Receiver Operating Characteristic (ROC) Curve Analysis

A useful tool to evaluate
the effectiveness of beach posting policies

Kelly Coughlin
Massachusetts Water Resources Authority
Boston, MA

Receiver Operating Characteristic (ROC) Curves

- Developed in 1940's to make sense of radio signals, used by radar receivers to analyze radar images during World War II
- Beginning in the 1970's, recognized as useful for interpreting medical test results
- Simple to use

Overview

1. Description of ROC analysis
2. Walk-through constructing an ROC curve
3. Implications for monitoring and management

Beach Management Concerns

- Need to regulate beach usage on a daily basis
- Measuring bacteria is informative, but a day late; results are not available in real-time
- Bacteria counts fluctuate, a measurement made the previous day may not accurately reflect current conditions

A perfect test…or not

Actual Water Quality

Advisory Status	Fails to meet swimming standards	Meets Swimming Standards
Swimming Advisory	True Positive	
No Swimming Advisory		True Negative

Effectiveness of previous day's *Enterococcus* to trigger swimming advisories

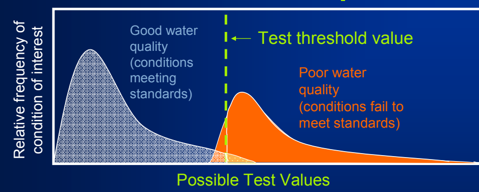
Beach	Average # of days in a season	True Positives (actual alarm)	False negatives (missed alarm)	False positives (false alarm)
City Point Beach, South Boston	70 days	0 days	2 days	2 days
Constitution Beach, East Boston	70 days	1-2 days	2 days	2 days
Wollaston Beach, Quincy	70 days	6 days	13 days	13 days

Results from 2000 – 2005 swimming seasons. Previous day's *Enterococcus* count of 104 cfu/100 mL triggers a swimming advisory.

Comparing the effectiveness of other swimming advisory triggers

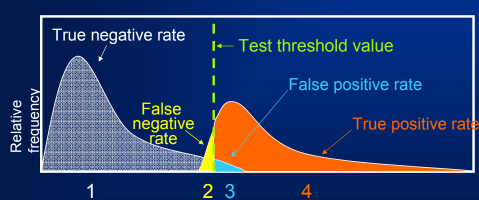
- Set as a given that water quality conditions are measured by EPA swimming standards
- We selected rainfall as an alternative test since runoff is a known cause of elevated bacteria counts at Harbor beaches
- How does a rainfall threshold compare to previous day's *Enterococcus* in effectiveness?

Testing to identify two conditions



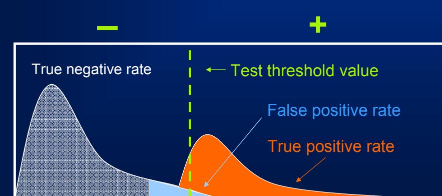
- Two actual conditions exist that the test attempts to identify
  - GOOD WATER QUALITY
  - POOR WATER QUALITY
- A test uses a single value to identify the condition to which a randomly selected sample belongs
- Test variables can include previous day's bacteria, rainfall

The position of a given test threshold value creates four regions



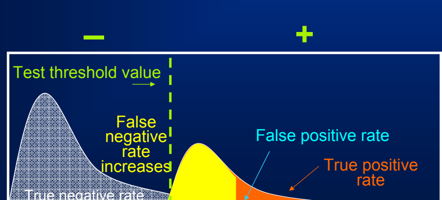
Selecting a test threshold value

Changing the test threshold can increase the true positive rate at the expense of increasing the false positive rate:
more hits, but more false alarms



Selecting a test threshold

Likewise, changing the test threshold in the other direction can decrease the false positive rate at the expense of decreasing the true positive rate:
fewer false alarms, but also fewer hits



ROC analysis allows you to quantify the management tradeoff of a test: beach access vs. protection of public health

Statistical Perspective	Management Perspective	Public Perspective
False positive	False alarm	Crying wolf
False negative	Missed alarm	Wolf in sheep's clothing
True positive	Actual alarm	Wolf in plain sight
True negative	No alarm	No wolf

