Probability Review

Probabilities apply to processes with unpredictable outcomes ("random experiments")

The probability of a particular result, or outcome, measures the tendency of the process to produce that result.

Probability model
(A mathematical representation of the Process)

   (1) Random variable X (the result, or outcome)
   (2) Sample Space S (Set of all possible outcomes)
   (3) Probability distribution over S

When a probability model like this represents the experiment, an "event" is represented by a set of the points in S.

The probability of an event (set) A, P(A), is the sum of probabilities of all the points that are in A.

Bayes' Theorem

Example:

Suppose we select one student at random from those registered for this class and determine the number of teeth in that person's head. The result of this process will be a number -- call it X. This is our <u>Random</u> <u>Variable</u>.

The sample space is S={0,1,2,...,30,31,32}   ?

Let P(X=0) be the proportion of students with no teeth
       P(X=1) be the proportion of students with one tooth
       P(X=2) etc...

The event "selected student has at least 26 teeth" is represented by the set

       A = {26, 27, 28, 29, 30, 31, 32}

And the probability of this event is

       P(A) = P(26) + P(27) + ... + P(32)   (Why?)

The event "Selected student has an even number of teeth" is represented by the set B = {0,2,4, ..., 30, 32}. Its probability is:

$$P(B) = P(0) + P(2) + ... + P(32)$$

The event "Selected student has at least 26 teeth <u>or</u> has an even number of teeth" is represented by the set

A or B = {26, 27, 28, 29, 30, 31, 32 , 0, 2, ..., 22, 24}

Its probability is

$$P(A) + P(B) - P(AB) = [P(26) + P(27) + ... + P(32)]$$

$$+ [P(0) + P(2) + ... + P(32)]$$

$$- [P(26) + P(28) + P(30) + P(32)]$$

Properties of Probabilities

For event A in sample space S.

- $0 \leq P(A) \leq 1$

- $P(S) = 1$

- $P(A) = 1 - P(A^c)$

- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

If $AB = \varnothing$ then $P(AB) = 0$.
(The intersection of A and B is the empty set –
     A and B are "mutually exclusive" or "disjoint"),

- If $P(A \text{ and } B) = P(A)P(B)$, then A and B are "independent events"

- The conditional probability of A, given B, is defined as $P(A|B) = P(A \text{ and } B)/P(B)$

- $P(A \text{ and } B) = P(A|B)P(B)$
    $= P(B|A)P(A)$

Bayes' Theorem

- If A and B are independent, then

$$P(A|B) = P(A \text{ and } B)/P(B) = P(A)P(B)/P(B) = P(A)$$

This says that the probability of A and the probability of A given B, are the same. The probability of A is unaffected by B.

On the other hand, if $P(A|B) \neq P(A)$ then A and B are <u>not</u> independent events. The occurrence of B changes the probability that A will occur.

- If A and B are <u>disjoint</u> (mutually exclusive) events, that both have a positive probability of occurring, then they are <u>not</u> <u>independent</u>.

To show this, simply note that $AB = \varnothing$, so

$$P(A \text{ and } B) = P(\varnothing) = 0 \neq P(A)P(B)$$

Alternatively (in terms of conditional probabilities) if A and B are disjoint and B occurs, then A cannot, so that

$$P(A|B) = 0 \neq P(A)$$

Therefore A and B are not independent.

Problem:
(source: Parade Magazine 7/27/97 – Ask Marilyn)

"A woman and a man (unrelated) each have two children. At least one of the woman's children is a boy, and the man's older child is a boy. Do the chances that the woman has two boys equal the chances that the man has two boys?"

Marilyn says: "The chances that the woman has two boys are 1 in 3 and the chances that the man has two boys are 1 in 2."

Many people write in to tell Marilyn that she is horribly wrong and a disgrace to the human race. Obviously the chances are equal. Who is correct?

Bayes' Theorem

To answer the question we need to set up some notation. For any family, the probability of a boy on one birth is ½, and births are independent.

