Question Sheet 1.

EMa2 : APPLIED STATISTICS

20% of bolts defective $(p = 0.2)$.
A sample of 4 is taken. $(n = 4)$ and the distribution of events is binomial.

$P(X=0) = q^4 = (0.8)^4 = 0.4096$ (0 defective).

$P(X=1) = {}^4C_1\, p q^3 = \frac{4!}{1!\,3!}\, 0.2\,(0.8)^3 = 0.4096$ (1 defected)

$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$

$= 0.4096 + 0.4096 + \frac{4!}{2!\,2!}(0.2)^2(0.8)^2 = 0.9728.$

The probabilities on the rhs are added as the events are mutually exclusive.).

. Vibration test - 8 M/C on test. 80% known to survive after a large number of experiments.

$\therefore$ We have a binomial distribution - $n=8$, $p=0.8$.

$$P(X=3) = {}^8C_3\,(0.8)^3(0.2)^5 = \frac{8!}{5!\,3!}(0.8)^3(0.2)^5$$

$$= 0.0092. \quad \text{(remote)}$$

$$P(X=6) = {}^8C_6\,(0.8)^6(0.2)^2 = \frac{8!}{2!\,6!}(0.8)^6(0.2)^2$$

$$= 0.2936. \quad \text{(quite likely)}.$$

Coin tosses

Let S denote the event that no head is thrown.

then $P(S=1) = \frac{1}{2}$, Survival after ONE toss

$P(S=2) = \left(\frac{1}{2}\right)^2$, " " TWO tosses

and $P(S=n) = \left(\frac{1}{2}\right)^n$. " " N " tosses

We require $\left(\frac{1}{2}\right)^n < 0.01$, i.e. 99% chance of head.

whence $n=7$. $(2^7=128)$.

7. Of 800 cars, 64 are defective in a production.
Then the probability, p, of a defective at random = 0.08.
We wish to know the probability of 9 or more defectives out of a sample of 50 at random.

note. $p < 0.1$, n (sample) > 25. Use the Poisson approximation to the binomial distribution.

$np = 0.08 \times 50 = 4 \quad \therefore \mu = 4.$

$$\therefore P(X \geq 9) \doteqdot \sum_{x=9}^{50} e^{-4} \cdot \frac{\mu^x}{x!}$$

$$= e^{-4}\left\{\frac{4^9}{9!} + \frac{4^{10}}{10!} + \cdots\right\}$$

or preferably

$$1 - \sum_{x=0}^{8} P(x) = 1 - e^{-4}\left\{1 + \frac{4}{1!} + \frac{4^2}{2!} + \cdots + \frac{4^8}{8!}\right\}$$

$$= 0.0167.$$

. The 'MEAN NUMBER' is two/six months. n is not given but can safely be assumed to be large; on the other hand p is small. Then $np = \mu = 2$, and we can assume a Poisson distribution.

$$\therefore P(X=x) = \frac{e^{-2} 2^{x}}{x!}$$

In particular.

$$P(X=0) = e^{-2} = 0.1353$$

$$P(X=1) = 2 . e^{-2} = 0.2707 \quad \text{etc.}$$

Cosmic particle strikes. — $\rho = 0.02/\text{sec}$.

In one minute the mean number of strikes is

$60 \times 0.02 = 1.20$.

The probability distribution of strike numbers is thus

Poissonian with $\mu = 1.2$.

Then

$$P(X=0) = e^{-1.2} = 0.3012.$$

. Spare part usage. (random failure, events independent)
Mean usage (μ). 1 per month - Poisson

$$P(X=3) = \frac{e^{-1} \cdot 1^3}{3!} = 0.0613.$$

To ensure that the level of spares is such that there is less than 1% chance of running out in a one month period we must ensure that the probability of cover is > 0.99, i.e.

$$P(X = 0, 1, 2, \ldots n) > 0.99.$$

i.e. $P(X=0) + P(X=1) + \ldots P(X=n) > 0.99.$

$P(X=0) = e^{-1} = 0.3679$
$P(X=1) = 1 . e^{-1} = 0.3679$
$P(X=2) = \frac{1}{2} . e^{-1} = 0.1840$
$P(X=3) = \frac{1}{3} . P(X=2) = 0.0613$
$P(X=4) = \frac{1}{4} . P(X=3) = 0.0153$
$P(X=5) = \frac{1}{5} P(X=4) = 0.0031$

Clearly, if 5 or more spares are carried the chance is < 0.01 of being caught without a spare.

3. γ rays are emitted on average 4.2 millisec.

