Basic Statistics for the Pediatric Hospitalist

Jack McDonnell, MD

Variables

-**continuous**

-*BP, LDL, height*

-categorical

-**ordinal**

-*likert scale*

-**nominal**

-*sex, race*

-ratio

-meaningful "0" value

-division/multiplication and subtraction/addition makes sense

-*response time*

-interval

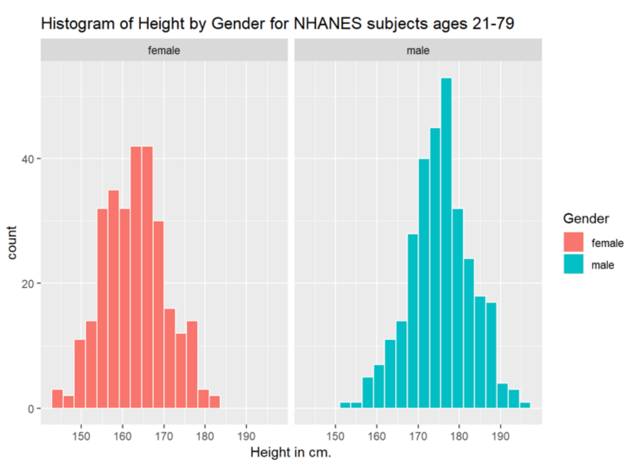
-no meaningful "0" value

-only subtraction/addition makes sense

-*temperature*

Distribution

-normal (parametric)



-mean (average)

-standard deviation (how far are points from the mean?)

$$\text{Standard Deviation} = s = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n-1}}$$

-skew (tail behavior)

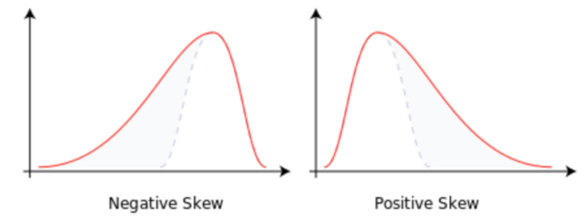
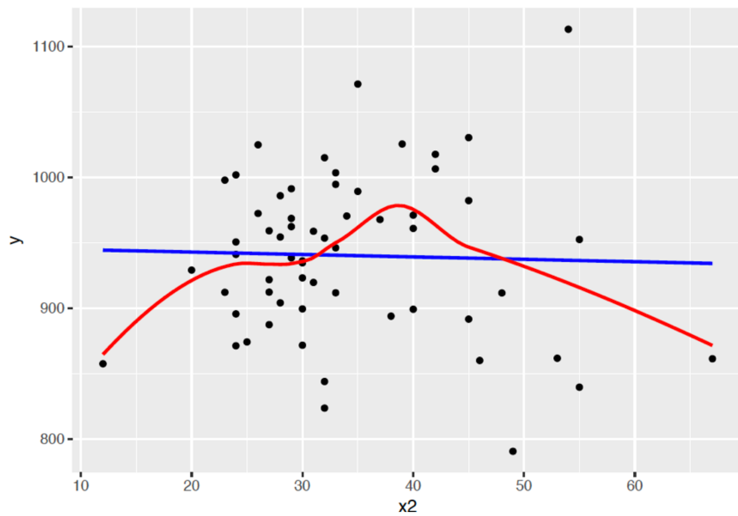


Figure 5.3: Negative (Left) Skew and Positive (Right) Skew

-nonnormal (nonparametric)



Hypothesis Testing

The Two Types of Hypothesis Testing Errors

–	H_A is true	H_0 is true
Test Rejects H_0	Correct Decision	Type I Error (False Positive
Test Retains H_0	Type II Error (False Negative)	Correct Decision

- A Type I error can only be made if the null hypothesis is actually true.
- A Type II error can only be made if the alternative hypothesis is actually true.

-significance level (alpha) is **probability of a type I error**

-probability of avoiding a type II error (1-beta) is **power**

P-values

-statisticians HATE these
-WHY?