Our sample space is   S={(0,0), (0,1), (1,0), (1,1)}

Let our event be
A={older birth is a boy}   ={(0,1),(1,1)}
C={Exactly one boy in two births}   ={(1,0),(0,1)}
D={Exactly two boys in two births}   ={(1,1)}

So that P(A)=1/2, P(C)=1/2, and P(D)=1/4.

- We are given that the man's older child is a boy. What is the probability of two boys, given the older is a boy?

$$P(D|A) = ?$$

We know that

$$P(D|A) = P(D \text{ and } A)/P(A)$$

$$=P(D)/P(A)=(1/4)/(1/2)=1/2$$

The probability that the man has two boys is ½.

- We are also given that the woman has at least one boy. What is the probability of two boys, given at least one boy?

The event "At least one boy" is the set

$\{(1,0),(0,1),(1,1)\}$ = C or A

So the question is to find P(D|{C or A})

$$P(D|C \text{ or } A) = P(D \text{ and } \{C \text{ or } A\})/P(C \text{ or } A)$$

$$= P(D)/P(C \text{ or } A) = (1/4)/(3/4) = 1/3$$

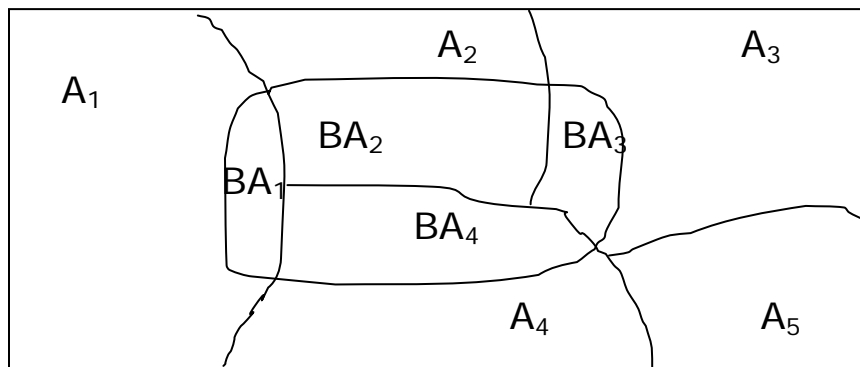The probability that the woman has two boys is only 1/3. (Marilyn is correct.)

Bayes' Theorem

First we need to be familiar with the Law of Total Probability?

Suppose the sample space is divided into any number of disjoint sets, say A_1, A_2, ..., A_n, so that $A_i \cap A_J = \emptyset$ and $A_1 \cup A_2 \cup ... \cup A_n = S$

In this case we can write

$P(B) = P(BA_1 \cup BA_2 \cup ... \cup BA_n)$

$\quad\quad = P(B \text{ and } A_1) + P(B \text{ and } A_2) + ... + P(B \text{ and } A_n)$

$\quad\quad = P(B|A_1)P(A_1) + P(B|A_2)P(A_1) + ... + P(B|A_n) P(A_n)$

Or more generally:

(LOTP) $\quad\quad P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$

Bayes' Theorem

Example:

Suppose that we only have two disjoint sets A_1, A_2 so that $A_1 \cup A_2 = S$

Then by the Law of Total Probability we have

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$
$$= P(B \text{ and } A_1) + P(B \text{ and } A_2)$$
$$= P(B)$$

(because S is only divided by two sets)

By itself, the Law of Total Probability is not very interesting. However, in conjunction with the law of conditional probability, we have:

$P(A_1|B)$ $= P(A_1 \text{ and } B) / P(B)$

$= P(B|A_1)P(A_1)/P(B)$

$= P(B|A_1)P(A_1) / (P(B|A_1)P(A_1) + P(B|A_2)P(A_2))$

or in general

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^{n} P(B | A_i)P(A_i)}$$

Formally stated **Bayes' Theorem** says

For mutually disjoint sets, A_1, A_2, ..., A_n that comprise the total sample space ($A_1 \cup A_2 \cup ... \cup A_n = S$)

We have:

$$P(A_k \mid B) = \frac{P(B \mid A_k)P(A_k)}{\displaystyle\sum_{i=1}^{n} P(B \mid A_i)P(A_i)}$$

In its simplest form, for two events A and B, we have

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}$$

Example

Suppose that 5% of men and 0.25% of women are color-blind in a population that consists of equal numbers of men and women. A person is chosen at random and that person proves to be color-blind. What is the probability that the person is male?

