6.780 SEMICONDUCTOR MANUFACTURING SPRING 2003
Problem Set #2 Solutions

$\boxed{\text{Problem 1}}$ Comparison of Left/Right Tube Temperatures.

Part a: We have 10 observations of left and right end temperatures of an LPCVD tube. We know there is some "natural" variation in tube end temperature. So our question – "are the tube end temperatures significantly different (to 95% confidence" is basically asking if the observed differences can be explained as chance measurements, or if they indicate a true underlying difference.

There are several ways to attack this problem. In all of these we have to wrestle with the question of estimating what the within group variance is, since we don't have prior knowledge or data for temperature variance.

First, we can calculate statistics on the tube ends separately. For the left end, we have $\bar{x}_\ell$; for the right $\bar{x}_r$ and s_r^2 as estimates:

$$\bar{x}_\ell = 605.85 \qquad \bar{x}_r = 605.48$$

$$s^2_\ell = \frac{\sum_i (x_\ell - \bar{x}_\ell)^2}{n_\ell - 1} = 1.7917 \qquad s^2_r = 1.2396$$

$$s_\ell = 1.3385 \qquad s_r = 1.1134$$

where $n_\ell = n_r = 10$ measurements at left and right ends.

While s_ℓ^2 and $\bar{x}_\ell$ are estimates of the left tube end variance and mean, we will want to reason about the sampling distributions $\bar{x}_\ell, \bar{x}_r$ which are t-distributed, e.g.

$$\frac{\bar{x}_\ell - \mu_\ell}{s_r/\sqrt{n_\ell}} \sim t_{n_\ell - 1} = t_9$$

Since $\bar{x}_r$ is also t distributed, the linear combination (in this case the difference in means $\bar{x}_\ell - \bar{x}_r$) will also be t distributed. To get the finest resolution in $\bar{x}_\ell - \bar{x}_r$, we want the tightest estimate for the underlying process variance we can get.

We assume we can "pool" the data to get a good tight estimate for within-end variance. We should check this assumption and NOT do this if we find that evidence exists that the variances are different (we check this assumption in part b).

$$S_p^2 = \frac{(n_e - 1) S_e^2 + (n_r - 1) S_r^2}{(n_e - 1) + (n_r - 1)} = \frac{9(1.3385)^2 + 9(1.1134)^2}{18}$$

which can also be thought of as

$$S_p^2 = \frac{\sum (x_e - \bar{x}_e)^2 + \sum (x_r - \bar{x}_r)^2}{n-2} = 1.5156 = S_p^2$$

or $1.231 = S_p$

Using this "within-end" variance estimate, we can now form the sampling distribution for our differences in mean. I like to think about this explicitly using the fact that variances of (independent) r.v.'s add. So our estimate for the $Var\{\bar{x}_e - \bar{x}_r\} = Var\{\bar{x}_e\} + Var\{\bar{x}_r\}$, or

$$S^2_{\bar{x}_e - \bar{x}_r} = S^2_{\bar{x}_e} + S^2_{\bar{x}_r} = \frac{S_p^2}{n_e} + \frac{S_p^2}{n_r}$$

$$= S_p^2 \left(\frac{1}{10} + \frac{1}{10} \right) = \frac{S_p^2}{5} = \frac{1.5156}{5} = 0.3031$$

We form our test statistic under the null hypothesis as

$$t = \frac{(\bar{x}_e - \bar{x}_r) - \cancel{(\mu_e - \mu_r)}^{\to 0}}{S_{\bar{x}_e - \bar{x}_r}} \quad \text{since } \mu_e = \mu_r \text{ under null hypothesis}$$

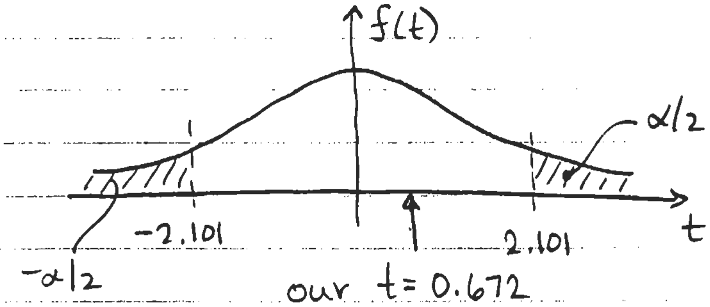
$$= \frac{0.37}{\sqrt{0.3031}} = \frac{0.37}{0.551} = 0.672$$

