11.2 Inferences for σ's, 2 Populations

Study Ch. 11.2, #51, 63-69, 73

11.2 Inferences for σ's, 2 Populations

Procedures that assume = σ's

1. Pooled-t test
2. Regression Analysis
3. ANOVA

Previous Techniques to Check for =σ's

1. Box Plots: visually compare spread of data
2. Residual Plot: visually look for non-random pattern that suggests different distances from x-axis
3. Compare s for each population. If any s is 2 or more times any other s, assume that σ's are different.

If the populations really had identical standard deviations, what is the chance of observing as large a discrepancy among sample standard deviations as occurred in the data (or an even larger discrepancy)?

11.2 Inferences for σ's, 2 Populations

Need a more analytical approach

If the populations have identical standard deviations,

what is the chance of observing as large a discrepancy among sample standard deviations as occurs in the data?

11.2 Inferences for σ's, 2 Populations

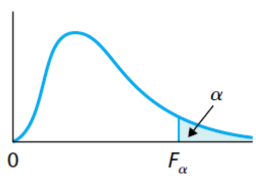
More analytical approach

geogebra

Use the F- Distribution with Hypothesis Test

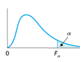
F - Distribution

1. Ratio of variations
   - ANOVA $F = \dfrac{MSTR}{MSE}$
   - Two σ's Test $F = \dfrac{s_1^2}{s_2^2}$
2. Total area under curve = 1
3. Starts at 0.
4. Right-skewed.



F-Distribution Table
- 8 pages long
- df numerator across top
- df denominator on sides, with
  α = 0.10, 0.05, 0.025, 0.01, 0.005

TABLE VIII
Values of F_α



dfd	α	dfn 1	2	3	4	5	6	7	8	9
	0.10	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	0.05	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
1	0.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	0.01	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	0.005	16211	20000	21615	22500	23056	23437	23715	23925	24091
	0.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	0.05	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
2	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	0.01	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	0.005	198.50	199.00	199.17	199.25	199.30	199.33	199.36	199.37	199.39
	0.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	0.05	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
3	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	0.01	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	0.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
	0.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	0.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
4	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	0.01	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	0.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14

11.2 Inferences for σ's, 2 Populations

What F - value can we expect?

1. s_1 is best estimate of σ_1, s_2 is best estimate of σ_2

2. If $\sigma_1 = \sigma_2$ then $F = \frac{s_1^2}{s_2^2}$ should be close to 1

   If $\sigma_1 < \sigma_2$ then $F = \frac{s_1^2}{s_2^2}$ should be < 1

   If $\sigma_1 > \sigma_2$ then $F = \frac{s_1^2}{s_2^2}$ should be >1

3. Since we expect variation in sample stdev's, we do NOT expect that the sample F value =1 exactly, not even if $\sigma_1 = \sigma_2$

4. Use hypothesis test to decide how much less than or greater than 1 F needs to be to decide between null and alternative hypotheses.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page ©Gertrude Battaly, 2014 Class Notes Homework

11.2 Inferences for σ's, 2 Populations

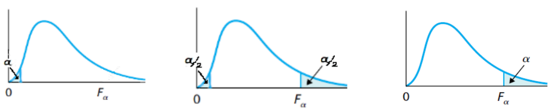
Two Standard Deviation F-Test

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND or large samples

H_0: $\sigma_1 = \sigma_2$

H_a: $\sigma_1 < \sigma_2$ or $\sigma_1 \neq \sigma_2$ or $\sigma_1 > \sigma_2$

Step 2: Decide α



Step 3: Compute $F = \frac{s_1^2}{s_2^2}$

Step 4: Find CV(s) OR Find p-value:

left-tailed p = Fcdf (0, Ftest, dfn, dfd)
2-tailed p = 2*Fcdf (0, Ftest, dfn, dfd) if F <1
or p = 2(1 - Fcdf (0, Ftest, dfn, dfd)) if F >1
right-tailed p = 1- Fcdf (0, Ftest, dfTR, dfE)

