(8 points)

5. **A study included 200 men and 280 women to see whether people differ by gender in their sensitivity to smells. Use the data to test to claim that the proportion of sensitive women is greater than the proportion of sensitive men. Do not analyze this as a contingency table.**
(Use $\alpha = 0.05$ for this test.)

Gender	Men	Women
Sensitive	20	42
Insensitive	180	238
Total	200	280

claim: $p_w > p_m$

$(p_w - p_m) > 0$

$H_0: (p_w - p_m) \le 0$

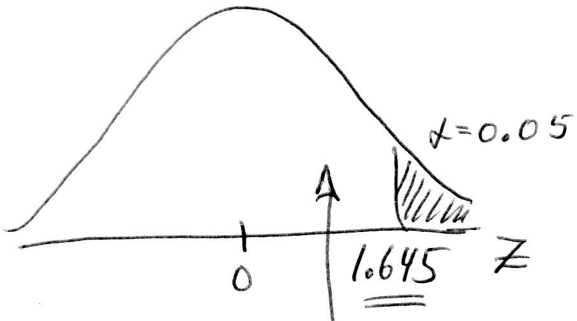
$H_1: (p_w - p_m) > 0$

$\alpha = 0.05$ right tail

$\hat{p}_m = \frac{20}{200} = 0.10$ $\qquad$ $\hat{p}_w = \frac{42}{280} = 0.15$

$\bar{p} = \frac{20 + 42}{200 + 280} = 0.1292$

$\bar{q} = 0.8708$



Test Statistic (with $\bar{p}$)

$$\frac{(\hat{p}_w - \hat{p}_m) - 0}{\sqrt{\frac{\bar{p}\bar{q}}{n_w} + \frac{\bar{p}\bar{q}}{n_m}}}$$

Do Not reject Ho!

$$= \frac{(0.15 - 0.10) - 0}{\sqrt{\frac{(0.1292)(0.8708)}{280} + \frac{(0.1292)(0.8708)}{200}}} = \frac{0.05}{0.0311} = \boxed{1.61}$$

(8 points)

5. **A study included 200 men and 280 women to see whether people differ by gender in their sensitivity to smells. Use the data to test to claim that the proportion of sensitive women is 0.02 greater than the proportion of sensitive men. Do not do this as a contingency table.**
(Use α = 0.05 for this test.)

Gender	Men	Women
Sensitive	20	42
Insensitive	180	238
Total	200	280

$\hat{p}_m = 20/200 = 0.10$ | $\hat{p}_w = \frac{42}{238} = 0.15$

$\hat{q}_m = 0.90$ | $\hat{q}_w = 0.85$

$n_m = 200$ | $n_w = 280$

claim: $p_w = p_m + 0.02$

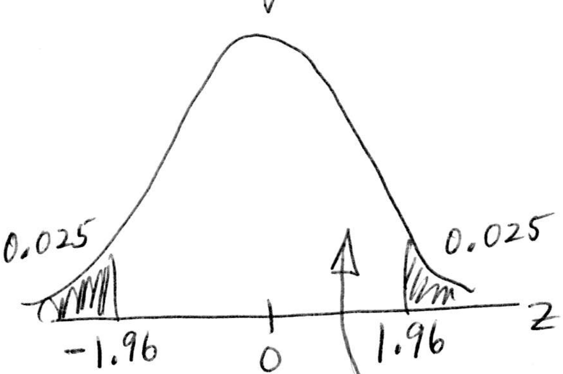
$(p_w - p_m) = 0.02$

$H_0 : (p_w - p_m) = 0.02$

$H_1 : (p_w - p_m) \neq 0.02$

$\alpha = 0.05$ 2 tails

critical region



$\boxed{\text{Do Not reject } H_0\text{:}}$

test statistic (Not using $\bar{p}$)

$$\frac{(\hat{p}_w - \hat{p}_m) - (p_w - p_m)_0}{\sqrt{\frac{\hat{p}_w \hat{q}_w}{n_w} + \frac{\hat{p}_m \hat{q}_m}{n_m}}}$$

$$= \frac{(0.15 - 0.10) - 0.02}{\sqrt{\frac{(0.15)(0.85)}{280} + \frac{(0.10)(0.90)}{200}}} = \frac{0.03}{0.030} = \boxed{1.00}$$

5

(8 points; 8 minutes)

1. **A random sample of incoming freshmen was selected. Of the men, 22 chose sociology as a major and 191 did not. Of the women, 38 chose sociology and 219 did not. Use these results to prepare a 95% confidence interval for the difference in the proportions of all incoming men and women that will choose sociology as their major. Then use your results to answer this question:**

Is it reasonable to claim that women and men choose Sociology in the same proportion?

YES NO **Why?** If women and men choose sociology in the same proportion, then the difference $(p_m - p_w) = 0$, which is in the reasonable range.

$$95\%\ CI(p_m - p_w) = (\hat{p}_m - \hat{p}_w) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_m \hat{q}_m}{n_m} + \frac{\hat{p}_w \hat{q}_w}{n_w}}$$

$$\hat{p}_m = \frac{22}{(22+191)} = \frac{22}{213}$$

$$= 0.1033$$

$$\hat{q}_m = 0.8967$$

$$n_m = 213$$

$$\hat{p}_w = \frac{38}{257} = 0.1479$$

$$\hat{q}_w = 0.8521$$

$$n_w = 257$$

$$\alpha = 1 - \text{confid}$$

$$= 1 - 0.95$$

$$= 0.05$$

2 tails.

$$Z_{\alpha/2} = 1.96$$

$$= (0.1033 - 0.1479) \pm 1.96\sqrt{\frac{(.1033)(.8967)}{213} + \frac{(.1479)(.8521)}{257}}$$

$$= (-0.0446) \pm 0.0596$$

$$= [-0.1042 < (p_m - p_w) < 0.015]$$

The confidence interval is the "reasonable range for the true difference $(p_m - p_w)$.

6

(8 points - 10 minutes)

1. **The data at the right represent two different acid treatments in an indistrial process. Use the information to construct a 98% confidence interval for $\mu_1 - \mu_2$. (Assume that $\sigma_1^2 = \sigma_2^2$)**

Treat.	Mean	Std. Dev.	N
Acid 1	11.620	1.185	10
Acid 2	10.865	1.117	14

Acid 1	Acid 2
11.3	10.3
10.9	11.1
11.0	11.0
11.2	10.0
11.2	9.5
10.9	12.4
13.0	10.6
11.0	11.8
14.5	8.9
11.2	11.9
	10.4
	12.4
	9.8
	12.0

since ($\sigma_1^2 = \sigma_2^2$), do 2 things:
① pool the variances $\Rightarrow S_p^2$
② add the degrees of freedom

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

$$= \frac{(10-1)(1.185)^2 + (14-1)(1.117)^2}{(10-1)+(14-1)}$$

$$= 1.312$$

$$df = (10-1)+(14-1) = 22$$

$\alpha = 1 - \text{confidence}$
$= 1 - 0.98 = 0.02$
in 2 tails

$t_{\alpha/2} = 2.508$
22 d.f.

$$98\%\ CI(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

$$= (11.620 - 10.865) \pm 2.508\sqrt{\frac{1.312}{10} + \frac{1.312}{14}}$$

$$= 0.755 \pm 1.189$$

$$= [-0.434 < \mu_1 - \mu_2 < 1.944]$$

7

Statistics 1
Istructor: L. Larsen   Name: ______________________   Spring Semester 2002
Tu/Th 7:00 - 8:50 p.m.

(7 points; 8 minutes)

11. A manufacturer of ceramic tiles compares two different curing temperatures to see which one makes the tiles stronger. Use the test data below to construct a 95% confidence interval for the difference between the population mean for 800 degrees and the population mean for 1200 degrees. Experts advise you that the temperature difference is likely to cause a difference in the variability of the strength of the bricks.

