STATISTICS MA1972
SUMMER 2006 - **SOLUTIONS**

SECTION A

A1 **a)** The distribution of the sample mean is also normal with mean = 4 and variance = $0.6^2/16 = 0.0225$; i.e.

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N(4, 0.0225)$$

b) $\mathbf{P(\overline{X} > 3.8)} = P\left(Z > \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right) = P\left(Z > \frac{3.8 - 4}{0.6/\sqrt{16}}\right) = P\left(Z > \frac{-0.2}{0.15}\right)$

$= P(Z > -1.33) = = 1 - P(Z < -1.33) = 1 - 0.0918$
$= \mathbf{0.9082}$

A2 The rv X for passengers who arrive is B(180, 0.9). Use normal approximation with $\mu = np = 180(0.9) = 162$ $\sigma^2 = npq = 16.2$, $\sigma = 4.025$

a) P(more passengers than seats) = $P(X > 170) = 1 - P(X \le 170)$

$$= 1 - P\left(Z < \frac{170.5 - 162}{4.025}\right) = 1 - P(Z < 2.11) = 1 - 0.9826 = 0.0174$$

There is a **1.7%** probability that there are unsatisfied passengers.

b) P($5 \le X \le 10$ unfilled seats) = P($160 \le X \le 165$ filled seats)

$$= P\left(\frac{159.5 - 162}{4.025} \le Z \le \frac{165.5 - 162}{4.025}\right) = P(-0.62 \le Z \le 0.87)$$

$= 0.8078 - 0.2676 = 0.5402$

There is a **54**% probability that between 5 and 10 seats are unfilled

A3

a) Since f(x) is a probability density function

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 c x^2 + x \;\; dx = 1$$

But $\int_0^1 c x^2 + x \;\; dx = \left[\frac{cx^3}{3} + \frac{x^2}{2}\right]_0^1 = \frac{c}{3} + \frac{1}{2} = 1 \Rightarrow$ **c=1.5**

b) Mean of X is $\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)\,dx = \int_0^1 1.5x^3 + x^2 dx = \left[\frac{1.5x^4}{4} + \frac{x^3}{3}\right]_0^1$

$= \frac{3}{8} + \frac{1}{3}$ **= 0.7083(** $\frac{17}{24}$ **)**

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 1.5x^4 + x^3 dx = \left[\frac{1.5x^5}{5} + \frac{x^4}{4}\right]_0^1 = \frac{3}{10} + \frac{1}{4}$ **= 0.55**

$V(X) = E(X^2) - E^2(X) = 0.55 - 0.7083^2 =$ **0.0483**

A4 If n = 40, df = 39 $S^2 = 16.4$

For a 95% CI $L_{0.025} = 23.65$, $U_{0.025} = 58.12$

The 95% CI for σ^2 is $= \left[\frac{(n-1)s^2}{\mathbf{U2.5\%}}, \frac{(n-1)s^2}{\mathbf{L2.5\%}}\right] = \left[\frac{(39)16.4}{58.12}, \frac{(39)16.4}{23.65}\right]$

$= [11.0048, 27.0444]$

SECTION B

B1.a) (i) $\bar{X} = 53.4$ days S = 24.8 days

The approx 95% C I is $\bar{x} \pm z_{\alpha/2}\frac{S}{\sqrt{n}} = \bar{x} \pm 1.96\frac{S}{\sqrt{n}}$

$$= 53.4 \pm 1.96\frac{24.8}{\sqrt{80}} = 53.4 \pm 1.96(2.77)$$

$$= 53.4 \pm 5.43 \qquad \mathbf{=[\ 47.97 , 58.83]\ days}$$

<u>We are approximately 95% confident that the mean overdue time is between 48 and 58.8 days</u>

(ii) Candidiates should mention the CLT. For this sample size of 80 the distribution of sample means will be approximatley normal

b) (i) $n_1 = 250$, $p_1 = 42/250 = 0.168$, $n_2 = 200$, $p_2 = 48/200 = 0.24$

$$\hat{\pi} = \frac{n_1p_1 + n_2p_2}{n_1 + n_2} = \frac{90}{450} = 0.20$$

H_0: There is no difference in the proportion of accounts paid within 30 days between the two schemes. $\pi_1 = \pi_2$

