Class #13 Monday 28 February 2011

What did we cover last time?

- Parameter Fitting
  - Maximum Likelihood
- Random number generation
- Hypothesis Testing Terms
  - Parametric vs Nonparametric
  - Sampling Distribution
  - Null Hypothesis
  - Test statistic
  - Test levels, p levels
  - Type I and Type II errors
  - One- and Two- sided tests
- Hypothesis Testing Types
  - Student's t-test

Practicalities of Student's t-test

❋(A) One-sample t-test, very popular since based on the Gaussian. Student's t-test, based on the t distribution:

❋ Pooled Variance T-test, difference of means

$$tstat = \frac{|ave_1 - ave_2|}{sqrt(\text{var}(1/n_1 + 1/n_2))}$$

$$\text{var} = \frac{(n_1 - 1)\text{var}_1 + (n_2 - 1)\text{var}_2}{DOF}$$

$$DOF = n_1 + n_2 - 2$$

Practical t-test equations

❋Correlation t-test

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad r = correlation\ coefficient$$

❋ Quenouille to reduce DOF, lag 1-3 correlations

$$n_{effective} = \frac{n}{\left(1 + 2(r_1 r_1^{'} + r_2 r_2^{'} + r_3 r_3^{'})\right)}$$

4.2 Parametric Tests continued...

❋(E) Goodness of Fit (Sec 5.2.5 Wilks)

❋ Null Hypothesis - Data drawn from given distribution. Evidence positive then we confirm the null hypothesis. Many tests - will discuss a few

❋ Chi squared test (χ^2)

❋ Data histogram compared with PDF, works best with discrete data.

$$\chi^2 = \sum_{classes} \frac{(\#Observed - \#Expected)^2}{\#Expected}$$

$$= \sum_{classes} \frac{(\#Observed - n\Pr\{data\ in\ class\})^2}{n\Pr\{data\ in\ class\}} \qquad (5.14)$$

❋ Avoid classes with small numbers (<5 avoided).

❋ DOF (#classes - #of parameters fit -1), 1-sided

❋(E) Goodness of Fit (Sec 5.2.5 Wilks) continued...

❋ Chi squared test (χ^2) continued...

❋ EXAMPLE: Jan Ithaca Precip.

❋ parameters (2 each type),

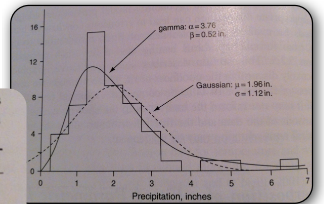


TABLE 5.1 The χ^2 goodness-of-fit test applied to gamma and Gaussian distributions for the 1933–1982 Ithaca January precipitation data. Expected numbers of occurrences in each bin are obtained by multiplying the respective probabilities by $n = 50$.

Class	<1"	1 — 1.5"	1.5 — 2"	2 — 2.5"	2.5 — 3"	≥3"
Observed #	5	16	10	7	7	5
Gamma:						
Probability	0.161	0.215	0.210	0.161	0.108	0.145
Expected #	8.05	10.75	10.50	8.05	5.40	7.25
Gaussian:						
Probability	0.195	0.146	0.173	0.173	0.132	0.176
Expected #	9.75	7.30	8.65	8.90	6.60	8.80

❋ 6 classes, 2 parameters, ==> 3 DOF

❋ 0.161 * 50 = 8.05, fill in rest similarly, take difference & square $(8.05-5)^2$, divide and sum up

❋ Get χ^2=5.05 Gamma and χ^2=14.96 Gaussian

❋ Gamma not rejected at 90%, Gaussian rejected 1%

4.2 Parametric Tests continued...

❋(F) Kolmogorov-Smirnov Test

❋ Compares empirical and theoretical CDFs

❋ Null hypothesis is that data drawn from particular distribution

❋ Works as long as parameters are not estimated from data sample used in test. Test statistic is Empirical CDF minus theoretical CDF. Critical values K 1.224, 1.358, & 1.628 for 90, 95 and 99% levels.

$$D_n = \max_x \left| F_n(x) - F(x) \right| \quad (5.15)$$

$$C_\alpha = \frac{K_\alpha}{\sqrt{n} + 0.12 + 0.11/\sqrt{n}} \quad (5.16)$$

4.2 Parametric Tests continued...

❋(F) Kolmogorov-Smirnov Test continued...

