ISQS 5347 Midterm 1 Solutions.

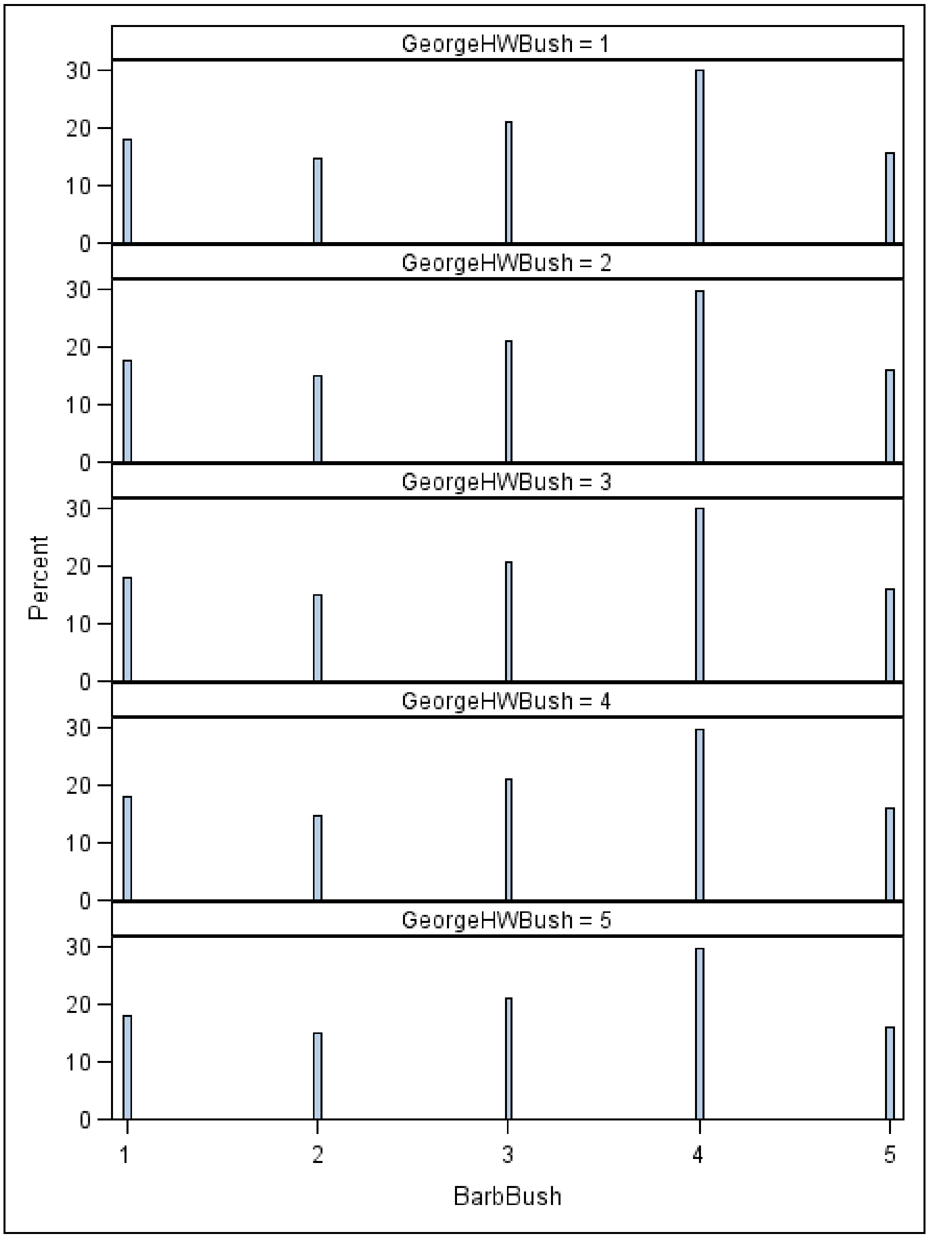
Instructions: closed book and notes; no electronic devices. Points (out of 100) in parentheses.

1. A. (7) Let X and Y be random variables. Give the mathematical definition of independence of X and Y in terms of conditional distributions.

Solution: If $p(y \mid X \in S_1) = p(y \mid X \in S_2)$ for all sets S_1 and S_2, then X and Y are independent.

