1. [*15 marks*] Twenty-five measurements of quantities *x* and *y* are measured and reported to the right, along with the means and sample standard deviations. Are the sample means "significantly different"? Support your answer with a few sentences and simple calculations.

point #	x	y
1	10.8	9.7
2	15.3	5.4
3	14.3	9.9
4	6.4	15.8
5	15.0	9.3
6	13.5	1.8
7	16.2	5.8
8	0.7	14.0
9	16.4	2.5
10	23.9	14.9
11	31.7	15.3
12	1.1	14.1
13	12.6	12.3
14	11.5	10.6
15	12.7	11.4
16	14.3	25.2
17	19.4	13.4
18	13.6	6.7
19	16.2	5.3
20	4.8	9.6
21	16.8	10.9
22	9.2	5.2
23	10.2	6.0
24	19.3	8.7
25	27.5	0.2
sample mean	14.1	9.8
standard dev	7.1	5.4

The standard deviations (7.1 and 5.4) are estimates of the errors in each measurement, but we want the uncertainty in the mean of 25 measurements. Because the random errors will partially cancel, the standard deviation of the mean will decrease a $N^{-0.5}$.

That is, we should characterize the means as

$$\bar{x} = 14.1 \pm \frac{7.1}{\sqrt{25}} = 14.1 \pm 1.4$$

$$\bar{y} = 9.8 \pm \frac{5.4}{\sqrt{25}} = 9.8 \pm 1.1$$

This 67% confidence intervals do not overlap, by a wide margin, and even the 2σ (95% confidence intervals) do not overlap: **THE MEANS ARE SIGNIFICANTLY DIFFERENT**

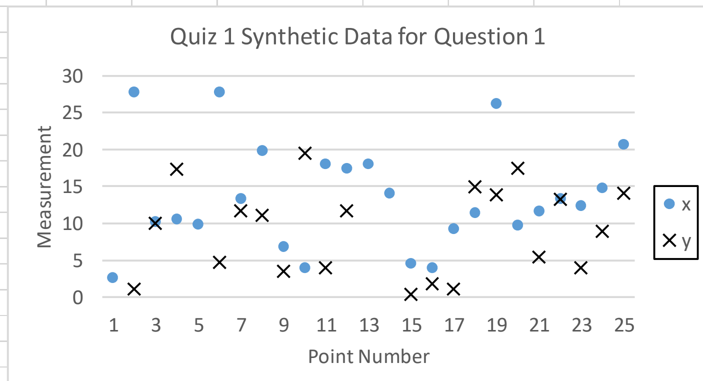
Discussion of question 1

A key concept here is that the means are “significantly different” if the observed difference is unlikely to be due to chance. This means that the difference is larger than the sum of the error estimates, or graphically that the error bars do not overlap. Making an assessment of this sort is worth 8/15, even if the next key part is wrong.

The second key part (and maybe the most important idea in the question) is that the standard deviation of the mean decreases with the sample size. This could be the single most important idea you will take from the class, and apparently the majority of the class didn’t have this clear on test day. Make sure you understand this!

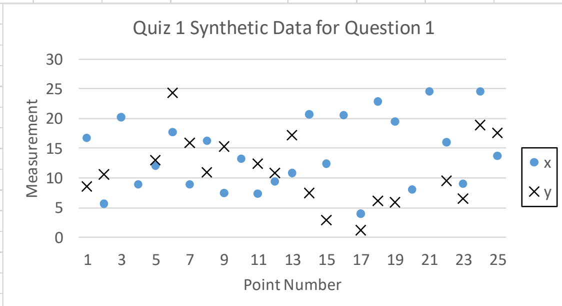
The numbers for this quiz problem were generated using random normal deviates, and every time the set is updated, you get a different set of scattered points. The graph at the right shows another set. It is different from the one on the test, but the sample means are pretty close, as are the SDOM values. Notice that the raw data itself is pretty scattered! Simply glancing at the pattern, it is not clear that there is a substantial difference – but that is why we do statistical calculations.

The next page shows four more “realizations” of the data – hopefully convincing you that the SDOM works as it should.

