Dis 12: Inference about Two Populations

Related textbook chapter: 13

Ch 13 handout and solution offered by Dr. Pac can be accessed here: Handout Solution This handout incorporates reviews with all exercises from the handout given by Dr. Pac.

1 Motivation

- Last week, we talked about how to conduct hypothesis testings on parameters obtained from one population.
- This week, we will look at how to conduct hypothesis for parameters obtained from two different populations.
- One immediate question: how to test something like

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

where both sides of the equation contain something unknown (μ_1 on the left hand side, μ_2 on the right hand side)?

- Solution: we can rewrite the above hypotheses as the following:

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

Now the left hand side is something unknown related to the population parameters & waiting to be tested, and the right hand side is a concrete number to test the unknown against.

- Similar to last discussion about the one population case, we will discuss how to test / compare the following three sets of population parameters:
 - 1. The population means $(\mu_1 \mu_2)$
 - 2. The population variances $(\frac{\sigma_1^2}{\sigma_2^2})$
 - 3. The population success proportions ($p_1 p_2$, from two binomial experiments)

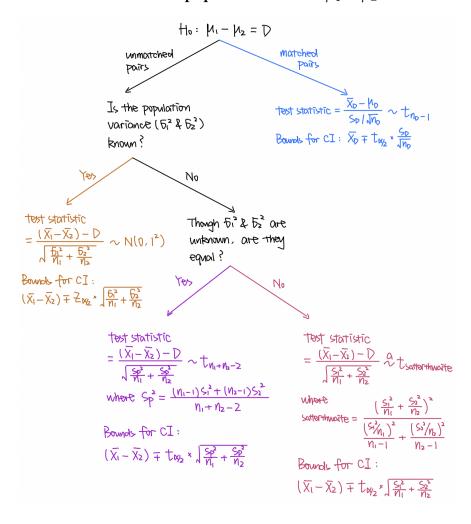
2 General Approach

- The general testing approach that we will take is very similar to what we talked about last discussion:
 - If using the rejection region & test statistic method:
 - 1. Find out what distribution the test statistic follows.
 - 2. Set up significance level & the appropriate sizes for the tail-ends of the distribution. Find cutoff values to construct rejection region.
 - 3. Calculate test statistic from the given sample, and see if it falls within the rejection region.

- * If test statistic falls within the rejection region, then we reject H_0 under the specified significance level.
- * If test statistic doesn't fall within the rejection region, then we fail to reject H_0 under the specified significance level.
- If using the confidence interval method:
 - 1. Find out what distribution the test statistic follows.
 - 2. Construct (1α) level of confidence interval based on the confidence level (1α) and the associated cutoffs from the distribution.
 - 3. Check if the hypothesized null falls within the (1α) confidence interval:
 - * If the hypothesized null falls within the (1α) confidence interval, then we fail to reject H_0 at α significance level.
 - * If the hypothesized null doesn't fall within the (1α) confidence interval, then we reject H_0 at α significance level.

3 Inference about Two Populations

3.1 About the difference between two population means ($\mu_1 - \mu_2$)



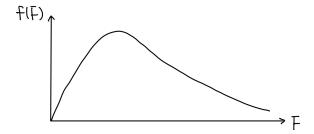
3.2 About two population variances $(\frac{\sigma_1^2}{\sigma_2^2})$

Ho:
$$\frac{{\overline{b_1}}^2}{{\overline{b_2}}^2} = 1$$

Numerator DOF (\mathcal{V}_1)

Test statistic = $\frac{{S_1}^2}{{S_2}^2} \sim \overline{f_{n_1-1}}, n_2-1$

- We are introduced to a new distribution: F distribution
 - F distribution usually looks like the following:



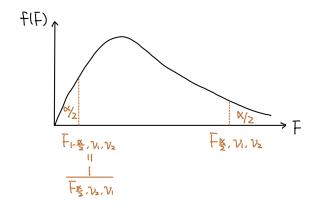
It looks very similar to the Chi-squared distribution. (In fact, F distribution is actually obtained by dividing a Chi-squared distribution by a different Chi-squared distribution.)

