This can also be solved by χ^2 test ↓ (on last page)

STAT 250 Practice Problems Chapters 12, 13, 14

1. Use the data below to test at the 5% level whether the likelihood of developing breast cancer is different for the hormone group than the placebo group. $\alpha = 0.05$

$\neq$

	Yes	No	Total
Hormone	166	8340	8506
Placebo	124	7978	8102
Total	290	16318	16608

$H_0: P_{hormone} = P_{placebo}$

$H_a: P_{hormone} \neq P_{placebo}$

$\hat{P}_{hormone} = \frac{166}{8506} = 0.0195$

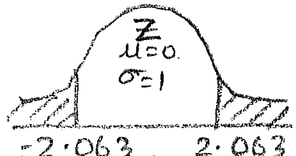
$\hat{P}_{placebo} = \frac{124}{8102} = 0.0153$

$\hat{P}_c$ (assume $P_h = P_p$)

$= \frac{290}{16608} = 0.0175$

$$Z = \frac{(\hat{P}_h - \hat{P}_p) - 0}{\sqrt{\hat{P}_c(1-\hat{P}_c)\left(\frac{1}{n_h} + \frac{1}{n_p}\right)}} = \frac{(0.0195 - 0.0153) - 0}{\sqrt{0.0175(0.9825)\left(\frac{1}{8506} + \frac{1}{8102}\right)}}$$

$Z = 2.063$



p-value $= 2(0.019)$
$= 0.038 < \alpha$

Reject H_0 and conclude likelihood for developing breast cancer is different

For two groups

2. Use the data and output below to test at the 5% level whether Religious preference and opinion on premarital sex are associated. If the variables are independent, what is the expected cell count of Protestants who say premarital sex is never wrong?

H_0: religious preference and opinion are independent

H_a: religious preference and opinion are dependent

observed

Religion	Always	Almost Always	Sometimes	Never	Total
Protestant	221	54	98	288	661
Catholic	45	17	54	179	295
Jewish	2	1	8	18	29
None	15	10	32	164	221
Other	20	7	12	41	80
Total	303	89	204	690	1286

Expected for never & protested

$= \frac{(661)(690)}{1286}$

$= 354.66$

$$\text{expected value} = \frac{(\text{row total})(\text{column total})}{(\text{grand total})}$$

Chi-Square Test			
SUMMARY		Alpha	0.05
Count	*Rows*	*Cols*	*df*
1286	5	4	12
CHI-SQUARE			
	chi-sq	*p-value*	*x-crit*
Pearson's	108.1346	1.4E-17	21.02607
Max likeli	116.6153	2.92E-19	21.02607

$\chi^2_{df=12} = 108.1346$

p-value $= 0$

Reject H_0

Religious preference and opinion on premarital sex are associated

$H_0: \mu_d - \mu_e = 0$

3. In a study comparing diet versus exercise, the 42 men on a diet lost an average of 7.2 kg with a standard deviation of 3.7 kg. The 47 men on an exercise regimen lost an average of 4.0kg with a standard deviation of 3.9kg. Conduct a hypothesis test at the .05 level to determine if there is a difference in mean weight loss for the two groups.

$H_a: \mu_d - \mu_e \neq 0$ ← two sided

Assumptions : $n_1 \geq 30$, $n_2 \geq 30$ and random mess.

(diet) Group 1 — Group 2 (exercise)

$\bar{x}_1 = 7.2$ — $\bar{x}_2 = 4$

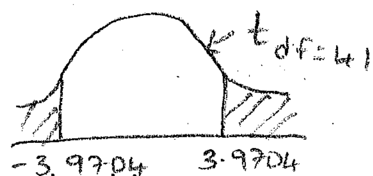
$S_1 = 3.7$ — $S_2 = 3.9$

$n_1 = 42$ — $n_2 = 47$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(7.2-4)-0}{\sqrt{\frac{3.7^2}{42} + \frac{3.9^2}{47}}}$$

$df = \min(42, 47) - 1$
$= 42 - 1$
$= 41$

$t = 3.9704$



p-value $= 0.000282 < \alpha$, reject H_0

4. The forearm lengths for 9 men are given below. Is there evidence that mean forearm length is longer than the 25cm guideline used by a garment manufacturer? Test at the 5% level. 25.5, 24, 26.5, 25.5, 28, 27, 23, 25, 25. The sample average is 25.5 and the sample standard deviation is 1.5207.

population standard deviation (σ) unknown hence use t test

$\bar{x} = 25.5$ $\quad \sum x^2 = 5870.75$

$n = 9$

$$S = \sqrt{\frac{1}{n-1}\left[\sum x^2 - n\bar{x}^2\right]} =$$

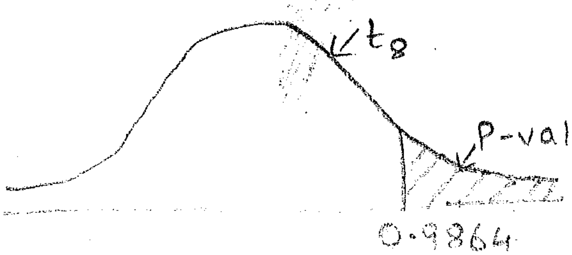
$$= \sqrt{\frac{1}{8}\left[5870.75 - (9 * 25.5)\right]}$$

