Statistical Methods in Natural Resources Management

ESRM 304

Statistical Methods in Natural Resources Management

I. Estimating a Population Mean

II. Comparing two Population Means

III. Reading Assignment

I. Estimating a Population Mean

A. Simple Random Sampling (SRS)

B. Mean, variance, standard dev., std. error

C. Estimating Reliability with Confidence

I. Estimating a Population Mean

A. Simple Random Sampling (SRS)

1) The real 'workhorse' of statistical methods
   - All other methods have their 'roots' in SRS
2) Every possible combination of *n* units is equally probable
   - How to do this?
3) Units may be selected w/ or w/o replacement

I. Estimating a Population Mean

A. Simple Random Sampling - Example

- A 250-acre forest was sampled for cordwood volume per acre in trees larger than 5" DBH
- Sample units were chosen to be 1/4-acre, circular plots
- Sample size was chosen to be $n = 25$
- Sampling without replacement will be used

7	10	7	4	7
8	8	8	7	5
2	6	9	7	8
6	7	11	8	8
7	3	8	7	7
			Total =	175

I. Estimating a Population Mean

B. Mean, variance, standard dev., std. error

- ✓ Cordwood volume on the i^{th} plot will be named y_i

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{25}\left(7+8+2+\cdots+7\right) = \frac{175}{25} = 7 \ cds \ per \ 1/4 - acre \ plot$$

Naturally, there are four 1/4-ac plots per acre, so
There are 28 cds per acre on average in the forest.

Total volume in the forest …

$$\hat{Y} = \text{(cds/acre)(no. of acres)} = 28\left(250\right) = 7{,}000 \ cds$$

$$\hat{Y} = N\bar{y} = 1{,}000\left(7\right) = 7{,}000 \ cds$$

I. Estimating a Population Mean

B. Mean, variance, standard dev., std. error

$$s_y^2 = \frac{\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{2} y_i\right)^2}{(n-1)}$$

$$= \frac{\left(7^2 + 8^2 + \cdots + 7^2\right) - \frac{1}{25}(175)^2}{(25-1)}$$

$$= \frac{1{,}317 - 1{,}225}{24} = 3.8333\ (cds/qtr.acre)^2$$

I. Estimating a Population Mean

B. Mean, variance, standard dev., std. error

$$s_y = \sqrt{s_y^2} = \sqrt{3.8333} = 1.9579 \; cds$$

Standard Error has two variants …

For sampling with replacement:

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}} = \frac{1.9579}{\sqrt{25}} = 0.3916 \; cds$$

For sampling without replacement:

$$s_{\bar{y}} = \frac{s_y}{\sqrt{n}}\sqrt{\left(1 - \frac{n}{N}\right)} = \frac{1.9579}{\sqrt{25}}\sqrt{\left(1 - \frac{25}{1{,}000}\right)} = 0.3916\left(0.975\right) = 0.3867 \; cds$$

I. Estimating a Population Mean

C. Estimating Reliability with Confidence

- By itself, the mean does not tell us everything we need to know
- What remains is knowing how reliable the sample mean is for estimating the population mean
  - Standard Error comes to our assistance here
- We use a 'Confidence Interval' to tell us the probability of being within a certain range of the population mean
- Accuracy and Precision interact at this stage

I. Estimating a Population Mean

C. Estimating Reliability with Confidence

- For large samples, we can appeal to the Normal dist'n
  - A 95% CI for the population mean is given by

$$\text{Estimate} \quad \pm \quad 2\left(\text{Standard Error of the Estimate}\right)$$

- For small samples we rely on "Student's *t*" dist'n
  - A CI for the population mean is given by

$$\text{Estimate} \quad \pm \quad t\left(\text{Standard Error of the Estimate}\right)$$

  - The value of *t* we choose depends on how confident we wish to be that we have "covered" the population mean within our range of estimated values (interval) n

C.

