Technical Problem Solving 2014-2015

This exam is not guaranteed to be error free, but attempts to cover the main topics of the event. In particular, the following topics are covered:

- Newton's law of cooling
- Model fitting
- Residuals
- Correlation coefficient
- Outliers
- Mean, median mode
- Standard deviation
- Normal distribution
- Random/systematic error

Please post comments/questions/corrections on scioly.org forums under 2015 "Technical Problem Solving C", direct link here:
http://www.scioly.org/phpBB3/viewtopic.php?f=186&t=5917

If there are any corrections, they will be posted there.

This was developed for Santa Clara County Regional Science Olympiad.

The solutions may not have the correct significant figures.

Technical Problem Solving (KEY)

Division C
March 21, 2015
Bay Area Regional Science Olympiad

Team name: ____
Team number: ____
Team member name #1: ____
Team member name #2: ____

Time: 50 minutes

Scoring:

- 200 points total, 100 points for each of two stations
- each question is worth 10 points
- partial/full credit will be assigned based on both correctness of answers and procedure leading to answers, so please show all work and provide appropriate units and significant figures for answers
- specifying measurement or calculation uncertainty (for example, 1.1 cm **+/- 0.1 cm**) is not required unless specifically asked for in the question
- ties will be broken at random

Materials you may use:

- up to two calculators
- Two 8.5" x 11" sheets of paper with information on both sides

Equations:

Newton's law of cooling:

T(t) = Ts+(To-Ts)*exp(-kt)

Notes

- Please turn phones off
- This may be a difficult test, so do your best
- Good luck!

Station #1

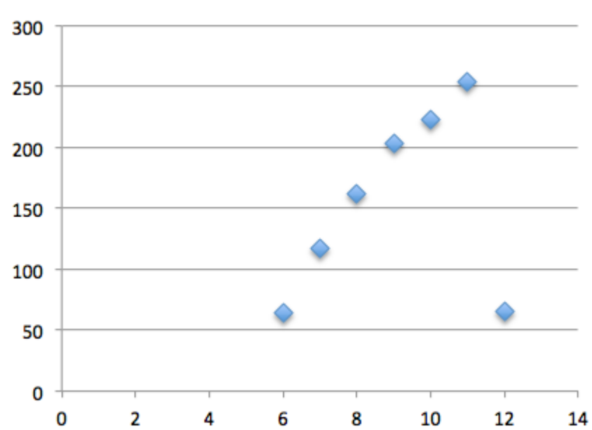
Storyline: Someone was found immobilized shortly after they had started baking a piece of salmon in their oven. When investigators arrived, they noted the oven was at 450F and that the temperature of the piece of salmon was 375F.

Your lab oven only goes up to 350F so you set it at that and let it warm up. You stick your temperature probe in the similarly sized piece of salmon, initially at room temperature (70F), and start the temperature recording at time t = 6 minutes. You then place the piece of salmon in the oven, and come back about 5 minutes later and pull out the temperature probe. At this point you stop the temperature recording.

Here is the data:

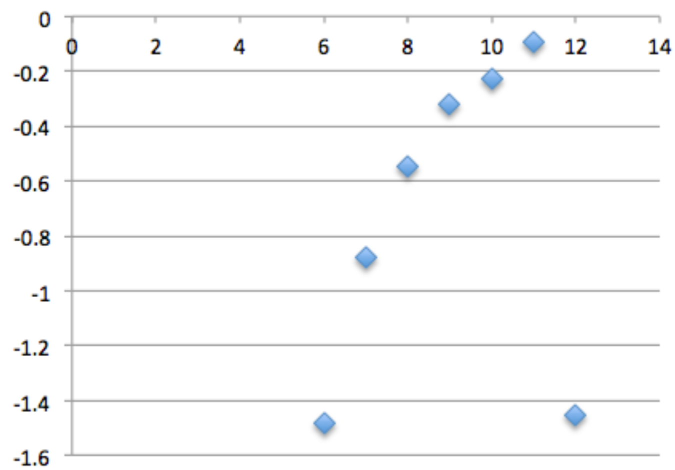
time (m)	temperature (F)
6	**63.7**
7	**116.6**
8	**162.5**
9	**203.2**
10	**222.7**
11	**254.4**
12	**65.5**