$= 1.5207$

$df = n-1$
$= 9-1$
$= 8$

$H_0: \mu = 25$
$H_a: \mu > 25$ ← one sided

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{25.5 - 25}{1.5207/\sqrt{9}} = 0.9864.$$



P-value $= 0.1764 > \alpha$, Fail to reject H_0

No evidence to conclude mean forearm length is longer than 25 cm

5. A test of manual dexterity was given to students to determine if they perform better with their dominant hand than with their non-dominant hand. The data are presented below. Conduct a test at the 5% level. $\mu_{difference} = \mu_{dominant} - \mu_{non\ dominant}$

paired t test

both observations come from same source

Participant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Mean	sd
Dominant	22	19	18	17	15	16	16	20	17	15	17	17	14	20	26	17.93	3.10
Nondominant	18	15	13	16	17	16	14	16	20	15	17	17	16	18	25	16.87	2.83
Difference	4	4	5	1	-2	0	2	4	-3	0	0	0	-2	2	1	1.07	2.43

$\leftarrow s_{difference}$

$\hat{x}_{difference}$

$\bar{x}_d = 1.07$

$s_d = 2.43$

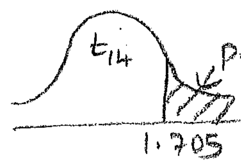
$n = 15$

$H_0: \mu_{difference} = 0$

$H_a: \mu_{difference} > 0$

$df = 15 - 1 = 14$

$$t = \frac{\bar{x}_d - 0}{s_d / \sqrt{n}} = \frac{1.07 - 0}{2.43/\sqrt{15}} = 1.705$$



p-value = 0.055 $> \alpha$

FTR H_0

No evidence to conclude students perform better with dominant hand.

6. The population of eligible voters in a region is 45.1% non-Hispanic white, 6.7% Black or African American, 28.8% Hispanic, 15.9% Asian, and 3.5% other. A sample of actual voters had the following racial/ethnic breakdown. Does the distribution of race/ethnicity for actual voters match that of eligible voters? Conduct a test at the .05 level.

$\chi^2 = \frac{(O-E)^2}{E}$

5 categories

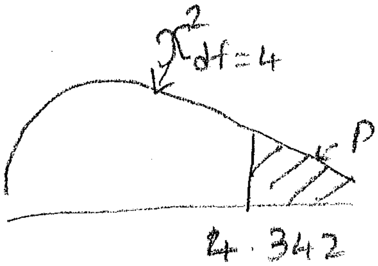
	obs	expected	
White	93	0.451(178) = 80.278	$\frac{(93 - 80.278)^2}{80.278} = 2.016$
Black or African American	13	0.67(178) = 11.926	0.097
Hispanic	43	0.288(178) = 51.264	1.332
Asian	24	0.159(178) = 28.302	0.654
Other	5	0.035(178) = 6.23	0.243
Total	178		$\chi^2 = 4.342$

$df = 5 - 1 = 4$

H_0: distribution of actual voters match that of eligible voters

vs

H_a: distribution of actual voters do not match that of eligible voters



P-value = 0.3617 $> \alpha$ Fail to Reject H_0

No evidence to conclude that distribution of race for actual voters do not match that of eligible voters.

Q1] Chi Squared (χ^2) test

$$\text{Expected} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

Observed

	Y	N	T
H	166	8340	8506
P	124	7978	8102
T	290	16318	16608

	Y	N	T
H	148.53	8357.47	8506
P	141.47	7960.53	8102
T	290	16318	16608

H_0: variables are independent

H_a: variables are dependent

$$\chi^2 = \frac{(O-E)^2}{E}$$

	Y	N
H	2.05	0.04
P	2.16	0.04

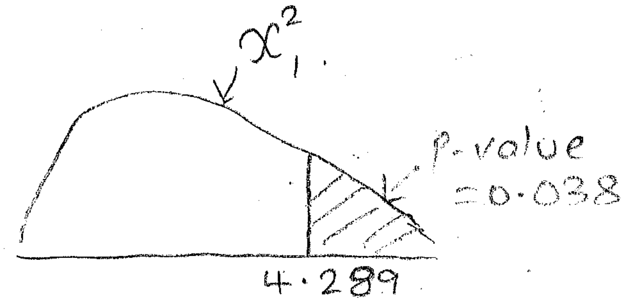
$$0.04 \rightarrow \frac{(8340 - 8357.47)^2}{8357.47}$$

$\chi^2 = 4.29$

$df = (\#\text{rows} - 1)(\#\text{columns} - 1)$

$= (2-1)(2-1)$

$= 1$



pvalue < α, reject H_0.

Evidence to conclude variables are dependent.