→ so $\sigma^2_{800} \neq \sigma^2_{1200}$

(1) do not pool variances

(2) use smaller of the 2 deg. of freedom

Tile Strength At Temperature	
800	1200
85	105
109	105
82	103
89	100
106	102
118	102
99	107
104	113
	100

$\bar{x}$ =	99.000	104.111
s =	12.649	4.076
n =	8	9

$df = \boxed{7}$ 8

→ smaller d.f.

$\alpha = 1 - 0.95$
$= 0.05$ in 2 tails

$t_{\alpha/2} = 2.365$

$$95\%\ CI(\mu_{1200} - \mu_{800}) =$$

$$(\bar{X}_{1200} - \bar{X}_{800}) \pm t_{\alpha/2}\sqrt{\frac{S^2_{1200}}{n_{1200}} + \frac{S^2_{800}}{n_{800}}}$$

$$= (104.111 - 99.000) \pm 2.365\sqrt{\frac{(4.076)^2}{9} + \frac{(12.649}{8}}$$

$$= 5.111 \pm 11.05$$

$$= [-5.94 < (\mu_{1200} - \mu_{800}) < 16.16]$$

(8 points; 10 minutes)

6. Use the data below to construct a 90% confidence interval for the difference between the two population means from which the random samples were selected. Previous experiments have shown that the variability in Vitamin D levels is similar for both men and women.

$\hookrightarrow \sigma_m^2 = \sigma_W^2$, so:

① pool variances

② add deg. of freedom

Vitamin D in Blood		
	Women	Men
	34.8	60.7
	31.3	43.6
	38.6	41.1
	46.7	39.8
	41.6	49.9
	47.7	36.1
	44.3	48.4
	38.9	33.8
	51.6	50.6
	26.9	25.7
	42.3	63.4
	27.7	32.8
		27.5
		45.0
		45.8
		60.2
Sample mean =	39.37	44.03
Sample stand. dev. =	7.58	11.03
Sample count =	12	16

$$S_p^2 = \frac{df_W \cdot S_W^2 + df_M \cdot S_M^2}{df_W + df_M}$$

$$= \frac{(11)(7.58)^2 + (15)(11.03)^2}{11 + 15} = 94.5$$

$df = 11 + 15 = 26$

$\alpha = 1 - 0.90 = 0.10$ in 2 tails

$t_{\alpha/2} = 1.706$

$$90\%\, CI(\mu_M - \mu_W) = (\bar{x}_m - \bar{x}_W) \pm \cancel{1.706}\, t_{\alpha/2} \sqrt{\frac{S_p^2}{n_M} + \frac{S_p^2}{n_W}}$$

$$= (44.03 - 39.37) \pm 1.706 \sqrt{\frac{94.5}{16} + \frac{94.5}{12}}$$

$$= 4.66 \pm 6.33$$

$$= \boxed{[-1.67 < (\mu_M - \mu_W) < 10.99]}$$

Based on your results, is it reasonable to claim that men in general have more vitamin D in their blood than women do? For credit, you must explain why?

Yes No Why? If men have more vitamin D in their blood than do women, then $(\mu_m - \mu_w)$ will be positive, and positive values are included in the CI, which is the reasonable range.

9

(8 points - 10 minutes)

2. Use the information given here to test the claim that hot treatments require one more day for recovery on average than cold treatments. Use a significance level of 0.05 for this test. (Assume that σ_1^2 is not equal to σ_2^2)

	Recovery Time (Days) Treatment	
	Cold	Hot
	19	20
	19	16
	10	7
	11	10
	10	7
	8	8
	16	12
	11	11
	11	
	19	
Mean	13.4	11.4
Std. Dev.	4.35	4.60
n	10	8

df = 9 7

$\sigma_1^2 \neq \sigma_2^2$, so

① do not pool variances

② use the smaller of the two deg. of freedom

claim: $\mu_H = \mu_c + 1$

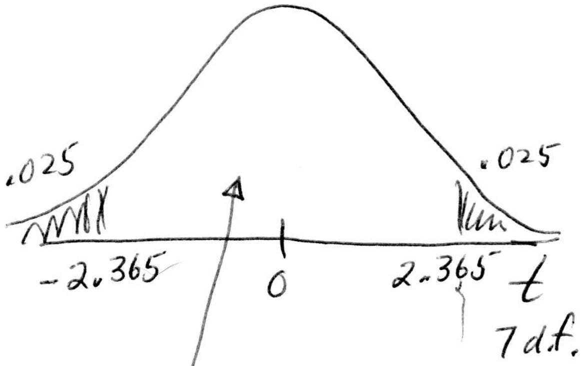
$(\mu_H - \mu_c) = 1$

$H_0: (\mu_H - \mu_c) = 1$

$H_1: (\mu_H - \mu_c) \neq 1$

$\alpha = 0.05$ in 2 tails

critical region



Test Statistic

$$\frac{(\bar{x}_H - \bar{x}_c) - (\mu_H - \mu_c)_0}{\sqrt{\frac{s_H^2}{n_H} + \frac{s_c^2}{n_c}}}$$

$$= \frac{(11.4 - 13.4) - 1}{\sqrt{\frac{(4.60)^2}{8} + \frac{(4.35)^2}{10}}} = \frac{-3}{2.13} = \boxed{-1.408}$$

$\boxed{\text{Do Not reject } H_0}$

(8 points : 12 minutres)

1. **Do carpool lanes save commute time?**
Transportation planners want to know whether carpool lanes save time for commuters. An experiment was carried out using 14 commuters who do not carpool and 11 commuters who do carpool. Use the summary of the survey results below to test the claim that carpool use does NOT save time, so the mean commute time for carpoolers is greater than or equal to the mean time for those who do not carpool. (Use α = 0.01 for the test and assume that variation is similar for both populations.)

	Time in Minutes	
	Regular Lanes	Carpool Lane
$\bar{x}$ =	44.8	41.0
s =	15.1	13.6
n =	14	11

$df = 13 + 10 = 23$

$t = -2.500$

$\sigma_R^2 = \sigma_C^2$, so

① pool variances

② add degrees of freedom

$$S_p^2 = \frac{(n_R-1)S_R^2 + (n_C-1)S_C^2}{(n_R-1)+(n_C-1)}$$

$$= \frac{(14-1)(15.1)^2 + (11-1)(13.6)^2}{(14-1)+(11-1)}$$

$$= 209.3$$

claim: $\mu_c \geq \mu_R$: $(\mu_c - \mu_R) \geq 0$

$H_0: (\mu_c - \mu_R) \geq 0$

$H_1: (\mu_c - \mu_R) < 0$

$\alpha = 0.01$ left tail 23 d.f.

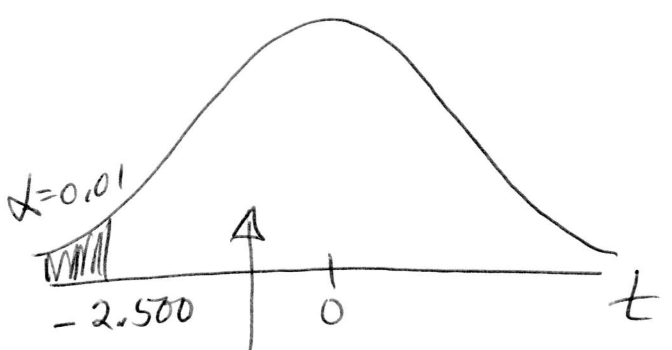
Test Statistic

$$\frac{(\bar{x}_c - \bar{x}_R) - (\mu_c - \mu_R)_0}{\sqrt{\frac{S_p^2}{n_c} + \frac{S_p^2}{n_R}}}$$

$$= \frac{(41.0 - 44.8) - 0}{\sqrt{\frac{209.3}{11} + \frac{209.3}{14}}}$$

$$= \frac{-3.8}{5.83} = \boxed{-0.652}$$

critical region



Do not reject H_0!