- A F distribution is denoted as the following:

$$F_{\nu_1,\nu_2}$$

Here, v_1 is the numerator degree of freedom, and v_2 is the denominator degree of freedom.

• The F distribution table is usually given for a right tail area only – how does one find the appropriate left tail cutoff value, if we are performing a two-tail test?

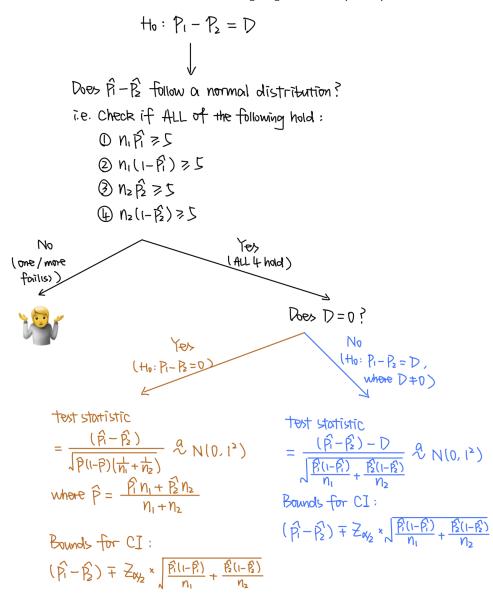


Solution: the left tail cutoff value can be found as right tail cutoff value from a slightly different F distribution (switching the numerator and denominator degree of freedom)

$$F_{1-\frac{\alpha}{2},\nu_1,\nu_2} = \frac{1}{F_{\frac{\alpha}{2},\nu_2,\nu_1}}$$

3

3.3 About the difference between two success proportions $(p_1 - p_2)$



4 Exercises

- 1. You are conducting a poll for a hat company in preparation for the upcoming holiday season. In your samples of 296 men and 282 women, 26 percent of men and 34 percent of women say they plan to purchase a hat.
 - (a) Using a 5% significance level, test whether women are more likely than men to purchase a hat.

(b) Construct and interpret women and men who p	a 95% confidence inter lan to purchase a hat.	val for the difference	e between the prop	ortion of

(c)	If women are more than 5 percentage points more likely to purchase a hat than men, then the store will run a special female targeted promotion. Would you advise them to run the promotion? (Use a 5% significance level.)

2.	Continuing with the previous example, the hat company now wants to get a sense of how many
	hats men and women will purchase. So they commission a new survey, this time asking respondents
	how many hats they plan to purchase. They survey 200 men and 200 women, with men planning to
	purchase an average of 2.2 hats and women purchasing to purchase an average of 1.8 hats in your
	sample. You estimate a sample standard deviation of 2.3 hats for men and 2.0 hats for women.

(a)	Assuming the population variances are unequal, test whether the mean number of planned hat
	purchases are equal for men and women. (Use a 5% significance level.)

(b) Construct and interpret a 95% confidence interval for the difference between the mean planned number of hat purchases for women and men.

(c)	Repeat previous two questions while assuming equal population variance.
(d)	Using a 5% significance level, test whether the population variance of planned hat purchases for men and women are equal.

- 3. Now suppose that the survey in Exercise 2, rather than being a simple random sample of men and a simple random sample of women, the survey had a paired design. In particular, the data collection was paired using the respondents' zip codes. As before, you have data for 200 women and 200 men. Using this data, you compute the paired mean difference and standard deviation, which are: $\bar{X}_D = -0.4$, and $s_D = 2$.
 - (a) Construct a hypothesis test to determine whether these populations have different means. (Use a 5% significance level.)

(b) Construct and interpret a 95% confidence interval for the difference between the mean planned number of hat purchases for women and men.