a. P(none in 6/m sec). Mean, $\mu = \frac{6}{4.2} = 1.429$

$P(X=0) = e^{-1.429} = 0.2397$

b. In 8 m/sec, $\mu = \frac{8}{4.2} = 1.905$

P(At least one) $= 1 - P(X=0)$

$= 1 - e^{-1.905} = 1 - 0.1489 = 0.8511.$

c. 5 in 16 m/sec. $\mu = \frac{16}{4.2} = 3.810.$

$P(X=5) = e^{-3.810} \times \frac{(3.810)^5}{5!} = 0.1482.$

Cumulative distribution function –

$$F(x) = 1 - e^{-0.2x} \quad , x \geq 0.$$

$$= 0 \quad , x < 0.$$

a. p.d.f. – derivative, namely –

$$F'(x) = 0.2e^{-0.2x}. \quad (f(x) \text{ say}).$$

b. Probability that the aeroplane is more that 10 minutes late

i)

$$P(X \geq 10) = \int_{10}^{\infty} f(x)\,dx = F(\infty) - F(10)$$

$$= e^{-0.2 \times 10} = e^{-2} = 0.1353.$$

). Time to failure of a component T has pdf

$$f_T(t) = \alpha e^{-\alpha t} \quad , t \geq 0$$

$$= 0 \quad , t < 0.$$

$$P(k \leq X \leq k+1) = \int_k^{k+1} \alpha e^{-\alpha t}\,dt = -e^{-\alpha(k+1)} + e^{-\alpha k}$$

$$= e^{-k\alpha}(1 - e^{-\alpha}).$$

Mean time to failure $= \int_0^{\infty} \alpha t e^{-\alpha t}\,dt = \left[\alpha t \cdot -\frac{1}{\alpha} e^{-\alpha t}\right]_0^{\infty} + \int_0^{\infty} e^{-\alpha t}\,dt$

$\frac{1}{\alpha}$. ∴ $P(X \geq 200)$, given $\alpha = \frac{1}{100}$, $= \int_{200}^{\infty} e^{-\frac{t}{100}}\,dt = e^{-2}$.

Normal Distribution

Question Sheet 2

Missile misses,

Mean -15, Variance $25\ m^2$.

μ: $\qquad \sigma^2$.

$SD = 5m$.

$Z = \frac{X-\mu}{\sigma} = \frac{X+15}{5}$ $\quad$, $\therefore P(X<-20) = P\left(Z < \frac{-20+15}{5}\right)$

$= P(Z<-1)$

$= 0.159$.

. Lifetime T: $\mu = 200$, $\sigma = 9$.

$P(T>220) = P\left(Z > \frac{220-200}{9} = \frac{20}{9}\right)$ $\qquad \frac{T-200}{9} = Z$

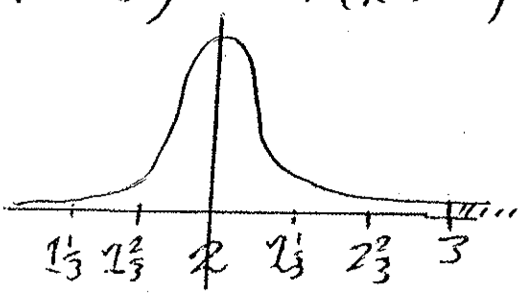
$= 0.0132$.

$P(T>220 \mid T>210) = \frac{P(T>220)}{P(T>210)} \doteq \frac{P(Z>\frac{20}{9})}{P(Z>\frac{10}{9})} \doteq 0.1$.

X: $\mu = 2$, $\sigma = \frac{1}{3}$.

$P(X>3) = P\left(Z > \frac{3-2}{1/3}\right) = P(Z>3) = 0.001$.

$P(2<X<3) = P(X>2) - P(X>3) = 0.5 - 0.001 = 0.499$.



90% of tubes are to last 80 hours. Mean = 100.

We want $P(X > 80) = 0.1$ $\quad (X = \sigma Z + \mu)$

$$\therefore P(\sigma Z + 100 > 80) = P\left(Z > -\frac{20}{\sigma}\right) = 0.1.$$

This tallies with $P(Z > -1.28) = 0.1$ for $N(0,1)$

whence $\sigma = \frac{20}{1.28} = 15.6.$

6% of cotter pins are defective.

So want to $P(\geq 10)$ are defective.

- Use the normal approximation to the binomial distribution.

$p = 0.06$, $np = 6 \; (= \mu)$, $npq = 6 \times 0.94 = 5.6 \; (= \sigma^2)$

$\therefore \sigma \doteqdot \sqrt{5.6} \doteqdot 2.37.$

$$\therefore P(X > 10) = P\left(Z > \frac{10-6}{2.37}\right) = P(Z > 1.7) = 0.04.$$

200 coin tosses. (presumably unbiassed).

$\mu = np = 200 \times 0.5 = 100.$ $\quad \sigma^2 = npq = 200 \times \frac{1}{2} \times \frac{1}{2} = 50.$

$\therefore \sigma \doteqdot 7.07.$

a. $P(80 < X < 120) = P\left(|Z| < \frac{20}{7.07}\right) = 0.9962.$

b. $P(X < 90) = P\left(Z < \frac{-10}{7.07}\right) = 0.0687.$

ctd.

c. $P(85 < X < 115)$

$= P\left(|Z| < \frac{15}{7.07}\right) = 0.9716.$

$\therefore P\left(|Z| > \frac{15}{7.07}\right) = 0.0284.$

'. $P(100 \text{ heads})$ — use the exact binomial form. (*)

$P(X=100) = {}^{200}C_{100}\, p^{100} q^{100} = {}^{200}C_{100} \cdot \left(\frac{1}{2}\right)^{200}$

This is very difficult to evaluate so alternatively use —

$P(99.5 < X < 100.5)$ — Continuous analogue

$= P\left(|Z| < \frac{0.5}{7.07}\right) = 0.0564.$

(*) Effectively here we are using a continuous approximation (normal distribution) to a discrete distribution (binomial).