11

(6 points - 8 minutes)

5. Use the information given here to test the hypothesis that the two samples (X and Y) come from "populations" for which σ_x^2 is greater than σ_y^2. test: $\sigma_x^2 > \sigma_y^2$

(Use α = 0.025)

H₀: $\sigma_x^2 \le \sigma_y^2$

H₁: $\sigma_x^2 > \sigma_y^2$

$\alpha = 0.025$, all in the tail.

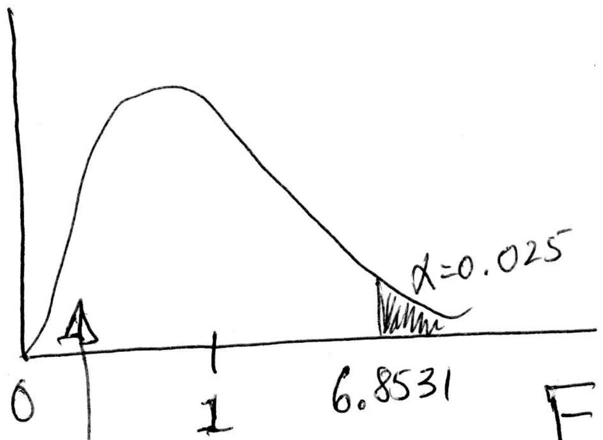
	Sample X	Sample Y
	103.3	43.3
	105.0	17.4
	101.5	21.2
	114.2	58.7
	102.4	37.7
	103.1	40.1
	94.2	
	114.7	
Mean	104.8	36.4
Std. Dev.	6.79	15.15
n	8	6
df	7	5

$$F = \frac{S_x^2}{S_y^2}$$

since H₁: says σ_x^2 is larger.

$$= \frac{(6.79)^2}{(15.15)^2}$$

$$= 0.2$$



7 d.f. Numer.

5 d.f. denom.

Do Not reject H₀.

(7 points; 8 minutes)

7. A produce manager at a super market wants to compare the uniformity of grapes from two suppliers. The manager takes a random sample of grapes from each supplier and tests the grapes for their sugar content. Use the results shown below to test the claim that grapes from supplier A are more uniform (less variable) sugar content than grapes than grapes from supplier B.

(Use a 0.05 significance level.)

Claim: $\sigma_A^2 < \sigma_B^2$

H0: $\sigma_A^2 \geq \sigma_B^2$

H1: $\sigma_A^2 < \sigma_B^2$

S_A^2 in denominator

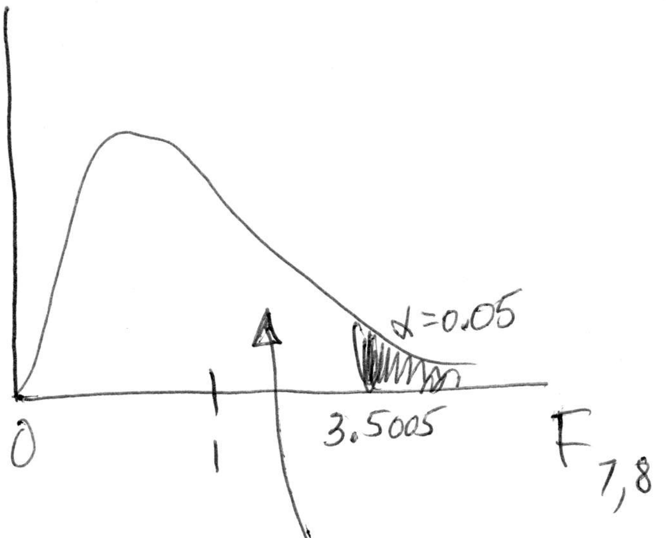
S_B^2 in numerator

α all in the one tail

Supplier	
A	**B**
6.92	3.71
3.75	3.94
8.95	2.45
3.91	2.66
8.35	3.96
1.63	9.49
4.99	6.38
9.04	5.78
	8.86

	A	B
$\bar{x}$ =	5.943	5.248
s =	2.773	2.571
n =	8	9
df =	7 (Numer)	8 (denom)

$$\frac{S_A^2}{S_B^2} = F = \frac{(2.773)^2}{(2.571)^2} = 1.163$$



Do not reject Ho:

(8 points - 8 minutes)

8. **For each of the correlation coefficients given below, enter the letter of a graph from the following pages that best corresponds to the given correlation value.**

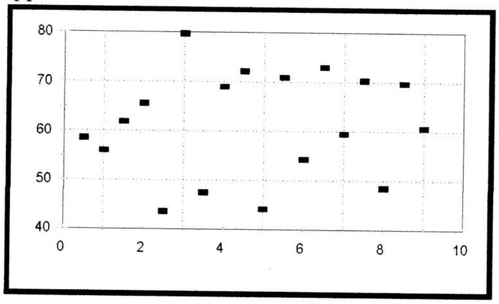
There may be NONE or, perhaps, MORE THAN ONE GRAPH for any correlation.

Correlation Coefficient	Corresponding Graph
$\rho = 1.0$	D and I
$\rho = -0.70$	E
$\rho = 0.90$	C
$\rho = 0.70$	F
$\rho = -0.50$	H

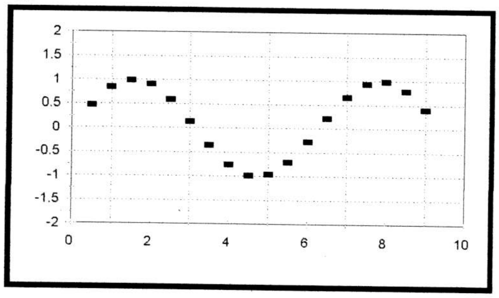
Correlation Coefficient	Corresponding Graph	
$\rho = -0.90$	J	
$\rho = 0.50$	G	
$\rho = 0.00$	A and B	
$\rho = 5.32$		NONE
$\rho = -1.0$		None