Solution:
Let A be the event "selected person is male" and let B be the event "selected person is color-blind."
We want the conditional probability of A, given B.

We are given that $\quad P(B|A) = 0.05$
$\qquad\qquad\qquad\qquad P(B|A^c) = 0.0025$ and
$\qquad\qquad\qquad\qquad P(A) = 0.5$.

Thus

$$P(A|B) = P(A \text{ and } B)/P(B)$$

$$= P(B|A)P(A) / \{P(B|A)P(A) + P(B|A^c)P(A^c)\}$$

$$= (0.05)(0.5) / \{(0.05)(0.5) + (0.0025)(0.5)\}$$

$$= (0.05) / (0.0525)$$

$$= 0.95238$$

Here B represents strong evidence supporting A vs A^c, i.e., male vs female. Before B is observed, the probability ratio is $P(A)/P(A^c) = \tfrac{1}{2} / \tfrac{1}{2} = 1$. Observing B increases it to $P(A|B)/P(A^c|B) = 0.95238/(1-0.95238) = 20$.

Important Aside:

Bayes' theorem shows how to "turn the conditional probabilities around".

It is a simple fact, which has been made controversial because of attempts to apply probability theory to problems where A represents a scientific hypothesis (call it H_1) and B represent a body of observed data (call it D for data).

It say in those problems, if you know the probability of observing D when then hypothesis H_1 is true $P(D|H_1)$ and the probability when it isn't, $P(D|\text{not } H_1)$, then if you also can assign a probability, $P(H_1)$, to the truth of H_1 before D is observed, you can calculate the probability $P(H_1|D)$ that hypothesis H_1 is true, given the data D.

The controversial part concerns when and how one might determine the "prior" probability that hypothesis H_1 is true, $P(H_1)$. Some have argued that when you have no knowledge of whether H_1 is true of not you should use $P(H_1) = 1/2$. This has been strongly criticized.

How? Because:
$$P(H_1|D) = P(D|H_1)P(H_1)/\{P(D)\}$$
$$= P(D|H_1)P(H_1)/\{P(D|H_1)P(H_1) + P(D|\text{not } H_1)\, P(\text{not } H_1)\}$$

Diagnostic tests

Sensitivity, Specificity, Positive and Negative Predictive Value are all related via Bayes' Theorem. We use Bayes' theorem every time we calculate these values from a 2x2 table, even though it does not feel like it.

Example:
The probability (prevalence) of disease XYZ in a population of interest is 10%. A test is developed to detect the disease in its early stages. When the test was applied to the whole population 170 tested positive. Of those who tested positive, only 80 were confirmed to have the disease XYZ.

	XYZ	Not XYZ	
Test +	80	90	170
Test -	20	810	830
	100	900	1000

From the table is it easy to see that:

1) P(XYZ)=10%

2) P(Test+| XYZ) = (80/1000)/(100/1000)=80/100 = 80%
3) P(Test- | not XYZ) = (810/1000)/(900/1000) = 90%

4) P(XYZ | Test +) = (80/1000)/(170/1000)=80/170= 47%
5) P(not XYZ | Test -) = (810/1000)/(830/1000) = 97.6%

Bayes' Theorem

These quantities have special names:

1) Prevalence
2) Sensitivity = P(Test+ | Disease)
3) Specificity = P(Test- | No Disease)
4) Positive predictive value = P(Disease | Test+)
5) Negative predictive value = P(No Disease | Test-)