Table 2.—*Distribution of* t

df	Probability								
	.5	.4	.3	.2	.1	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	$1.821	63.657	536.619
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707

I. Estimating a Population Mean

C. Estimating Reliability with Confidence

- For an estimate made from 25 units (24 df), a 95% confidence interval is given by

$$\bar{y} \pm t \cdot s_{\bar{y}} = 7 \pm 2.064(0.3867)$$

$$= 6.20 \text{ to } 7.80 \; cds \; per \; 1/4-ac$$

Expanding that to an acre …

$$[4]7 \pm 2.064([4]0.3867)$$

$$= 24.80 \text{ to } 31.20 \; cds \; per \; acre$$

II. Comparing Two Population Means

A. Estimating two means, two variances
B. Estimating variance and standard error of the *difference* between the two means
C. Two methods to test the difference

II. Comparing Two Population Means

A. Estimating two means, two variances

Example -

We are interested if our lowland hardwood forest produced the same volume (ft^3/acre) as our upland hardwood forest over a given time period

We sample ten, randomly chosen, one-acre plots in each forest type

II. Comparing Two Population Means

A. Estimating two means, two variances

The data (y_1 = lowland; y_2 = upland):

Lowland Hardwoods		Upland Hardwoods	
520	760	420	180
710	890	210	260
770	810	290	320
840	580	350	270
630	860	540	200
Total =	7,370	Total =	3,040

$$\bar{y}_1 = \frac{1}{10}(7,370) = 737\ ft^3 / acre \qquad \bar{y}_2 = 304\ ft^3 / acre$$

II. Comparing Two Population Means

A. Estimating two means, two variances

$$s_1^2 = \frac{\left(520^2 + 710^2 + \cdots + 860^2\right) - \frac{1}{10}\left(7,370\right)^2}{\left(10-1\right)} = 15,556.6667$$

$$s_2^2 = \frac{\left(420^2 + 210^2 + \cdots + 200^2\right) - \frac{1}{10}\left(3,040\right)^2}{\left(10-1\right)} = 12,204.4444$$

II. Comparing Two Population Means

B. Estimating variance and standard error of the *difference* between the two means

First, estimate a common (pooled) variance…

$$s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{(n_1 - 1) + (n_2 - 1)}$$

$$= \frac{15{,}556.6667(10-1) + 12{,}204.4444(10-1)}{(10-1)+(10-1)}$$

$$= 13{,}880.5556\ (ft^3\ per\ acre)^2$$

II. Comparing Two Population Means

B. Estimating variance and standard error of the *difference* between the two means

… then, the standard error of the difference…

$$s_{\bar{y}_1 - \bar{y}_2} = \sqrt{\frac{s_p^2 (n_1 + n_2)}{(n_1)(n_2)}} = \sqrt{\frac{13{,}880.5556(10+10)}{(10)(10)}}$$

$$= 52.6888\ ft^3\ per\ acre$$

II. Comparing Two Population Means

C. Two methods to test the difference

First, the Confidence Interval method…

$$\text{Estimate} \quad \pm \quad t\left(\text{Standard Error of the Estimate}\right)$$

For a 95% CI … (recall we have 18 *df*)

$$\bar{y}_1 - \bar{y}_2 \;\pm\; t \cdot s_{\bar{y}_1 - \bar{y}_2}$$

$$= \quad 737 - 304 \;\pm\; 2.101\left(52.6888\right)$$

$$= \quad 322.3 \;\text{ to }\; 543.7 \;\; \text{ft}^3 \; per\ ac.$$

on

C.

Table 2.—*Distribution of* t

df	Probability								
	.5	.4	.3	.2	.1	.05	.02	.01	.001
1	1.000	1.376	1.963	3.078	6.314	12.706	$1.821	63.657	536.619
2	.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	.741	.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	.727	.920	1.156	1.476	2.015	2.571	3.365	4.032	6.859
6	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499	5.405
8	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707

II. Comparing Two Population Means

C. Two methods to test the difference

<u>Second</u>, the *t* test (Hypothesis Testing) method…

- The hypothesis we actually test states "there is no difference between the means" (the "null")
- To run the test, we choose a probability of incorrectly rejecting the notion of "no difference" … let's say 5%

$$t_{obs} = \frac{\bar{y}_1 - \bar{y}_2}{s_{\bar{y}_1 - \bar{y}_2}} = \frac{737 - 304}{52.6888} = 8.218$$

Is this "more extreme" than a 5% *t* with 18 *df*? … YES!

III. Reading Assignment

A. "Elementary Statistical Methods for Foresters," by Freese, p. 1-13; 24-27

B. "READ This! Study Tips," by Rose