B. (8) Draw graphs showing the conditional distributions when the variables are independent.

Solution: The graphs will have exactly the same appearance. Same center, same spread, same skewness. They should be exactly the same. For example, in the George and Barbara Bush example, the distributions might look like this when the responses are independent:



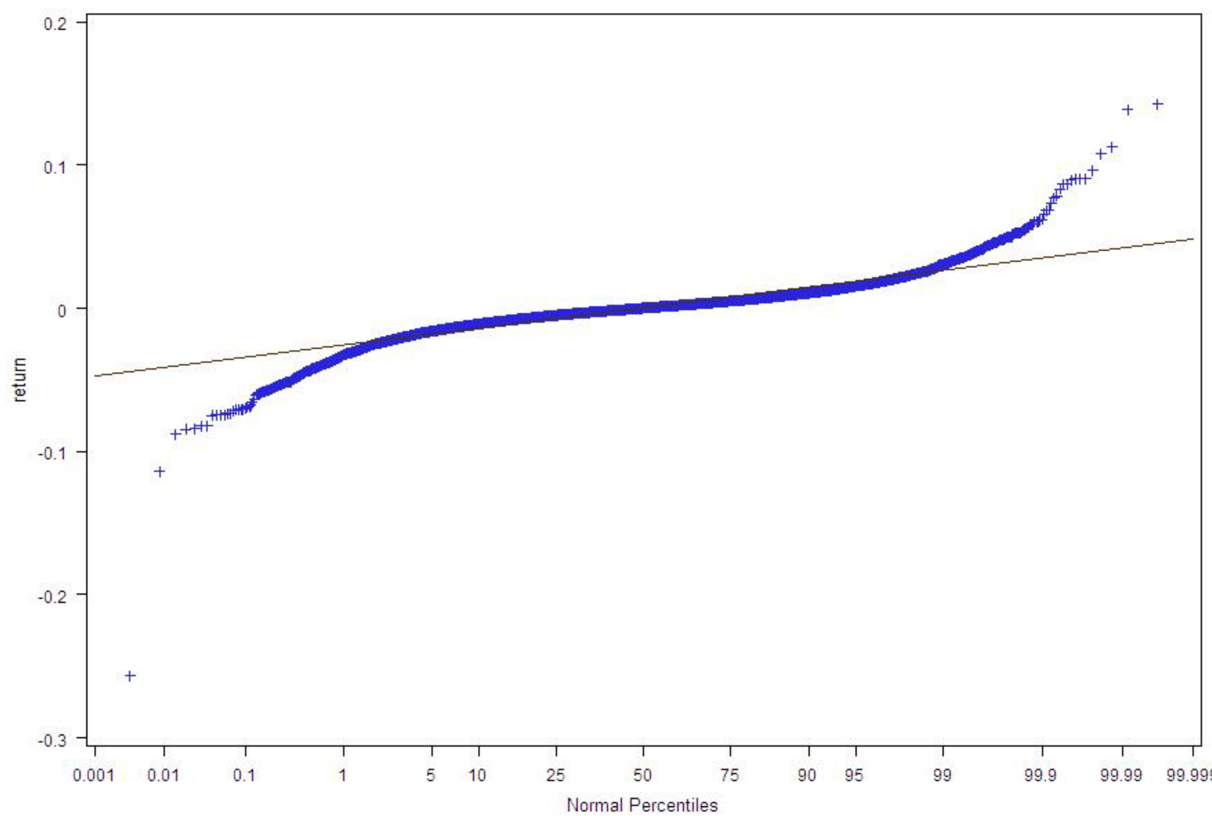
2. (10) What is in the data set work.try? Be specific.

```
data try;
   do y = 1 to 10 by 1;
   output;
   end;
run;
```

Solution: A list of the 10 numbers 1,2,…,10, all in a column labeled “y”.

3. (15) The concern about whether a statistical pattern is "explainable by chance alone" has appeared several times. Using one example taken either from class, or from the book, or from homework, describe what it means for the pattern(s) shown in a statistical graph or graphs to be "explainable by chance alone."

   Solution: Pick the q-q plot for the DJIA returns. Deviations from a straight line in the q-q plot suggest deviation from normality. The question is whether the curvature in the following plot is explainable by chance alone:



   The issue is that, even when the data are produced by a normal distribution, the graph will deviate from a perfect straight line *by randomness alone:* one random sample from a normal distribution, being random, is different from another random sample from the same normal distribution. By simulating data from a normal distribution, we can see how much deviation from the perfect straight line is expected by chance alone. In the case of the DJIA, deviations as extreme as those shown cannot happen by chance alone with that large a sample size: deviations from a straight line are very minor when you simulate data from a normal distribution with that sample size.

