STATISTICS MA1972
SUMMER 2007 - **SOLUTIONS**

SECTION A

A5 $\overline{X} = 8.7$ mins S = 2.3 mins

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

$= 8.7 \pm 1.96\frac{2.3}{\sqrt{150}} = 8.7 \pm 1.96(0.19)$

$= 8.7 \pm 0.37$ mins or **[8.3 , 9.1] minutes**

We are 95 % confident that average time taken to perform eye tests for all patients at this practice ais between 8.3 to 9.1 minutes

A5 T is normally distributed with

Mean = $E(T) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ...+E(X_n)$

$= \mu + \mu + ... + \mu \quad = \mathbf{n\mu}$

Variance = $V(T) = V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ...+V(X_n)$

$= \sigma^2 + \sigma^2 + ... + \sigma^2 \quad = \mathbf{n\sigma^2}$

So T is $N(n\mu, n\sigma^2)$

A7

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

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 cx^{\alpha-1}\,dx = 1$$

$$\text{But } \int_0^1 cx^{\alpha-1}\,dx = c\left[\frac{x^\alpha}{\alpha}\right]_0^1 = \frac{c}{\alpha} = 1 \Rightarrow \mathbf{c = \alpha}$$

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

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^1 \alpha x^{\alpha+1}\,dx = \left[\frac{\alpha x^{\alpha+2}}{\alpha+2}\right]_0^1 = \frac{\alpha}{\alpha+2}$$

$$V(X) = E(X^2) - E^2(X) = \frac{\alpha}{\alpha+2} - \left[\frac{\alpha}{\alpha+1}\right]^2 = \frac{\alpha}{(\alpha+2)(\alpha+1)^2}$$

c) $P(X > \mu) = P\left(X > \frac{\alpha}{\alpha+1}\right) = \int_{\frac{\alpha}{\alpha+1}}^{1} f(x)\,dx = \int_{\frac{\alpha}{\alpha+1}}^{1} \alpha x^{\alpha-1}\,dx = \left[x^\alpha\right]_{\frac{\alpha}{\alpha+1}}^{1} = 1 - \left(\frac{\alpha}{\alpha+1}\right)^\alpha$

A8 (a) State hypotheses

H_0: There is no association between age at graduation and employment status.

H_1: There is an association between age at graduation and employment status.

(b) The p value is 0.000, and so is < 0.001 The p value is very small. We reject H_0 and conclude that there is overwhelming statistically significant evidence that graduate's employment status depends on their age at graduation.

SECTION B

B3.a) $\sigma = 0.008$ litres

i) If $\mu = 1.005$ litres $P(X > 1.01) = P(Z > \frac{1.01 - 1.005}{0.008}) = P(Z > 0.625) = 1 - 0.7340$

$= 0.266$ **So 26.6% cartons overflow**

ii) $P(X < 0.990) = P(Z < \frac{0.990 - 1.005}{0.008}) = P(Z < -1.875) = 0.0303$

So 3.03% cartons will contain less than 0.990 litres

b) The distribution of $\bar{x} \approx N\left(\sum \frac{xi}{16}, \left(\frac{\sigma}{\sqrt{16}}\right)^2\right) = N\left(\frac{16x1.005}{16}, \left(\frac{0.008}{\sqrt{16}}\right)^2\right)$

$= N(1.005, 0.002^2)$

P(mean contents of 16 cartons < 1.000) = $P(\bar{x} < 1.000)$

$= P(Z < \frac{1.000 - 1.005}{0.002}) = P(Z < -2.5) = 0.0062$

There is a 0.62% chance that the mean contents of 16 cartons are less than 1 litre

c)

i) The time to failure has an exponential distribution with mean $1/\lambda = 250$ hours so $\lambda = 1/250$

P(weighing machine fails after one month(448hours)) = $P(X \geq 448) = e^{-448/250} = e^{-1.792}$ $= 0.16666$

