MA1972
JUNE 2008
SOLUTIONS to statistics questions

A5 a) $\sum_i x_i = \mathbf{261}$ Mean = $\dfrac{\sum_i x_i}{n} = \dfrac{261}{10} = \mathbf{26.1}$

b) $\sum_i x_i^2 = 6961$

The standard deviation = $\sqrt{\dfrac{\sum_i x_i^2 - \dfrac{\left(\sum_i x_i\right)^2}{n}}{n-1}} = \sqrt{\dfrac{6961 - \dfrac{(261)^2}{10}}{9}} =$

$\sqrt{\dfrac{6961 - 6812.1}{9}} = \sqrt{\dfrac{148.9}{9}} = \sqrt{16.5444} = \mathbf{4.0675}$

A6 $E[x] = \mu \qquad Var[x] = \sigma^2$

($E[nx] = nE[x] = \underline{\underline{n\mu}}$

$$Var[nx] = E\left[(nx)^2\right] - \left(E[nx]\right)^2$$
$$= n^2 E\left[(x)^2\right] - n^2\left(E[x]\right)^2$$
$$= n^2 Var[x] = \underline{\underline{n^2\sigma^2}}$$

A7 $f(x) = \begin{cases} \dfrac{c}{x^4} & x \geq 100 \\ 0 & x < 100 \end{cases}$

(a). $\int_{100}^{\infty} \dfrac{c}{x^4} dx = 1 \quad ie \quad c\left[\dfrac{x^{-3}}{-3}\right]_{100}^{\infty} = c\left[\dfrac{1}{3(100)^3}\right] = 1$

$$\therefore \quad c = 3(100)^3 = \underline{\underline{3(10^6)}}$$

(b) $E[x] = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{100} x(0)dx + \int_{100}^{\infty} \dfrac{3(10^6)}{x^3} dx$

$$= 0 + 3(10^6)\left[\frac{1}{-2x^2}\right]_{100}^{\infty} = 0 + 3(10^6)\left[0 + \frac{1}{2(100^2)}\right]$$

$$= \frac{3(10^6)}{2(10^4)} = \underline{\underline{150}}$$

$E[x^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = 0 + \int_{100}^{\infty} \dfrac{3(10^6)}{x^2} dx = 3(10^6)\left[\dfrac{1}{-x}\right]_{100}^{\infty}$

$$= \frac{3(10^6)}{10^2} = \underline{3(10^4)}$$

$$Var[x] = E[x^2] - (E[x])^2 = 3(10^4) - [(1.5)(10^2)]^2 = 10^4[3 - (1.5)^2] = \underline{7500}$$

A8 $\quad n = 100 \qquad p = \dfrac{59}{100} = 0.59 \qquad \therefore\ q = 0.41$

Since $n > 30$ the proportion of students in debt is $\approx$ normal

The 95% CI for π is $\quad p \pm z\sqrt{\dfrac{p(q)}{n}} = 0.59 \pm 1.96\sqrt{\dfrac{0.59(0.41)}{100}}$

$\qquad = 0.59 \pm 0.096$

$\underline{\underline{ie\ [0.494, 0.686]}}$

A9 $\quad \sigma$ is known = 16 $\qquad$ n = 15 $\qquad$ s = 13.03

H_0: Urban area standard deviation in IQ is the same as the whole country $\quad \sigma = 16$

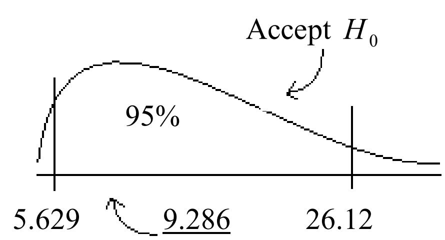
H_1: Urban area standard deviation in IQ is not the same as whole country $\quad \sigma \neq 16$

At 5% level – Critical values are from χ^2 distribution $\qquad \chi^2_{L2.5\%} = \underline{\underline{5.629}}$

$\qquad \chi^2_{U2.5\%} = \underline{\underline{26.12}}$

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{14(169.8)}{256} = \underline{\underline{9.286}}$$