A

r=0



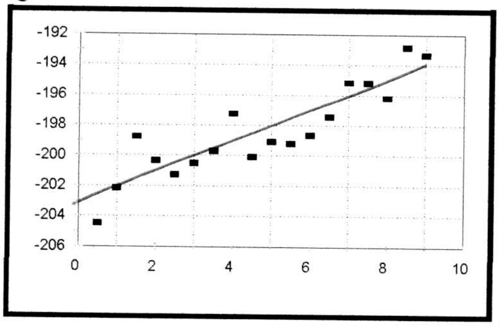
B



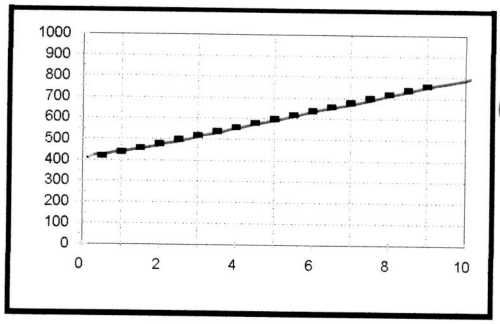
r=0

C

+



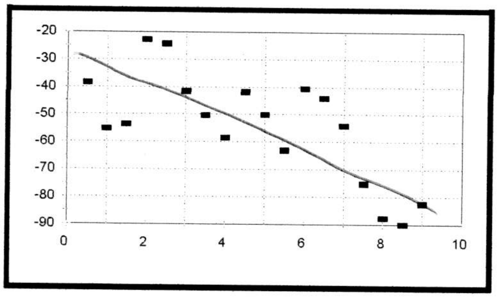
D



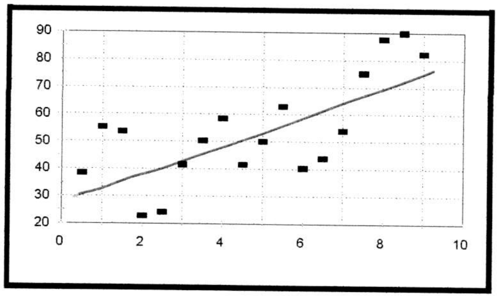
r= +1

E

−



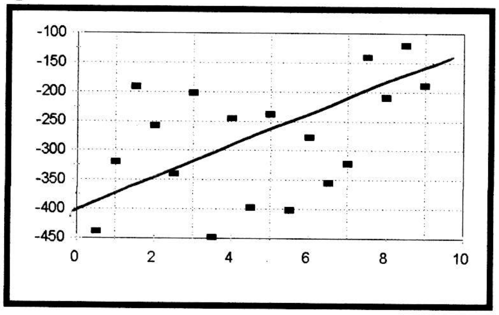
F



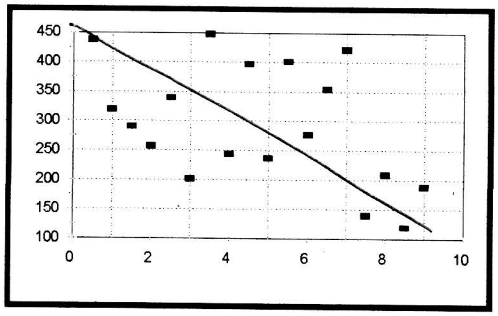
+

G

+


-100
-150
-200
-250
-300
-350
-400
-450
0
2
4
6
8
10

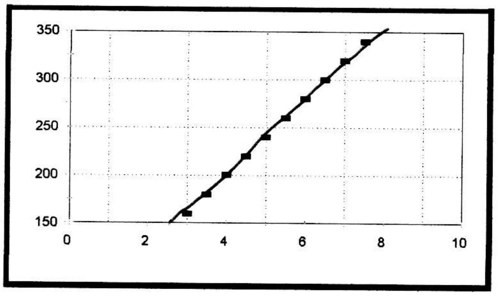
H


450
400
350
300
250
200
150
100
0
2
4
6
8
10

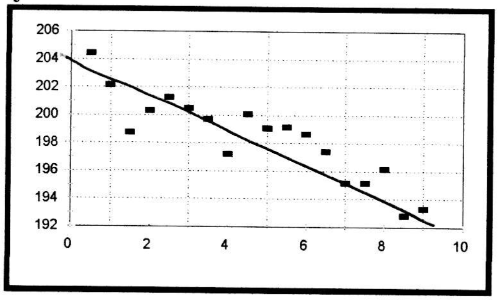
−

I

r=+1


350
300
250
200
150
0
2
4
6
8
10

J


206
204
202
200
198
196
194
192
0
2
4
6
8
10

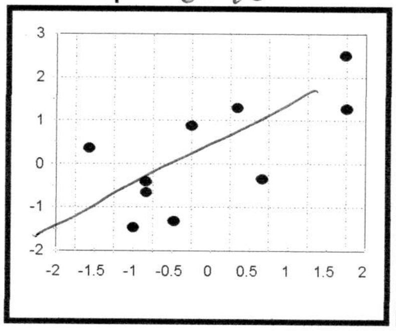
−

(3 points, 5 minutes)

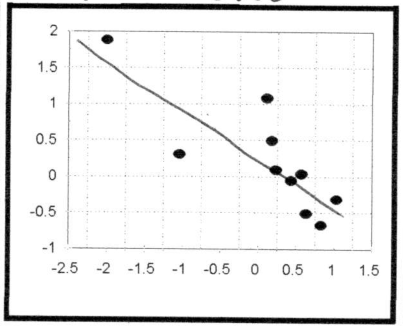
8. **Assign one of the following correlation coefficients to each of the graphs to the right. (or state r = none if no correlation coefficient seems appropriate for a graph)**

r
0.00
-0.80
0.70

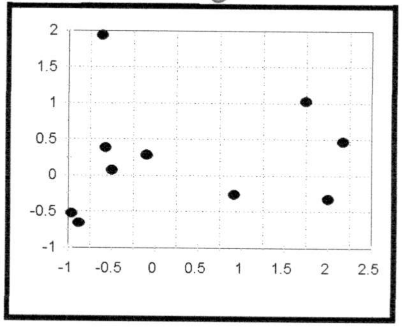
r = 0.70



r = -0.80



r = 0.00

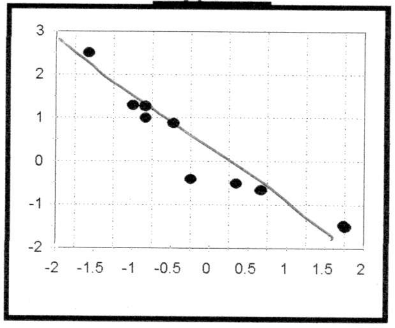


(3 points, 5 minutes)

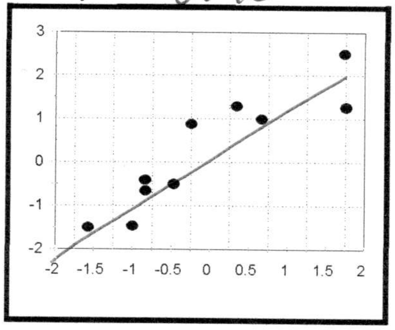
9. **Assign one of the following correlation coefficients to each of the graphs to the right. (or state r = none if no correlation coefficient seems appropriate for a graph)**

r
1.30
0.90
-1.10

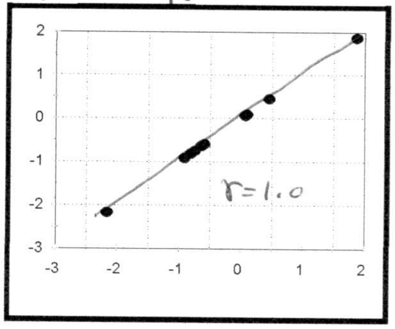
r = none



r = 0.90



r = None

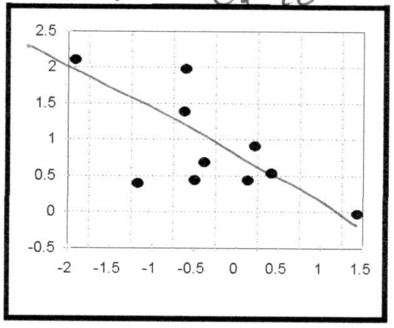


(3 points, 5 minutes)

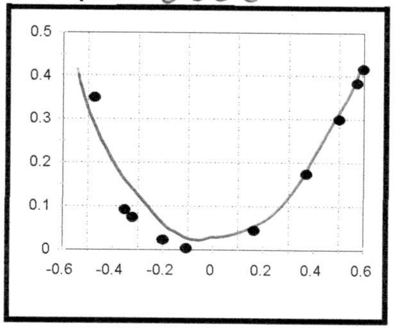
10. **Assign one of the following correlation coefficients to each of the graphs to the right. (or state r = none if no correlation coefficient seems appropriate for a graph)**

r
-0.70
0.00
1.00

r = -0.70

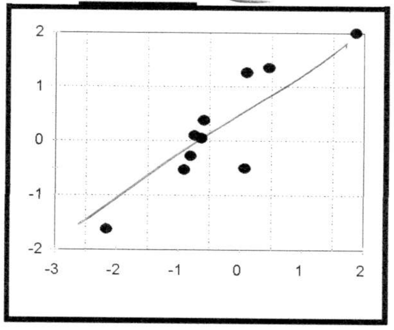


r = 0.00



ups cancel downs

r = none



Not = 1.0

16

(8 points)