We compare this to the 95% confidence Interval for what we would expect to observe for t (I use 18 degrees of freedom since I used up two of the 20 in calculating S_p^2):

$$t_{crit} = t_{\alpha/2, 18} = t_{0.025, 18} = 2.101$$

Since $-t_{crit} \le t \le +t_{crit}$

$$-2.101 \le 0.672 \le 2.101$$



we conclude that there is no convincing evidence (to 95% confidence) that the means are different. //

- Alternative approach: another way to do this problem is to compute the 95% c.i. (confidence interval) on each of the means, and determine if these intervals overlap (if they do, then there is no evidence the means are different), e.g.

$$\mu_\ell = 605.85 \pm t_{0.025,9} \frac{s_\ell}{\sqrt{10}} \quad \text{and} \quad \mu_r = 605.48 \pm \overset{2.262}{t_{0.025,9}} \frac{s_r}{\sqrt{10}}$$

Note that here we use the left/right sampling std. devs. separately. If we believe the variances are different left to right, then this approach has the advantage that we can use each of the left and right variance estimates to test the means. In this case

$$\mu_\ell = 605.85 \pm 0.96 \quad \text{and} \quad \mu_r = 605.48 \pm 0.79$$

which overlap.

On the other hand, if we do pool the results we get a BETTER resolution and confidence interval by looking at

$$\mu_\ell - \mu_r = (\bar{x}_\ell - \bar{x}_r) \pm \overset{2.101}{t_{0.025,18}}\, s_p \sqrt{\frac{1}{n_\ell} + \frac{1}{n_r}} = 0.37 \pm 1.157$$

Since this c.i. (which is identical to our t-test) includes 0.0, we conclude means may be equal.

Part b Find the 90% c.i. on ratio of Left/Right variances. Here we use the F distribution assuming $x_l, x_r \sim$ Normal

$$s_l^2 = 1.7917 \quad , \quad s_r^2 = 1.2396$$

$$\frac{s_l^2}{s_r^2} F_{1-\alpha/2, n_l-1, n_r-1} \le \frac{\sigma_l^2}{\sigma_r^2} \le \frac{s_l^2}{s_r^2} F_{\alpha/2, n_l-1, n_r-1}$$

Noting that $\quad F_{1-\alpha/2, u, v} = \dfrac{1}{F_{\alpha/2, u, v}} \quad$ we need only look up

$$F_{0.05, 9, 9} = 3.18$$

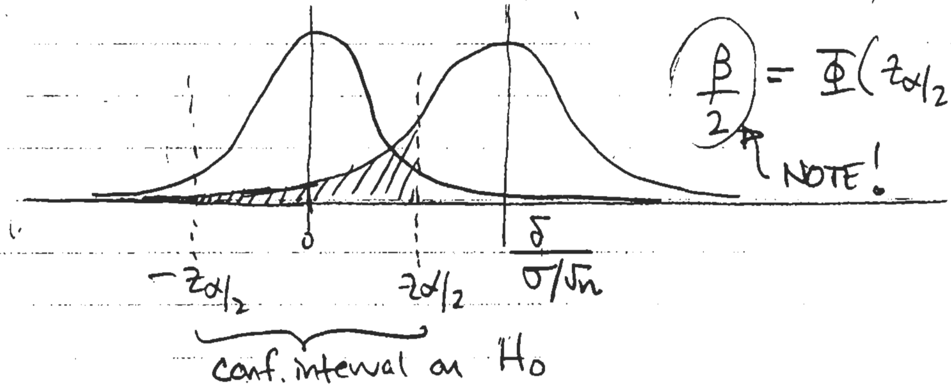
so

$$\frac{1.7917}{1.2396}\left(\frac{1}{3.18}\right) \le \frac{\sigma_l^2}{\sigma_r^2} \le \frac{1.7917}{1.2396}(3.18)$$

$$0.4545 \le \frac{\sigma_l^2}{\sigma_r^2} \le 4.5964 \quad //$$

Since this interval includes $\underline{1}$, we can conclude that there is no reason to think (to 90% confidence) that the Left and Right variances are different.

