University of Waterloo STAT 221 – W. H. Cherry

Assignment 5 Outline Solution

A5 – 1. Model: Let the random variable Y_i represent the Rockwell hardness of an equiprobably-selected specimen; we let $E(Y_i) = \mu$, where model parameter μ represents $\bar{\mathbf{Y}}$, the average respondent (or study) population Rockwell hardness of this steel.

For the random variable $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, as a consequence of the Central Limit Theorem, $\bar{Y}$ has approximately a normal distribution, so we can write: $\bar{Y} \div N(\mu, \sigma/\sqrt{n})$,

where σ represents $\mathbf{S}$, the standard deviation of the respondent (or study) population Rockwell hardness of this steel.

We assume that the method of selecting the sample of 50 steel specimens was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o**:** $\mu = 64$ units; *i.e.*, the model mean [the respondent (ot study population average] *is* the *lower* end of the range of values claimed by the manufacturer of the steel.

H_a**:** $\mu < 64$ units; *i.e.*, the model mean [the respondent (or study) population average] is *below* the lower end of the range of values claimed by the manufacturer.

Dis. meas.: By standardizing in the model for $\bar{Y}$, we obtain: $\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \div N(0,1)$.

Observed: $\bar{Y} = 62$ units;
expected: $\mu = 64$ units (under the null hypothesis);
we are given that $\sigma = 8$ units; also, $n = 50$.
Value of discrepancy mesasure: $\frac{62-64}{8/\sqrt{50}} \simeq -1.767\,767$.

***P*-value:** $\Pr[N(0,1) \le -1.767\,767] \simeq 0.038\,550 \simeq 0.039$.

Decision: Because the significance level of about 0.039 *is* less than the specified value of $\alpha = 0.05$, we reject the null hypothesis and decide that the true average Rockwell hardness of this particular type of steel *is* less than the value of at least 64 units claimed by the manufacturer. [Hardness values in the null hypothesis for μ *above* the lower end of the range (*i.e.*, above 64 units would lead to even *more* extreme values of the discrepancy measure and so would also be rejected with the specified value of $\alpha = 0.05$.]

A5 – 2. Model: Let the random variable W_i represent the weight (in kg) of an equiprobably-selected battery; we let $E(W_i) = \mu$, where model parameter μ represents $\bar{\mathbf{Y}}$, the average respondent (or study) population battery weight.

For the random variable $\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W_i$, as a consequence of the Central Limit Theorem, $\bar{W}$ has approximately a normal distribution, so we can write: $\bar{W} \div N(\mu, \sigma/\sqrt{n})$,
where σ represents $\mathbf{S}$, the s.d. of the respondent (or study) population battery weight for the production run.

We assume that the method of selecting the sample of 40 batteries was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o**:** $\mu = 30$ kg; *i.e.*, the model mean [the respondent (or study) population average] *is* the value in the specifications.

H_a**:** $\mu < 30$ kg; *i.e.*, the model mean [the respondent (or study) population average] is *smaller* than the specifications.

Dis. meas.: By standardizing in the model for $\bar{W}$, we obtain: $\frac{\bar{W}-\mu}{\sigma/\sqrt{n}} \div N(0,1)$.

Observed: $\bar{w} = 29.6$ kg;
expected: $\mu = 30$ kg (under the null hypothesis);
we are given that $\sigma = 0.55$ kg; also, $n = 40$.
Value of discrepancy measure: $\frac{29.6-30}{0.55/\sqrt{40}} \simeq -4.599\,677$.

***P*-value:** $\Pr[N(0,1) \le -4.599\,677] \simeq 0.000\,002\,116 \simeq 0.000\,002$.

Decision: Because the significance level of about 0.000 002 *is* (much) less than the specified value of $\alpha = 0.01$, we reject the null hypothesis and decide that the manufacturing process for the batteries *was* out of control during the production run.

A5 – 3. (a) The alternative hypothesis H_a**:** $\mu \ne 5$ mg was chosen presumably because HWC regards as serious *either* too *much* codeine ($\mu > 5$ mg) *or* too *little* codeine ($\mu < 5$ mg) in a tablet.