$\therefore$ $\underline{\underline{\text{Accept } H_0}}$



<u>There is no evidence to suggest that the variability of IQ's in urban areas is different from the whole country.</u>

B3

a) State hypotheses

H_0: The new lacquer provides on average the same protection against rust $\mu_1 = \mu_2$

H_1: The new lacquer provides added protection against rust $\mu_2 > \mu_1$ (one tail)

State level of significance e.g. let $\alpha = 5\%$ then critical value = 1.64

$$\textbf{Test Statistic} \quad Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{730 - 480}{\sqrt{\frac{80^2}{256} + \frac{100^2}{50}}} = \frac{250}{\sqrt{25 + 200}}$$

$$= \frac{250}{\sqrt{225}} = \frac{250}{15} = \underline{\mathbf{16.667}}$$

Since Z > 1.64 reject H_0 and accept H_1, there is a significant result.

CONCLUSION: **there is evidence that the new lacquer provides added protection against rust.**

b)

	OBSERVED					EXPECTED		
Firm	No Satisfied Customers	No Not Satisfied	Total		Firm	No Satisfied Customers	No Not Satisfied	Total
A	576	24	600		A	565	35	600
B	558	42	600		B	565	35	600
C	580	20	600		C	565	35	600
D	546	54	600		D	565	35	600
	2260	140	2400			2260	140	2400

We have a 4 X 2 contingency table problem, and can test the hypotheses,

H_0: The coach firms are the same service

H_1: There is a difference between the coach firms.

If $\alpha = 5\%$ $\chi^2_{3 \times 1} = 7.815$

$$\chi^2 = \frac{\sum_i (O_i - E_i)^2}{E_i} = 1/565(11^2 + 7^2 + 15^2 + 19^2) + 1/35(11^2 + 7^2 + 15^2 + 19^2)$$

$= 756/565 + 756/35 = 1.3380 + 21.6 = \mathbf{22.938}$

So we reject Ho, <u>there is strong evidence that the firms have different levels of satisfaction amongst its customers</u>.

B4 a) British $n_1 = 20$ $\bar{X}_1 = 3.7$ $S_1 = 0.6$

American $n_2 = 15$ $\bar{X}_2 = 4.2$ $S_2 = 0.9$

State hypotheses

H_0: British & American companies spend on average the same on R&D $\mu_1 = \mu_2$

H_1:British companies do not spend as much on R&D as American companies $\mu_2 > \mu_1$ (one tail)

We will assume that the distribution of expenditure 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. = $n_1 + n_2 - 2 = 33$, and pooled sample standard deviation, S, where

$$S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{19(0.6)^2 + 14(0.9)^2}{20+15-2} = \frac{19(0.36) + 14(0.81)}{33} = \mathbf{0.55}$$

Then the critical value of t for $\alpha = 5\%$ with d.f. = $(n_1 + n_2 - 2) = 33$ is $t_{.05} = \mathbf{-1.692}$.

So we reject H0 if $T < -1.692$

Since $S^2 = 0.55$. $\mathbf{S = 0.7416}$

$$\text{The test statistic is} \quad T = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{3.7-4.2}{0.7416\sqrt{\frac{1}{20}+\frac{1}{15}}} = \mathbf{-1.97}$$

$-1.97 < t_{.05} = -1.692$, so we **reject Ho**

We conclude that there is evidence to support the view that British companies spend less on R & D than American companies.

b (i) We wish to test the hypotheses

Ho:The variances in expenditure are the same $\sigma_1^2 = \sigma_2^2$

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

At $\alpha = 5\%$ significance level. The right-tailed critical value of F with $v_2 = (15 - 1) = 14$ and $v_1 = (20 - 1) = 19$ $U_{0.025} = 2.65$.

$$\text{The test statistic is } F = \frac{s_2^2}{s_1^2} = \frac{(0.9)^2}{(0.6)^2} = \mathbf{2.25}$$

We accept Ho:$\sigma_1^2 = \sigma_2^2$.

Thus, there is no evidence to suggest unequal variances.

(ii) There is no evidence against equal variances, therefore the t test is valid.