1. Plot the temperature as a function of time



y axis label: temperature(F)
x-axis label: time (minutes)

2. We know from Newton's Law of Cooling that temperature does not change linearly with time. In order to use linear regression, we need to transform the data. Plot the ln(temperature/(350F-70F)) as a function of time (minutes), where ln(x) denotes the natural logarithm function.



y axis label: ln(temperature(350F-70F))
x-axis label: time (minutes)

Note: As a physical model, this quantity plotted does not relate directly to Newton's low of cooling, since we should have the following relationship:

$$\ln((\text{temperature} - 350F)/(70F - 350F)) = -k*t$$

This was a mistake on my part when I generated the data. However, for data-fitting purposes and for the rest of this question, we'll use this. That is, although the models used in this question may not have an accurate physical connection to the data, the modeling and analysis techniques here still apply.

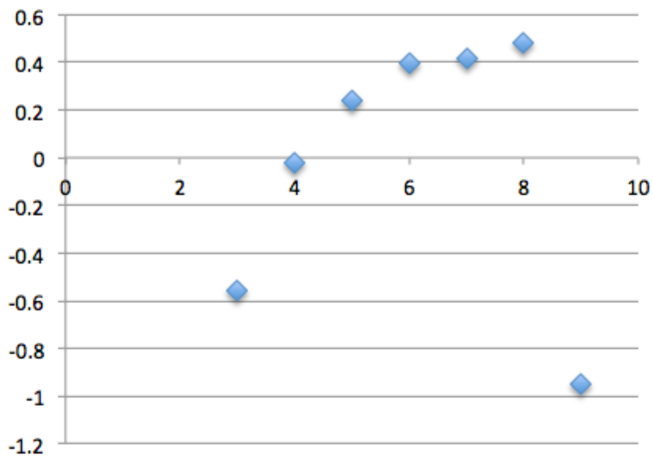
You hand this data over to your two colleagues who come up with two models that describe the data, where Y is the quantity plotted in #2 and t is time in minutes and units have been omitted.

model A: $Y = 0.0699t - 1.3433$
model B: $Y = 0.1875t - 2.1005$

They disagree about who has the better model, so you decide to do some analysis to find out.

3. Compute the residual for each time point using the log of the temperature as defined in #2 and model A. Then plot these residuals.

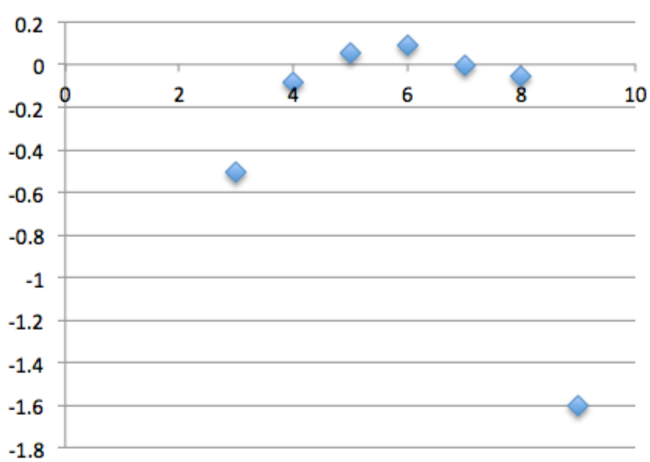
The residual shows how far from the actual data are a given model's predictions.



y axis label: residual
x-axis label: time (minutes)

What you should notice is that the average residual is roughly around zero. This is because this model is a linear regression on all points.

4. Compute the residual for each time point using the log of the temperature as defined in #2 and model B. Then plot these residuals.



y axis label: residual
x-axis label: time (minutes)

What you should notice is that the model fits all of the points extremely well except for the last point. This model was fit with the last outlier point excluded

5. Compute the R^2-values for each model including all time points. When is it possible for the R^2 to be negative?

0.06987 for A, -0.4513 for B, see http://en.wikipedia.org/wiki/Coefficient_of_determination#Definitions

negative when the model is worse than simply using the average of all measurements. See http://stats.stackexchange.com/questions/12900/when-is-r-squared-negative

6. Which model has the higher R^2-value, and what does this mean?
Model A, that it better represents all data points

7. Which model is more informative for the investigation?
Model B since it matches the relevant data points (the ones not including the outlier) better. Also, we know that during the data collection, the temperature probe was pulled out before the recording was stopped, which produced the outlier.