3. **Use the data provided below to test the claim that the population correlation (ρ) between vehicle speed (X) and miles per gallon (Y) is less than zero. Assume the data are randomly selected.** (Use $\alpha = 0.01$ and do not use Table A.6)

Observation	Speed	Miles per Gallon
1	28	32
2	35	27
3	50	28
4	75	25

claim: $\rho < 0$

$H_0: \rho \geq 0$

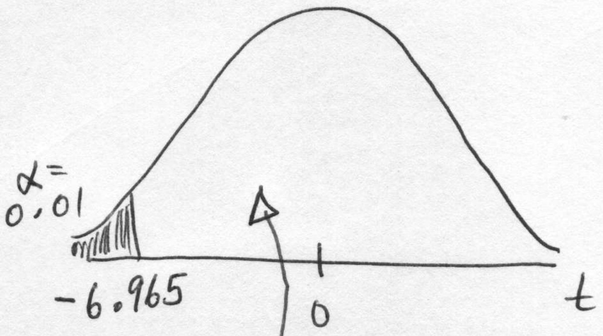
$H_1: \rho < 0$

$\alpha = 0.01$ left tail

$n = 4 \quad df = n - 2 = 2$

$t = -6.965$

$r = -0.8056$



$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$= \frac{-0.8056}{\sqrt{\frac{1-(-0.8056)^2}{4-2}}}$$

$$= \frac{-0.8056}{0.4189} = -1.923$$

Do Not reject H_0!

(7 points; 8 minutes)

2. Market research concerning spending patterns found a sample correlation of 0.73 between X=purchase price of house and Y=purchase price of automobile for a sample of 10 families. Use these results to test the claim that the prices paid for houses and cars are positively correlated for the population of all families.
(Use a 0.10 significance level for this test.)

$n = 10 \quad df = n - 2 = 8$

$r = 0.73$

Claim: $\rho > 0$

H₀: $\rho \le 0$

H₁: $\rho > 0$

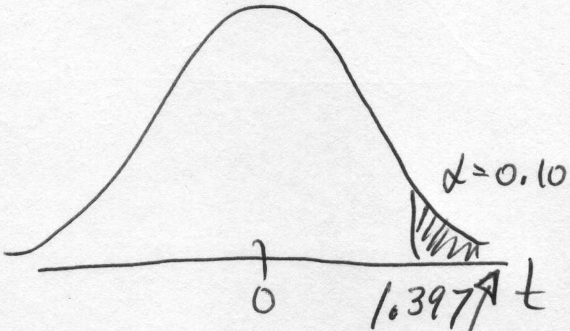
$\alpha = 0.10$ right tail

Test Statistic

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$= \frac{0.73}{\sqrt{\frac{1-(0.73)^2}{10-2}}}$$

$$= \frac{0.73}{0.2416} = \boxed{3.022}$$

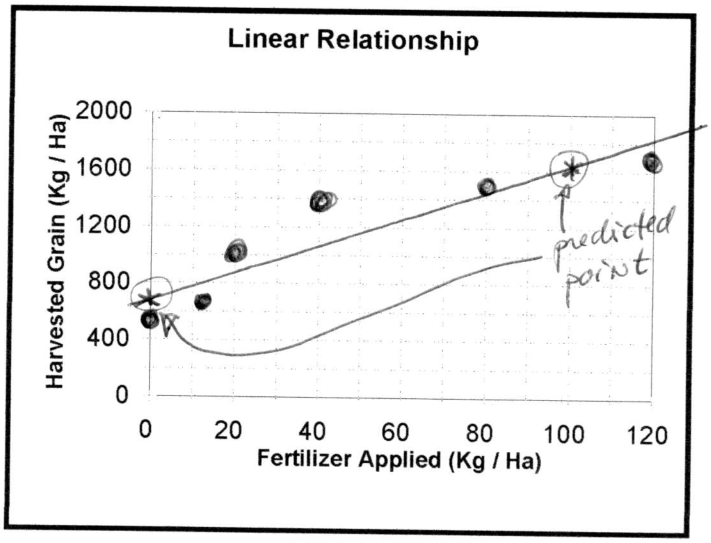


Reject Ho:

(10 points - 15 minutes)

10. Use the data given below to answer questions (a) through (i).

Test Area	(X) Fertilizer Applied (Kg / Ha)	(Y) Harvested Grain (Kg / Ha)
1	0	540
2	10	610
3	20	1018
4	40	1420
5	80	1548
6	120	1731



(a) Plot the data on the coordinate axes. ✓

(b) Determine the equation of the regression line and write ~~in~~ it in the space to the right: ✓

Equation for line: $\hat{y} = 701.6 + 9.84X$

(c) If a farmer used 30 Kg of fertilizer per hectare, how much grain should be expected? 997

$= 701.6 + 9.84(30) = 997$

(d) What is the linear correlation between fertilizer applied and grain harvested? 0.913 $= r$

(e) What is the value of the total variation in Y, the amounts of grain harvested? 1249807.5

$= S_y^2(n-1)$

(f) What fraction of the total variation in Y is explained by the regression line? 0.833 $= r^2$

(g) What is the value of the explained variation in Y? 1041089.6 $= r^2(\text{total})$

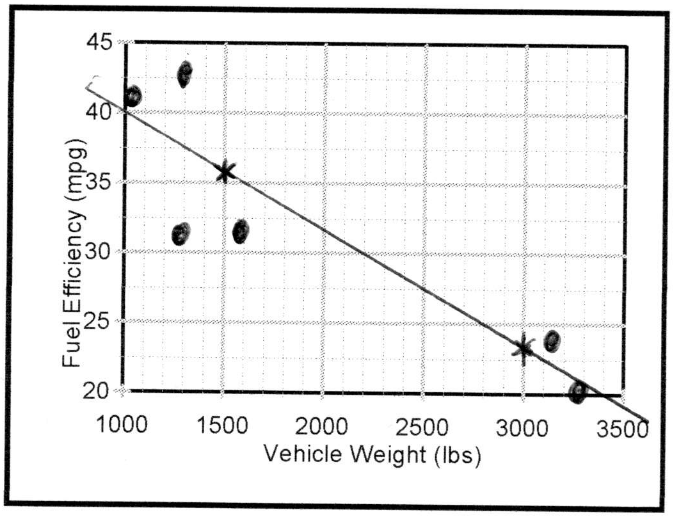
(19)

(14 points - 15 minutes)

4. Use the data shown for vehicle weight and fuel efficiency (miles per gallon of fuel) to answer the following questions.

Data:

Vehicle	Fuel Efficiency (mpg)	Weight (lbs)
1	41.7	1009
2	20.1	3275
3	24.0	3148
4	31.4	1269
5	32.9	1578
6	43.4	1206



(a) Plot the points on the graph ✓

(b) What is the equation for the straight line that best predicts fuel efficiency (y) given vehicle weight (x)? $\hat{y} = 47.94 - 0.0082(x)$

(c) Plot the line on the graph. ✓

(d) Based on ~~you~~ your results, what is the predicted fuel efficiency for a vehicle that weighs 2500 lbs? $(\hat{y} | x = 2500) = 27.45$

$\hat{y} = 47.94 - 0.0082(2500)$

(e) What proportion of the total variation in Y does your line explain? $0.8158 = r^2$

(f) For the total variation in fuel efficiency (Y):

The expression is: $\Sigma(y - \bar{y})^2$ The value is: $430.455 = s_y^2(n-1)$

(g) For the explained variation in fuel efficiency (Y):

The expression is: $\Sigma(\hat{y} - \bar{y})^2$ The value is: $351.184 = (\text{total})(r^2)$

(h) For the unexplained variation in fuel efficiency (Y):

The expression is: $\Sigma(y - \hat{y})^2$ The value is: 79.271 = total − explained

(i) For the Standard error of estimate:

The expression is: $\sqrt{\dfrac{\Sigma(y-\hat{y})^2}{n-2}}$ The value is: 4.452

$$\sqrt{\frac{79.271}{6-2}} = 4.452$$

21

(8 points - 10 minutes)

4. A chain of shoe stores compares the sales for different shoe styles at their NEW California stores to the proportions during the last ten years at their New York stores. Use the data for 400 randomly selected sales in California to test the claim than California consumers buy the same styles of shoes in the same proportions as people in New York. (Use a Type I error rate of 0.05 for this test)

Shoe Style	New York Rate	Observed Counts in Calif.	Expected Counts	$\frac{(O-E)^2}{E}$
Flats	0.15	50	60 = (0.15)(400)	1.67
High heels	0.25	60	100	16
Athletic shoes	0.35	140	140	
Sandals	0.1	60	40	
Hiking Boots	0.1	60	40	
Platform Shoes	0.05	30	20	
		N = 400		> 17.67

H_0: CA consumers buy shoes of the same styles in the same proportions as do NY consumers.