Part c Sample size to confirm 1°C deviation with $\beta = 10\%$, given $\sigma = 1°C$. We'll assume $\alpha = 0.05$ as in part a. Here we might have a positive temperature deviation of $\delta = 1°C$, <u>or</u> a negative deviation of $\delta = -1°C$, so we must split β risk across both cases. For a positive deviation, $\beta/2$ is shaded area:



$$\frac{\beta}{2} = \Phi\left(z_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_{\alpha/2} - \frac{\delta}{\sigma/\sqrt{n}}\right)$$

NOTE!

where $\sigma = 1°C$

$n = 10$

$\alpha/2 = 0.025$

$\beta/2 = 0.05$

We can note that for small $\beta/2$, the part of the integration to the "left" of $-z_{\alpha/2}$ is negligible, so a very good approximation is the area under the shifted distribution to left of $z_{\alpha/2}$:

$$\beta/2 \approx 1 - \Phi\left(\frac{\delta\sqrt{n}}{\sigma} - z_{\alpha/2}\right)$$

Here $z_{0.025} = 1.96$, so we have $0.05 = 1 - \Phi\left(\frac{1 \cdot \sqrt{n}}{1} - 1.96\right)$

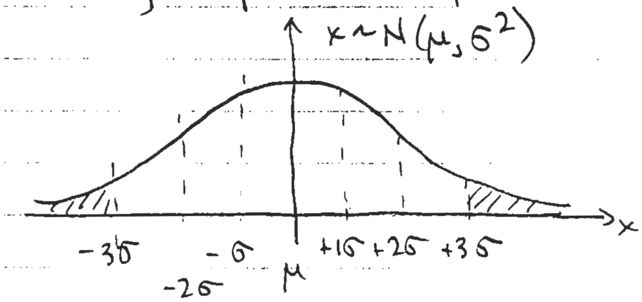
or $0.95 = \Phi(\sqrt{n} - 1.96)$. Using tables for $\Phi(z)$ we get $\Phi(z) = 0.95$ for $z \approx 1.645$, or

$\sqrt{n} - 1.96 = 1.645$

$\sqrt{n} = 3.605 \implies n = 12.99$. So we need 13 samples //

$\boxed{\text{Problem 2}}$ Normally distributed x, monitored with $\pm 3\sigma$ chart.

<u>Part a</u> Probability a point will plot outside control limits



$$\Pr\{x > \mu + 3\sigma\} + \Pr\{x < \mu - 3\sigma\}$$

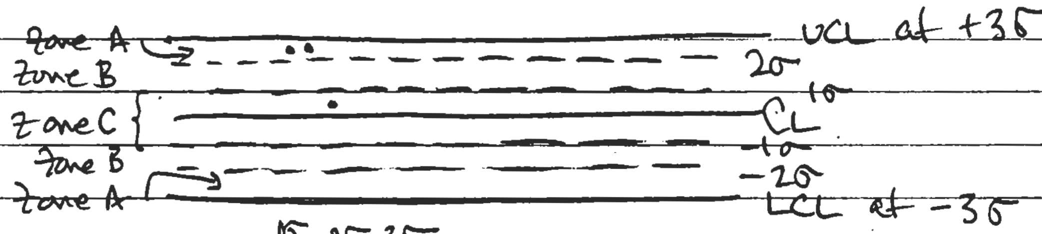
$= 1 - \Phi(3) + \Phi(-3)$ but $\Phi(-z) = 1 - \Phi(z)$

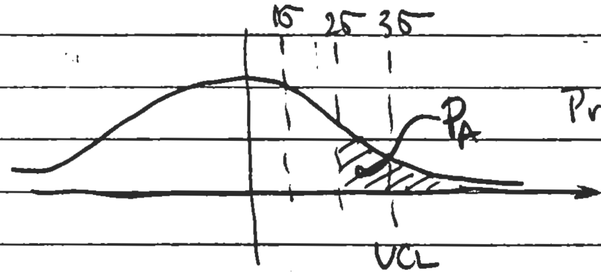
$= 2(1 - \Phi(3))$

$= 2(1 - 0.99865) = 0.0027$

<u>Part b</u> Probability that 2 out of 3 consecutive points lie on same side of center line outside the $2\text{-}\sigma$ line (WECO rule #2).