To quantify the tradeoffs, measure sensitivity and specificity of a test

Advisory Status		Actual Water Quality: Fails to meet swimming standards	Actual Water Quality: Meets Swimming Standards
	Swimming Advisory	True Positive	False Positive (False Alarm)
	No Swimming Advisory	False Negative (Missed Alarm)	True Negative

- Sensitivity, or the true positive rate (TPR), is the proportion of cases correctly identified as meeting a certain condition
- Specificity, or the true negative rate (TNR), is the proportion of cases correctly identified as *not* meeting a certain condition. This is also equivalent to 1 - false positive rate.

ROC analysis is an ideal technique to quantify the tradeoffs of test sensitivity and specificity

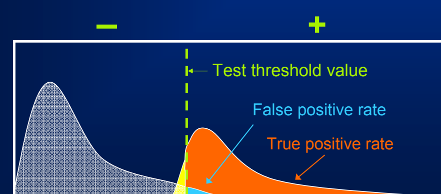
Construction of an ROC Curve

Before beginning an ROC analysis, assemble a dataset that includes:

- Historical beach water quality results (the more the better)
- Indicator variable of interest, for which a threshold is to be developed or evaluated
- Use the water quality results to calculate the true positive and false positive pairs associated with each value of your indicator variable

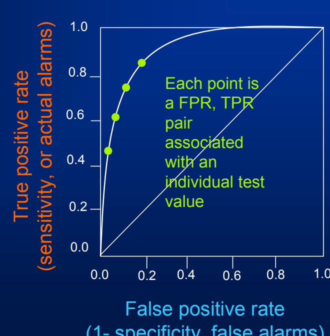
Construction of an ROC Curve

An ROC curve is constructed using a true positive rate and false positive rate pair for each possible threshold value of the test



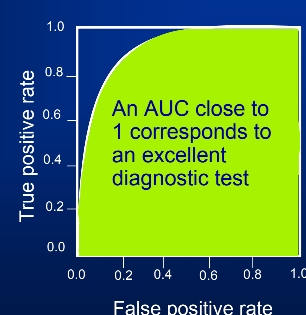
Construction of ROC Curves

Each point on the curve is created by plotting the unique true positive rate (TPR) and false positive rate (FPR) associated with each unique test value



Area Under the ROC Curve (AUC)

- The area under the ROC curve (AUC) is a common metric that can be used to compare different tests (indicator variables)
- An AUC is a measure of test accuracy



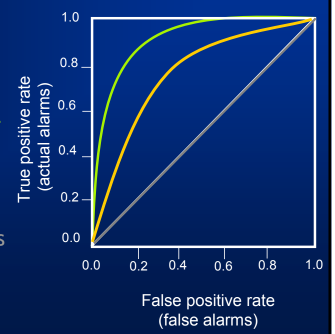
Area Under the ROC Curve (AUC)

- An AUC close to 0.5 corresponds to an poor diagnostic test

Area Under the ROC Curve (AUC)

The larger the AUC, the more accurate the test

- 0.90 - 1.0 Excellent
- 0.80 - 0.90 Good
- 0.70 - 0.80 Fair
- 0.60 - 0.70 Poor
- < 0.60 Worthless



Using ROC curves to evaluate beach water quality variables

Using ROC analysis for beach water quality variables

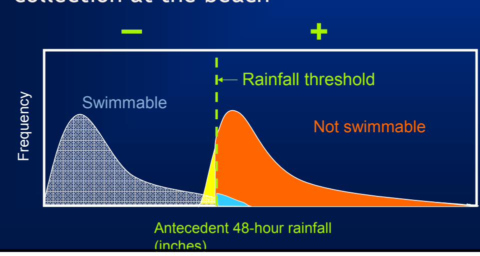
- Two conditions of water quality that the test will identify:
  1) swimmable
  2) not swimmable

Using ROC analysis with beach water quality variables

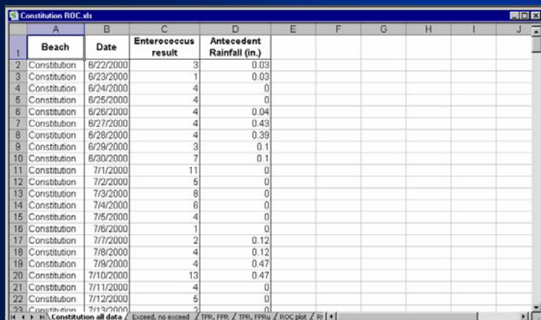
- First select a test variable to identify the water quality condition: previous day's *Enterococcus* or *E. coli*, rainfall, turbidity, etc.