(*continued overleaf*)

Assignment 5 Outline Solution (continued 1)

A5 – 3. (cont.) (b) A type I error (rejecting a true null hypothesis) corresponds to deciding that the average codeine weight per tablet is significantly *different* from (either under or over) 5 mg when *in fact* it is effectively 5 mg. The consequence might be to prohibit sale of the drug when in fact it meets specifications; this would be unfair to the company making the drug.

A type II error (accepting a false null hypothesis) corresponds to deciding that the average codeine weight per tablet *is* effectively 5 mg when *in fact* it is significantly different from (either under or over) 5 mg. The consequence might be permit the sale of a drug which does not meet specifications; this could endanger users of the drug by giving them either too high or too low a dose of codeine.

A type II error is likely more serious in this situation.

A5 – 4. (a) The alternative hypothesis H_a**:** $\pi > 0.1$ was chosen because, from a consumer's perspective, it is only too *high* (*not* too low) a proportion of defective flares that is a matter for concern.

(b) A type I error (rejecting a true null hypothesis) corresponds to deciding that the proportion of defective flares is significantly higher than the manufacturer claims when *in fact* the claim is correct. The consequence would be to wrongly take the manufacturer to court, with all the unnecessary expense.

A type II error (accepting a false null hypothesis) corresponds to deciding that the proportion of defective flares *is* effectively 10% as claimed by the manufacturer, when *in fact* the proportion is significantly higher. The consequence would be that people purchasing the flares would be getting a product that might place them in danger when it did not function properly.

Both types of error are highly undesirable in this situation.

A5 – 5. (a) Let the random variable W_i represent the weight (in mg) of an equiprobably-selected white sucker larva one day posthatch; we use the model $W_i \sim N(6.9, 2.6)$.

We want: $\Pr(W_i > 8.5) = \Pr(\frac{W-\mu}{\sigma} > \frac{8.5-6.9}{2.6}) = \Pr[N(0,1) > 0.615\,385] = 0.269\,150 \simeq 0.27.$

(b) For the *average* of 25 measurements, we use the model $\overline{W} \sim N(6.9, 2.6/\sqrt{25})$ $[\overline{W} = \frac{1}{25}\sum_{i=1}^{25} W_i]$.

We want: $\Pr(\overline{W} > 8.5) = \Pr(\frac{\overline{W}-\mu}{\sigma/\sqrt{\text{n}}} > \frac{8.5-6.9}{0.52}) = \Pr[N(0,1) > 3.076\,923] = 0.001\,046 \simeq 0.001.$

(c) The calculation in (b) shows that there is only about one chance in 1,000 that the average weight for a sample of 25 of the larvae one day posthatch would exceed 8.5 mg; the actual observation of such an average therefore most likely indicates (*i.e.*, provides good evidence) that the larvae *are* older than one day.

(d) **Model:** Let the random variable W_i represent the weight (in mg) of an equiprobably-selected larva; we use the model $W_i \sim N(\mu, \sigma)$, where model parameter μ represents $\overline{\mathbf{W}}$, the average weight, and σ represents $\mathbf{S}$, the standard deviation of the weights of *all* larvae in the respondent (or study) population.

If the random variable $\overline{W} = \frac{1}{\text{n}}\sum_{i=1}^{\text{n}} W_i$, then: $\overline{W} \doteatom N(\mu, \sigma/\sqrt{\text{n}})$.

We also assume that the method of selecting the sample of 25 larvae was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o**:** $\mu = 6.9$ mg; *i.e.*, the value of the model mean [the average weight of the respondent (or study) population of larvae] *is* consistent with an age of one day posthatch.

H_a**:** $\mu > 6.9$ mg; *i.e.*, the value of the model mean [the average weight of the respondent (or study) population of larvae] is too *high* for an age of one day posthatch.

Dis. meas.: By standardizing in the model for $\overline{W}$, we obtain: $\frac{\overline{W}-\mu}{\sigma/\sqrt{\text{n}}} \sim N(0,1)$.

Observed: $\overline{w} = 8.5$ mg;
expected: $\mu = 6.9$ mg (under the null hypothesis);
we are given that $\sigma = 2.6$ mg; also, $\text{n} = 25$.