Probability table for a standard normal distribution (z \geq 0)

TABLE **3** (Continued)

		\								
	0	Z								
	$P(-\infty < Z < z$	z)								
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Probability table for a t-distribution

TABLE **4**Critical Values of the Student *t* Distribution



egrees of				4	
reedom	t _{.100}	t _{.050}	t _{.025}	t _{.010}	t _{.005}
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.343	1.753	2.131	2.602	2.947
				2.583	2.947
16	1.337	1.746	2.120		
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.306	1.690	2.030	2.438	2.724
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
55			2.009		2.668
	1.297	1.673		2.396	
60	1.296	1.671	2.000	2.390	2.660
65	1.295	1.669	1.997	2.385	2.654
70	1.294	1.667	1.994	2.381	2.648
75	1.293	1.665	1.992	2.377	2.643
80	1.292	1.664	1.990	2.374	2.639
85	1.292	1.663	1.988	2.371	2.635
90	1.291	1.662	1.987	2.368	2.632
95	1.291	1.661	1.985	2.366	2.629
100	1.290	1.660	1.984	2.364	2.626
110	1.289	1.659	1.982	2.361	2.621
120	1.289	1.658	1.980	2.358	2.617
130	1.288	1.657	1.978	2.355	2.614
140	1.288	1.656	1.977	2.353	2.611
150	1.287	1.655	1.976	2.351	2.609
160	1.287	1.654	1.975	2.350	2.607
170	1.287	1.654	1.973	2.348	2.607
180	1.286	1.653	1.973	2.347	2.603
190	1.286	1.653	1.973	2.346	2.602
200	1.286	1.653	1.972	2.345	2.601
00	1.282	1.645	1.960	2.326	2.576