1. Very rarely does a situation emerge in which a *p* value can be available in which looking at the associated confidence interval isn't far more helpful for making a comparison of interest.
2. The use of a *p* value requires making at least as many assumptions about the population, sample, individuals and data as does a confidence interval.
3. Most null hypotheses are clearly not exactly true prior to data collection, and so the test summarized by a *p* value is of questionable value most of the time.
4. No one has a truly adequate definition of a *p* value, in terms of both precision and parsimony. Brief, understandable definitions always fail to be technically accurate.
5. Bayesian approaches avoid some of these pitfalls, but come with their own issues.
6. Many smart people agree with me, and use *p* values sparingly.

Common Statistical Tests (DO NOT MEMORIZE!!!)

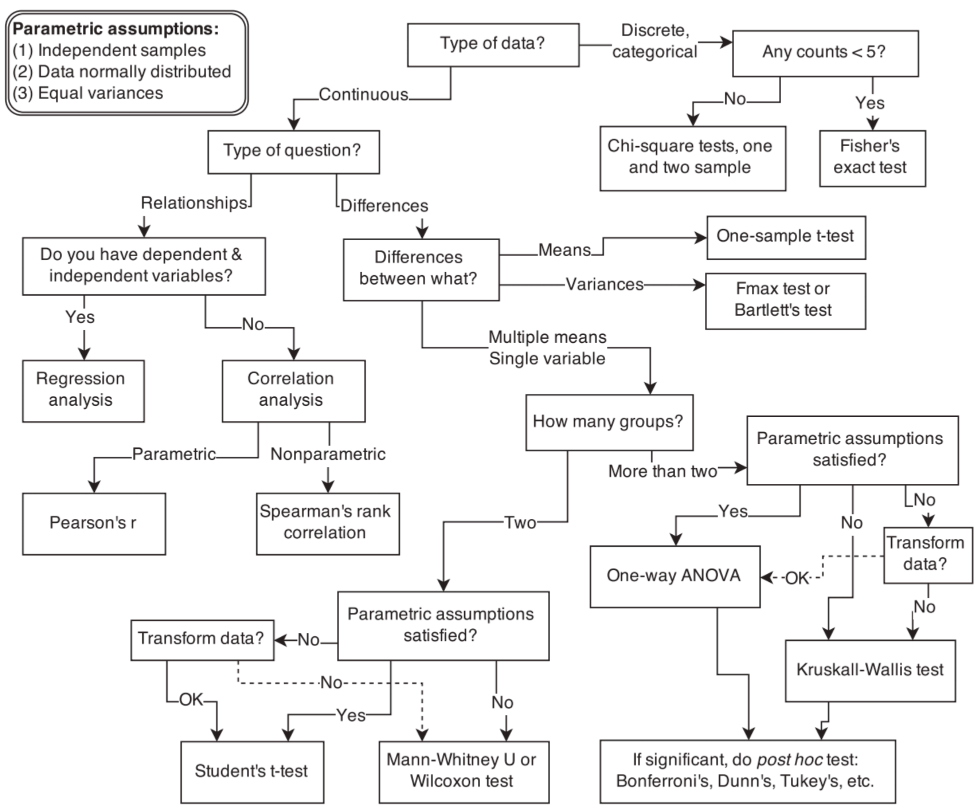


FIGURE 1.1. Example decision tree, or flowchart, for selecting an appropriate statistical procedure. Beginning at the top, the user answers a series of questions about measurement and intent, arriving eventually at the name of a procedure. Many such decision trees are possible.

-Tests worth knowing a bit about

- **-T Test** (parametric) **/Wilcoxon** (nonparametric)
- **-Chi Square /Fisher's Exact** (very small cell frequencies)
- **-ANOVA** (think T-test for >2 groups)

Models – THE BIG THREE

1. Linear regression
   a. Think: *continuous outcome*

```
summary(lm(recov.score ~ dose, data = hydrate))
```

```
Call:
lm(formula = recov.score ~ dose, data = hydrate)

Residuals:
    Min      1Q  Median      3Q     Max
-22.336  -7.276   0.063   8.423  23.903

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    63.90       3.97   16.09   <2e-16 ***
dose            4.88       2.17    2.25    0.031 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.2 on 34 degrees of freedom
Multiple R-squared:  0.129, Adjusted R-squared:  0.104
F-statistic: 5.05 on 1 and 34 DF,  p-value: 0.0313
```

