Chapter 11

(33) $X_1, \dots, X_n \sim N(\mu_X, \sigma_X^2)$ $Y_1, \dots, Y_m \sim N(\mu_Y, \sigma_Y^2)$ where Xs and Ys are independent samples.

- Under $H_0: \sigma_X = \sigma_Y$ argue that $\frac{S_X^2}{S_Y^2} \sim F_{(n-1, m-1)}$

We know that $\frac{(n-1)S_X^2}{\sigma_X^2} \sim \chi^2_{(n-1)}$ and $\frac{(m-1)S_Y^2}{\sigma_Y^2} \sim \chi^2_{(m-1)}$ (page 197)

and that S_X^2, S_Y^2 are independent, then $\dfrac{\frac{(n-1)S_X^2}{\sigma_X^2(n-1)}}{\frac{(m-1)S_Y^2}{\sigma_Y^2(m-1)}} \sim F_{(n-1, m-1)}$

For $U \sim \chi^2_{(m)}$ and $V \sim \chi^2_{(n)}$ $\Rightarrow \frac{U/m}{V/n} \sim F_{(m,n)}$

$\Rightarrow \frac{\sigma_Y^2}{\sigma_X^2} \frac{S_X^2}{S_Y^2} \sim F_{(n-1, m-1)}$ but under H_0 $\sigma_X = \sigma_Y \Rightarrow \frac{S_X^2}{S_Y^2} \sim F_{(n-1, m-1)}$

a) Construct rejection regions for 1 and 2 sided tests of H_0

One sided case: $H_0: \sigma_X = \sigma_Y$, $H_A: \sigma_X > \sigma_Y$ $\Leftrightarrow$ $H_0: \frac{\sigma_X}{\sigma_Y} = 1$, $H_A: \frac{\sigma_X}{\sigma_Y} > 1$

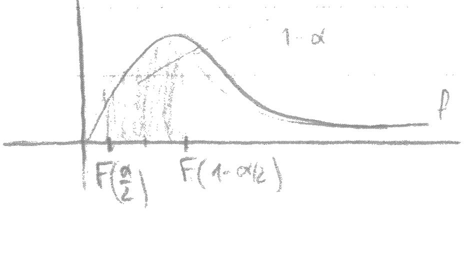
Therefore a large value of $\frac{S_X^2}{S_Y^2}$ indicates rejection of H_0

$\Rightarrow$ we reject H_0 at significance level α if $\frac{S_X^2}{S_Y^2} > F_{(n-1, m-1)}(\alpha)$

Two sided case: $H_0: \frac{\sigma_X}{\sigma_Y} = 1$, $H_A: \frac{\sigma_X}{\sigma_Y} \neq 1$

Large or small values of $\frac{S_X^2}{S_Y^2}$ indicate rejection of H_0

$\Rightarrow$ we reject H_0 at significance level α if $\frac{S_X^2}{S_Y^2} < F_{(n-1, m-1)}(\alpha/2)$ or if $\frac{S_X^2}{S_Y^2} > F_{(n-1, m-1)}(1 - \alpha/2)$



b) Confidence interval for $\frac{\sigma_x^2}{\sigma_y^2}$

We know that $P\left[F_{(n-1,m-1)}(1-\frac{\alpha}{2}) < \frac{\sigma_y^2}{\sigma_x^2} \frac{S_x^2}{S_y^2} < F_{(n-1,m-1)}(\frac{\alpha}{2}) \right] = 1-\alpha$

$$\Rightarrow \mathbb{P}\left[\frac{\frac{S_x^2}{S_y^2}}{F_{(n-1,m-1)}(\alpha/2)} < \frac{\sigma_x^2}{\sigma_y^2} < \frac{\frac{S_x^2}{S_y^2}}{F_{(n-1,m-1)}(1-\alpha/2)} \right] = 1-\alpha$$

Therefore $\left[\frac{\frac{S_x^2}{S_y^2}}{F_{(n-1,m-1)}(\alpha/2)} , \frac{\frac{S_x^2}{S_y^2}}{F_{(n-1,m-1)}(1-\frac{\alpha}{2})} \right]$ is a $100(1-\alpha)\%$ CI for $\frac{\sigma_x^2}{\sigma_y^2}$

c) Apply results to Example A (page 423)

$S_A = 0.024$ $\Rightarrow$ $\hat{\theta} = \frac{S_A^2}{S_B^2} = 0.5994$

$S_B = 0.031$

$n = 13$ $\quad \alpha = 0.05$ $\quad F_{(13-1,8-1)}(0.05/2) = 0.2773$

$m = 8$ $\quad F_{(13-1,8-1)}(1-0.05/2) = 4.6658$

$$CI: \left[\frac{0.5994}{4.6658} , \frac{0.5994}{0.2773} \right] = (0.1285, 2.1617)$$

$\Rightarrow$ do not reject H_0

P-value: $\mathbb{P}(\hat{\theta} \geq 0.5994 \mid H_0) = 1 - F_{12,7}(0.5994) = 0.7920 = 1-\frac{\alpha}{2}$

$\Rightarrow$ $1-\frac{\alpha}{2} = 0.7920$

$\frac{\alpha}{2} \doteq 0.2080$

$\alpha = 0.4160$

(36) $\bar{X} = 85.26 \qquad \bar{D} = \bar{X} - \bar{Y} = 0.44$

$\bar{Y} = 84.82$

$S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = 449.2754$

$S_Y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 464.5660$

$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y}) = 446.1987$

$S_{\bar{D}}^2 = \frac{1}{n} (S_X^2 + S_Y^2 - 2S_{XY}) = \frac{1}{15} (449.2754 + 464.5660 - 2 \cdot 446.1987) = 1.4296$