Value of discrepancy measure: $\frac{8.5-6.9}{2.6/\sqrt{25}} \simeq 3.076\,923.$

***P*-value:** $\Pr[N(0,1) \geq 3.076\,923] \simeq 0.001\,046 \simeq 0.001.$

Answer: Using a one-sided z-test for one mean, the null hypothesis is rejected at the 1% level ($P \simeq 0.001$); thus, the data provide highly statistically significant evidence that the larvae *are* older than one day posthatch.

(*continued*)

Assignment 5 Outline Solution (continued 2)

A5 – 6. Because we are dealing with *measurements* and the population standard deviation (σ) is *known*, this question involves a z-test for one mean.

(a) **Model:** Let the random variable L_i represent the critical dimension (in mm) of an equiprobably-selected crankshaft; we let $E(L_i) = \mu$, where model parameter μ represents $\bar{\mathbf{L}}$, the average respondent (or study) population crankshaft critical dimension.

For the random variable $\bar{L} = \frac{1}{n}\sum_{i=1}^{n} L_i$, as a consequence of the Central Limit Theorem, $\bar{L}$ has approximately a normal distribution, so we can write: $\bar{L} \dot{\sim} N(\mu, \sigma/\sqrt{n})$,

where σ represents $\mathbf{S}$, the s.d. of the respondent (or study) population crankshaft critical dimension for the production run.

We assume that the method of selecting the sample of 16 crankshafts was such as to make the limitation imposed by samlpe error acceptable in the investigation context.

H_o**:** $\mu = 224$ mm; *i.e.*, the model mean *is* consistent with the process for manufacturing crankshafts being centred on target.

H_a**:** $\mu \neq 224$ mm; *i.e.*, the model mean is *in*consistent with the process for manufacturing crankshafts being centred on target.
[We use a *two*-sided test because the statement of the question gives no indication that departure from the critical dimension in *one* direction (too ihgh or too low) was of greater concern.]

Dis. meas.: By standardizing in the model for $\bar{L}$, we obtain: $\frac{\bar{L} - \mu}{\sigma/\sqrt{n}} \dot{\sim} N(0, 1)$.

Observed: $\bar{l} = 3{,}584.031/16 = 224.001\,9375$ mm;
expected: $\mu = 224$ mm (under the null hypothesis);

we are given that $\sigma = 0.060$ mm; also $n = 16$. [The sample s.d. of 0.0618 mm, is quite close to the value given for σ, consistent with the assumption that the sample of crankshafts *was* selected by an acceptable method.]

Value of discepancy measure: $\frac{224.001\,9375 - 224}{0.060/\sqrt{16}} \simeq 0.129\,1\dot{6}$.

***P*-value:** $\Pr[N(0, 1) \geq 0.129\,1\dot{6}] = 2 \times \Pr[N(0, 1) \geq 0.129\,1\dot{6}] \simeq 2 \times 0.448\,613 \simeq 0.90$.

Answer: Using a two-sided z-test for one mean, the null hypothesis is *not* rejected at the 5% (or even at the 80%) level ($P \simeq 0.90$); thus, the data provide no statistically significant evidence that the process for manufacturing the crankshafts is not centred on its target; *i.e.*, the process *does* appear to be correctly centred.

(b) An approximate 90% confidence interval for model parameter μ representing $\bar{\mathbf{L}}$, the average critical dimension of the respondent (or study) population of camshafts, is:

$\bar{l} \pm z_\alpha^* \times \frac{\sigma}{\sqrt{n}}$ which is: $224.001\,9375 \pm 1.6449 \times \frac{0.060}{\sqrt{16}}$ $\Longrightarrow$ (223.977 264, 224.026 661) mm
or about (223.977, 224.027) mm.

*A5 – 7. **Dis. meas.:** $Z = \frac{\bar{Y} - 65}{\sigma/\sqrt{n}} = \frac{\bar{Y} - 65}{5/\sqrt{50}} \sim N(0, 1)$; $z > 1.8$ corresponds to: $\bar{y} > 65 + 1.8 \times \frac{5}{\sqrt{50}} \simeq 66.273$°C.

(a) Pr(type I error) $\equiv \alpha = \Pr[N(0, 1) > 1.8] \simeq 1 - 0.9641 = 0.0359 \simeq 3.6\%$.