2. Logistic regression
   a. Think: *binary outcome*

```
logit admit gre gpa i.rank

Iteration 0:   log likelihood = -249.98826
Iteration 1:   log likelihood = -229.66446
Iteration 2:   log likelihood = -229.25955
Iteration 3:   log likelihood = -229.25875
Iteration 4:   log likelihood = -229.25875

Logistic regression                               Number of obs   =        400
                                                  LR chi2(5)      =      41.46
                                                  Prob > chi2     =     0.0000
Log likelihood = -229.25875                       Pseudo R2       =     0.0829

------------------------------------------------------------------------------
       admit |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gre |   .0022644    .001094     2.07   0.038     .0001202    .0044086
         gpa |   .8040377   .3318193     2.42   0.015     .1536838    1.454392
             |
        rank |
          2  |  -.6754429   .3164897    -2.13   0.033    -1.295751   -.0551346
          3  |  -1.340204   .3453064    -3.88   0.000    -2.016992   -.6634158
          4  |  -1.551464   .4178316    -3.71   0.000    -2.370399   -.7325287
             |
       _cons |  -3.989979   1.139951    -3.50   0.000    -6.224242   -1.755717
------------------------------------------------------------------------------
```

```
logit , or

Logistic regression                               Number of obs   =        400
                                                  LR chi2(5)      =      41.46
                                                  Prob > chi2     =     0.0000
Log likelihood = -229.25875                       Pseudo R2       =     0.0829

------------------------------------------------------------------------------
       admit | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gre |   1.002267   .0010965     2.07   0.038      1.00012    1.004418
         gpa |   2.234545   .7414652     2.42   0.015     1.166122    4.281877
             |
        rank |
          2  |   .5089309   .1610714    -2.13   0.033     .2736922    .9463578
          3  |   .2617923   .0903986    -3.88   0.000     .1330551    .5150889
          4  |   .2119375   .0885542    -3.71   0.000     .0934435    .4806919
------------------------------------------------------------------------------
```

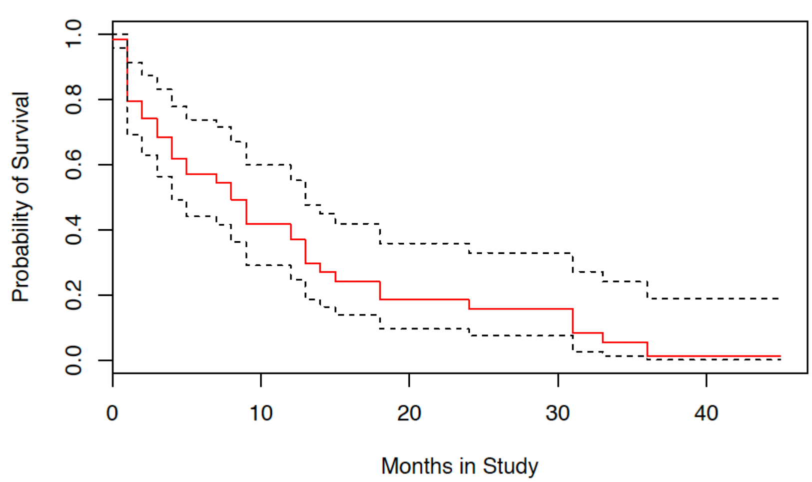
3. Cox proportional hazards
    a. Think: *time to event/censoring*

```
modB <- coxph(Surv(months, alive==0) ~
          age + pblasts + pinf + plab + maxtemp, data=leukem)
modB
```

```
Call:
coxph(formula = Surv(months, alive == 0) ~ age + pblasts + pinf +
    plab + maxtemp, data = leukem)

            coef exp(coef) se(coef)     z      p
age      0.03308   1.03363  0.01016  3.26 0.0011
pblasts  0.00945   1.00950  0.01396  0.68 0.4983
pinf    -0.01710   0.98304  0.01224 -1.40 0.1625
plab    -0.06600   0.93613  0.03865 -1.71 0.0877
maxtemp  0.15545   1.16818  0.11198  1.39 0.1651

Likelihood ratio test=18.5  on 5 df, p=0.00241
n= 51, number of events= 45
```