Using ROC analysis with beach water quality variables

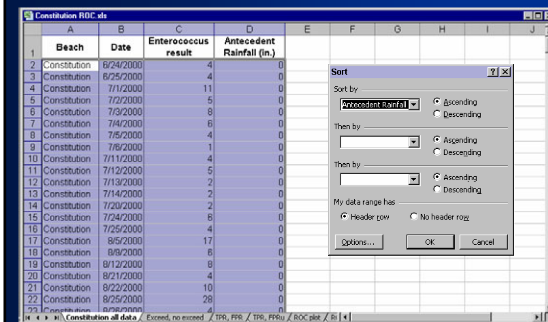
- 48 hour antecedent rainfall, i.e. the rain that has fallen in the 48 hours prior to sample collection at the beach



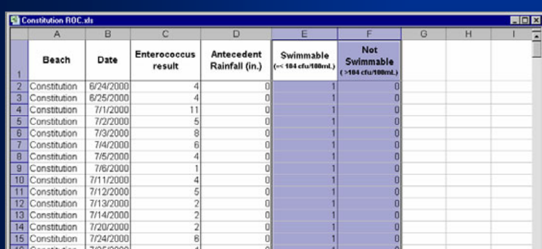
1) Enter beach data into Excel, include indicator variable to test



2) Sort rows by indicator variable, in ascending order

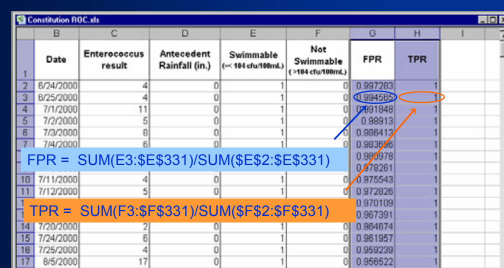


3) Create two new columns that correspond to outcome you are testing for (“swimmability”), with values 0 or 1



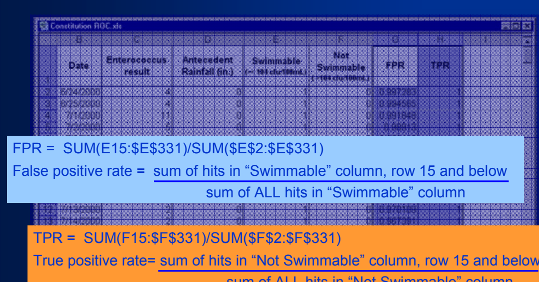
“Swimmable” for this example is defined as compliance with EPA swimming standard, single sample limit for *Enterococcus*, marine waters (104 col/100 mL)

4) Calculate the true positive rate and false positive rate for each unique value of the test

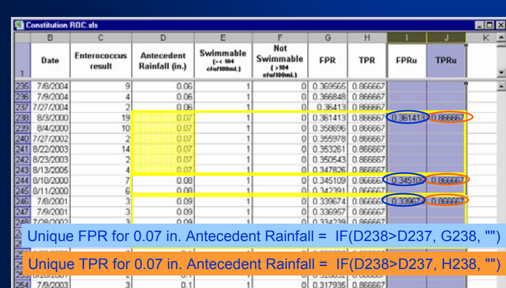


“Swimmable” for this example is defined as compliance with EPA swimming standard, single sample limit for *Enterococcus*, marine waters (104 col/100 mL)