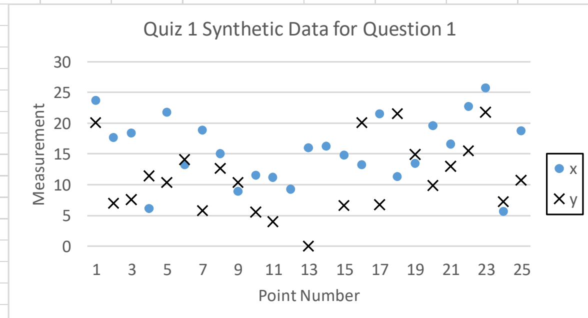


sample mean	13.6	7.2
standard dev	7.1	7.0
sample size N	25	25
stdev of mean	1.4	1.4

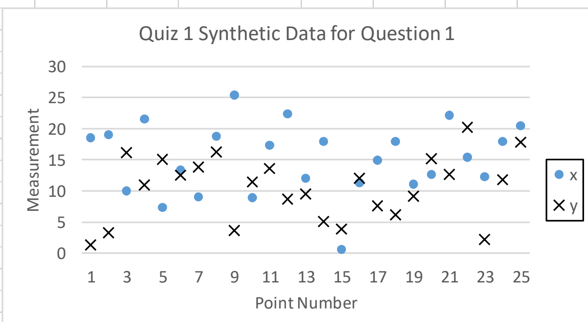
Question 1 continued



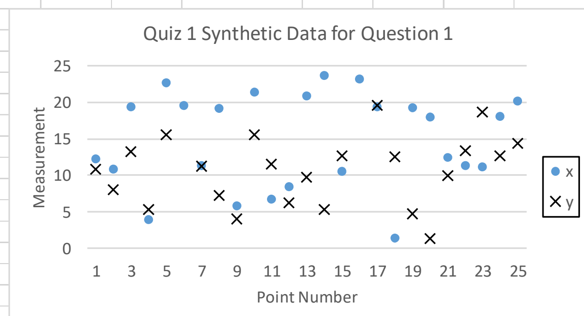
sample mean	14.0	7.7
standard dev	6.1	8.3
sample size N	25	25
stdev of mean	1.2	1.7



sample mean	15.7	10.1
standard dev	5.4	6.7
sample size N	25	25
stdev of mean	1.1	1.3



sample mean	15.1	10.4
standard dev	5.7	5.2
sample size N	25	25
stdev of mean	1.1	1.0



sample mean	14.9	9.7
standard dev	6.5	5.6
sample size N	25	25
stdev of mean	1.3	1.1

Marking notes for question 1

- Half of the class at least assessed the difference by comparing with the standard deviations. If carefully and logically done – 8/15 marks (you get exactly the wrong result, don't do this again.
- If you uses the standard deviations but concluded that the difference was significant (without doing the SDOM calculation), there is something logically very wrong 7/15.
- Calculating the standard deviations of the mean, but concluding that the differences are NOT significant – 8/15. Although we did not specify the exact confidence level to use, this case is not close and nobody uses 3σ intervals.
- Possibly a few people concluded that the difference was significant because 14.1 is "significantly" larger than 9.8. No marks for this unless there is an appeal to engineering design, where typically 40% difference is of practical importance 7/15. There is little justification to take this approach in this quiz because there was no engineering context provided.
- A few people had fairly creative and mysterious interpretations of the question, such that they computed means of sub-samples of the data and showed that they were similar to the means of the 25 measurements. 2 to 4 marks if there seems to be some internal consistency in the calculations.