(b) Pr(type II error) $\equiv \beta = \Pr[N(66.5, \frac{5}{\sqrt{50}}) < 66.273] = \Pr[N(0, 1) < -0.321\,320] \simeq 0.373\,984 \simeq 37\frac{1}{2}\%$.

(c) Pr(type II error) $\equiv \beta = \Pr[N(70, \frac{5}{\sqrt{50}}) < 66.273] = \Pr[N(0, 1) < -5.271] \simeq 6.782 \times 10^{-8} \simeq 10^{-7}$.

(d) The diagrams are given overleaf on page 0.13d; they are based on the diagrams given in the statement of the question and the calculations above in (a), (b) and (c).

(e) If $\bar{y} = 66.2$°C, $z = \frac{66.2 - 65}{5/\sqrt{50}} = 1.697\,056 \simeq 1.70$, ***P*-value:**: $\Pr[N(0, 1) > 1.697\,056] = 0.044\,843 \simeq 4\frac{1}{2}\%$.
Because this z value of just less than 1.70 does *not* lie in the rejection region $z > 1.8$ (and the P-value is *above* $\alpha = 3.6\%$), we do *not* reject the null hypothesis; thus, the data are consistent with (continued overleaf on page 0.13d)

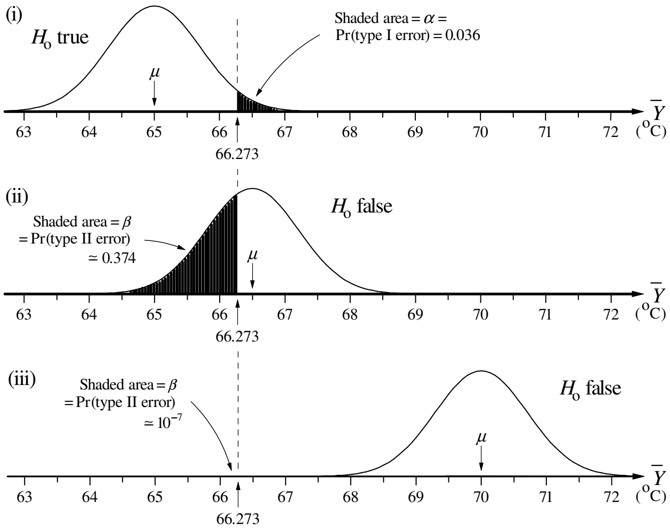
A5 – 8. Because we are dealing with *measurements* and the population standard deviation (σ) is *un*known [and so is *estimated* by the *sample* standard deviation (s)], this question involves a t-test for the mean (μ) of a normal model, where model parameter μ represents $\bar{\mathbf{C}}$, the average blood cholesterol level in the respondent (or study) population of these men.

Assignment 5 Outline Solution (continued 3)

***A5 – 7. (cont.)** (e) the generating station meeting the requirement for the temperature of its cooling water discharge (on the basis of a one-sided, one-sample z-test for a model mean and the rejection region $z > 1.8$).

A type II error may have been made in reaching this decision – such an error is quite likely when μ = 66.5°C, say [see diagram (ii) at the right] and does not become reasonably *im*probable until μ is around 68°C.

(d)



A5 – 8. (a) **Model:** Let the random variable C_i represent the blood cholesterol level (in mg/dl) *after* treatment with cholestyramine of an equiprobably-selected man with a cholesterol level initially above 265 mg/dl; we assume that such blood cholesterol levels are normally distributed and so use the model $C_i \sim N(\mu, \sigma)$, where σ represents **S**, the s.d. of these men's cholesterol levels in the respondent (or study) population.

If the random variable $\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$, then: $\overline{C} \sim N(\mu, \sigma/\sqrt{n})$.

We assume that the method of selecting the sample of 100 men was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o: $\mu = 210$ mg/dl; *i.e.*, the value of the model mean is consistent with cholestyramine treatment *having* the desired effect of lowering these men's blood cholesterol levels to essentially 'normal'.

H_a: $\mu > 210$ mg/dl; *i.e.*, the value of the model mean is *in*consistent with cholestyramine treatment having the desired effect of lowering these men's blood cholesterol levels to essentially 'normal'.

[We use a *one*-sided test because the statement of the question makes it clear that *failure* of the treatment is when cholesterol levels remain too *high*.]