In actual data science applications, it's important to study the data closely and to "clean it up" before starting to fit any models or do any statistical analysis.

8. Pick the more informative model based on #7. Assume this model gives you the coefficient "k" in Newton's law of cooling – that is, assume k = 0.0699 for model A and k = 0.1875 for model B. Determine how many minutes prior to the arrival of the investigators was the piece of salmon placed into the oven. You may assume that the oven was preheated to 450F and that the initial temperature of the salmon was the same as in the lab test.

Using model B we get that k = 0.1875.

T(t)=375 = 450-(450-70)*EXP(-0.1875*t)

t = 8.65 minutes (23.2 minutes for the other model)

9. Name 3 factors that physically impact heating rates in these situations that have not been factored in to our lab experiment.
*** chemical changes during the cooking process**
*** changes in density**

*** changes in oven temperature from the salmon or opening the door to take measurements**
*** other reasonable explanations**

10. You learn that the temperature probe used in your lab experiments was always 5F higher than the true temperature. However the probe used at the crime scene was still accurate. What type of measurement error is this? How would it qualitatively change your estimate in #8?

Systematic error. It would not change because the "k" that you estimate in the lab only depends on relative temperatures, and not the absolute temperature.

Station #2

You take these temperature measurements (F) at time zero:

32, 56, 102, 101, 102, 98, 105, 103

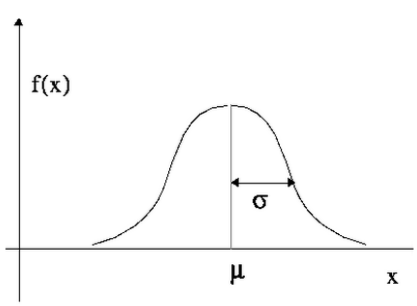
1. Compute the mean, mode and median of the data

mean : ____________**87 F (87.4 F)**____________
mode : ____________**102 F** ____________
median : ____________**102 F (101.5 F)**____________

2. Compute the (population) standard deviation of the data

Standard deviation: __________**26 F (25.8 F)** ____________

3. Unrelated to #1 and #2, draw an ideal normal distribution, and label the mean, standard deviation and label the axes



At a later time exactly one hour later, you take these measurements (units of F):

26, 87, 238, 241, 252, 237, 240, 241

4. Compute the mean, mode and median of the data

mean : __________**195 F (195.25 F)**__________________
mode : _____________**241 F**________________
median : __________**239 F**___________________

5. Using the mean temperatures, determine how long after the initial measurement will the temperature reaches 400F, assuming the ambient temperature is 500F.

k = 0.303, t = 4.68 hours

6. Do the same, but use the median instead of the mean.
k = .423, t = 3.27 hours

7. You learn that your temperature probe takes time to start up and reads a much lower temperature when it is initially first powered on, making the first two measurements in each sequence inaccurate. Unfortunately, this happened both times you took the measurements. Without necessarily doing any math, explain which of your answers, to question #5 or #6, is probably more accurate given this information?

Answer to #6 because the median isn't skewed by the two outliers.

8. What is the systematic error in this experiment?

Temperature probe reading low values initially.

9. What is the random error in this experiment and describe how it could be reduced through statistical analysis?

Fluctuations in measurements after the first two. Can be reduced by averaging.

10. You find a wormhole behind your bookshelf and end up in a world where instead of the rate of change of temperature of a body being proportional to the difference between its own and the surrounding temperature (Newton's law of cooling), it is proportional to the square of that difference. Derive Newton's law of cooling on this new planet.

dT/dt = k(T-Ts)^2
dt/(T-Ts)^2 = kdt

integrating both sides
1/(T-Ts) +c = -kt
T = Ts - 1/(kt-c)

The sign on k and c can (and might be) be flipped