Question Sheet 3

1. The pdf of Z is:

$$f_Z(z) = z, \quad 0 \le Z \le 1$$
$$= 2 - z, \quad 1 \le Z \le 2.$$

a. $P(Z \geq 1.5) = \int_{3/2}^{2} (2-z)\, dz = 1/8$
$= 0.125$

b. $Var[Z] = E[Z^2] - \{E[Z]\}^2$

$$= \int_0^2 z^2 f_Z(z)\, dz - 1$$

$$= 1/6.$$

2. $E[aX] = \int_{-\infty}^{\infty} ax\, f_X(x) dx = a \int_{-\infty}^{\infty} x f_X(x) dx = a E[X].$

$Var[aX] = \int_{-\infty}^{\infty} a^2 (x-\mu)^2 f_X(x) dx = a^2 Var[X].$

3. a. $\mu_1 = 3, \mu_2 = 4;\ \sigma_1^2 = 4,\ \sigma_2^2 = 9.$

b. note $X_1 - X_2 \sim N(\mu_1 - \mu_2,\ \sigma_1^2 + \sigma_2^2)$

because $-X_2 \sim N(-\mu_2, +\sigma_2^2)$.

Question Sheet 9/4

Lamps. $-\mu = 800$, $\sigma = 60$. If a sample of size 64 is chosen, then $E(\bar{X}) = 800$, $var(\bar{x}) = 60^2/64$.

$\therefore$ $Z = \dfrac{\bar{X} - 800}{60/8}$ is $N(0,1)$.

$P(790 < X < 810)$

$P\left(\dfrac{790-800}{60/8} < Z < \dfrac{810-800}{60/8}\right)$

i.e. $P\left(-\frac{4}{3} < Z < \frac{4}{3}\right) = 0.8164$.

2. Castings. $\mu = 5.02N$. $\sigma = 0.30N$. If a sample of 100 is chosen, $E(\bar{x}) = 5.0$ N,

$var(\bar{x}) = \frac{0.09}{100}$, $\therefore sd(\bar{x}) = \frac{.3}{10} = 0.03$.

a. $P(4.96 < \bar{x} < 5.0) = P\left(\frac{4.96-5.02}{0.03} < Z < \frac{5.00-5.02}{0.03}\right)$

$= P(-2 < Z < -\frac{2}{3}) = 0.4772 - 2476 = 2296$.

b. $P(\bar{x} > 5.1) = P\left(Z > \frac{5.10-5.02}{0.03}\right) = P(Z > 2.667)$

$= 0.0038$.

Soft drink machine. $\mu = 207$ ml, $\sigma = 15$ ml.

The machine is thought to be operating satisfactorily if $\overline{X}$ lies between $\mu_{\overline{x}} \pm 2\sigma_{\overline{x}}$.

$\mu_{\overline{x}}$, or $E(\overline{x}) = \mu = 207$.

$\sigma_{\overline{x}}, = \dfrac{\sigma}{\sqrt{n}}$, n sample size, $= \dfrac{15}{3} = 5$.

$\therefore \quad \mu_{\overline{x}} \pm 2\sigma_{\overline{x}} = 207 \pm 2.5$, i.e. $(197, 217)$.

If a sample gives a mean of 219 the machine needs adjustment.

7. Heights of 1000 students are $N(174.5, 6.9^2)$.
200 random samples of 25 are drawn.
For each sample $E(\bar{x}) = \mu = 174.5$. $\sigma^2(\bar{x}) = \frac{6.9^2}{25}$,
or $\sigma(\bar{x}) = \frac{6.9}{5} = 1.38$.

Between 172.5 and 175.8, how many samples do we have?

$P(172.5 < \bar{x} < 175.8)$

$= P\left(\frac{172.5 - 174.5}{1.38} < Z < \frac{175.8 - 174.5}{1.38}\right)$

i.e. $P(\underset{-1.45}{\cancel{-1.232}} < z < 0.942)$

$= 0.3903 + 0.3267 = 0.\cancel{7170}$ .7534

∴ The number of samples with means in this range is –
$200 \times 0.7170 = \cancel{141}.150$.

With mean less than 172.0, we have –
$P(172.0 < x) = P\left(\frac{172.0 - 174.5}{1.38} < z\right)$

$= P(-1.812 < z)$

$= 0.035$

The number is $200 \times 0.035 = 7$.

Average fuel Consumption.

If $\overline{X_1}$ of size n_1 is taken from a population $N(\mu_1, \sigma_1^2)$
and $\overline{X_2}$ of size n_2 " " " " " $N(\mu_2, \sigma_2^2)$

then $\overline{X_1} - \overline{X_2}$ is distributed $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$.

In this case A: $\mu_1 = 20$ mpg, $\sigma_1 = 6$ mpg.
B: $\mu_2 = 25$ mpg, $\sigma_2 = 5.5$ mpg.

$\mu_1 - \mu_2 = -5$, $\quad \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{6^2}{40} + \frac{5.5^2}{30} = 0.9 + 1.01$
$= 1.91$.

$\overline{X_1} - \overline{X_2}$ is $N(-5, 1.91)$. $\therefore P$(A's fuel consumption exceeds B)
is $P(\overline{X_1} - \overline{X_2} < 0) = P\left(Z < \frac{5}{\sqrt{1.91}}\right) = P(Z < 3.618)$
$\doteqdot 1$

Sample 1: $n_1 = 100$, $\mu_1 = 50$, $\sigma_1 = 8$.