Measures of Association and Effect

-we've seen both odds ratios and risk ratios above, because they feature in regression model interpretation
-in general, for ratios:
    -ratio >1 = increased probability or odds of something happening
    -ratio 1 = probability/odds neither increased nor decreased
    -ratio <1 = decreased probability or odds of something happening

OR and RR, part 3

If we started with a defined population, assessed exposure and subsequently collected incident cases…

N=400	Cases	Controls	Total
Exposed	120	100	220
Not exposed	80	100	180
Total	Total cases	Total controls	

$$OR_{disease} = \frac{120/100}{80/100} = 1.5 \qquad RR = \frac{120/220}{80/180} = 1.23$$

Disease not rare… so OR does not approximate RR well

-Case control studies use ODDS
-Cohort studies tend to use RELATIVE RISK
-RELATIVE RISK is always better, if you can swing it
    -but for rare diseases OR is fairly equivalent

Diagnostic Tests

Sensitivity and specificity

		Truth/Gold Standard	
		Has disease	Does not have disease
Test result	Positive	80	100
	Negative	20	800
		100	900

Sensitivity = 80/100 = 80%
Specificity = 800/900 = 89%

Predictive value

		Truth		
		Has disease	Does not have disease	
Test result	Positive	80	100	180
	Negative	20	800	820
		100	900	

Positive predictive value = 80/180 = 0.44 or 44%
Negative predictive value = 800/820 = 0.98 or 98%

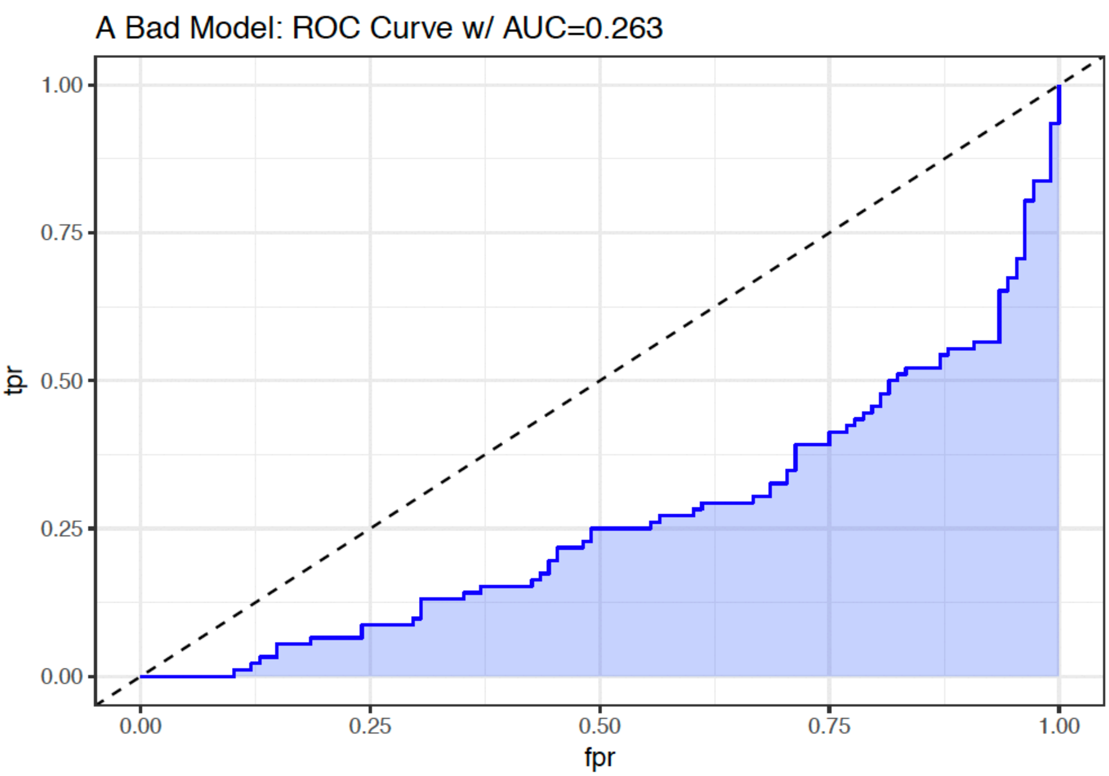
REMEMBER: sensitivity = “PID” = positive in disease, specificity = “NIH” = negative in health