Bayes' theorem is used implicitly in the 2x2 table:

Let D+ represent the event "having disease XYZ"
and T+ represent the event "testing positive for XYZ"

Bayes' theorem tells us that

$$P(D+|T+) = \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+) + P(T+|D-)P(D-)}$$

or

$$PPV = \frac{sens \times prev}{sens \times prev + (1 - spec) \times (1 - prev)}$$

And in our example

PPV = (0.8)(0.1)/((0.8)(0.1)+(0.1)(0.9)) = 47%

Thus, to calculate the PPV, we would need the sensitivity, specificity, and prevalence.

Or

To calculate sensitivity, we would need the PPV, NPV and prevalence.

Notice that both calculations depend on the prevalence!

In an objective situation such as a diagnostic test, the prevalence can always, in theory, be specified because we can learn about the prevalence.

Example:

An insurance company has three types of customers -- high risk, medium risk, and low risk. Twenty percent of its customers are high risk, 30% are medium risk, and 50% are low risk. Also, the probability that a high risk customer has at least one accident in the current year is 0.25, while the probability for medium risk is 0.16, and for low risk it is only 0.10. If a randomly selected customer has at least one accident during the year, what is the probability that he is in the high risk group?

Solution:
Let A_1 be "high risk group"
      A_2 be "medium risk", and
      A_3 be "low risk"
      B   be the event "has at least one accident"

We are asked to find $P(A_1|B)$, given that

$$P(A_1) = 0.20, \quad P(A_2) = 0.30, \quad P(A_3) = 0.50, \text{ and}$$

$$P(B|A_1) = 0.25 \quad P(B|A_2) = 0.16, \quad P(B|A_3) = 0.10$$

Bayes' Rule gives the solution:

$$
\begin{aligned}
P(A_1|B) \quad &= \quad P(B|A_1)P(A_1) \, / \, \{P(B|A_1)P(A_1) + \\
&\qquad\qquad\qquad\qquad\qquad P(B|A_2)P(A_2) + \\
&\qquad\qquad\qquad\qquad\qquad P(B|A_3)P(A_3)\} \\[1em]
&= \quad (0.25)(0.20) \, / \, \{(0.25)(0.20) + \\
&\qquad\qquad\qquad\qquad\qquad (0.16)(0.30) + \\
&\qquad\qquad\qquad\qquad\qquad (0.10)(0.50)\} \\[1em]
&= \quad (0.05) \, / \, (0.05 + 0.048 + 0.05) \\[1em]
&= \quad 0.338
\end{aligned}
$$

Expanded Diagnostic Test

MRI is often used to asses whether a tumor might be cancerous. After looking at a scan, the tumor is graded on the following scale:

1 = definitely not cancerous
2 = probably not cancerous
3 = inconclusive
4 = probably cancerous
5 = definitely cancerous

The following table presents some Fake data for an experiment design to assess the accuracy of MRI for grading tumors.

	MRI Tumor Assessment					Total
	1	**2**	**3**	**4**	**5**	
Malignant	7	13	22	45	91	178
Benign	78	56	60	5	2	201
Total	85	69	82	50	93	379

- How do we assess the accuracy of this test?

MRI Tumor Assessment						Total
	1	**2**	**3**	**4**	**5**	
Malignant	7	13	22	45	91	178
Benign	78	56	60	5	2	201
Total	85	69	82	50	93	379

If this were a 2x2 table we could calculate the sensitivity and specificity. In fact we can do something analogous here by collapsing the above table into a series of 2x2 table.

If the patient received a 4 or 5 MRI score, we'll say that they are eligible for surgery. The properties of such an assessment can be summarized in the following table:

MRI Tumor Assessment		Total
	1-3	**4-5**	
Malignant	42	136	178
Benign	194	7	201
Total	236	143	379

Sensitivity = P(4 or 5|M)   = 136/178 = 0.764
Specificity = P(1,2, or 3|B) = 194/201 = 0.9652

Bayes' Theorem

But the cutoff was somewhat arbitrary, why not 1-2 versus 3-5 just to be sure?