Dis. meas.: By standardizing in the model for $\overline{C}$ and estimating σ by s (the *sample* standard deviation), we obtain: $\frac{\overline{C} - \mu}{s/\sqrt{n}} \sim t_{n-1}$.

Observed: $\overline{c} = 228$ mg/dl;

expected: $\mu = 210$ mg/dl (under the null hypothesis;

we are given that $s/\sqrt{n} = 12$ mg/dl; also, n = 100.

Value of discrepancy measure: $\frac{228 - 210}{12} = 1.5$.

***P*-value:** $\Pr[t_{99} \geq 1.5] \simeq 0.068\,398 \simeq 0.07$.

Answer: Using a one-sided t-test for the mean of a normal model, the null hypothesis is *not* rejected at the 5% level ($P \simeq 0.07$); thus, the data provide no statistically significant evidence that, after treatment with cholestyramine, the average blood cholesterol level of men with levels initially above 265 mg/dl any longer differs from the national average blood cholesterol level of 210 mg/dl, so that the cholestyramine treatment *does* appear to be effective.

(b) A 95% confidence interval for μ representing $\overline{\mathbf{C}}$, the respondent (or study) population average blood cholesterol level after treatment with cholestyramine for men with cholesterol levels initially above 265 mg/dl, is:

$\overline{c} \pm t^*_{99} \times \frac{s}{\sqrt{n}}$ which is: $228 \pm 1.984 \times 12 \implies (204.2, 251.8)$ mg/dl
or about (204, 252) mg/dl.

A5 – 9. Because we are dealing with *measurements* and the population standard deviation (σ) is *un*known [and so is *estimated* by the *sample* standard deviation (s)], this question involves a (one-sample) t-test for the mean (μ) of a normal model, where model parameter μ represents $\overline{\mathbf{H}}$, the average respondent (or study) population soil pH after the effluent treatment.

(*continued*)

Assignment 5 Outline Solution (continued 4)

A5 – 9. (a) **Model:** Let the random variable H_i represent the pH after treatment with the effluent mixture of an equiprobably-selected sample of the highly alkaline soil; we assume that such soil pHs are normally distributed and so use the model $H_i \sim N(\mu, \sigma)$,
where σ represents **S**, the s.d. of pH variation in the soil which comprises the respondent (or study) population.
If the random variable $\overline{H} = \frac{1}{\mathrm{n}}\sum_{i=1}^{\mathrm{n}} H_i$, then: $\overline{H} \sim N(\mu, \sigma/\sqrt{\mathrm{n}})$.
We assume that the method of selecting the five soil samples was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o**:** $\mu = 8.75$ pH units; *i.e.*, the value of the model mean is *in*consistent with the treatment with the effluent mixture reducing the pH [of the respondent (or study) population] of the highly alkaline soil.

H_a**:** $\mu < 8.75$ pH units; *i.e.*, the value of the model mean is *consistent* with the treatment with the effluent mixture reducing the pH [of the respondnet (or study) population] of the highly alkaline soil.
[We use a *one*-sided test because the statement of the question makes it clear that the treatment with effluent is successful only when it *reduces* the soil pH.]

Dis. meas.: By standardizing in the model for $\overline{H}$ and estimating σ by s (the *sample* standard deviation), we obtain: $\dfrac{\overline{H} - \mu}{s/\sqrt{\mathrm{n}}} \sim t_{\mathrm{n}-1}$.
Observed: $\overline{h} = 8.00$ pH units;
expected: $\mu = 8.75$ pH units (under the null hypothesis);
we are given that $s = 0.60$ pH units; also, n = 5.
Value of discrepancy measure: $\dfrac{8.00 - 8.75}{0.60/\sqrt{5}} \simeq -2795\ 085$.

***P*-value:** $\Pr[t_4 \leq -2.796\ 085] \simeq 0.024\ 528 \simeq 0.024$.

Answer: Using a one-sided t-test for the mean of a normal model, the null hypothesis *is* rejected at the 5% level ($P \simeq 0.024$); thus, the data provide statistically significant evidence that, after the treatment with the effluent mixture, the pH of the highly alkaline soil really *is* reduced, we hope by a practically important amount.

