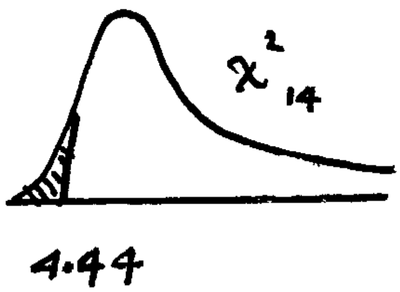
MSc Foundation Block
Block B
Statistical Methods
Solutions to Exercises

1. Grouping of values on specific numbers (especially noticeable around 0). Possibly original values integers in °F and transformed to $(x-32)\times\frac{5}{9}$ °C.
(Has implications for the real accuracy of the data.)

2. (a) $H_0 : \mu = 180 \;\; v \;\; H_1 : \mu \neq 180$
$\overline{x} = 178.46 \qquad s = 2.2535$
$t = -2.65 \qquad$ c.f. t_{14} distribution under H_0.
$p = P(|T| > 2.65)$
So $0.01 < p < 0.02$ (R gives $p = 0.019$)
i.e. **some evidence against H_0 (mean appears < 180).**
95% CI for μ is $178.46 \pm 2.145 \times \sqrt{\frac{2.2535^2}{15}} = (177.21, 179.71)$.

(b) $H_0 : \sigma \geq 4 \qquad H_1 : \sigma < 4$
$s^2 = 5.0783$
$\frac{(n-1)s^2}{\sigma_0^2} = \frac{5.0783\times 14}{16} = 4.44$

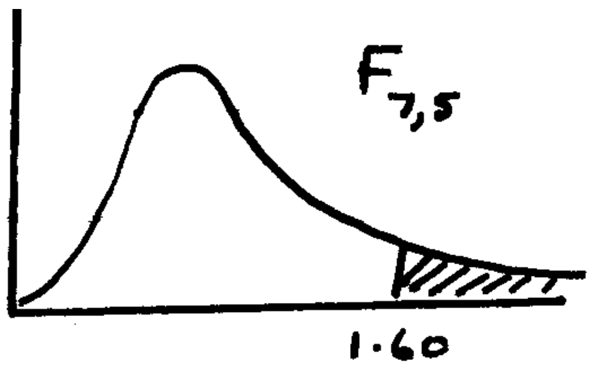


$p = P\left(\chi^2_{14} < 4.44\right)$
(here small values of test statistics make us doubt H_0).
From Neave $0.005 < p < 0.01$
i.e. **reject view expressed that procedure is unsatisfactory**.

3.
$$n_1 = 8 \quad \overline{x} = 61 \quad s_1^2 = 221.4 \quad n_2 = 6 \quad \overline{y} = 49 \quad s_2^2 = 138.0$$

(a) $H_0 : \sigma_1^2 = \sigma_2^2 \;\; v \;\; H_1 : \sigma_1^2 \neq \sigma_2^2$

$f = \frac{221.4}{138.0} = 1.60$



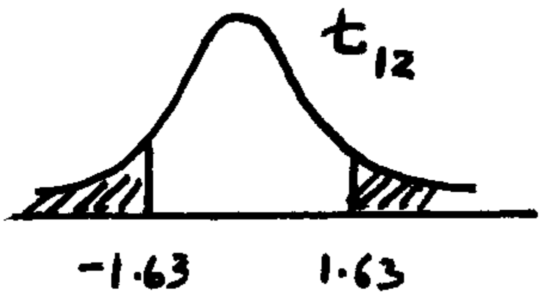
$p = 2 \times P(F_{7,5} > 1.60)$

$>> 2 \times 0.1$ since $F_{7,5;0.9} = 3.37$
(R gives $2 \times (1 - 0.6873) = 0.625$)
i.e. **no evidence to reject** $\sigma_1^2 = \sigma_2^2$.

(b) $H_0 : \mu_1 = \mu_2$ v $H_1 : \mu_1 \neq \mu_2$ under assumption of $\sigma_1^2 = \sigma_2^2$

$s^2 = \frac{1550+690}{8+6-2} = 186.7$

$t = \frac{61-49}{\sqrt{186.7\left(\frac{1}{8}+\frac{1}{6}\right)}} = 1.63.$



From Neave $0.10 < p < 0.15$
[R gives $p = 2 \times 0.0645 = 0.129$]
i.e. **little evidence to reject** H_0 **- equal means**.