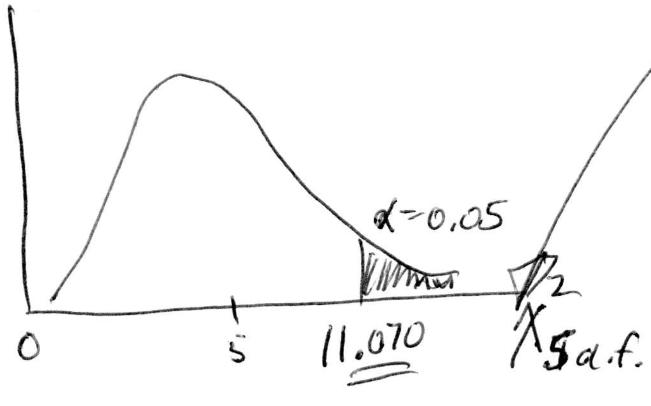
H_1: Not so!

$\alpha = 0.05$ right tail of χ^2 distribution with 6-1 d.f. = 5 d.f.

test statistic

$$\sum \frac{(O-E)^2}{E} = \text{at least } 17.67$$

Reject Ho!



(22)

(10 points - 15 minutes)

3. The proportions of people in the U.S. that prefer 5 different kinds of entertainment are shown in the table below. A local survey of 500 people found 60 people who prefer movies, 200 who prefer to watch TV, 90 who like to listen to music, 30 who prefer dancing, and 120 that prefer to play sports. Test the claim that the true local proportions are the same as the national rates.
(Use a 0.05 significance level for the test)

Claim: local proportions = National proportions

Ho: local proportions = National proportions

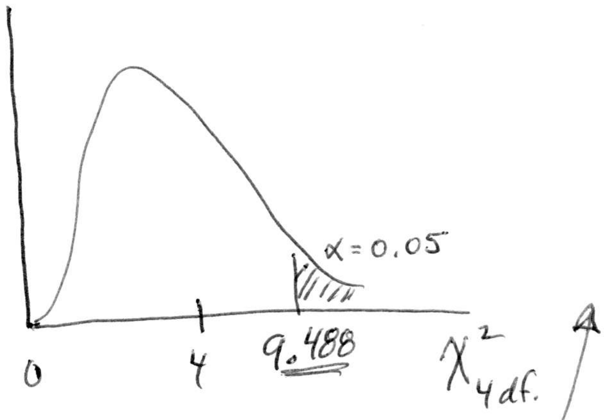
H1: local proportions ≠ national proportions

$\alpha = 0.05$ right tail

$df = k-1 = 5-1 = 4$

"O"

Preferred Entertainment	National Rates	Local Survey Count	"E"
Watching Movies	0.15	60	75 = (0.15)(500)
Watching Television	0.35	200	175
Listening to Music	0.10	90	50
Dancing	0.10	30	50
Playing Sports	0.30	120	150
Survey Total =		500	



$$\frac{(O-E)^2}{E} = \frac{(60-75)^2}{75} = 3$$

$$\frac{(200-175)^2}{175} = 3.6$$

$$= 32$$

already too big.

$$\sum \frac{(O-E)^2}{E} > 32$$

reject Ho!

24

(8 points)

7. **A company that markets sodas does a survey of consumer preferences. Use the data to test whether the two age-groups (in general, not these specific individuals) have the same proportions that prefer each soda.** (use $\alpha = 0.025$ for the test)

Counts in categories arranged according to more than one factor: "contingency table"

Number of People Who Prefer

	Coke	Pepsi	7-Up	
< 21 years old	20	10	10	40
21 or older	10	15	15	40
	30	25	25	$\boxed{80}$

$\boxed{\text{observed counts}}$

claim: the two age groups prefer the different sodas in the same proportions.

expected counts $= \dfrac{(\text{row total})(\text{column total})}{\text{grand total}}$

15	12.5	12.5
15	12.5	12.5

H_0: homogeneous proportions

H_1: proportions are not homogeneous

$\alpha = 0.025$ right tail

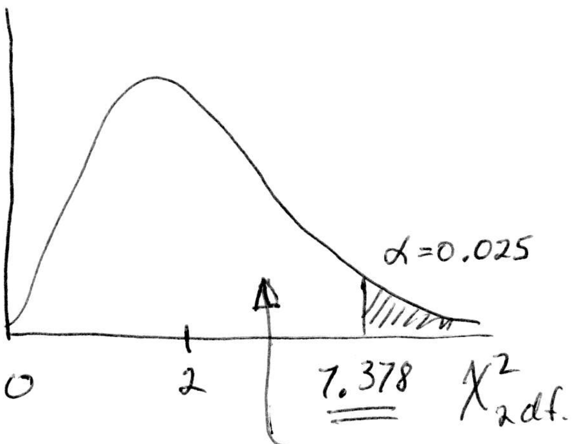
$df = (r-1)(c-1) = (2-1)(3-1) = 2$

critical $\chi^2_{2\,df} = 7.378$

$(O-E)^2/E$

1.67	0.5	0.5
1.67	0.5	0.5

critical region



$\sum \frac{(O-E)^2}{E} = 5.33$

$\boxed{\text{Do not reject } H_0\text{!}}$

(8 points - 10minutes)

11. Use the data in the contingency table to test the claim that 2-year old boys and girls choose toys in the same proportions when placed in an observation room. (Use $\alpha = 0.05$ for this test)

Toy chosen	Gender		
	Boys	Girls	Total
Ball	40	20	60
Doll	5	25	30
Bell	5	5	10
Total	50	50	100

claim: girls and boys choose the different toys in the same proportions

H_0: homogeneous proportions

H_1: not H_0:

$\alpha = 0.05$ right tail

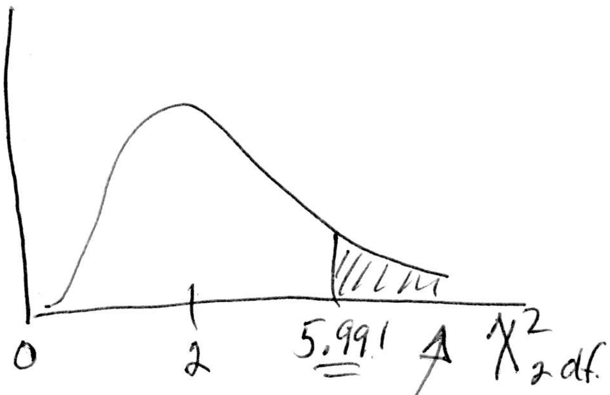
$df = (r-1)(c-1) = (3-1)(2-1) = 2$

critical Chi-square value $= 5.991$

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

30	30
15	15
5	5

$= \frac{(60)(50)}{100}$



$(O-E)^2/E$

3.33	3.33
6.67	6.67
0	0

$= \frac{(20-30)^2}{30}$

$$\sum \frac{(O-E)^2}{E} = 20$$

reject H_0:

26

(10 points - 20 minutes)

4. Use the data in the table to test the idea that the use of some "slang" terms is independent of age. The data represent a stratified random sample of 400 people from Los Angeles.
(Use $\alpha = 0.025$ for this test)

Most used Slang Term	Age Group 10 to 20	21 to 40	41 to 60	> 60	Total
I'm like ...	90	50	10	0	150
totally	10	40	40	10	100
far out	0	10	50	90	150
Total	100	100	100	100	400

counts in categories arranged according to more than one factor. "contigency table"

The idea (claim): Slang usage and age group are independent

H_0: independence of the row factor (slang term) and column factor (age)

H_1: dependence.