Probability table for a F-distribution, with A=0.025

NUMERATOR DEGREES OF FREEDOM 72

8	1018	39.5	13.9	8.26	6.02	4.85	4.14	3.67	3.33	3.08	2.88	2.73	2.60	2.49	2.32	2.25	2.19	2.13	2.09	2.00	1.94	1.88	1.83	1.79	0.70	1.04	1.55	1.48	1.44	1.40	1.37	1.35	1.31	1.28	1.26	1.25	1.23
200	1016	39.5	13.9	8.29	6.05	4.88	4.18	3.70	3.37	3.12	2.92	2.76	2.63	2.55	2.36	2.29	2.23	2.18	2.13	2.05	1.98	1.92	1.88	1.84	1.75	1.69	1.60	1.54	1.50	1.47	1.44	1.42	1.39	1.36	1.35	1.33	1.32
180	1015	39.5	13.9	8.29	6.05	4.89	4.18	3.71	3.37	3.12	2.92	2.77	2.64	2.53	2.36	2.29	2.23	2.18	2.13	2.05	1.99	1.93	1.88	1.84	1.76	1.70	1.61	1.55	1.51	1.47	1.45	1.43	1.40	1.37	1.35	1.34	1.33
160	1015 1	39.5	13.9	8.30	90.9	4.89	4.18	3.71	3.38	3.13	2.93	2.77	2.64	2.54	2.37	2.30	2.24	2.19	2.14	2.06	1.99	1.94	1.89	1.85	/:-	1.70	1.62	1.56	1.52	1.48	1.46	1.44	1.41	1.38	1.36	1.35	1.34
140	1015 1	39.5	13.9	8.30	90.9	4.90	4.19	3.72	3.38	3.13	2.94	2.78	2.65	2.34	2.37	2.31	2.25	2.19	2.15	2.07	2.00	1.94	1.90	1.86	: :	1./1	1.63	1.57	1.53	1.49	1.47	1.45	1.42	1.39	1.38	1.36	1.35
120	1014 10	39.5	13.9	8.31	6.07	4.90	4.20	3.73	3.39	3.14	2.94	2.79	2.66	2.33	2.38	2.32	2.26	2.20	2.16	2.08	2.01	1.95	1.91	1.87	1.79	1.72	1.64	1.58	1.54	1.51	1.48	1.46	1.43	1.41	1.39	1.38	1.37
100	1013 10	39.5	14.0	8.32	90.9	4.92	4.21	3.74	3.40	3.15	2.96	2.80	2.67	2.50	2.40	2.33	2.27	2.22	2.17	2.09	2.02	1.97	1.92	1.88	1.80	1.74	1.66	1.60	1.56	1.53	1.50	1.48	1.45	1.43	1.42	1.40	1.39
06		39.5	14.0	8.33	60.9	4.92	4.22	3.75	3.41	3.16	2.96	2.81	2.68	2.37	2.40	2.34	2.28	2.23	2.18	2.10	2.03	1.98	1.93	1.89	 	1.75	1.67	1.61	1.57	1.54	1.52	1.50	1.47	1.45	1.43	1.42	1.41
80	1013	39.5	14.0 1	8.33	6.10	4.93	4.23	3.76	3.42	3.17	2.97	2.82	2.69	2.30	2.42	2.35	2.29	2.24	2.19		2.05			_		1.76		1.63	1.59	1.55	1.53	1.51	1.48	1.46	1.45	1.43	1.42
20	1 1012	39.5		8.35	6.11	4.94	4.24	3.77	3.43	3.18	2.99	2.83	2.70	2.50	2.43	2.36	2.30	2.25	2.20				1.96	1.92	1.84	1.78	1.70	1.64	1.60	1.57	1.55	1.53	1.50	1.48	1.47	1.46	1.45
09	1101 0	39.5		8.36	6.12	4.96	4.25	3.78	3.45	3.20	3.00	2.85	2.72	2.51	2.45	2.38		2.27									1.72				1.58			1.51	1.50	1.48	1.47
50 60 70 80	3 1010	39.5	14.0 1				4.28	3.81						2.54			2.35									1.83			. 99'1	. 69.1	. 19.1	. 65.1	. 95.	.55	.53	.52	.51
	1008			_			_	3.82 3	_			_		2.65	_				2.27 2								1.77						_	.57	.55 1	.54	.53
45	1007	5 39.5	_		~			_		_				2.67 2		_		2.33 2									1.80					.64			9		.56 1
40	1006	5 39.5	14.0		_								_																	_	_	_		_	-	_	.60 1.
35	1004	39.5	_				6 4.33	9 3.86					2.80				4 2.41									06.1 4		_	_			_			_	_	_
30	1001			~				_	~					2.73													9 1.87					4		_	•	_	9
28	1000		_	8.48																							1.89			_			1.7	_	1.68	_	3 1.66
26	666	39.5	14.1	8.49	6.26	5.10	4.35	3.93	3.59	3.34	3.15	3.00	2.87	2.68	2.60	2.54	2.48	2.43	2.35	2.31	2.25	2.19	2.15	2.11	2.04	1.98	1.91	1.86	1.82	1.79	1.77	1.76	1.73	1.72	1.70	1.69	1.68
24	266	39.5	14.1	8.51	6.28	5.12	4.41	3.95	3.61	3.37	3.17	3.02	2.89	2.79	2.63	2.56	2.50	2.45	2.41	2.33	2.27	2.22	2.17	2.14	2.06	10.7	1.93	1.88	1.85	1.82	1.80	1.78	1.76	1.74	1.73	1.72	1.71
22	995	39.5	14.1	8.53	6.30	5.14	4.44	3.97	3.64	3.39	3.20	3.04	2.92	2.01	2.65	2.59	2.53	2.48	2.43	2.36	2.30	2.24	2.20	2.16	2.09	2.03	1.96	1.91	1.88	1.85	1.83	1.81	1.79	1.77	1.76	1.75	1.74
7_	-	2	3	4	2	9	7	8	6	10	=	12	13	4 r	16	17	18	19	20	22	24	56	28	30	35	45	50	09	20	80	90	100	120	140	160	180	200