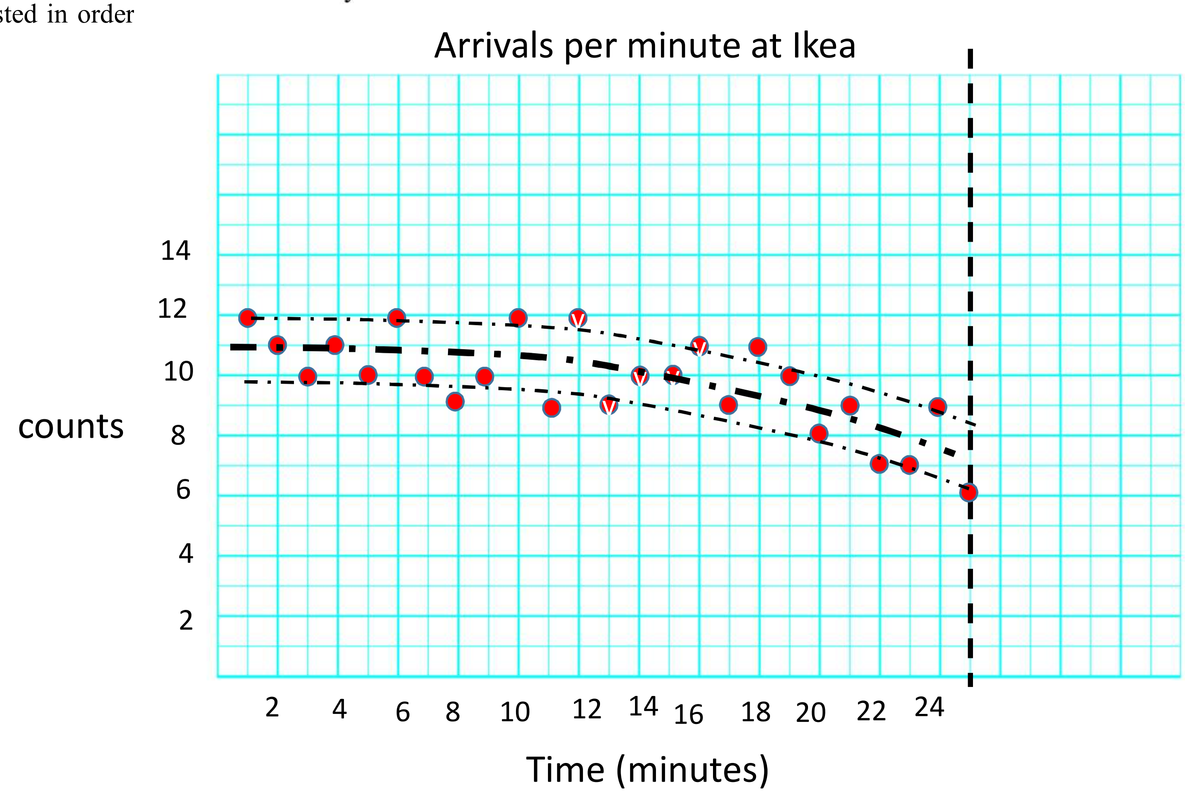
Question 2

[*15 marks*] In an entry-level job at Ikea, you are asked to count the number of customers per minute entering the store over a 25-minute period. They want to know whether the arrival process is Poisson, that is, customers arrive randomly, independent of when other customers arrive. Here are your data, listed in order from the 1st to the 25th minute.

```
12, 11, 10, 11, 10,
12, 10,  9, 10, 12,
 9, 12,  9, 10, 10,
11,  9, 11, 10,  8,
 9,  7,  7,  9,  6
```

A time series makes a lot of sense as the first graph to plot because the ORDER of the data might make a difference. For this reason, 7 marks for a time series, 5 marks for a histogram or other plausible plot. The graphs should have at least enough labelling so that there is no doubt about what is plotted.

Are arrivals random (that is, described by a Poisson process)? Explain how you know, supporting your answer with a plot of the data below and any computed statistics that you think would be useful.



Question 2 continued

The arrival process does NOT appear to be Poisson and there are 2 reasons for this:

a) There is an obvious downward trend in arrival rates, so the number of counts is not independent of when the interval occurs.
b) The expected standard deviation of random counts should be N^0.5, and for our case the typical counts ~10, so expect σ=3.2. Our data is more tightly banded, with a deviation of about 1 count around the trend (see plot). This is too smooth for a Poisson process!

Answer (yes/no) 2 marks for stating that this is not Poisson
Reasons:

6 marks for both explanations
5 marks for one of the 2 reasons

For example, students plotting a histogram instead of a times series would get 5 for the graph. If they compare the width to the distribution to root(N), then 5. This probably leads to the wrong conclusion about whether it is Poisson, so total 10/15, with adjustments up and down for clarity and logic.