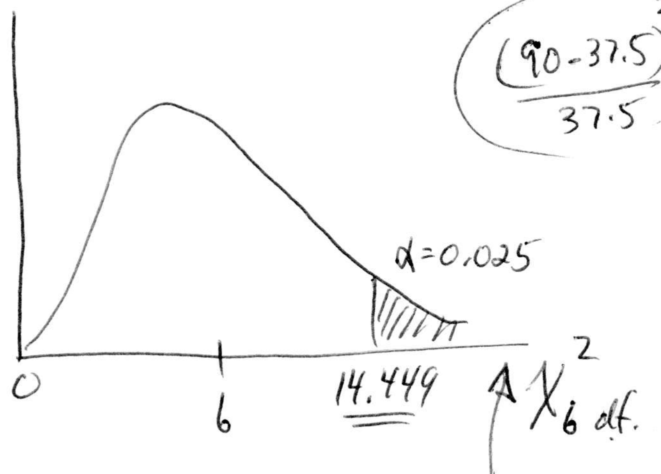
$\alpha = 0.025$ right tail

$df = (r-1)(c-1) = (3-1)(4-1) = 6$

critical Chi-square value = 14.449

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

37.5	37.5	37.5	37.5
25	25	25	25
37.5	37.5	37.5	37.5



$(O-E)^2/E$

$\dfrac{(90-37.5)^2}{37.5}$

73.5	4.17		

$\sum \dfrac{(O-E)^2}{E} > 73.5$

reject H_0:

27

(7 points - 8 minutes)

6. **Based on the data below, complete the Analysis of Variance Table by filling in the missing values. (Use $\alpha = 0.01$ for the appropriate hypothesis test.)**

Years to Train for Professions			
Railroad Engineer	Freighter Captain	Airline Pilot	Space Shuttle Pilot
2.0	4.0	3.0	6.0
2.8	1.5	5.5	2.8
3.2	2.0	4.0	5.0
	1.0	2.9	5.0
	3.4		3.6
	0.8		4.0
			3.2

$\bar{x}$ =	2.67	2.12	3.85	4.23
s =	0.61	1.31	1.21	1.15
n =	3	6	4	7

3 + 6 + 4 + 7 = $N = 20$

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: at least one μ is not the same as another

$\alpha = 0.01$ right tail

F distribution with
3 df for numerator
16 df for denominator
Critical F is not available in tables, but p-value is given, so

p-value = 0.0233 which is is greater than α, so

do not reject H_0!

Source	d.f.	Sum of Squares	Mean Square	F	p-value
Profession (Between groups)	3	16.85	5.617	4.167	0.0233
Error (Within groups)	16	21.56	1.348		
Total	19	38.41			

(8 points : 10 minutes)

2. **The Mayor of a city claims that the total fees paid by residents of similar sized cities are the same throughout the United States. Random samples of 20 residents in 26 cities of similar size are selected and the total fees they paid are determined (the data are on the following page). Use the data (as you may need to) to complete the analysis of variance table below. Then carry out a test of the Mayor's claim that average fees are the same for all 26 cities.**
(Use a significance level of 0.04 for this test.)

Analysis of Variance Table

Source	Degrees of Freedom	Sum of Squares	Mean Square	F	p-value
Cities	25	11557	462.28	1.713	0.0179
Error	494	133329	269.90		
Total	519	144886			

$N = (26)(20) = 520$

H_0: means (μ) for all the same for cities of this size

H_1: Not H_0:

$\alpha = 0.04$ right tail

F distribution.

critical F value
Not available in tables, but p-value is given

p-value = 0.0179
which is less than α, so
reject H_0:

(29)

Data on fees paid for 20 randomly selected residents in 26 cities in the U.S.

City	Fees paid									
1	$137.57	$126.49	$161.81	$148.41	$148.40	$141.25	$134.99	$139.89	$145.46	$147.16
	$154.02	$131.30	$134.11	$139.18	$125.14	$143.82	$140.38	$155.78	$142.53	$160.14
2	$174.82	$115.38	$140.20	$166.08	$160.67	$130.64	$117.01	$105.09	$112.96	$145.98
	$146.97	$145.57	$145.56	$128.07	$146.04	$129.74	$112.15	$137.65	$133.77	$130.23
3	$170.20	$148.37	$167.76	$148.62	$169.98	$108.22	$170.23	$125.31	$100.33	$128.98
	$143.01	$154.64	$135.23	$80.95	$144.34	$126.15	$150.19	$144.73	$122.04	$128.54
4	$143.83	$144.91	$159.57	$152.24	$134.45	$98.63	$127.29	$93.49	$120.79	$142.07
	$135.65	$123.39	$123.76	$159.54	$113.87	$128.78	$131.95	$108.60	$185.93	$141.42
5	$161.80	$134.43	$118.22	$145.90	$122.90	$125.09	$178.05	$141.45	$103.25	$105.38
	$127.42	$113.80	$140.49	$136.08	$134.48	$140.46	$124.76	$167.53	$128.84	$108.71
6	$120.99	$135.35	$133.74	$116.80	$144.35	$111.99	$125.01	$123.15	$107.62	$141.37
	$138.17	$123.86	$146.14	$141.48	$113.36	$144.81	$139.05	$153.24	$130.12	$135.37
7	$156.03	$137.98	$154.10	$166.55	$172.91	$137.92	$123.78	$143.36	$158.19	$137.73
	$154.34	$128.53	$127.27	$90.94	$130.65	$150.87	$126.26	$114.05	$150.01	$136.99
8	$146.92	$123.72	$140.98	$129.96	$105.59	$120.18	$133.32	$118.33	$106.14	$121.49
	$144.04	$135.30	$140.59	$144.42	$135.05	$127.75	$139.47	$123.95	$129.57	$147.87
9	$158.63	$126.17	$134.10	$129.84	$153.50	$168.41	$149.36	$144.24	$138.60	$123.43
	$133.30	$151.93	$151.12	$138.21	$147.95	$118.56	$132.98	$152.23	$138.42	$123.47
10	$137.21	$106.55	$159.60	$141.61	$105.70	$118.50	$143.87	$115.54	$129.20	$124.59
	$121.06	$147.72	$138.37	$123.15	$133.94	$131.84	$117.57	$182.01	$111.40	$133.38
11	$178.54	$121.61	$127.39	$139.00	$144.95	$136.60	$115.61	$178.28	$148.33	$140.63
	$148.72	$98.85	$142.52	$110.48	$126.00	$124.69	$143.24	$119.70	$168.19	$146.74
12	$144.34	$131.20	$143.21	$134.16	$147.34	$136.79	$143.08	$133.92	$140.17	$140.70
	$132.35	$142.07	$147.94	$139.21	$115.51	$145.02	$136.58	$125.53	$156.30	$141.13
13	$133.19	$127.20	$134.09	$122.15	$126.54	$135.46	$111.79	$156.51	$145.72	$144.65
	$159.25	$132.61	$144.52	$99.64	$145.64	$117.98	$131.90	$155.16	$152.52	$131.74
14	$144.27	$144.06	$119.42	$94.44	$168.23	$110.18	$129.69	$146.21	$143.63	$130.90
	$158.28	$135.96	$140.62	$112.79	$162.64	$146.21	$139.85	$111.59	$132.54	$145.29
15	$175.32	$118.92	$186.54	$120.93	$110.29	$170.68	$165.66	$115.07	$159.36	$113.40
	$128.19	$116.40	$134.74	$119.43	$136.78	$167.45	$151.62	$136.39	$164.69	$142.57
16	$138.09	$165.62	$138.41	$126.04	$124.92	$135.59	$118.45	$132.18	$136.18	$117.10
	$135.76	$146.83	$109.82	$107.25	$132.65	$132.97	$135.34	$123.86	$137.70	$117.19
17	$148.96	$145.08	$154.68	$162.18	$139.39	$147.45	$141.88	$148.54	$134.09	$142.77
	$134.74	$141.32	$153.18	$154.89	$148.37	$132.35	$138.96	$139.26	$140.80	$143.62
18	$126.41	$125.37	$138.63	$132.19	$164.43	$161.18	$165.23	$141.50	$122.17	$163.27
	$125.01	$125.17	$153.56	$122.04	$163.28	$126.36	$125.53	$138.30	$152.47	$158.71
19	$147.31	$120.94	$141.74	$138.47	$131.20	$122.19	$119.92	$153.09	$112.76	$123.35
	$118.89	$114.87	$123.29	$127.08	$149.92	$124.43	$118.20	$124.67	$147.30	$134.95
20	$154.35	$153.85	$111.12	$154.42	$107.91	$127.55	$128.16	$154.13	$133.67	$143.86
	$122.75	$114.68	$115.02	$114.54	$96.90	$155.75	$134.22	$130.75	$146.37	$157.17
21	$130.18	$145.32	$118.34	$109.37	$118.76	$128.08	$149.20	$112.54	$137.70	$132.46
	$139.98	$127.99	$135.92	$127.93	$125.34	$124.47	$138.62	$143.43	$112.06	$109.49
22	$148.38	$137.55	$138.58	$139.59	$115.75	$136.78	$124.13	$136.37	$158.21	$146.14
	$133.44	$139.41	$143.19	$138.73	$133.91	$143.93	$141.10	$144.42	$114.53	$129.74
23	$123.91	$139.28	$137.32	$144.72	$119.38	$155.70	$125.38	$137.16	$137.51	$149.52
	$125.93	$130.85	$121.13	$134.81	$155.48	$169.72	$135.69	$139.24	$125.87	$147.28
24	$127.88	$112.11	$132.40	$175.64	$143.22	$143.22	$143.39	$124.56	$126.92	$136.45
	$139.44	$150.68	$165.91	$127.48	$166.44	$165.07	$142.40	$145.52	$169.69	$166.83
25	$105.38	$150.61	$157.03	$127.41	$120.20	$154.60	$125.67	$151.50	$126.83	$130.99
	$145.41	$116.79	$146.21	$156.32	$124.58	$151.88	$131.76	$125.05	$149.85	$125.69
26	$151.86	$135.62	$138.73	$124.02	$112.79	$145.95	$139.48	$119.06	$117.00	$114.78
	$128.35	$109.97	$115.80	$143.21	$150.33	$129.95	$153.50	$149.56	$123.16	$122.23