❋ EXAMPLE: January Ithaca Precipitation again

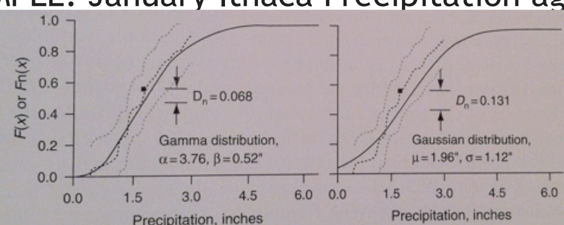


FIGURE 5.5 Illustration of the Kolmogorov-Smirnov D_n statistic as applied to the 1933–1982 Ithaca January precipitation data, fitted to a gamma distribution (a) and a Gaussian distribution (b). Solid curves indicate theoretical cumulative distribution functions, and black dots show the corresponding empirical estimates. The maximum difference between the empirical and theoretical CDFs occurs for the highlighted square point, and is substantially greater for the Gaussian distribution. Grey dots show limits of the 95% confidence interval for the true CDF from which the data were drawn (Equation 5.16).

❋Table 5.2 Lists critical values for Gamma distribution for a given alpha, when fit parameters are based on data. Use alpha=infinity for Gaussian, C_α for confidence intervals.

4.2 Parametric Tests continued...

❋(G) 2 Sample Kolmogorov-Smirnov Test

❋Use Eq 5.15 to test the null hypothesis that two samples came from the same distribution.

$$D_n = \max_x \left| F_n(x_1) - F_m(x_2) \right| \quad (5.17)$$

$$D_s > \left[-\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \ln\left(\frac{\alpha}{2}\right) \right] \quad (5.18)$$

❋ Eqns for test statistic (5.17) and Critical Value (5.18). 1 and 2 refer to the two samples.

❋(H) Shapiro-Wilk Test (variant called Filliben test)

❋ Preferable test for gaussian distribution. Test statistic is correlation for the points along the Gaussian Q-Q plot. Table 5.3 gives critical values.

❋(I) Likelihood ratio test (Section 5.2.6 Wilks)

❋ Test works for whether two parts of record from same distribution. Example 5.6

4.3 Nonparametric Tests

❋(A) Classical tests for location

❋ May be appropriate if we think parametric assumptions not met or don't have sampling distribution.

❋ 5 Elements of Hypothesis testing still apply.

❋ (A1) Wilcoxon-Mann-Whitney rank-sum test

❋ For two independent samples (diff of means), Null hypothesis-data have been drawn from same distribution. Serial correlation needs to be considered. Sum of ranks (R), Mann-Whitney U statistic. crit values, Null dist. of U statistic is gaussian with mu and sigma.

$$U_1 = R_1 - \frac{n_1}{2}(n_1+1) \quad (5.22a)$$

$$U_2 = R_2 - \frac{n_2}{2}(n_2+1) \quad (5.22b)$$

$$\mu_U = \frac{n_1 n_2}{2} \quad (5.23a)$$

$$\sigma_U = \left[\frac{n_1 n_2 (n_1+n_2+1)}{12}\right]^{1/2} \quad (5.23b)$$

4.3 Nonparametric Tests cont...

❋ Example 5.7 Cloud Seeding Experiment. Suspect that cloud seeding reduced lightning. Very non-gaussian so use Mann-Whitney Test.

❋n1=12, n2=11

❋19.25 & 69.45 averages

❋ pool points

❋ rank points

❋ $(n!)/[n_1!\ n_2!]$ total number possible combinations

TABLE 5.5 Counts of cloud-to-ground lightning for experimentally seeded and nonseeded storms. From Baughman *et al.* (1976).

Seeded		Unseeded	
Date	Lightning strikes	Date	Lightning strikes
7/20/65	49	7/2/65	61
7/21/65	4	7/4/65	33
7/29/65	18	7/4/65	62
8/27/65	26	7/8/65	45
7/6/66	29	8/19/65	0
7/14/66	9	8/19/65	30
7/14/66	16	7/12/66	82
7/14/66	12	8/4/66	10
7/15/66	2	9/7/66	20
7/15/66	22	9/12/66	358
8/29/66	10	7/3/67	63
8/29/66	34		

4.3 Nonparametric Tests

❋ Example 5.7 cont...

❋ pool and rank 23 pts.

❋ R_1=108.5, R_2=167.5

❋ U_1=108.5-6*(12+1)

❋ U_1=30.5

❋ Possible=1,352,078

μ_u = (12)(11)/2=66

σ_u = [ ] = 16.2

Z=(30.5-66)/16.2=-2.19

Table B.1 p=0.014, 1.4% are smaller than Ho, so it is rejected.

TABLE 5.6 Illustration of the procedure of the rank-sum test using the cloud-to-ground lightning data in Table 5.5. In the left portion of this table, the $n_1 + n_2 = 23$ counts of lightning strikes are pooled and ranked. In the right portion of the table, the observations are segregated according to their labels of seeded (S) or not seeded (N) and the sums of the ranks for the two categories (R_1 and R_2) are computed.