Question 3

[15 marks] The turbulent convective heat transfer coefficient h in a pipe is known to be modelled accurately by the correlation

$Nu=0.023\ Re^{0.8}\ Pr^{0.33}$ (for cooling)

where $Nu=hd/k$, $Re=\rho\ V\ d/\mu$ and $Pr=\mu\ C_p/k$. In an experiment we measure the following:

Heat transfer coefficient $h = 200 \pm 10$ W/m^2/C

Pipe diameter $d = 0.01 \pm 0.0002$ m

Fluid conductivity $k = 0.04 \pm 0.002$ W/m/C

Viscosity $\mu = 20 \times 10^{-6}$ Pa-s $\pm$ 3%

Density $\rho = 3 \pm 0.1$ kg/m^3

Velocity $V = 5$ m/s $\pm$ 4%

a) Nu=hd/k=50 Re=ρVd/μ=7500

To get the uncertainty, it is easier to work with relative uncertainties because the expressions for Nu and Re are products of simple powers

$$\left(\frac{\sigma_{Nu}}{Nu}\right)^2 = \left(\frac{\sigma_h}{h}\right)^2+\left(\frac{\sigma_d}{d}\right)^2+\left(\frac{\sigma_k}{k}\right)^2 = \left(\frac{10}{200}\right)^2+\left(\frac{.0002}{.01}\right)^2+\left(\frac{.002}{.04}\right)^2 = 0.0054$$

$$\left(\frac{\sigma_{Nu}}{Nu}\right) = 0.073 \quad \sigma_{Nu} = 3.7 \quad Nu = 50 \pm 5$$

a) Compute the uncertainty in Nu and Re.

$$\left(\frac{\sigma_{Re}}{Re}\right)^2 = \left(\frac{\sigma_\rho}{\rho}\right)^2+\left(\frac{\sigma_V}{V}\right)^2+\left(\frac{\sigma_d}{d}\right)^2+\left(\frac{\sigma_\mu}{\mu}\right)^2 = \left(\frac{.1}{3}\right)^2+(0.04)^2+\left(\frac{.0002}{.01}\right)^2+(0.04)^2 = 0.00401$$

$$\left(\frac{\sigma_{Re}}{Re}\right) = 0.063 \quad \sigma_{Re} = 475 \quad Re = 7500 \pm 500$$

6 marks for part a, with correct answers even if there a couple extra decimal points.

Question 3 cont.

b) Compute Pr and its uncertainty.

$$Pr = \sqrt[0.33]{\frac{Nu}{0.023\ Re^{0.8}}} = 5.15$$ (probably more decimals then warranted, but keep these for now)

$$Pr = constant\ x\ Nu^3 Re^{-2.4}$$

$$\left(\frac{\sigma_{Pr}}{Pr}\right)^2 = \left(3\frac{\sigma_{Nu}}{Nu}\right)^2 + \left(2.4\frac{\sigma_{Re}}{Re}\right)^2 = \left(3\ (.073)\right)^2 + \left(2.4(.063)\right)^2$$

$$\left(\frac{\sigma_{Pr}}{Pr}\right) = \sqrt{0.0479 + 0.0299} = 0.266 = 27\%$$

$$\sigma_{Pr} = 1.4 \qquad Pr = 5.2 \pm 1.4$$

5 marks for part b (2 for Pr and 3 for the uncertainty). No need to have both relative AND absolute uncertainties, and no need to do the calculations using relative uncertainties – it is just a lot neater to do so.

c) Is it a good idea to determine Pr by determining Nu, Re and using the correlation as we have done above? Why?

No, this is a poor approach, as can be seen from the 27% uncertainty. This results from many measurements going into the calculation and also because Pr will depend on a high power (3 or 2.4) of Nu and Re – essentially this triples the errors.

4 marks for part c (1 for the correct answer (yes/no), 3 for the reason, and possibly fraction of a bonus mark for the suggestion of a better method)

Question 3 commentary and marking

- The class did very well on this question. Up to one bonus mark for the following:
  - Careful presentation of significant figures
  - Explanation that we are assuming that the errors in the different parameters are independent (therefor allowing addition by quadrature)
  - Particularly good discussion for part c.
- If you did not assume independent errors, this problem would be a real mess, but 8/15 for starting correctly on this path, stating a reason why you are doing it this way. If you managed to slog through to the end on this path, 12/15.