30

(3 points; 2 minutes)

11. **Some people do not believe their driving is affected by drinking beer. A sample of 62 such sceptics takes part in an experiment, with the results listed below. Use the information given to complete the Analysis of Variance Table and test the idea that the mean number of errors is unaffected by the number of beers consumed.**

Number of Errors Committed While Driving Test Course
After Consuming the Indicated Number of Beers

	0	1	2	3	4	5	6
	7	2	15	12	39	43	27
	7	8	3	14	17	38	21
	0	1	20	23	15	10	36
	3	5	19	18	12	12	46
	5	10	12	12	12	17	60
	5	9	15	15	36	28	56
	5	7	4	19		45	26
	5	6		7		15	29
		7		22		35	46
		8				21	27
		5				43	
$\bar{x}$ =	4.6	6.2	12.6	15.8	21.8	27.9	37.4
s =	2.26	2.79	6.75	5.19	12.32	13.44	13.71
n =	8	11	7	9	6	11	10

N = 62

Use alpha=0.025 with the "Traditional Approach" for this test.

$H_0: \mu_1 = \mu_2 = \cdots = \mu_7$

H_1: at least one μ is different

$\alpha = 0.025$ F distribution right tail
6 d.f. for numerator
55 d.f. for denominator
critical F = 2.6274
(for 60 d.f. denominator in Table A-5)

SUMMARY

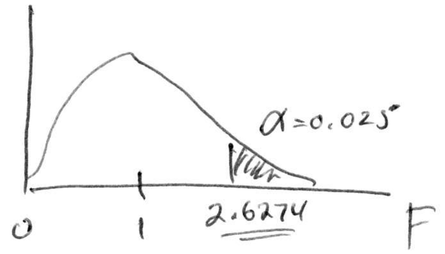
Groups	Count	Sum	Average	Variance
0 Beers	8	37	4.6	5.13
1 Beers	11	68	6.2	7.76
2 Beers	7	88	12.6	45.62
3 Beers	9	142	15.8	26.94
4 Beers	6	131	21.8	151.77
5 Beers	11	307	27.9	180.69
6 Beers	10	374	37.4	188.04

7 groups

N = 62

Analysis of Variance Table

Source of Variation	*SS*	*df*	*MS*	*F*
Treatments (beers)	8134.576	6	1355.8	15.34
Error	4860.924	55	88.38	
Total	12995.5	61		



reject H_0!

31

(8 points : 10 minutes)

1. **Use the data below to complete the Analysis of Variance Table and test the claim that all of the 1998 Chevy Nova cars have the same gas mileage today.** (Use a 0.05 significance level for the test.)

Car	Test 1	2	3	4	Sample Size	Mean	Standard Deviation
Car 1	20.33		20.63	17.00	3	19.32	2.015
Car 2	19.93	20.06	17.52		3	19.17	1.430
Car 3	17.53	18.50	17.10	20.87	4	18.50	1.685
Car 4	19.54	17.81	20.81	17.91	4	19.02	1.434
Car 5	20.39	20.33	18.56		3	19.76	1.040
Car 6	19.14	17.29	17.01	20.04	4	18.37	1.460
Car 7	19.77	20.60	19.08	19.96	4	19.85	0.626
Car 8	17.85	17.72	18.45		3	18.01	0.389
Car 9	19.10	17.09	17.45		3	17.88	1.072

$H_0: \mu_1 = \mu_2 = \cdots = \mu_9$

H_1: at least one μ is different

$\alpha = 0.05$ right tail

F distribution
8 df. numerator
22 df. denominator
Critical F = 2.3965

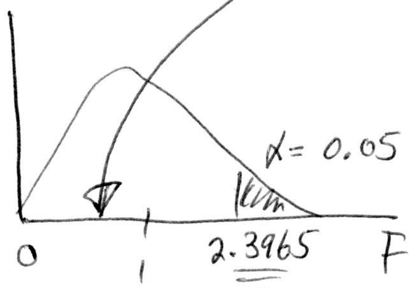
Total N	Overall Mean	Overall St. Dev.	Pooled St. Dev.
31	18.88	1.331	1.335

S_p : $S_p^2 = \text{MS(error)}$
$= (1.335)^2$
$= 1.782$

Analysis of Variance Table

Source	Deg. of Freedom	Sum of Squares	Mean Square	F
Cars	8	13.943	1.743	0.9781
Error	22	39.204	1.782	
Total	30	53.147		

$(N-1) * (\text{overall st.dev})^2$



Do not reject H_0!

32