H_1: There is an increase in the proportion of accounts paid within 30 days with 1% incentive. $\pi_1 < \pi_2$

<u>One tailed test, 5% critical value is – 1.645</u>

Calculate the test statistic, $Z = \dfrac{0.168 - 0.24}{\sqrt{0.20(0.80)(\frac{1}{250} + \frac{1}{200})}} = \dfrac{-0.072}{\sqrt{(0.16)(0.009)}}$

$$= \frac{-0.072}{\sqrt{(0.00144)}} = \frac{-0.072}{0.03795} = \mathbf{-1.90}$$

We take the numerical value of Z = -1.90 < -1.645 , so H_0 is rejected.

<u>We conclude that there is evidence to suggest that the proportion of invoices paid within 30 days has been increased as a result of the 1% discount incentive</u>

(ii) Since Ho was rejected, the 90 % confidence interval for π_2 is

$$0.24 \pm 1.645\sqrt{\frac{p2(1-p2)}{n}} = 0.24 \pm 1.645\sqrt{\frac{0.24(0.76)}{200}}$$

$$= 0.24 \pm 1.645(0.0302) \qquad = 0.24 \pm 0.0497$$

So the proportion of invoices paid within 30 days has increased to between **19.03% to 24.97%**

B2 a) i) Brand A $n_1 = 15$ $\bar{X}_1 = 255$ $S_1 = 8.7$

Brand B $n_2 = 15$ $\bar{X}_2 = 271.33$ $S_2 = 11.24$

ii) State hypotheses

H_0: Mean distances are the same for both brand of golf club $\mu_1 = \mu_2$

H_1: Mean distance for brand B is greater than brand A $\mu_1 < \mu_2$ (one tail)

We will assume that the distribution of distances is normal and the population variances are unknown but equal. Since the sample sizes are small we will use the t test with d.f. = $n_1 + n_2 - 2 = 28$, and pooled sample standard deviation, S, where

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)} = \frac{14(8.7)^2 + 14(11.24)^2}{14 + 14 - 2} = \frac{14(75.69) + 14(126.34)}{28} = \frac{2828.42}{28} = \mathbf{101.015}$$

Then the critical value of t for $\alpha = 5\%$ with d.f. = 28 is $t_{.05} = \mathbf{-1.701}$.

So we reject H0 if $T < -1.701$

Since $S^2 = 101.015$. **S = 10.05**

The test statistic is $$T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{255 - 271.33}{10.05\sqrt{\frac{1}{15} + \frac{1}{15}}} = \frac{-16.33}{10.05(0.3651)} = \frac{-16.33}{3.6692} = \mathbf{-4.45}$$

$-4.45 < t_{.05} = -1.701$, so we **reject Ho**

<u>We conclude that there is evidence that the distances reached with brand B are greater than for brand A</u>.

There is evidence to support the customer's view that brand B clubs could improve his game. Is the extra cost worth it? – looking for sensible comments here.

b (i) We wish to test the hypotheses

Ho: The variance in distances are the same for both brands $\sigma_1^2 = \sigma_2^2$

H1: The variance in distances are not the same for each brand $\sigma_1^2 \neq \sigma_2^2$

At $\alpha = 5\%$ significance level. The right-tailed critical value of F with $v_2 = (15 - 1) = 14$ and $v_1 = (15 - 1) = 14$ $U_{0.025} = 2.86$.

The test statistic is $F = \frac{s_2^2}{s_1^2} = \frac{(11.24)^2}{(8.7)^2} = \mathbf{1.669}$ We accept Ho: $\sigma_1^2 = \sigma_2^2$.

<u>Thus, there is no evidence to suggest unequal variances.</u>

(ii) <u>NONE the samples sizes are EQUAL so the t test is robust against departure from equal variances anyway.</u>