This problem as written is somewhat ambiguous, and can be interpreted in multiple ways. Looking at pg. 18 in the estimation lecture notes and the WECO rule chart there, we'll assume the intent is to require that all three points are on same side of CL, and two or three of the three points lie in zone A or beyond:

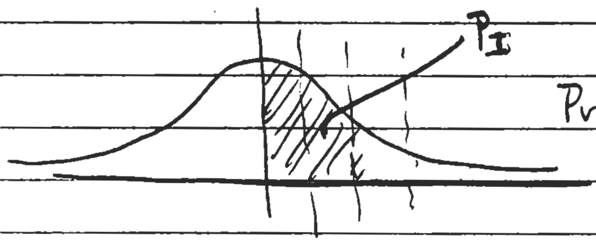




$Pr\{\text{zone A above CL or beyond}\} = P_A = 1 - \Phi(2)$

$= 1 - .97725$

$= 0.02275$



$Pr\{\text{above CL but not in zone A or beyond}\} = P_{II} = 1/2 - P_A$

So the possible ways we could trigger endpoint by rule 2 is:

$$Pr\{\text{trigger}\} = 2\Big[\overbrace{P_A^3}^{\text{all 3 in zone A or beyond}} + \underbrace{3\,P_A^2 \cdot P_{II}}_{\text{Exactly 2 in zone A and 1 in zone II}}\Big]$$

because the same probabilities apply to 2/3 points in zone A or beyond BELOW the CL

$$= 2\left[(0.02275)^3 + 3\,(0.02275)^2(0.5 - 0.02275)\right]$$

$$= 2\left[1.18 \times 10^{-5} + 74.1 \times 10^{-5}\right] = 0.001506$$

so 3 in a row in zone A or beyond is relatively unlik

NOTE: See next page for an explanation of possible errors.

- One CLOSE but not quite right approach:

A valid interpretation of the rule is that 2/3 points must lie on same side of C_L and be $\geq 2\sigma$ away; and the third point can lie anywhere.

So we might think

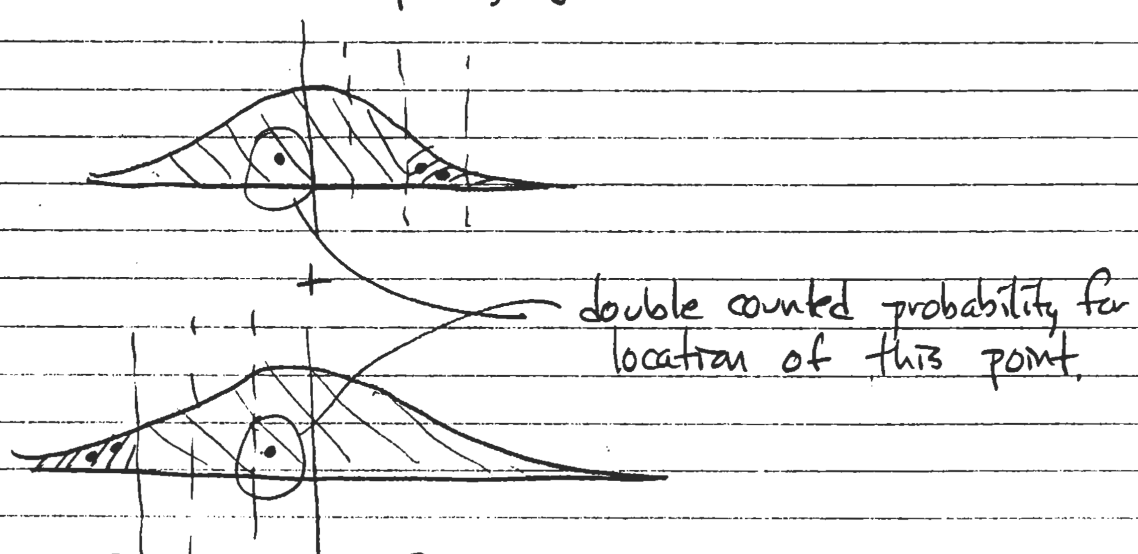
$$\Pr\{\text{trigger}\} = \binom{3}{2} P_A^2 (1-P_A) = 0.001517$$

This gives a very close # to what we saw before. The problem is that the formula doesn't match the above interpretation, as it really only captures the probability of triggering on the High side of C_L (2/3 above C_L).

However, if we do

$$\Pr\{\text{trigger}\} = 2\left[\binom{3}{2} P_A^2 (1-P_A)\right] = 0.003034$$

we make another subtle error: we double count possible locations for the "third" point, e.g.



- NOTE: $2\left[\binom{3}{2} P_A^2 \left(\frac{1}{2}\right)\right]$ also slightly overcounts – 3 combinations of P_A^3

↑ same side of C_L