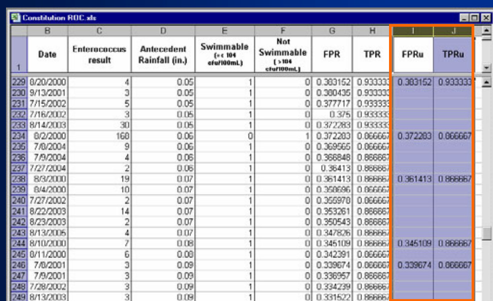
4) Calculate the true positive rate and false positive rate for each row



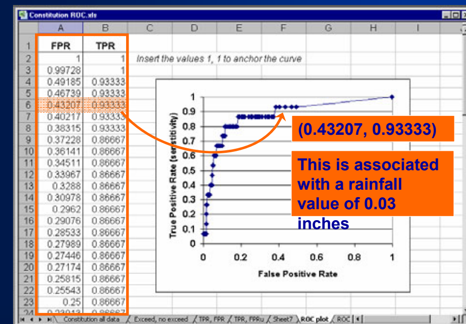
5) Select unique false positive, true positive pairs for each individual test value



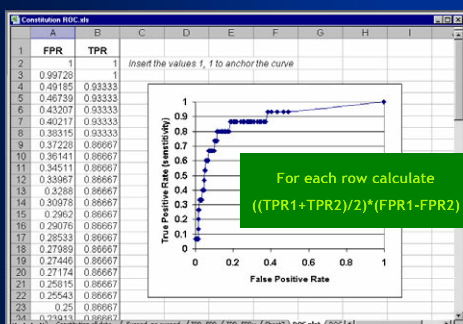
6) Copy columns of unique false positive, true positive pairs for each individual test value and paste into new spreadsheet



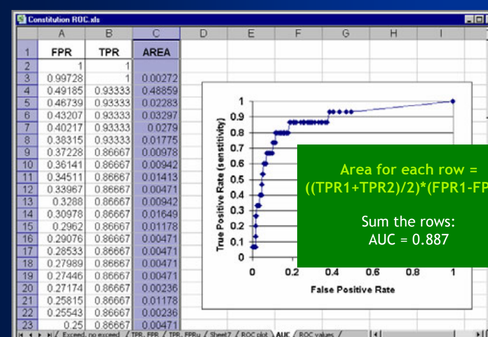
7) Plot false positive rates and true positive rate as an x-y scatter plot



8) Calculate the area under the curve (AUC)



8) Calculate AUC using trapezoid rule



You're done! Now evaluate your test(s)

a. Identify an optimal threshold

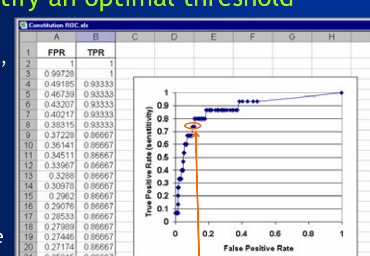
For a given test, the ROC curve allows you to select an optimal threshold (with a true positive rate and false alarm rate of your choice), among all possible *test thresholds*

b. Compare AUC values for different tests

The AUC value allows you to compare the performance of different tests, allowing you to select the best performing test among a group of *tests*.

ROC Benefit #1: For a given test, you can identify an optimal threshold

- Want a high TPR, a low FPR, AND a realistic threshold
- We selected a 75% sensitivity rate to compare the performance of beach water quality variables



75% sensitivity (TPR = 0.75) corresponds roughly to the 75% confidence levels for bacteria indicator limits used by EPA in the WQC for bacteria.

ROC Benefit #1: For a given test, you can identify an optimal threshold

- Sort the original spreadsheet by TPR values, look for TPR of 0.75
- Corresponding rainfall value for TPR of 0.75 is between 0.52 and 0.55 inches of rain