" 2: $n_2 = 400$, $\mu_2 = 40$, $\sigma_2 = 12$.

We wish $P(\overline{X}_1 - \overline{X}_2 > 8)$.

$\overline{X}_1$: $E(\overline{X}_1) = \mu_1 = 50$, $\sigma_{\overline{X}_1} = \sigma_1/\sqrt{n_1} = 0.8$.

$\overline{X}_2$: $E(\overline{X}_2) = \mu_2 = 40$, $\sigma_{\overline{X}_2} = \sigma_2/\sqrt{n_2} = 0.6$

$\overline{X}_1 - \overline{X}_2$ is distributed $N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

$$= N(10, 0.8^2 + 0.6^2) = N(10, 1).$$

$\therefore P(\overline{X}_1 - \overline{X}_2 > 8) = P\left(Z > \frac{8-10}{1}\right) = P(Z > -2)$

$= 0.9772$.

$P(|\overline{X}_1 - \overline{X}_2| > 12) = P(\overline{X}_1 - \overline{X}_2 > 12) + P(\overline{X}_2 - \overline{X}_1 > 12)$

$= P\left(Z > \frac{12-10}{1}\right) + \underbrace{P\left(Z < \frac{-12-10}{1}\right)}_{\text{(negligible)}}$

$P(|Z| > 2) = 1 - 0.9772 = 0.0228$.

RV X, pdf. $f_X(x) = 2xe^{-x^2}$,

$= 0$ elsewhere.

$$\therefore \quad F(x) = \int_0^x 2te^{-t^2}\,dt$$

$$= 1 - e^{-x^2}.$$

$\therefore \; w = 1 - e^{-x^2}$,

and $e^{x^2} = \dfrac{1}{1-w}$

i.e. $x^2 = \ln\left\{\dfrac{1}{1-w}\right\} = F^{-1}(w)$.

$\therefore \; x = \frac{1}{2}\ln\left\{\dfrac{1}{1-w}\right\}$.

Taking the values in the notes, namely 0.03, 0.47, 0.43, 0.73, 0.86 we have

0.0152, 0.3174, 0.2811, 0.6547, 0.9831.

3ƒ.2.3.2.

Question Sheet 10/5

1. X_1, X_2 and X_3 are $\chi^2(1)$, $\chi^2(5)$ and $\chi^2(10)$, resp.

a. $X_1 + X_2$ is $\chi^2(6)$.

b. $X_1 + X_3$ is $\chi^2(11)$.

2. $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 (\mu_1, \sigma_2^2)$

$\Rightarrow X_1 - X_2$ is $N(0, \sigma_1^2 + \sigma_2^2)$.

$\Rightarrow \dfrac{X_1 - X_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}$ is $N(0,1)$.

$\therefore \dfrac{(X_1 - X_2)^2}{(\sigma_1^2 + \sigma_2^2)}$ is $\chi^2(1)$.

i.e. mean $= 1$, variance $= 2$.

a. (1) $\chi^2_{p\%} = 19.02 \ (\nu = 9)$, $p = 2.5\%$.

(2) $\chi^2_{p\%} = 24.43 \ (\nu = 40)$, $p = 97.5\%$.

b. (1) $\chi^2_{1/2\%} = x$, $(\nu = 29)$, $= 52.34$

(2) $\chi^2_{99\%} = x$, $(\nu = 4)$, $= 0.297$

. X_1 and X_2 are independent r.v.'s. $\chi^2(5)$ and $\chi^2(4)$ resp.

a. $P(X_1 < 6.25) = 0.9$

b. $P(X_1 < 0.115) = 0.01$

c. $P(X_2 > 14.86) = 0.005$

d. $P(X_1 + X_2 > 14.07) = 0.05$. $(\chi^2(7))$.

Sample size 20 is taken from a Normal Distribution having a variance 28 and sample variance s^2.

$\therefore (n-1)\dfrac{s^2}{\sigma^2} \sim \chi^2(n-1)$.

i.e. $\dfrac{19}{28} s^2 \sim \chi^2(19)$

The rest follows from the Answer Sheet.

3/1.38 2. Student 't' distribution.

Question Sheet 11/6

For $T(12)$, outside -3.93 to 3.93 lies 0.2% of the distribution. (3.09 for the Normal Case).

From $T(7)$, 90% of the distribution lies within $t = \pm 1.895$ (1.64 Normal).

If $Y \sim T(10)$,

a. $P(Y > 2.76) = 0.01$. (one side).

$\therefore P(Y > 3.17) = 0.005$.

$\therefore P(2.76 < Y < 3.17) = 0.005$.

. Sample size 5 from a normal population yeilds -
12.7, 13.3, 12.9, 13, 13.1. $\therefore \bar{x} = 13.0$.

$s^2 = 0.05$.

$\dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ is distributed as $T(n-1)$, i.e. $T(4)$.

$\therefore \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$ lies between -2.132 and 2.132 with 90% certainty.

i.e. $-2.132 \leq \dfrac{\bar{x} - \mu}{s/\sqrt{n}} \leq 2.132$.

i.e. $12.79 \leq \mu \leq 13.21$ with 90% Certainty.