There is a 17% chance that a weighing machine does not fail during a working month.

ii) P(weighing m/c fails within one week (112hours)) = $P(X \leq 112) = 1 - P(X \geq 112)$

$= 1 - e^{-112/250} = 1 - e^{-448}$

$= 1 - 0.6389 = 0.3609$

There is a 36% chance that a weighing machine fails within one week.

d)

i) n= 400 receipts P(error) = 32/400 = 0.08

Ho: Sample proportion is same as National Average $\pi = 0.05$

H1: Sample proportion is higher than National Average $\pi > 0.05$

accuracy deteriorated

$\alpha = 1\%$ so critical value (one tail) = 1.645

Test Statistic $Z = \frac{p - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{0.08 - 0.05}{\sqrt{\frac{0.05(0.95)}{400}}} = \frac{0.03}{0.010897} = 2.753$

Since Z > 1.645, we reject Ho.

There is evidence that accuracy has deteriorated

ii) To eliminate error to ≤ 1%, let n be the sample size such that

$\text{Error} = 1.645\sqrt{\frac{0.08(0.92)}{n}} \leq 0.01$

Then $\sqrt{n} \geq \frac{1.645}{0.01}\sqrt{0.08(0.92)} = 44.62$ so $n \geq (44.62)^2 = 1990.9$

So the sample size should be at least 1991 till receipts.

B4 a) Alternative $n_1 = 20$ $\overline{X}_1 = 501.7$ $S_1 = 10.116$

Current $n_2 = 20$ $\overline{X}_2 = 495.3$ $S_2 = 4.485$

We wish to test the hypotheses

Ho: The variances in time to failure are the same $\sigma_1^2 = \sigma_2^2$

H1: The variances in time to failure are not the same $\sigma_1^2 \neq \sigma_2^2$

At $\alpha = 5\%$ significance level. The right-tailed critical value of F with $v_1 = (20 - 1) = 19$ and $v_2 = (20 - 1) = 19$ $U_{0.025}$ is approx 2.56

The test statistic for the F test is: $F = \frac{s_1^2}{s_2^2} = \frac{(10.116)^2}{(4.485)^2} = \frac{102.333}{20.115} = 5.087$

Since the test statistic is outside the critical values, we reject Ho: $\sigma_1^2 = \sigma_2^2$.

Thus, there is evidence to suggest unequal variances.

b)

i) The first T test uses the pooled estimate of the population variance S^2 with the test statistic , $\mathbf{T} = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ ~ t_{38} under H_0, but is only valid for <u>two normal populations which have the same variance</u>. There is no evidence to suggest equal variances, so it is inappropriate to use the pooled variance. The boxplots do suggest that it is appropriate to assume that both populations are normal. Hence it is appropriate to use the second T test

$$T = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{26.19} \text{ under } H_0.$$

ii)

H_0: Both suppliers have the same mean time to failure $\mu_1 = \mu_2$

H_1: Alternative company has higher mean time to failure than Current $\mu_1 > \mu_2$ (one tail).

The test statistic T = 2.587, and the two tail p-value = 0.016, so the p-value of this one tail test is 0.008. (< 1%). So at the 5% level of significance we reject H_0, and conclude that the DVDs from the new company last longer. So there is evidence that the alternative supplier has a higher mean time to failure than the current supplier. So the company SHOULD SWITCH to the new supplier.

iii) The 95% confidence interval for the difference in mean time to failure between the two suppliers is $(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \pm t_{26,0.025}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$ $= 6.40 \pm 2.056\sqrt{\frac{10.116^2}{20} + \frac{4.485^2}{20}}$

$= 6.40 \pm 2.056\sqrt{5.1165 + 1.0075}$ $= 6.40 \pm 5.08$ **= [1.32,11.48]hours.**

So we are 95% confident that alternative supplier's DVDs last between 1.32 to 11.48 extra hours compared to the current supplier.