(b) A 99% confidence interval for μ (representing $\overline{\mathbf{H}}$, the average pH of the treated soil) is:
$\overline{h} \pm t_4^* \times \frac{s}{\sqrt{\mathrm{n}}}$ which is: $8.00 \pm 4.604 \times \frac{0.60}{\sqrt{5}} \implies$ (6.765, 9.235) pH units
or about $(6.7_6,\ 9.2_4)$ pH units.

NOTE: The original pH of 8.75 units lies *within* this approximate 99% confidence interval because the *one*-sided test in (a) showed significance at the 5% level but *not* at the 0.5% level.

A5 – 10. Because there is a *natural* pairing of the two times for each of the 25 participants in the investigation, this two-sample situation reduces to *one* sample involving the *differences* between each participant's two times.

Because we are dealing with *measurement* differences and the population standard deviation (σ_d) is *un*known [and so is *estimated* by the the standard deviation (s_d) of the sample of 25 differences], this question involves a paired t-test for the mean (μ_d) of a normal model, where model parameter μ_d represents $\overline{\mathbf{D}}$, the average (left – right) difference in time between the two repetitions of the tasks for members of the respondent (or study) population.

Important considerations in the Plan and Execution stages of the investigation include:

- selecting participant *probabilistically* (*e.g.*, using appropriate systematic selecting) from the pool;
- ensuring that the protocol (*e.g.*, instructions, seating arrangements, familiarization time with the apparatus) is implemened *identically* for each participant;
- asking the participants who have done the tasks *not* to discuss their experience with others in the pool of students waiting to take part in the investigation;
- presenting the two repetitions to each participant in an *equiprobably-selected order*, with the constraint that, across the whole investigation, as near as possible to equal numbers of participants should do each thread direction first.

(b) **Model:** Let the random variable D_i represent the (left – right) difference in time between the two repetitions of the tasks for an equiprobably-selected participant; we assume that such time differences are normally distributed and so use the model $D_i \sim N((\mu_d, \sigma_d)$, where σ_d represents $\mathbf{S}_d$, the s.d. of the time differences in the respondent (or study) population.

Assignment 5 Outline Solution (continued 5)

A5 – 10. (b) **Model:** If the random variable $\overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$, then: $\overline{D} \sim N((\mu_d, \sigma_d/\sqrt{n})$.

We assume that the method of selecting the sample of 25 students was such as to make the limitation imposed by sample error acceptable in the investigation context.

H_o: $\mu_d = 0$ seconds; *i.e.*, the value of the model mean [the respondent (ot study population average] *is* consistent with *no* difference in the average time to complete the task for the two thread directions.

H_a: $\mu_d > 0$ seconds; *i.e.*, the value of the model mean [the respondent (ot study population average] is consistent with right-handed people finding the right-hand thread *quicker* to use.

[We use a *one*-sided test because the investigation's expectation, as described in the alternative hypothesis, is stated in part (b) of the question.]

(c) **Dis. meas.:** By standardizing in the model for $\overline{D}$ and estimating σ_d by s_d (the *sample* standard deviation), we obtain: $\dfrac{\overline{D} - \mu_d}{s_d/\sqrt{n}} \sim t_{n-1}$.

Observed: $\overline{d} = 333/25 = 13.32$ seconds;

expected: $\mu_d = 0$ seconds (under the null hypothesis);

we find $s_d = \sqrt{\dfrac{17{,}061 - (333)^2/25}{24}} \simeq 22.935\ 998$; also $n = 25$.

Value of discrepancy measure: $\dfrac{13.32 - 0}{22.936/\sqrt{25}} \simeq 2.903\ 732$.

***P*-value:** $\Pr[t_{24} \geq 2.903\ 732] \simeq 0.003\ 896 \simeq 0.004$.

Answer: Using a one-sided paired t-test for the mean of a normal model, the null hypothesis *is* rejected at the 1% level ($P \simeq 0.004$); thus, the data provide highly statistically significant evidence that right-handed people *do* find the right-hand thread easier (*i.e.*, quicker) to use than the left-hand thread.

A5 – 11. (a) is (3), two *independent* samples of children taught using the two text segments;

(b) is (2), *matched pairs* – each child is compared with itself under the two teaching methods;

(c) is (1), a *single* sample whose average is compared with the known value of the concentration;

(d) is (3), two *independent* samples of analyses by the new and old methods.