1. Consider the following DNA coding sequence:

  GATTACACGTGCCTTGGA

a) Suppose the DNA sequence was transcribed into RNA. Give the resulting RNA sequence. (Assume that the provided DNA sequence is the coding sequence rather than the template sequence when answering this question.)

b) What sequence of amino acids would result if a ribosome were to translate the RNA sequence from part (a)?

c) Suppose the 12th base of the DNA sequence was changed from a C to an A. What sequence of amino acids would result if a ribosome were to translate the RNA sequence coded for by this mutated DNA sequence?

2. Suppose the figure below represents a two-color microarray. Assume this slide was spotted in the manner discussed in class.
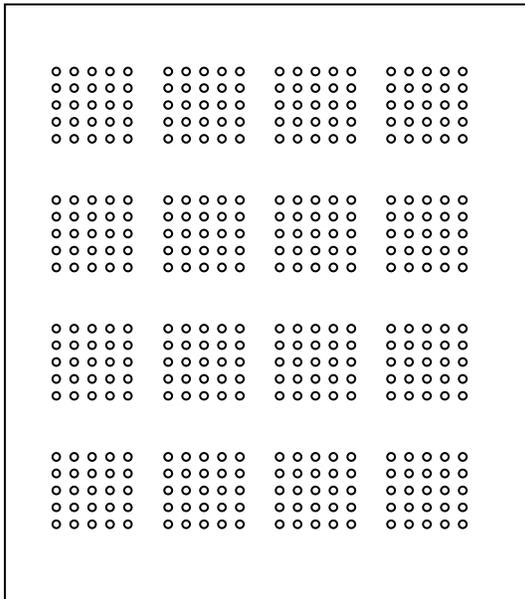
a) What is, most likely, the maximum number of pins that were used to print this slide?

b) Color in any two spots that were printed at the same time.

c) Circle any one set of spots that were printed with a single pin.

3. Use the symbolic notation discussed in class to describe the following microarray designs. You should draw one circle for each experimental unit. Each circle should be labeled to describe the treatment applied to the experimental unit. Each slide should be represented by an arrow connecting the experimental units hybridized on the slide. The direction of the arrow should indicate dye assignment (tail=Cy3=green and head=Cy5=red).

a) A total of 8 rats were randomly assigned to two treatments (A and B) with 4 rats in each treatment group. Two RNA extractions were made from each rat, and in each case one extraction was dyed with Cy3 dye and the other dyed with Cy5 dye. Each rat that received treatment A was paired with a rat that received treatment B. Two slides were used to measure gene expression for each pair. One slide compared the A and B samples using Cy3 dye for the A sample and Cy5 dye for the B sample. The other slide used the opposite dye assignment.

b)   An experiment was conducted to study gene expression in barley during the first 24 hours after inoculation with a fungus. Five plants were grown to two-weeks of age and randomly assigned to be harvested at either 0, 6, 12, 18, or 24 hours after inoculation (one plant for each time). The plants were simultaneously inoculated, and RNA was extracted from each plant at the assigned time. Two labeled samples (one Cy3 and one Cy5) were obtained from each plant. Plants harvested at adjacent time points were hybridized together on one slide with the Cy3 dye used for the earlier time point and the Cy5 dye for the later time point. In addition, samples from the time-0 plant and the time-24 plant were hybridized together on a slide with the time 0 sample dyed Cy5 and the time-24 sample dyed Cy3. The entire process was repeated three more times. In one of the repetitions, the dye assignment strategy was exactly as described above. In the other two repetitions, all dye assignments were reversed.

4. How many different images can be created with resolution 16 x 16 and color depth 3?

5. Suppose a two-color microarray experiment has been conducted to compare the effects of treatments A and B on gene expression soybean plants. A total of sixteen plants were assigned to the two treatments using a completely randomized design. Each treatment-A plant was randomly paired with a treatment-B plant to form 8 independent pairs. Samples from paired plants were hybridized to slides, with one slide for each pair. For 4 of the 8 slides, the treatment-A sample was dyed with the Cy3 dye, and the treatment-B sample was dyed with the Cy5 dye. The dye assignment was reversed for the other 4 slides. Data for one gene are provided below.

```
Slide  Dye  Treatment  Normalized Log Signal
  1     3       A               5.2
  1     5       B               7.9
  2     3       A               3.3
  2     5       B               6.5
  3     3       A               1.1
  3     5       B               5.9
  4     3       A               2.8
  4     5       B               7.6
  5     3       B               6.3
  5     5       A               7.6
  6     3       B               6.0
  6     5       A               6.2
  7     3       B               4.0
  7     5       A               4.1
  8     3       B               7.2
  8     5       A              10.7
```

a) Fit a model like that described for the two-color microarray example in the slides entitled "Introduction to Statistical Design and Analysis of Microarray Experiments" (available on our course web site). You may assume that the random effects are independent and normally distributed. Furthermore, you may assume that the slide random effects have a constant variance and that the residual random effects also have a constant variance. (If you do not know how to fit a model with random effects, you may assume that slide effects are fixed. For this dataset, you will get the same answer whether you treat slide effects as fixed or random.) Provide statistics for testing the significance of treatment and dye effects. State the degrees of freedom associated with each statistic and provide appropriate p-values along with brief conclusions.

b) Estimate the difference between treatment effects and provide a 95% confidence interval for the difference.

c) Reduce the pair of observations for each slide to one difference by subtracting each treatment-A observation from each treatment-B observation. Show how you can use these differences to obtain the test for treatment effects and the test for dye effects that you provided in part (a).

d) It is common to talk about fold changes in microarray data analysis because treatment effects are assumed to be multiplicative. For example suppose a gene expresses a level x when a plant is treated with a control substance and at level $1.7x$ when a plant is treated with a virus. Then the virus is said to cause a 1.7-fold change in expression. Multiplicative effects on the original scale become additive on the log scale. For example, compare $log\ x$ to $log\ (1.7x)$. Because $log\ (1.7x) = log\ 1.7 + log\ x$, we can see that the effect of the virus relative to control is to add $log\ 1.7$ to the log-scale expression. We analyze data on the log scale because linear model analysis is designed to detect additive treatment effects and because the assumptions of our linear model analysis are usually more closely satisfied with log-scale expression data. For interpretation purposes, it is often nice to convert an estimated difference (e.g., $log\ (1.7x) - log\ x = log\ 1.7$) back to a fold change (1.7). Usually, the fold change is defined as $e^{|\tau 1 - \tau 2|}$, where $\tau 1$ and $\tau 2$ denote the treatment effects associated with treatments 1 and 2, respectively. Convert your answers in (b) to an estimated fold change and a 95% confidence interval for the fold change. (See the next question for further discussion of fold change.)

6. Suppose $X_1,...,X_m$ are i.i.d. log normally distributed such that $U_i=log\ X_i$ is normal with mean μ_x and variance σ^2. Suppose $Y_1,...,Y_n$ are i.i.d. log normally distributed so that $V_i=log\ Y_i$ is normal with mean μ_y and variance σ^2. Further, suppose $X_1,...,X_m$ and $Y_1,...,Y_n$ are independent. Recall or note that if $log\ X$ is normal with mean μ and variance σ^2, then $E(X)=exp\{\mu+ \sigma^2/2\}$.

a) Determine $E(X_1/Y_1)$.

b) Determine $E(X_1)/E(Y_1)$.

c) Provide maximum likelihood estimators for the quantities in parts (a) and (b).