4. (15) A person may be a Musician, a Mathematician, neither, or both. Explain briefly the meaning of the equation

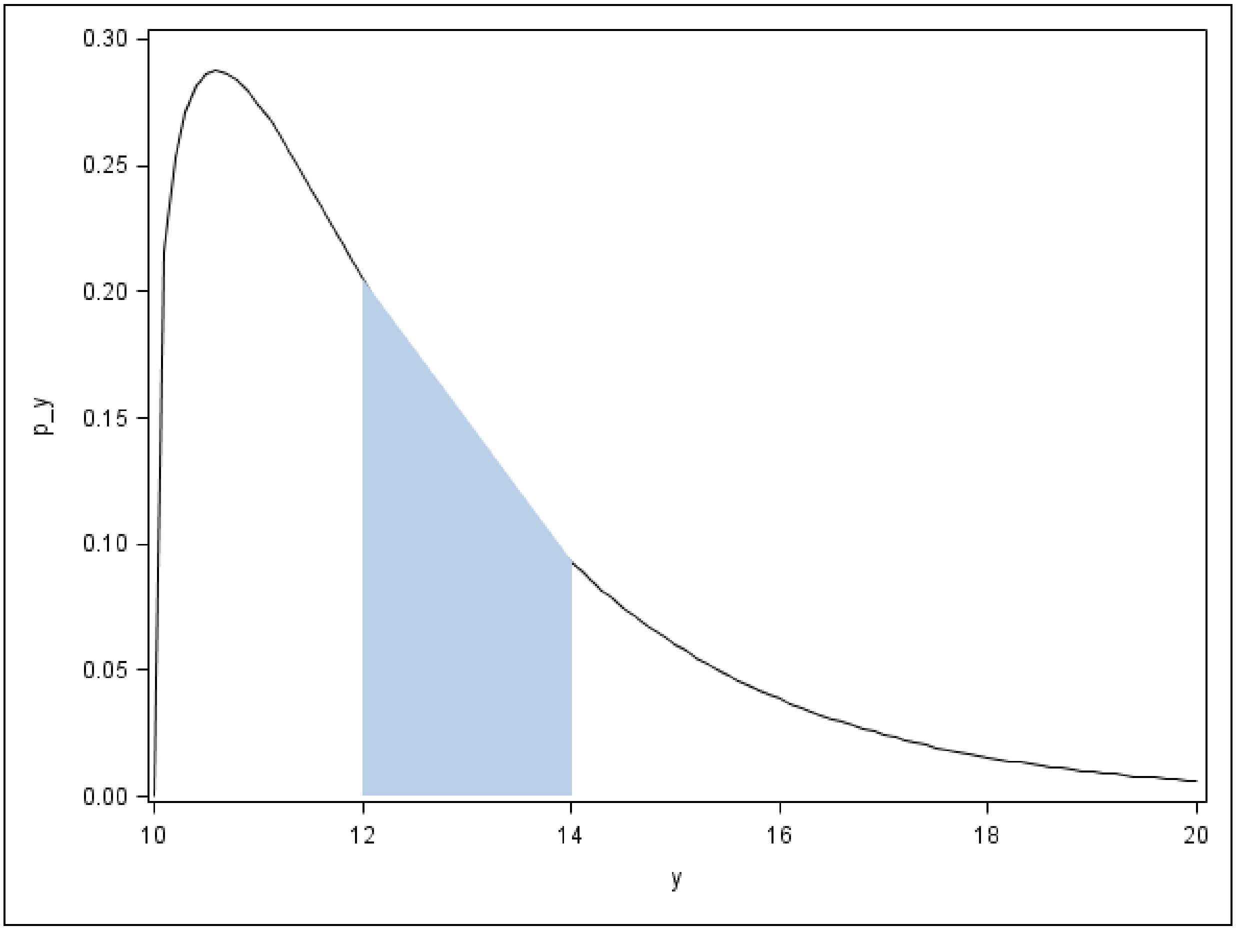
$$\Pr(\text{Musician} \mid \text{Mathematician}) = 0.25$$

to a person who doesn't know much about math or statistics.

Solution: In a room with 100 mathematicians, you can expect that about 25 of them are musicians.

5. Suppose Y has a generic continuous pdf $p(y)$ over the range $10<y<20$.
   A. (5) Draw a graph showing $\Pr(12 < Y < 14)$.

Solution:



B. (5) Express $\Pr(12 < Y < 14)$ using calculus notation.

Solution: $\Pr(12 < Y < 14) = \int_{12}^{14} p(y)dy$.

6. A. (5) What *assumptions* are needed for the Law of Large Numbers?

Solution: That the observations $Y_1, Y_2, \ldots$ are produced as *independent* and *identically distributed* observations from a pdf $p(y)$ having *finite mean* μ.