- We can calculate the properties for all score combinations!

	MRI Tumor Assessment					Total
	1	**2**	**3**	**4**	**5**	
Malignant	7	13	22	45	91	178
Benign	78	56	60	5	2	201
Total	85	69	82	50	93	379
Sensitivity*	1	0.96	0.89	0.76	0.51	
Specificity*	0	0.39	0.67	0.97	0.99	

*for having an MRI score that high or higher.

Example: Properties for basing surgery on a 2-5 MRI score.

Sensitivity $\quad$ P(2-5|M) = (13+22+45+91)/178 = 0.96
Specificity $\quad$ P(1|M) = 78/201 = 0.39

But instead of looking at a bunch of numbers, we can graph Sensitivity versus 1-Specificity to get a **Receiver Operator Characteristic (ROC curve)** for MRI assessment.

	MRI Tumor Assessment					Total
	1	**2**	**3**	**4**	**5**	
Malignant	7	13	22	45	91	178
Benign	78	56	60	5	2	201
Total	85	69	82	50	93	379
Sensitivity*	1	0.96	0.89	0.76	0.51	
Specificity*	0	0.39	0.67	0.97	0.99	

This data was entered as two columns and 379 rows. The first column is the MRI call for each individual (integer in [1,5]) and the second is the tumor status (1=malignant, 0=benign). Here are the Stata commands to do this:

```
. use rocex
. list in 1/10


     +---------------+
     |  mri   cancer |
     |---------------|
  1. |    1        1 |
  2. |    1        1 |
  3. |    1        1 |
  4. |    1        1 |
  5. |    1        1 |
     |---------------|
  6. |    1        1 |
  7. |    1        1 |
  8. |    2        1 |
  9. |    2        1 |
 10. |    2        1 |
     +---------------+
```
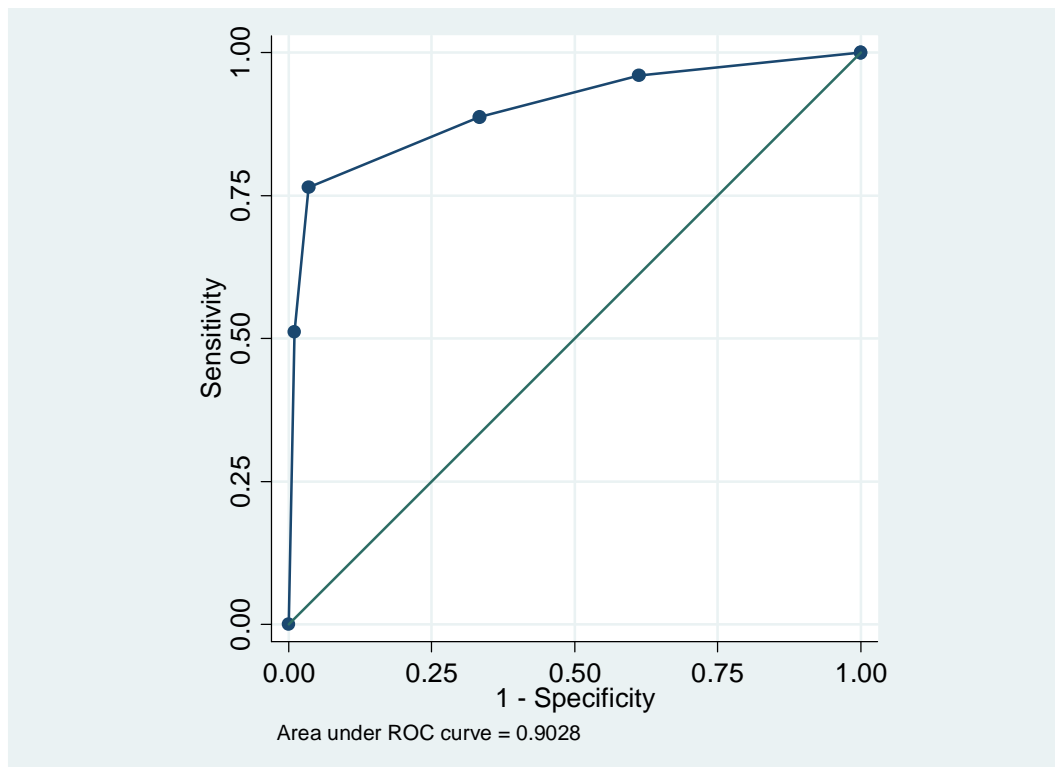
Bayes' Theorem

```
. roctab cancer mri, table summary detail graph aspectratio(1)
```