3/1.9.3 Confidence Intervals Question Sheet 13/8

$\bar{x}_{30} = 780$ hours. $\sigma = 40$ hours.

$\therefore \bar{x} - \frac{2.05\sigma}{\sqrt{n}} \le \mu \le \bar{x} - \frac{2.05\sigma}{\sqrt{n}}$, with 95% confidence.

i.e. $780 - \frac{2.05 \times 40}{\sqrt{30}} \le \mu \le \bar{x} + \frac{2.05 \times 40}{\sqrt{30}}$

Approx [765, 795]. To be within 10 hours of the true mean with 96% confidence, $\frac{2.05\sigma}{\sqrt{n}} \le 10$ or,

$n \geqslant \left(\frac{2.05\sigma}{10}\right)^2 \geqslant 67.24$. i.e. $n = 68$.

Soft drink machine. Drink dispensed has sd = 1.5 dcl.
95% confidence interval for the mean of all drinks dispensed

is $\bar{x} - \frac{1.96 \times 1.5}{\sqrt{n}} \le \mu \le \bar{x} + \frac{1.96 \times 1.5}{\sqrt{n}}$.

If $n = 36$ and $\bar{x} = 22.5$

then μ lies between [22.01, 22.99]

1. Sample from machine, diameters of cylindrical pieces.
1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01 and 1.03.

x	$x - \bar{x}$	$(x-\bar{x})^2$
1.01	.0044	0.02×10^{-3}
0.97	−.0355	1.26
1.03	.0244	0.58
1.04	.0344	1.19
0.99	−.0155	0.24
0.98	− 0255	0.65
0.99	−.0155	0.24
1.01	.0044	0.02
1.03	.0244	0.58
Σ. 9.05	0	4.78×10^{-3}

$\bar{x} = 1.0055$

$\therefore s^2 = \frac{1}{(n-1)} \sum_{i=1}^{9} (x_i - \bar{x})^2 = \frac{1}{8} \times 4.78 \times 10^{-3} = 0.598 \times 10^{-4}$.

$\therefore s = 0.0245$.

+ 99% Confidence interval for the mean diameter using the sample standard deviation (n=8) and the t distribution for 8.

$\bar{x} - t_8 \cdot \frac{s}{3} \leq \mu \leq \bar{x} + t_8 \cdot \frac{s}{3}$.

$t_8 = 3.36$.

$\frac{3.36 \times .0245}{3} = 0.02744$

$\therefore$ The confidence interval is $[0.9781, 1.0330]$

: Sample of 12 shearing pins.

$\bar{x} = 48.50$, $s = 1.5$.

$\therefore \bar{x} - t_{11}\frac{s}{\sqrt{12}} \le \mu \le \bar{x} + t_{11}\frac{s}{\sqrt{12}}$ (90% conf).

$t_{11}(90\%) = 1.80$.

$\therefore \mu$ lies between $[47.72, 49.28]$.

8 cigarettes, tar content mean = 18.6 mg, sd = 2.4 mg.

$\therefore$ A 99% confidence interval is

$\bar{x} - t_7\frac{s}{\sqrt{8}} \le \mu \le \bar{x} + t_7\frac{s}{\sqrt{8}}$, (99% conf.

$t_7 = 3.50$.)

μ lies between $[15.629, 21.570]$

Hypotheses Question Sheet 14/9

Distribution. $N(100, 20)$.

Population variance is known and population is NORMAL

We can consider the test statistic $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

irrespective of the sample size.

a. $n = 10, \bar{x} = 102$. $\therefore z = \frac{102-100}{\sqrt{20/n}} \doteq 1.414$

$\therefore$ accept hypothesis.

$n = 20, \bar{x} = 98$, $\therefore z = \frac{98-100}{\sqrt{1}} = -2$.

$\therefore$ reject hypothesis.

$n = 30, \bar{x} = 104$. $z = \frac{104-100}{\sqrt{2/3}} \doteq 4.90$

$\therefore$ reject hypothesis.

$n = 5, \bar{x} = 97$ $z = \frac{97-100}{\sqrt{20/5}} = -1.5$

$\therefore$ accept hypothesis.

∴ Population $\mu = 0.05$. Sample of 10 gives $\bar{x} = 0.053$, $s = 0.003$.

Use the 'T' statistic. $T = \dfrac{\bar{x}-\mu}{s/\sqrt{n}} = 3.1622$

$t_9(5\%) = 2.26$, $t_9(1\%) = 3.25$, ∴ Reject at 5% level but accept at 1% level.

" 'O' Level Exams. $\mu = 74.5$, $\sigma = 8.0$.

School sample 200 students – $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \dfrac{1.4 \times \sqrt{200}}{8} = 2.475$
($\bar{x} = 75.9$)

Yes. Significant at the 5% level.

Examination, mean mark 69. A sample of 10 students obtains a mean mark of 64 and standard deviation 6.6.

Use 'T'. $T = \dfrac{\bar{x}-\mu}{s/\sqrt{n}} = \dfrac{64-69}{6.6/\sqrt{10}} = -2.396$.

Yes. Significantly lower at the 5% level.

Sample of 100 from an assumed Normal Population gives $\bar{x} = 1570$, $s = 120$. Take $Z = \dfrac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, $(s \sim \sigma)$.