95% CI for $\mu_1 - \mu_2 : (61 - 49) \pm 2.179 \times \sqrt{186.7\left(\frac{1}{8} + \frac{1}{6}\right)} = 12 \pm 16.08$
$= (-4.08, 28.08)$

4.

0.0336; 0.0526; 0.1470[R gives 0.1471]; 0.0217; 0.6406;
0.0188; 0.0004; 0.0389[R gives 0.0390]; 0.0582; 0.1478

5. ($< p <$) gives bounds from Neave; [ ] gives value from R.

(a)

$(0.05 < p < 0.075)$	$[p = 0.063]$;	$(0.05 < p < 0.075)$	$[p = 0.072]$;
$(0.20 < p < 0.30)$	$[p = 0.284]$;	$(0.025 < p < 0.05)$	$[p = 0.033]$;
$(0.60 < p < 0.70)$	$[p = 0.640]$;	$(0.02 < p < 0.05)$	$[p = 0.030]$;
$(\simeq 0.001)$	$[p = 0.00099]$;	$(0.025 < p < 0.05)$	$[p = 0.041]$;
$(0.10 < p < 0.15)$	$[p = 0.100]$;	$(0.10 < p < 0.15)$	$[p = 0.148]$.

(b)

$(0.85 < p < 0.9)$	$[p = 0.872]$;	$(0.025 < p < 0.05)$	$[p = 0.044]$;
$(0.80 < p < 1.0)$	$[p = 0.968]$;	$(0.025 < p < 0.05)$	$[p = 0.037]$;
$(0.01 < p < 0.025)$	$[p = 0.017]$;	$(0.05 < p < 0.10)$	$[p = 0.057]$;
$(\simeq 0.20?)$	$[p = 0.218]$;	$(\simeq 0.925?)$	$[p = 0.921]$;
$(0.2 < p < 0.3)$	$[p = 0.208]$;	$(\simeq 0.9995?)$	$[p = 0.9998]$.

(c)

$(0.001 < p < 0.005)[p = 0.0013]$;	$f^{-1} = 3.125 \Rightarrow (0.1 < p < 0.2)[p = 0.139]$;
$(0.2 < p < 1)[p = 0.493]$;	$(0.001 < p < 0.005)[p = 0.0014]$;
$f^{-1} = 2.778 \Rightarrow (0.05 < p < 0.10)[p = 0.088]$;	$(0.05 < p < 0.1)[p = 0.076]$;
$(0 < p < 0.002)[p = 0.0015]$;	$(\simeq 0.02?)[p = 0.012]$;
$(0.05 < p < 0.1)[p = 0.070]$;	$(\simeq 0.05??)[p = 0.0047]$.

6. Matched pairs

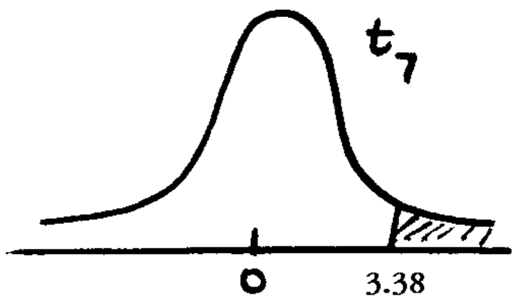
$B - A : 0.4 \;\; 0.0 \;\; 0.5 \;\; 0.5 \;\; 0.3 \;\; -0.1 \;\; 0.3 \;\; 0.2$

Define $\mu = \mu_B - \mu_A$

$H_0 : \mu = 0 \;\; v \;\; H_1 : \mu \neq 0$

$n = 8 \;\; \bar{x} = 0.2625 \;\; s = 0.21998$

$t = \frac{\bar{x}}{s/\sqrt{n}} = 3.38$



$\Rightarrow$ from Neave $\;\; 0.01 < p < 0.02$

[R gives $\;\; p = 0.012$]

Thus **fairly strong evidence to reject** $\mu = 0$, i.e. difference exists in means under treatments A and B; $\mu_B > \mu_A$ suggested, 95% CI for μ is (0.08,0.45).

