.0228)

P(X $\geq$ 1) = 1 - P(X = 0)
= 1 - $\frac{5!}{0!(5-0)!}(0.0228)^0(1-0.0228)^5$
= 1 - 0.8911
= 0.1089

d. P(x_{lower} < X < x_{upper}) = 0.60

P($\frac{x_{lower}-20}{5}$ < Z < $\frac{x_{upper}-20}{5}$) = 0.60

For symmetric distribution of probability (0.6) on both sides of μ,

P(Z < -0.84) = 0.2 and P(Z < 0.84) = 0.8

Therefore, $\frac{x_{lower}-20}{5}$ = -0.84 → x_{lower} = -0.84 × 5 + 20 = 15.8

$\frac{x_{upper}-20}{5}$ = 0.84 → x_{upper} = 0.84 × 5 + 20 = 24.2

e. As X ~ N(20,5^2), $\bar{X}$ ~ N(20, $\frac{5^2}{26}$)

P(18 $\leq \bar{X} \leq$ 23)

= P($\frac{18-20}{\frac{5}{\sqrt{26}}} \leq Z \leq \frac{23-20}{\frac{5}{\sqrt{26}}}$)

= P(-2.0396 $\leq Z \leq$ 3.0594)

= 0.99889 – 0.0207

= 0.9782

f. P($\bar{X}$ > μ + 2) < 0.05

1 - P($\bar{X} \leq \mu$ + 2) < 0.05

P($\bar{X} \leq \mu$ + 2) > 0.95

P(Z $\leq \frac{2}{10/\sqrt{n}}$) > 0.95

$\frac{\sqrt{n}}{5} \geq$ 1.645

n $\geq$ 67.6506 $\approx$ 68 (round up)

<u>Question 2</u> (25 marks)

a. min = 20 Q1 = 24 Q2 = 31 Q3 = 34 max = 47

The sample data set is right-skewed.

b. Normal population assumption is needed. This is because the sample is drawn from an unknown population distribution, and n = 10 <30, hence Central Limit Theorem is not applicable
With known $\sigma = 7$, Z-distribution is used in conducting inferential analysis

c. 95% CI for $\mu = 30.8 \pm 1.96 \times \frac{7}{\sqrt{10}}$ = [26.4614, 35.1386] µg/l
It is 95% confidence that the mean DMS odor threshold among all oenologists is between 26.4614 and 35.1386 ug/l.

d. H_0: $\mu \le 25$ H_1: $\mu > 25$
At $\alpha = 0.05$, reject H_0 if $Z > 1.645$
$Z = \frac{30.8-25}{\frac{7}{\sqrt{10}}} = 2.6202$
Reject H_0. There is sufficient evidence that the mean odor threshold for oenologists is higher than the published threshold, 25µg/l

e. p-value = $P(Z \geq 2.6202) = 1 - 0.9956 = 0.0044$

<u>Question 3</u> (25 marks)

a. P(F and O)
= P(O) - P(M and O)
= P(O) - P(O|M)×P(M)
= 0.34 - 0.35×0.76
= 0.074

b. P(M | O)
= P(M and O) / P(O)
= P(O|M)×P(M) / P(O)
= 0.35×0.76 / 0.34
= 0.7824

c. P(F and O) = 0.074
P(F)×P(O) = 0.24 × 0.34 = 0.0816
As P(F)×P(O) ≠ P(F and O), overweighted and gender is not statistically independent (or they are dependent).
-- any other reasonable method showing the two events are not independent also acceptable

d. n = 500 > 30
np = 500*0.34 = 170 > 5
n(1-p) = 500*(1 - 0.34) = 330 > 5
∴ Sampling distribution of p is approximately normal, $p \sim N(0.3, \frac{0.34*0.66}{500})$

85% CI for π = $p \pm Z\alpha/_2\sqrt{\frac{p(1-p)}{n}}$ = $0.34 \pm 1.44\sqrt{\frac{0.34(1-0.34)}{500}}$ = [0.3095, 0.3705]
We are 85% confidence that the population proportion of overweighted members are between 0.3095 and 0.3705.

e. When conducting a hypothesis test, π_0 should be used to compute the standard error. When $\pi_0 = 0.3$, which is smaller than p = 0.34, the standard error decreases, the confidence interval becomes narrower, or the Z-test statistics becomes more extreme, leading to a higher chance to reject the null hypothesis. Concluding there is sufficient evidence that the population proportion is not 0.3.

f. $n = \frac{Z^2 \times p \times (1-p)}{e^2} = \frac{(1.645)^2 \times 0.34 \times (1-0.34)}{0.02^2}$
$n = 1518.08 \approx 1519$

<u>Question 4</u>(25 marks)

a. $\hat{Y} = -0.5802 + 15.0352X$
When the number of copiers serviced increased by one, the service time is predicted to increase by 15.0352 minutes.

b. $r = \sqrt{R^2} = \sqrt{0.9575} = 0.9785$
There is a very strong positive linear relaionship between variable Y and X.

c. Covariance depends on the the units used to measure X and Y and thus usually cannot be directly compared for different variables. Correlation coefficient is a “standardized score” of the covariance and it is unit-free

d. 95.75% of the vairation in Y has been explained by the estiamated regression equation.

e. H_0: $\beta_1 \leq 0$ H_1: $\beta_1 > 0$
At $\alpha = 0.05$, reject H_0 if $t > 1.6811$ with df = 43
$t = \frac{15.0352}{0.4831} = 31.1223$
Since $t = 31.1223 > 1.6811$, we reject H_0.
There is sufficient evidence that there is a positive relationship between Y and X.

f. (i) $\hat{Y} = -0.5802 + 15.0352(1) = 14.455$ minutes

(ii) $\hat{Y} = -0.5802 + 15.0352(5) = 74.5958$ minutes

g. Part (ii)’s prediction is more justifiable because the value of X = 5 lies within the observed range.