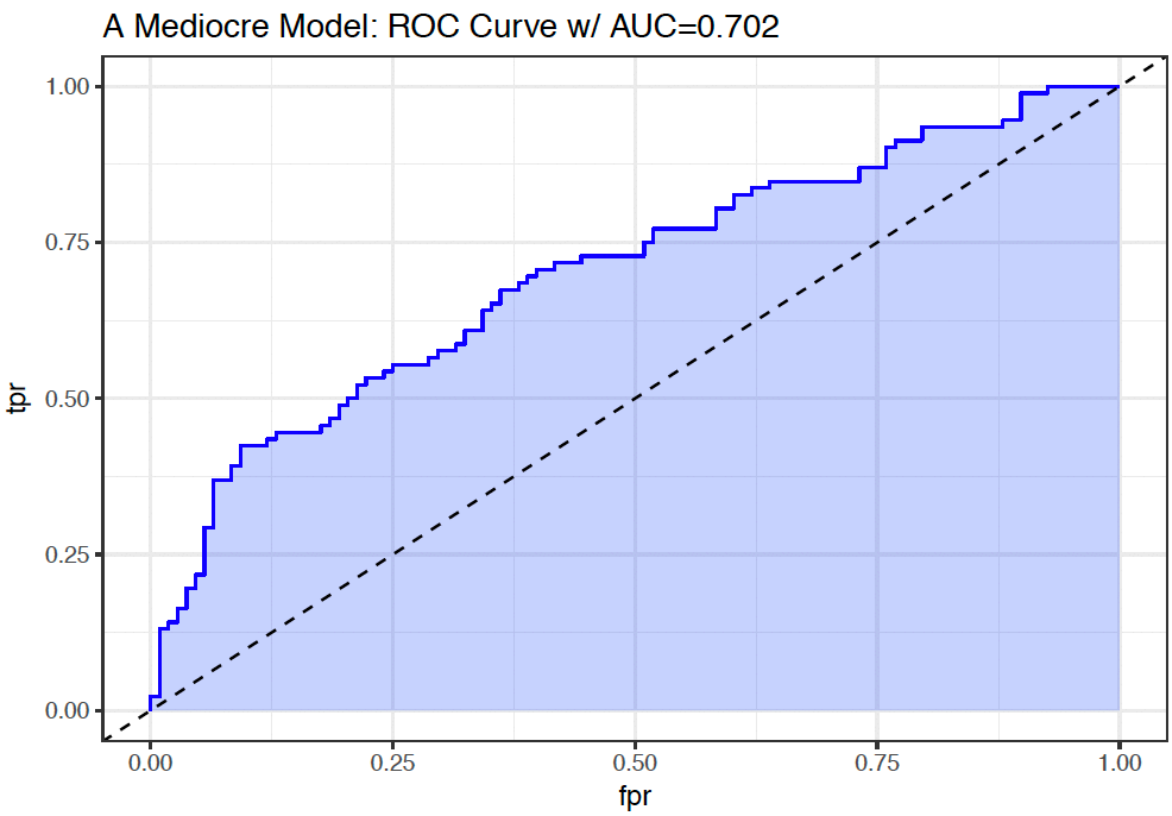
AUC Curves

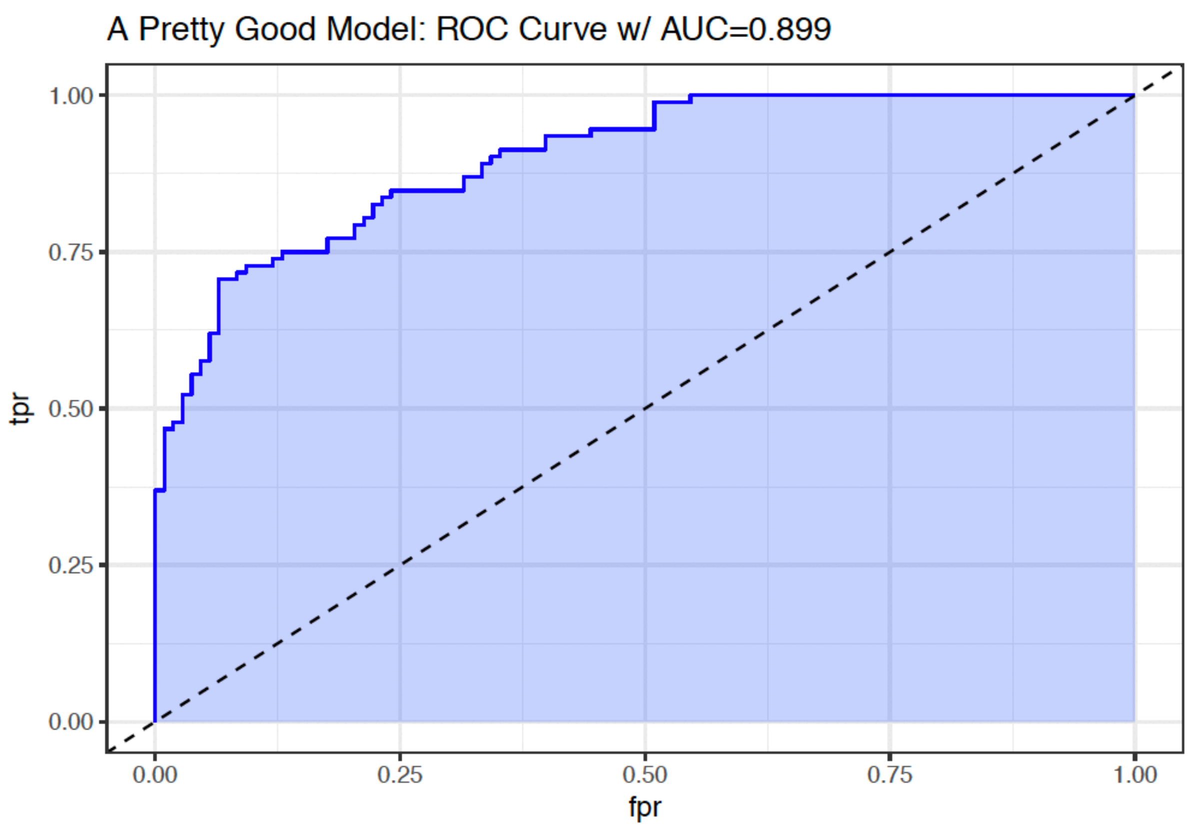
-generally pertain to logistic models, to show overall model performance (yes/no questions)

Sometimes people grasp for a rough guide as to the accuracy of a model's predictions based on the area under the ROC curve. A common thought is to assess the C statistic much like you would a class grade.

C statistic	Interpretation
0.90 to 1.00	model does an excellent job at discriminating "yes" from "no" (A)
0.80 to 0.90	model does a good job (B)
0.70 to 0.80	model does a fair job (C)
0.60 to 0.70	model does a poor job (D)
0.50 to 0.60	model fails (F)
below 0.50	model is worse than random guessing




A Mediocre Model: ROC Curve w/ AUC=0.702
tpr
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
fpr


A Pretty Good Model: ROC Curve w/ AUC=0.899
tpr
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
fpr

References

1. Dr. Thomas E. Love's *PQHS 431-432* Course Notes
2. Dr. Farren Briggs' *Introduction to Epidemiology* Course Notes from Fall 2017
3. UCLA Institute for Digital Research and Education website
4. Dr. Danielle Navarro's Learning Statistics with R
5. Dr. Richard McElreath's Statistical Rethinking: A Bayesian Course with Examples in R and Stan