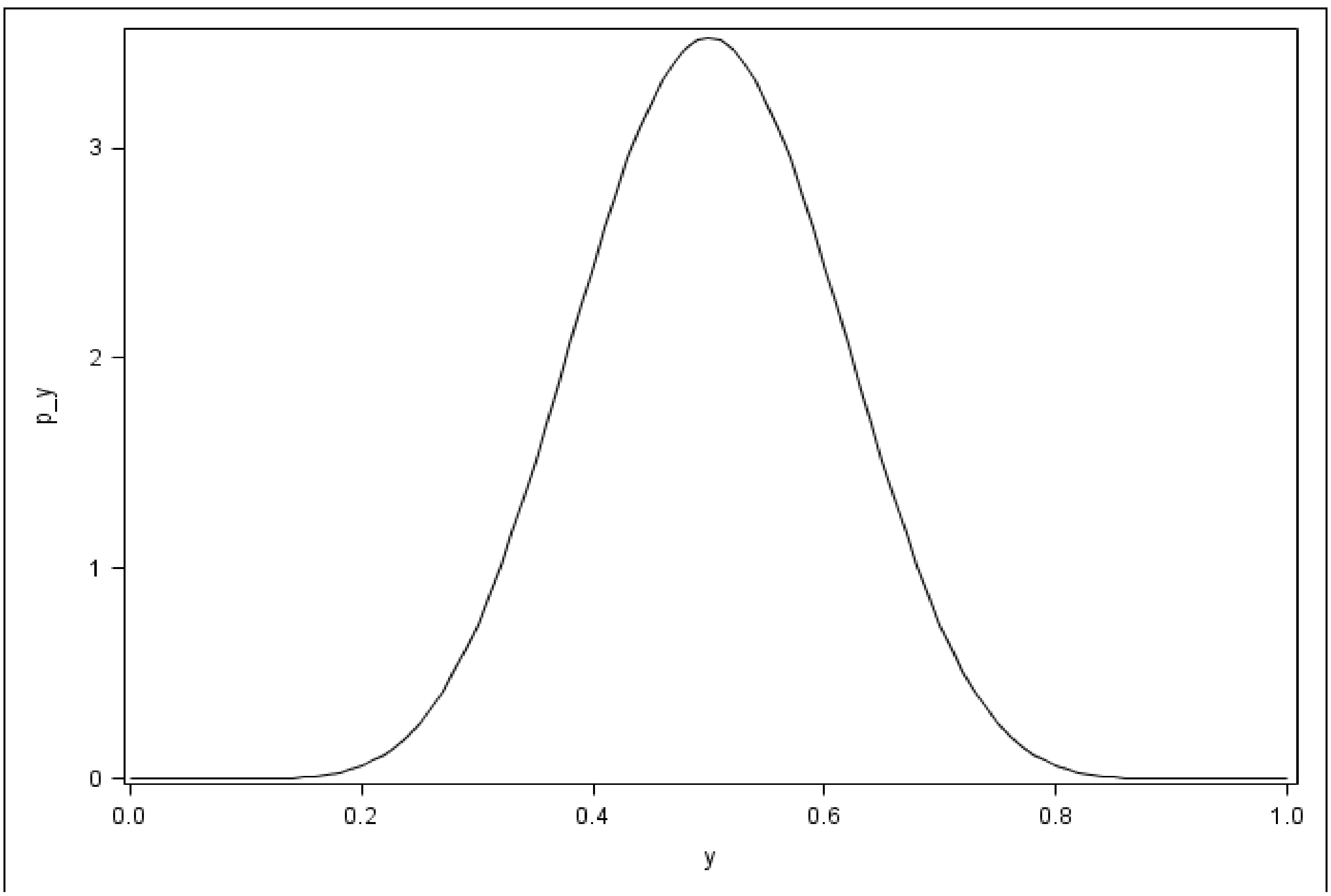
B. (5) What can happen when the assumptions are not true?

Solution: The sample average $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ might not converge to μ, the mean of the pdf that produced the data, as n gets larger. Or it might never converge at all, even for very large sample size n.

7. (10) Explain the meaning of the equation $\hat{\mu} = \bar{y}$ in a sentence or two.

Solution: The average of the data, $\bar{y}$, is an estimate of the mean μ of the pdf that produced the data.

8. (15) Model produces data. For the pdf $p(y)$ graphed below, produce, *from your mind*, an i.i.d. sample of n=10 observations $y_1, y_2, \ldots, y_{10}$.



Solution: 0.239, 0.551, 0.557, 0.403, 0.356, 0.778, 0.502, 0.433, 0.411, 0.801.