$H_0 : \mu = 1600$.

$H_1 : \mu > 1600$.

$Z = \dfrac{1570-1600}{120/\sqrt{100}} = -2.5$

∴ Accept at 1% level, but reject at 5% level.

[illegible] Contingency Tables (χ^2)

Question Sheet 15/10

200 coin tosses — 115 heads, 85 tails.

Assume H_0: $p = 0.5$. $\therefore$ $E = 100$ heads, 100 tails.

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{1}{100}\left(15^2 + 15^2\right) = 4.5.$$

$\chi^2_{(5\%)}$, $\nu = 1$, $= 3.84$. $\therefore$ Reject H_0.

Accidents	0	1	2	3	4	5
No of Miners	35	47	39	20	5	2

$$\bar{x} = \frac{35\times 0 + 47\times 1 + 39\times 2 + 20\times 3 + 5\times 4 + 2\times 5}{148}$$

$$= \frac{215}{148} = 1.453$$

itting a Poisson Distribution, taking $\bar{x}$ as mean, we have —

	0	1	2	3	4/5
P:	0.2339	0.3399	0.2469	0.1196	0.0560
E:	34.61	50.31	36.54	17.70	8.29
$-E$:	0.39	−3.31	2.46	2.30	1.29
$\frac{-E)^2}{E}$:	.0044	0.2178	0.1656	0.2988	0.2377

Regroup Contingencies < 5

$= 0.924$

This is a good fit at the 5% level. $\chi^2_{(4)}\ 5\% = 9.49$

Die tosses.

	1	2	3	4	5	6
O:	25	17	15	23	24	16
E:	20	20	20	20	20	20
O−E	5	−3	−5	3	4	−4
$\frac{(O-E)^2}{E}$	1.25	0.45	1.25	0.45	0.84	0.84

$\sum_{i=1}^{6} \frac{(O-E)^2}{E} = 5.08$, $\chi^2_{(5)}$ 5% = 11.07

∴ die is fair.

320 families − 5 children.

No. of BOYS	5	4	3	2	1	0
" GIRLS	0	1	2	3	4	5
No. of families 'O'	18	56	110	88	40	8
'E'	10	50	100	100	50	10
O−E	8	6	10	−12	−10	−2
$\frac{(O-E)^2}{E}$	6.4	0.72	1	1.44	2	0.4

ctd.

To evaluate the expected frequencies, we need the binomial probabilities, $p = 0.5$ (boy) to give us—

	5	4	3	2	1	0
	0	1	2	3	4	5
:	1/32	5/32	5/16	5/16	5/32	1/32.

$$\sum_{i=0}^{5} \frac{(O-E)^2}{E} = 11.96 \quad , \quad \chi^2_{\cancel{(4)}\,(5)}\ 5\% = \cancel{9.49}\ 11.07$$

here is bias towards boys. Reject the $H_0: p = 0.5$.

Question Sheet 16/11

1.

	Recover	Die	
Drug	20	15	35
No Drug	13	22	35
	33	37	70

Expectations:

Drug	$\frac{33 \times 35}{70} = 16.5$	18.5
No Drug	16.5	18.5

$$\chi^2 = \frac{(20-16.5)^2}{16.5} + \frac{(15-18.5)^2}{18.5} + \frac{(13-16.5)^2}{16.5} + \frac{(22-18.5)^2}{18.5}$$

$$= 2.81$$

As $\chi^2_{0.05;1} = 3.84$, accept H_0.

2.

	Maths	Science	Languages	
Rugby	27	92	32	151
Soccer	14	71	43	128
Cricket	35	130	56	221
	76	293	131	500

note $e_{11} = \frac{151 \times 76}{500} = 22.95$ etc.

$\sum (O-E)^2/E = 6.84 < \chi^2 = 9.49.$

Correlation

Question Sheet 17/12

1. a. $Cov[X,Y] = E[(X-E[X])(Y-E[Y])]$

If $Y = c$, then $E[Y] = \int c\,dy = c$, as $\int dy = 1$.

$\therefore Y - E[Y] = 0$, so $Cov[X,Y] = 0$.

b. If $Y = aX + b$, then $E[Y] = aE[X] + b$.

$\therefore E[(X-E[X])(Y-E[Y])]$

$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$

$= E[XY] - E[X]E[Y]$

ote 1. $E[XE[Y]] = E[Y]\int x f(x)\,dx = E[X]E[Y]$

2. $E[E[X]E[Y]] = E[X]E[Y]\int f(x)\,dx$ $\overset{\shortparallel}{1}$

If $Y = aX + b$, $E[X(aX+b)] = aE[X^2] + bE[X]$

$E[XY] - E[X]E[Y] = aE[X^2] + bE[X] - (aE[X]+b)E[X]$

$= a\{E[X^2] - E[X]^2\}$.

$E[(Y-E[Y])^2] = E[Y^2] - E[Y]^2$, from b. above –

$E[Y^2] = E[(aX+b)^2] = a^2E[X^2] + 2abE[X] + b^2$

$E[Y^2] - E[Y]^2 = a^2E[X^2] + 2abE[X] + b^2 - a^2E[X]^2 - 2abE[X]$

$= a^2\{E[X^2] - E[X]^2\}$ $\quad - b^2$

. ctd.

$$\therefore \sqrt{E[(X-E[X])^2]\,E[(Y-E[Y])^2]}$$

$$= \pm a\left\{E[X^2]-E[X]^2\right\}.$$

$$\therefore \rho = \frac{Cov(X,Y)}{\sqrt{var[X]\,var[Y]}} = \pm 1,$$ according to the sign of a.