7. Number died $X \sim Bi(30, p)$

$\hat{p} = \frac{8}{30} = 0.27$

(a) Using $\frac{X}{30} \sim N(p, \frac{p(1-p)}{30})$

$\Rightarrow$ 95% CI $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{30}} \Rightarrow (0.108, 0.425)$

or **solving quadratic** $(\frac{x}{30} - p)^2 < 1.96^2 \; \frac{p(p-1)}{30}$

i.e. $p^2(30 + 1.96^2) - p(2x + 1.96^2) + \frac{x^2}{30} < 0$

$x = 8 \Rightarrow (0.142 < p < 0.444)$

(b) Neave 1.2(a) $(0.125, 0.460)$.

8.

No. emitted x:	0	1	2	3	4	5	6	7	8
Observed:	57	203	383	525	532	408	273	139	45
Expected:	54.32	210.29	407.07	525.32	508.44	393.69	254.03	140.49	67.99

No. emitted x:	9	10	≥ 11
Observed:	27	10	6
Expected:	29.25	11.32	5.79

Expected under $(P_0(\mu)$. Use $\hat{\mu} = \frac{0\times 57+1\times 203+\ldots+13\times 1+14\times 1}{2608} = 3.871549$

$H_0 : X \sim P_0(\mu)$
Expected number $= 2608 \times \frac{e^{-\hat{\mu}}\hat{\mu}^x}{x!} \quad (x = 0, 1, 2, \ldots)$
$X^2 = 12.96$

compare with χ^2_{12-1-1}
↗ ↖
combined groups(since e_i low) estimate μ
$\Rightarrow 0.2 < p < 0.3$ i.e. **no evidence to reject Poisson model**

9.

	Years education			
	≤ 8	$9-12$	≥ 13	
Agree	83	255	76	414
Disagree	38	207	142	387
	121	462	218	801

$\hat{e}_{ij}$		
62.54	238.79	112.67
58.46	223.21	105.33

$X^2 = 40.83$ compare with $\chi^2_2 \Rightarrow p << 0.0005$.
Extremely strong evidence of association. Suggests that as 'education' increases, 'agreement' decreases.

10. a. Price depends on number of pages, therefore price is the response or dependent variable, and number of pages is the explanatory or independent variable.

b. Plot (omitted) shows price increasing roughly linearly with number of pages.

c. $\bar{x} = 4.55, \bar{y} = 16.7, s_{xx} = 26.925$ and $s_{xy} = 63.35$.
Therefore, $\hat{\beta} = \frac{s_{xy}}{s_{xx}} = \frac{63.35}{26.925} = 2.35$ and $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 16.7 - 2.35 \times 4.55 = 5.99$, yielding estimated/fitted line $y = 5.99 + 2.35x$.
Note the line should go through the point $(\bar{x}, \bar{y}) = (4.55, 16.7)$.

d. 30p for each extra 10 pages is equivalent to £3 for each extra 100 pages (in units of the original data). Therefore, the allegation is that $y = \alpha + 3x$. So we test $H_0 : \beta = 3$ versus $H_0 : \beta \neq 3$. Under H_0:

$$T = \frac{\hat{\beta} - 3}{\text{e.s.e.}(\hat{\beta})} \sim t_{10-2=8}$$

To calculate t_{obs} we require e.s.e.$(\hat{\beta}) = \sqrt{\widehat{\sigma^2}/s_{xx}}$,
where $\widehat{\sigma^2} = \frac{1}{n-2}\{s_{yy} - s_{xy}^2/s_{xx}\}$.
Here $s_{yy} = 206.1, \widehat{\sigma^2} = 7.131$ and e.s.e.$(\hat{\beta}) = 0.5146$. So

$$t_{obs} = \frac{2.35 - 3}{0.5146} = -1.263.$$

Therefore, the p-value is $2\text{P}(t_8 \leq -1.263) = 2\text{P}(t_8 \geq 1.263)$. From tables:

q	0.85	0.9
$t_{8,q}$	1.108	1.397

Therefore, $0.2 < p\text{-value} < 0.3$, and there is no evidence against H_0. The sample data provide no evidence, $0.2 < p\text{-value} < 0.3$, that the allegation that the publisher increases the price by 30p per 10 pages is untrue.