Pooled Data			Segregated Data			
Strikes	Seeded?	Rank				
0	N	1			N	1
2	S	2	S	2		
4	S	3	S	3		
9	S	4	S	4		
10	N	5.5			N	5.5
10	S	5.5	S	5.5		
12	S	7	S	7		
16	S	8	S	8		
18	S	9	S	9		
20	N	10			N	10
22	S	11	S	11		
26	S	12	S	12		
29	S	13	S	13		
30	N	14			N	14
33	N	15			N	15
34	S	16	S	16		
45	N	17			N	17
49	S	18	S	18		
61	N	19			N	19
62	N	20			N	20
63	N	21			N	21
82	N	22			N	22
358	N	23			N	23
		Sums of Ranks:	$R_1 = 108.5$		$R_2 = 167.5$	

4.3 Nonparametric Tests cont...

❋(A2) Wilcoxon signed-rank test for paired samples

❋ Example 5.8

Storm in 2 parts of US are correlated, Calc D_i, n'=21 since no 0, test statistic is T^+=78.5, T^-=152.5,

Null => Storm Freq is same in 2 places.

Estimate null Gaussian distribution, z=-1.29

Not rejected...

$$T_i = rank|D_i| \quad (5.24)$$

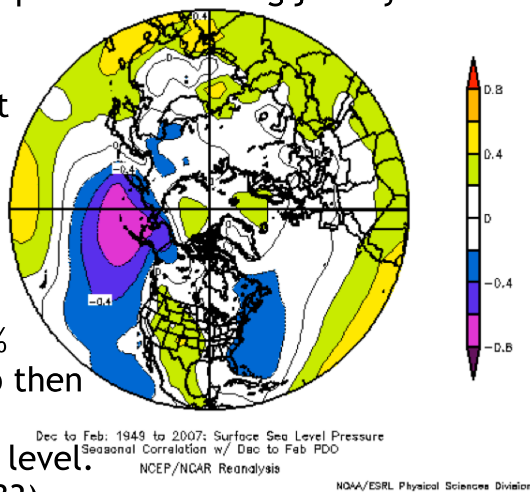
$$T^+ = \sum_{D_i>0} T_i \quad T^- = \sum_{D_i<0} T_i \quad (5.25)$$

TABLE 5.7 Illustration of the procedure of the Wilcoxon signed-rank test using data for counts of thunderstorms reported in the northeastern United States (x) and the Great Lakes states (y) for the period 1885–1905, from Brooks and Carruthers (1953). Analogously to the procedure of the rank-sum test (see Table 5.6), the absolute values of the annual differences, $|D_i|$, are ranked and then segregated according to whether D_i is positive or negative. The sum of the ranks of the segregated data constitute the test statistic.

Paired Data			Differences		Segregated Ranks	
Year	X	Y	D_i	Rank$\|D_i\|$	$D_i > 0$	$D_i < 0$
1885	53	70	−17	20		20
1886	54	66	−12	17.5		17.5
1887	48	82	−34	21		21
1888	46	58	−12	17.5		17.5
1889	67	78	−11	16		16
1890	75	78	−3	4.5		4.5
1891	66	76	−10	14.5		14.5
1892	76	70	+6	9	9	
1893	63	73	−10	14.5		14.5
1894	67	59	+8	11.5	11.5	
1895	75	77	−2	2		2
1896	62	65	−3	4.5		4.5
1897	92	86	+6	9	9	
1898	78	81	−3	4.5		4.5
1899	92	96	−4	7		7
1900	74	73	+1	1	1	
1901	91	97	−6	9		9
1902	88	75	+13	19	19	
1903	100	92	+8	11.5	11.5	
1904	99	96	+3	4.5	4.5	
1905	107	98	+9	13	13	
				Sums of Ranks:	$T^+ = 78.5$	$T^- = 152.5$

4.4 Field Significance

- Q: Is my correlation pattern occurring just by random chance?
- Correlation t test
- Count #N significant points versus total
- Monte Carlo way
  randomize SOI, get r 1000 times and then make histogram,
  Ask: Is the N in top 5% of my histogram? If so then correlation pattern is field significant at 5% level.
  Livesey and Chen (1983),
  Use 5.29 if little autocorrelation, Fisher transform Z



4.3 Nonparametric Tests cont...

- (B) Resampling tests
- Next time

4.3 Nonparametric Tests cont...

- (C) Permutation tests
- Blah Blah Blah

4.3 Nonparametric Tests cont...

- (D) The Bootstrap
- Blah Blah Blah

4.4 Field Significance and Multiplicity

- (A) The Bootstrap
- Blah Blah Blah

Chapter 5 Statistical Forecasting

- 5.1 Background
- Background
- 5.2 Linear Regression
- 5.2 Multiple Linear Regression
-