.a. Specimen test: Normal stress vs. Shear resistance.

x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
7.4	21.4	1.433	-3.367	-4.825	2.053	11.337
6.8	26.5	0.833	1.733	1.444	0.694	3.003
5.6	24.9	-0.367	0.133	-0.049	0.135	0.018
5.4	27.3	-0.567	2.533	-1.436	0.321	6.416
8.9	24.2	2.933	-0.567	-1.663	8.602	0.321
3.6	27.1	-2.367	2.333	-5.522	5.603	5.442
7.7	23.6	1.733	-1.167	-2.022	3.003	1.362
3.9	25.9	-2.067	1.133	-2.342	4.272	1.284
4.7	26.3	-1.267	1.533	-1.942	1.605	2.350
3.1	22.5	2.133	-2.267	-4.836	4.550	5.139
5.9	21.7	0.933	-3.067	-2.862	0.870	9.406
2.6	25.8	-3.367	1.033	-3.478	11.337	1.067
= 11.6	$\Sigma y_i =$ 297.2			-29.533 $\Sigma(x_i-\bar{x})(y_i-\bar{y})$	43.045 $\Sigma(x_i-\bar{x})^2$	47.146 $\Sigma(y_i-\bar{y})^2$

.967 $\bar{y}=24.767$

$$= \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\left(\Sigma(x_i-\bar{x})^2\,\Sigma(y_i-\bar{y})^2\right)^{1/2}} = \frac{-29.533}{\sqrt{43.045 \times 47.146}} = -0.6556$$

b. We wish to test $H_0 : \rho_0 = -0.5$ vs. $H_1 : \rho_0 \neq 0.5$ at the 5% level.

$$z = \frac{\sqrt{n-3}}{2}\left[\ln\left(\frac{1+r}{1-r}\right) - \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)\right]$$

$n = 12$, $r = -0.6556$, $\rho_0 = -0.5$.

$$\therefore\ z = 1.5 \ln\left(\frac{0.2080}{(1/3)}\right) = -0.7074.$$

$\therefore$ Accept H_0.

From [illegible] It follows that if X and Y are random variables with variances σ_X^2 and σ_Y^2 and covariance σ_{XY} then

$$\mathrm{Var}[aX+bY] = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\sigma_{XY}.$$

$\therefore$ If $Z = X+Y$, then $E[Z] = E[X]+E[Y] = \mu_X + \mu_Y$.

and – $\mathrm{Var}[Z] = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$

As – $\rho = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y}$, then $\mathrm{Var}[Z] = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$.

Bus Service – Running Times –

X: $\mu_X = 12$ mins. $\sigma_X = 1$ min.

Y: $\mu_Y = 20$ mins. $\sigma_Y = 2$ min. $(\rho = 0.5)$.

$\therefore$ $\mu_Z = 32$ mins. $\mathrm{Var}[Z] = 1^2 + 2^2 + 2\times 0.5 .1.2 = 7$.

$\therefore \sigma_Z = \sqrt{7}$. or $\sigma_Z^2 = 7$.

If the scheduled run time is 33 min, then

$$P(Z \geq 33) = P\left(z \geq \frac{1}{\sqrt{7}}\right) = 0.35$$

z is $N(0,1)$.

Regression

Question Sheet 18/13

Thickness of sheets (x)	Shear strength of sheets (Kg.) (y)
0.2	102
0.3	129
0.4	201
0.5	342
0.6	420
0.7	591
0.8	694
0.9	825
1.0	1014
1.1	1143
1.2	1219.

x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$ $\times 10^2$	$(x_i-\bar{x})(y_i-\bar{y})$
2	102	−0.5	−505.3	0.25	2553	252.7
3	129	−0.4	−478.3	0.16	2288	191.3
4	201	−0.3	−406.3	0.09	1651	121.9
5	342	−0.2	−265.3	0.04	704	53.1
6	420	−0.1	−187.3	0.01	351	18.7
7	591	0	−16.3	0	3	0
8	694	0.1	86.7	0.01	75	8.7
9	825	0.2	217.7	0.04	474	43.5
[illegible]	1014	0.3	406.7	0.09	1654	122.0
[illegible]	1143	0.4	535.7	0.16	2862	214.3
[illegible]	1219	0.5	611.7	0.25	3742	305.9
$=7.7$	$\Sigma y_i = 6680$			$\Sigma = 1.10$	$\Sigma = 16357$	$\Sigma = 1332.1$
$.7$	$\bar{y} = 607.3$					

ctd.