			mri				
cancer	1	2	3	4	5		Total
0	78	56	60	5	2		201
1	7	13	22	45	91		178
Total	85	69	82	50	93		379

Detailed report of Sensitivity and Specificity

Cutpoint	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 1)	100.00%	0.00%	46.97%	1.0000	
(>= 2)	96.07%	38.81%	65.70%	1.5699	0.1013
(>= 3)	88.76%	66.67%	77.04%	2.6629	0.1685
(>= 4)	76.40%	96.52%	87.07%	21.9390	0.2445
(>= 5)	51.12%	99.00%	76.52%	51.3792	0.4937
(> 5)	0.00%	100.00%	53.03%		1.0000

Obs	ROC Area	Std. Err.	-Asymptotic Normal-- [95% Conf. Interval]	
379	0.9028	0.0161	0.87133	0.93433

Area under ROC curve = 0.9028

An ROC curve visually displays the trade off between sensitivity and specificity. This can be quite useful for determining a cutoff for the diagnosis variable.
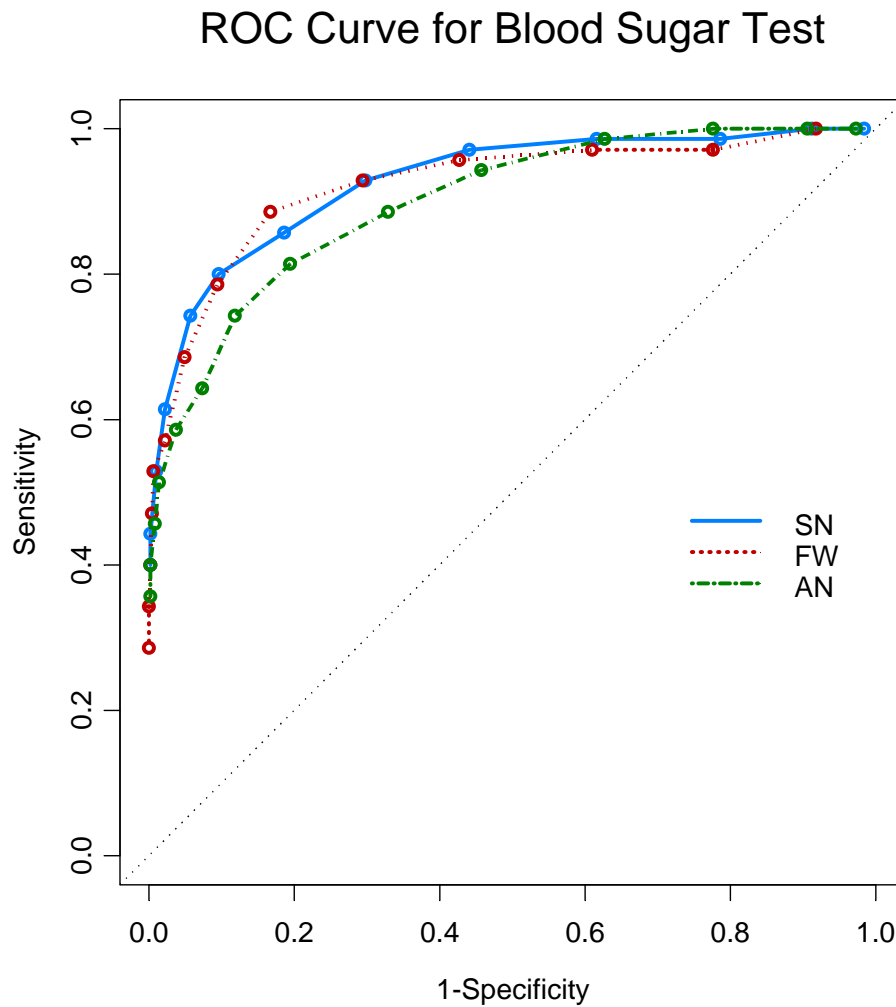
Example:
(Fisher and Van Belle p236)
Blood samples collected after a test meal, three different blood test gave the following data:

Blood sugar (mg/100ml)	Type of Test					
	Somogyi-Nelson		**Folin-Wu**		**Anthrone**	
	Sens	**Spec**	**Sens**	**Spec**	**Sens**	**Spec**
70	na	na	100	8.2	100	2.7
80	na	1.6	97.1	22.4	100	9.4
90	100	8.8	97.1	39.0	100	22.4
100	98.6	21.4	95.7	57.3	98.6	37.3
110	98.6	38.4	92.9	70.6	94.3	54.3
120	97.1	55.9	88.6	83.3	88.6	67.1
130	92.9	70.2	78.6	90.6	81.4	80.6
140	85.7	81.4	68.6	95.1	74.3	88.2
150	80.0	90.4	57.1	97.8	64.3	92.7
160	74.3	94.3	52.9	99.4	58.6	96.3
170	61.4	97.8	47.1	99.6	51.4	98.6
180	52.9	99.0	40.0	99.8	45.7	99.2
190	44.3	99.8	34.3	100	40.0	99.8
200	40.0	99.8	28.6	100	35.7	99.8

The corresponding ROC curves are

ROC Curve for Blood Sugar Test

- Which test looks most promising any why?

- How shall we compare these curves?

  1) Best operational point?
  2) Smoothest curve?
  3) Area under the curve?

Relative Risk and Odds Ratio

- The <u>relative</u> <u>risk</u> is a ratio of two probabilities

<u>Example:</u>
    P(disease | exposed) / P(disease | unexposed)

- The <u>odds</u> in favor of an event E are defined as:

    Odds in favor of E = P(E) / (1-P(E))

If P(E) =1/4 the odds are 1 to 3.

- An odds ratio is a ratio of two odds.

<u>Example:</u>

$$OR = \frac{P(D \mid E)/[1 - P(D \mid E)]}{P(D \mid unE)/[1 - P(D \mid unE)]} = \frac{P(D \mid E)P(notD \mid unE)}{P(D \mid unE)P(notD \mid E)}$$

is the odds ratio of disease for exposed versus unexposed.

When does the OR approximate the RR?

$$OR = \frac{P(D\,|\,E)\,/\,[1 - P(D\,|\,E)]}{P(D\,|\,unE)\,/\,[1 - P(D\,|\,unE)]} \approx \frac{P(D\,|\,E)}{P(D\,|\,unE)}$$

When both P(D|E) and P(D|unE) are small.

- How do we interpret an RR? An OR?

Aside: Independent Trials, Independent Events, and Independent Random Variables

If we repeat the trial or experiment, and if the outcome of each trial is not influenced by the outcomes of any of the others, then they are <u>independent</u> <u>trials</u>.

If the trials are independent, and if A_1 is an event that depends only on the result of the one trial, and if A_2 is an event that depends only on the result of another, then A_1 and A_2 are <u>independent</u> <u>events</u>.

If X_1, X_2, … represent the result of the different trials, and the trials are independent, then X_1, X_2, … are <u>independent</u> <u>random</u> <u>variables</u>.

Independent random variables can arise in other ways as well, but the prime example of independent random variables is this one — variables that represent the results of independent trials.