Step 5: Decide whether to reject H_0 or not

Step 6: Verbal interpretation

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page ©Gertrude Battaly, 2014 Class Notes Homework

11.2 Inferences for σ's, 2 Populations

An independent s.r.s of infants was taken. 10 infants were treated for pulmonary hypertension (PH), 25 infants were not treated (control). Head circumferences were measured:

Treated	33.9	33.4	37.9	32.5	36.3	35.1	34.5	31.3	32.9	34.2															
Control	35.2	33.4	34.3	31.8	31.6	35.6	31.3	33.1	34.1	31.9	36.7	33.5	32.4	35.2	31.9	35.1	35.8	35.1	34.8	32.8	36.0	36.3	33.6	34.5	34.0

Based on this data, at the 5% significance level, does a difference in variation exist between infants treated for PH and those not treated? *(Note: A normal probability plot is approximately linear. $s_1=1.907$, $s_2=1.594$)*

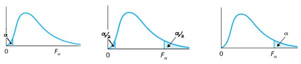
Two Standard Deviation F-Test

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND or large samples

H_0: $\sigma_1 = \sigma_2$

H_a: $\sigma_1 < \sigma_2$ or $\sigma_1 \neq \sigma_2$ or $\sigma_1 > \sigma_2$

Step 2: Decide α



Step 3: Compute $F = \dfrac{s_1^2}{s_2^2}$

Step 4: Find CV(s) OR Find p-value:

left-tailed p = Fcdf (0, Ftest, dfn, dfd)
2-tailed p = 2*Fcdf (0, Ftest, dfn, dfd) if F <1
or p = 2(1 - Fcdf (0, Ftest, dfn, dfd)) if F >1
right-tailed p = 1- Fcdf (0, Ftest, dfTR, dfE)

Step 5: Decide whether to reject H_0 or not

Step 6: Verbal interpretation

11.2 Inferences for σ's, 2 Populations

An independent s.r.s of infants was taken. 10 infants were treated for pulmonary hypertension (PH), 25 infants were not treated (control). Head circumferences were measured:

Treated	33.9	33.4	37.9	32.5	36.3	35.1	34.5	31.3	32.9	34.2															
Control	35.2	33.4	34.3	31.8	31.6	35.6	31.3	33.1	34.1	31.9	36.7	33.5	32.4	35.2	31.9	35.1	35.8	35.1	34.8	32.8	36.0	36.3	33.6	34.5	34.0

Based on this data, at the 5% significance level, does a difference in variation exist between infants treated for PH and those not treated? *(Note: A normal probability plot is approximately linear. $s_1=1.907$, $s_2=1.594$)*

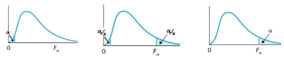
Two Standard Deviation F-Test

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND or large samples

H_0: $\sigma_1 = \sigma_2$

H_a: $\sigma_1 < \sigma_2$ or $\sigma_1 \neq \sigma_2$ or $\sigma_1 > \sigma_2$

Step 2: Decide α



Step 3: Compute $F = \dfrac{s_1^2}{s_2^2}$

Step 4: Find CV(s) OR Find p-value:

left-tailed p = Fcdf (0, Ftest, dfn, dfd)
2-tailed p = 2*Fcdf (0, Ftest, dfn, dfd)
right-tailed p = 1- Fcdf (0, Ftest, dfTR, dfE)

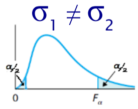
Step 5: Decide whether to reject H_0 or not