Question 4

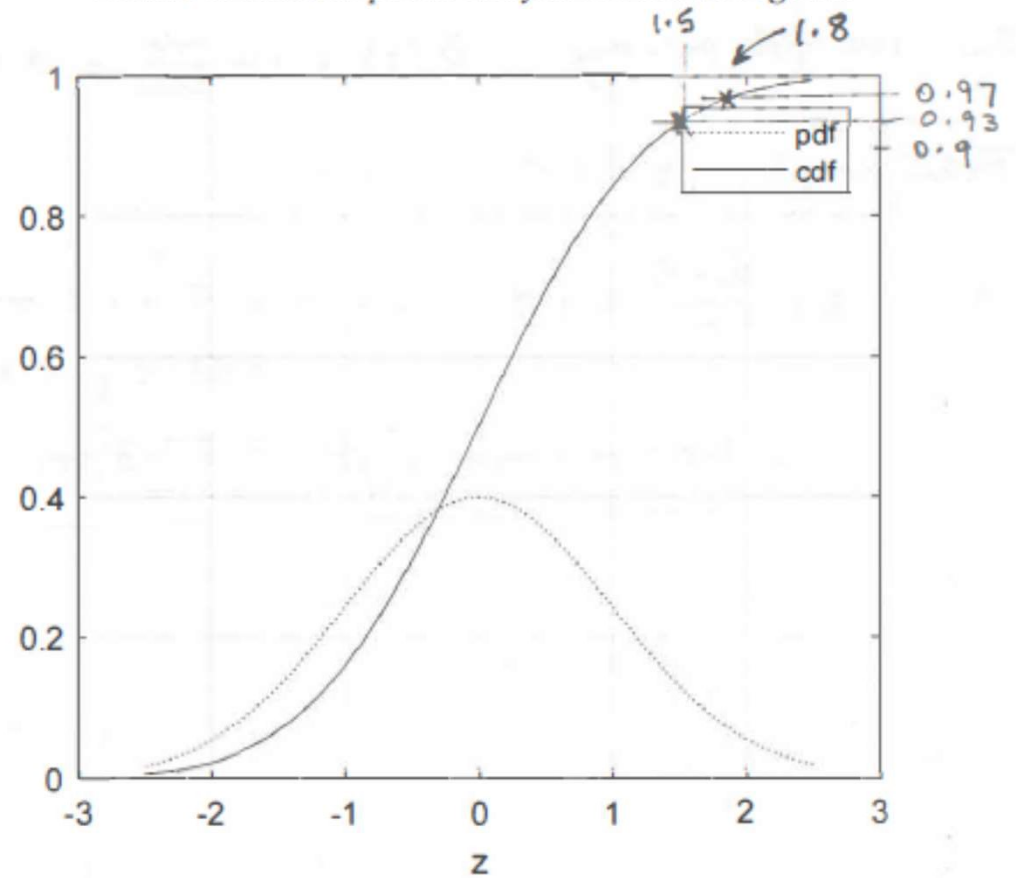
- Original hand-written solutions follow (a good example of the clarity and neatness desired). The solutions for the previous questions have much more commentary than we would have expected in your solutions.
- The two parts have approximately similar weights.
- From preliminary grading, it appears that a substantial number of people had difficulty reading the graph. This may have been due to "last-question rushing", but please make sure that this becomes easy for you. We believe that this graphical information is often more important than the equations.

4. [15 marks] The Weaver Creek Salmon Hatchery is a specially prepared area near Mission BC where they collect salmon fish to provide them with a safe and comfortable place to lay their eggs. In this way the fish will be more successful in reproducing and the size of the population can be enlarged. 2000 salmon swim by the entrance to the hatchery each day, and their lengths are distributed normally with a mean of 53.5cm and standard deviation of 2.5cm.

(a) If the hatchery sets a minimum acceptable fish length at 58cm, approximately how many fish could potentially be accepted each day?

(b) If the hatchery opens a new section that enables to it accept up to 144 fish each day, what revised fish minimum length limit would you recommend ?

You may use the normal pdf/cdf below for your evaluation.
Be sure to mark the points that you use on the diagram.



(a) *(A blank page for your writing)*

z-statistic: $z = \frac{x - \bar{x}}{\sigma} = \frac{58 - 53.5}{2.5} = \frac{4.5}{2.5} = 1.8$

From graph: $\Phi(1.8) \approx 0.97$.

$\rightarrow$ Probability of a fish > 58cm $= 1 - 0.97 = 0.03$

There are 2000 fish per day.

$\rightarrow$ number > 58 cm $= 2000 \times 0.03 = \underline{60 \text{ fish}}$

(b) For 144 fish per day, $\Phi(z) = 1 - \frac{144}{2000} = 0.93$

From graph: $z = 1.5$

$$z = \frac{x - \bar{x}}{\sigma} = 1.5 \rightarrow x = \bar{x} + 1.5\sigma$$

$$= 53.5 + 1.5 \times 2.5$$

New minimum length $= 57.3$ cm