$Y = aX + b$. Use $\hat{\alpha}$ and $\hat{\beta}$ for a and b. where—

$$\hat{\alpha} = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

$$\sum (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2, \quad \therefore \sum x_i^2 = \sum (x_i - \bar{x})^2 + n\bar{x}^2$$

$$= 1.10 + 11 \times 0.7 = 8.8.$$

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \bar{x} \sum y_i - \bar{y} \sum x_i + n\bar{x}\bar{y}$$

$$= \sum x_i y_i - n\bar{x}\bar{y}.$$

$$\therefore \sum x_i y_i = \sum (x_i - \bar{x})(y_i - \bar{y}) + n\bar{x}\bar{y}. \left(= \frac{\sum x_i \sum y_i}{n}\right)$$

$$= 1332.1 + \frac{7.7 \times 6680}{11} = 6008.2$$

$$\therefore \hat{\alpha} = \frac{6.49 \times 6680 - 7.7 \times 6008.2}{11 \times 1.1} \doteq -240.3$$

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{1332.1}{1.1} = 1211.$$

Volume V : 54.3 61.8 72.4 88.7 118.6 194.0

Pressure P : 61.2 49.5 37.6 28.4 19.2 10.1

Assuming $PV^{\gamma} = C$, where γ, C are Constants.

then - $\ln P + \gamma \ln V = \ln C$.

or $\ln P = -\gamma \ln V + \ln C$.

Set $Y = \ln P$ and $X = \ln V$, and obtain a regression line of the form $Y = \alpha + \beta X$

x_i	y_i	$x_i y_i$	x_i^2
995	4.114	16.435	15.960
124	3.902	16.092	17.007
282	3.627	15.531	18.335
485	3.346	15.007	20.115
775	2.955	14.110	22.801
268	2.313	12.185	27.752
929	20.257	89.360	121.970
(Σx_i)	(Σy_i)	$\Sigma x_i y_i$	Σx_i^2

f $\hat{\alpha}$ and $\hat{\beta}$ are estimates for a and b.

$$= \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} = \frac{121.97 \times 20.257 - 26.929 \times 89.360}{6 \times 121.970 - (26.929)^2}$$

$$= 9.679$$

$$\beta = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{6 \times 89.360 - 26.929 \times 20.254}{6 \times 121.970 - (26.929)^2}$$

$$= -1.405$$

$\gamma = -\beta, \quad \therefore \quad \gamma = 1.405$

$= \alpha, \quad \therefore \quad C = \exp(\alpha) \doteq 15{,}980.$

When $V = 100$, $P = CV^{-\gamma} \doteq 24.75$

(These answers are not exactly those on the pink sheet, showing the need for great accuracy

3. x: 0 1 2 3 4 5 6 7

y: 16 11 12 11 4 6 4 1

$\therefore \sum_{i=1}^{8} x_i = 28$. $\sum x_i^2 = 140$, $\sum x_i y_i = 145$

$\sum y_i = 65$. $\sum y_i^2 = 711$

$$a = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{8 \times 145 - 28 \times 65}{8 \times 140 - (28)^2} = -1.96$$

$$b = \frac{\sum y_i \sum x_i^2 - \sum x_i y_i \sum x_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{65 \times 140 - 145 \times 28}{8 \times 140 - (28)^2} = 15$$

$$a' = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2} = \frac{8 \times 145 - 28 \times 65}{8 \times 711 - (65)^2} = -0.45$$

$$b' = \frac{\sum x_i \sum y_i^2 - \sum x_i y_i \sum y_i}{n\sum y_i^2 - \left(\sum y_i\right)^2} = \frac{28 \times 711 - 145 \times 65}{8 \times 711 - (65)^2} = 7.17.$$

$r^2 = aa' = 0.882$. $r = -\sqrt{0.882} = -0.94$.

negative correlation is evident without further calculation).

Against $H_0: \rho_0 = 0$, we compute

$$z = \sqrt{\frac{n-3}{2}} \ln\left(\frac{1+r}{1-r}\right) = -3.88. \text{ (highly significant)}.$$

2. We wish to minimise

$$S = \sum_{i=1}^{n} (g(x_i) - y_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i - cx_i^2)^2$$

$$\frac{\partial S}{\partial a} = 0 \implies \sum y_i = na + b\sum x_i + c\sum x_i^2$$

$$\frac{\partial S}{\partial b} = 0 \implies \sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$$

$$\frac{\partial S}{\partial c} = 0 \implies \sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$$

With $\sum d_i = 1176$, $\sum d_i v_i = 64840$, $\sum d_i v_i^2 = 3{,}830{,}000$,
$\sum v_i = 270$, $\sum v_i^2 = 13900$, $\sum v_i^3 = 783{,}000$ $\sum v_i^4 = 46.750.000$

$$\therefore \quad 1176 = 6a + 270b + 13900c \quad — ①$$

$$64840 = 270a + 13900b + 783000c \quad — ②$$

$$3.830{,}000 = 13900a + 783000b + 46.750.000c \quad — ③$$

$45 \times ①$ $\quad 11920 = 1750b + 157{,}500c \quad — (4)$

$51.48 ②$ $\quad 492037 = 67428b + 1.441{,}160c \quad — (5)$

$38.53 ④$ $\quad 32759 = 372685c$

$$\therefore \quad c = 0.0879.$$

From (4) $b = -1.0996$ and from (1). $a = 41.85$.

$\therefore$ If $v = 45$ m/sec $\quad d = 170.37$ m.

As $v \to 0$, $d \to 41.85$ m., this contradicts the result that $d \to 0$. Formula $d = a + bv + cv^2$ should not be extrapolated outside this range.