Constitution ROC.xls

1	Date	Enterococcus result	Antecedent Rainfall (in.)	Swimmable (<= 104 cfu/100mL)	Not Swimmable (>104 cfu/100mL)	FPR	TPR
320	7/15/2004	14	0.44	1	0	0.141304	0.8
321	6/24/2001	5	0.46	1	0	0.138587	0.8
322	7/23/2002	2	0.46	1	0	0.13587	0.8
323	7/24/2002	2	0.46	1	0	0.133152	0.8
324	7/9/2000	4	0.47	1	0	0.130435	0.8
325	7/10/2000	13	0.47	1	0	0.127717	0.8
326	8/22/2002	5	0.47	1	0	0.125	0.8
327	8/10/2001	4	0.48	1	0	0.122283	0.8
328	8/11/2001	2	0.48	1	0	0.119565	0.8
329	8/23/2002	1	0.49	1	0	0.116848	0.8
330	8/14/2000	6	0.52	1	0	0.11413	0.8
331	7/31/2000	409	0.55	0	1	0.11413	0.733333
332	8/12/2004	8	0.55	1	0	0.111413	0.733333
333	8/6/2005	5	0.55	1	0	0.108696	0.733333
334	7/15/2000	10	0.56	1	0	0.105978	0.733333
335	8/1/2005	5	0.56	1	0	0.103261	0.733333
336	8/1/2000	25	0.57	1	0	0.100543	0.733333
337	8/2/2005	127	0.57	0	1	0.100543	0.666667
338	7/18/2000	4	0.6	1	0	0.097826	0.666667
339	7/19/2000	5	0.6	1	0	0.095109	0.666667
340	7/19/2003	2	0.6	1	0	0.092391	0.666667

Exceed, no exceed / TPR, FPR / TPR, FPRu / Sheet7 / ROC plot / AUC / ROC values

ROC Benefit #2: Compare different tests at a fixed sensitivity - how reasonable is the threshold?

Variable Tested	Value at 75% sensitivity (TPR = 0.75)	False Positive Rate
Previous Day's *Enterococcus*	10 cfu/ 100 mL	0.67
Antecedent Rainfall, 24 h	0.04 inches	0.31
Antecedent Rainfall, 48 h	0.53 inches	0.11
Antecedent Rainfall, 96 h	0.64 inches	0.27

ROC analysis of Constitution Beach

ROC Benefit #3: You can compare different thresholds for sensitivities (TPRs) and false alarms (FPRs)

	Previous Day's *Enterococcus* Threshold of 104 cfu/ 100 mL	Threshold of 35 cfu/ 100 mL	Threshold of 0 cfu/ 100 mL
Sensitivity (TPR) Ideally, this would be 100%	14 - 29%	29 - 57%	95 - 100%
False Alarm Rate (FPR) Ideally, this would be 0%	1 - 7%	10 - 49%	81 - 95%

Range of results for harbor beaches, including Constitution Beach, 2000 – 2004.

ROC Benefit #4: You can use AUC values to compare the effectiveness of different tests

- Antecedent rainfall, 24 hours
- Antecedent rainfall, 48 hours
- Antecedent rainfall, 96 hours
- Previous day's *Enterococcus*
- Combination of indicators

ROC Benefit #4: You can use AUC values to compare the accuracy of different tests

Variable Tested (Constitution Beach)	AUC (± Standard Error)
Previous Day's *Enterococcus* count	0.64 (± 0.029)
Antecedent Rainfall, 24 h	0.75 (± 0.024)
Antecedent Rainfall, 48 h	0.88 (± 0.039)
Antecedent Rainfall, 96 h	0.80 (± 0.053)

Conclusion: ROC Analysis Advantages for Beach Management

- Evaluates the overall ability of an indicator variable to correctly "classify" beach water quality as suitable or unsuitable for swimming
- Allows direct comparison of different indicator variables by a common metric (AUC)
- Facilitates the identification of a maximum threshold value for the indicator variable that produces a desired true positive rate and false positive rate

Conclusions of Boston Harbor analysis

- Previous day's *Enterococcus* is an inadequate indicator variable for determining beach usage on a daily basis at every beach
- Antecedent rainfall is usually a more accurate indicator variable and is available in real-time
- Daily monitoring for a prolonged period was necessary for ROC analysis

Acknowledgements

- Ann Michelle Morrison of University of New Hampshire
- Jim Shine and Brent Coull of Harvard School of Public Health
- Massachusetts Department of Conservation and Recreation, Division of Urban Parks