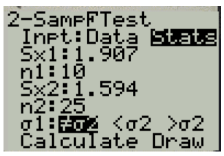
Step 6: Verbal interpretation

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND or large samples

H_0: $\sigma_1 = \sigma_2$ H_a: $\sigma_1 \neq \sigma_2$

Step 2: $\alpha = 0.05$

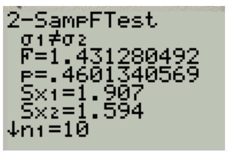




Step 3: Compute $F = \dfrac{s_1^2}{s_2^2} = \dfrac{1.907}{1.594} = 1.431$

Step 4: p = 0.4601

left-tailed p = Fcdf (0, Ftest, dfn, dfd)
2-tailed p = 2*Fcdf (0, Ftest, dfn, dfd) if F <1
or p = 2(1 - Fcdf (0, Ftest, dfn, dfd)) if F >1
right-tailed p = 1- Fcdf (0, Ftest, dfTR, dfE)



Step 5: $p = 0.4601 > 0.05 = \alpha$ $\therefore$ do NOT Reject H_0

Step 6: Based on this data, there is no difference in variation among head circumferences of infants treated for PH and those not treated.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page ©Gertrude Battaly, 2014 Class Notes Homework

11.2 Inferences for σ's, 2 Populations

Soil scientists have measured the arsenic concentration in the soil using two different methods. Ten independent simple random samples were taken using each of the two methods.

The scientists want to determine which methods results in more precise data. The more precise method would have a lower standard deviation since it would result in more consistent outcomes when measuring mean amounts of arsenic.

Data for the two methods includes:

Method	Mean (ppm)	s (ppm)	n
1	6.7	0.8	10
2	8.2	1.2	10

At a 5% significance level, is Method 1 more precise than Method 2? (A normal prob plot appears approx linear.)

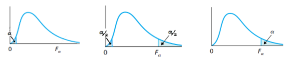
Two Standard Deviation F-Test

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND or large samples

H_0: $\sigma_1 = \sigma_2$

H_a: $\sigma_1 < \sigma_2$ or $\sigma_1 \neq \sigma_2$ or $\sigma_1 > \sigma_2$

Step 2: Decide α



Step 3: Compute $F = \dfrac{s_1^2}{s_2^2}$

Step 4: Find CV(s) OR Find p-value:

left-tailed p = Fcdf (0, Ftest, dfn, dfd)
2-tailed p = 2*Fcdf (0, Ftest, dfn, dfd)
right-tailed p = 1- Fcdf (0, Ftest, dfTR, dfE)

Step 5: Decide whether to reject H_0 or not

Step 6: Verbal interpretation

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page ©Gertrude Battaly, 2014 Class Notes Homework

11.2 Inferences for σ's, 2 Populations

Soil scientists have measured the arsenic concentration in the soil using two different methods. Ten independent simple random samples were taken using each of the two methods.

The scientists want to determine which methods results in more precise data. The more precise method would have a lower standard deviation since it would result in more consistent outcomes when measuring mean amounts of arsenic.

Data for the two methods includes:

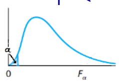
Method	Mean (ppm)	s (ppm)	n
1	6.7	0.8	10
2	8.2	1.2	10

At a 5% significance level, is Method 1 more precise than Method 2? (A normal prob plot appears approx linear.)

Step 1: Assumptions: 1. SRS, 2. Indep samples 3. ND

H_0: $\sigma_1 = \sigma_2$ H_a: $\sigma_1 < \sigma_2$

Step 2: $\alpha = 0.05$



Step 3: Compute $F = \dfrac{s_1^2}{s_2^2} = \dfrac{0.8}{1.2} = 0.444$

Step 4: p = 0.121 left-tailed p = Fcdf (0, Ftest, dfn, dfd)

Step 5: p = 0.121 > 0.05 = α ∴ do NOT Reject H_0

Step 6: Based on this data, conclude that there is no difference in variation among the two groups. Therefore, there is no difference in precision between the two methods.

Class Notes: Prof. G. Battaly, Westchester Community College, NY

Statistics Home Page ©Gertrude Battaly, 2014 Class Notes Homework

Attachments

Statistical Tables.pdf