$S_{\bar{D}} = 1.1957$

If independence assumed $S_{\bar{D}}^2 = \frac{1}{n} (S_X^2 + S_Y^2) = 60.9228$

$S_{\bar{D}} = 7.8053$

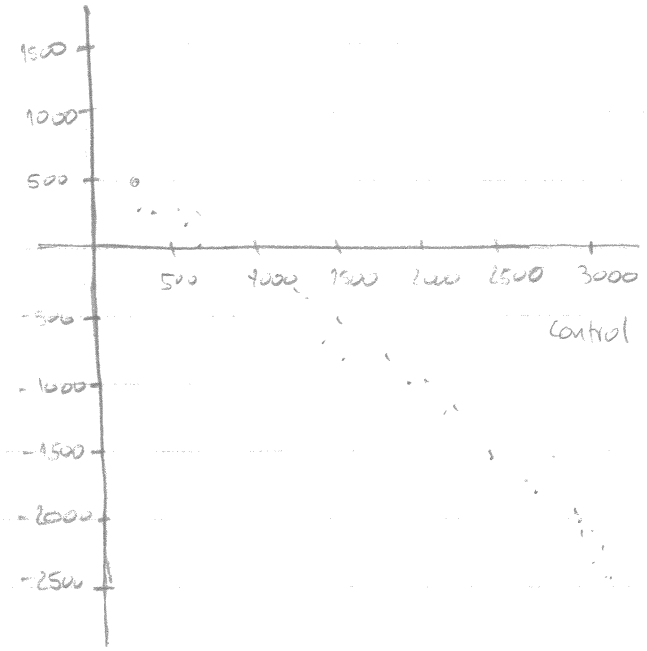
$H_0: \mu_X = \mu_Y \qquad t = \frac{\bar{D}}{S_{\bar{D}}} = \frac{0.44}{1.1957} = 0.3680$

$H_A: \mu_X \neq \mu_Y$

$t_{(14-1)}\left(\frac{0.05}{2}\right) = 2.1604$

$\Rightarrow$ No evidence to reject H_0

39) a) Plot the differences versus the control rate



Test	Control	Diff
676	88	588
206	570	-364
230	605	-375
256	617	-361
280	653	-373
433	2913	-2480
337	924	-587
466	286	180
497	1098	-601
512	982	-470
794	2346	-1552
428	321	107
452	615	-163
572	579	-7

There seems to be a relationship between them. The larger the value of the control, the larger the difference

b) Calculate $\bar{D} = \bar{T} - \bar{C}$, its std and CI

$\bar{D} = \bar{T} - \bar{C} = 434.2143 - 895.5 = -461.2857$

$S_{\bar{D}}^2 = \frac{1}{n}\left(S_T^2 + S_C^2 - 2S_{TC}\right) = 41020$

$S_T^2 = \frac{1}{n-1}\sum_{i=1}^{n}(T_i - \bar{T})^2 = 29007$ $\quad S_C^2 = \frac{1}{n-1}\sum_{i=1}^{n}(C_i - \bar{C})^2 = 621945$

$S_{TC} = \frac{1}{n-1}\sum_{i=1}^{n}(T_i - \bar{T})(C_i - \bar{C}) = 39339$

As $t = \frac{\bar{D} - \mu_D}{S_{\bar{D}}} \sim t_{(n-1)}$ then, a $100(1-\alpha)\%$ CI for μ_D is

$\bar{D} \pm t_{(n-1)}(\alpha/2) S_{\bar{D}} = -461.2857 \pm (+2.1604)(202.5330) = -461.2857 \pm 437.5460$
$= (-898.8317, -23.7397)$

c) $\tilde{D} = \frac{-373-364}{2} = -368.5$

$(X_{(k)}, X_{(n-k+1)})$ is CI for the median with $1 - \frac{1}{2^{n-1}} \sum_{j=0}^{k-1} \binom{n}{j}$ probability of coverage

For $k=4$ we have that $1 - \frac{1}{2^{14-1}} \sum_{j=0}^{3} \binom{n}{j} = 0.9426$

so $(X_{(4)}, X_{(14-4+1)}) = (X_{(4)}, X_{(11)}) =$
$(-587, -7)$ is a 94.26% CI interval for the median

d) We already know that H_0 is rejected because $0 \notin$ in the CI, but we still can do the test: $t = \frac{\bar{D} - \mu_0}{S_{\bar{D}}} \sim t_{(n-1)}$ Under H_0 $\mu_0 = 0$ so our test statistic is $t = \frac{461.2857}{202.5330} = 2.2776$

$t_{(14-1)}\left(\frac{0.05}{2}\right) = 2.16$ as $t > 2.16$ we reject the null hypothesis (Equivalently, the rejection region for H_0 is given by $|\bar{D}| > t_{(n-1)}(\alpha/2) S_{\bar{D}}$. In this case $t_{(n-1)}(\frac{\alpha}{2}) S_{\bar{D}} = (2.16)(202.533)$

$= 437.47$

so we reject)

|Diff|	Rank	Signed rank
588	11	11
364	6	-6
375	8	-8
361	5	-5
373	7	-7
2480	14	-14
587	10	-10
180	4	4
601	12	-12
470	9	-9
1557	13	-13
107	2	2
163	3	-3
7	1	-1

W_+ = sum of + signed rank

$= 11 + 4 + 2 = 17$

Critical value for $\alpha = 0.05$ is 21 and $W_+ < 21$ so we reject H_0

The small sample size suggests the non-parametric test is more suitable

(17) Let Y_{ijk} denote the kth observation in cell ij. The statistical model is: $Y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}$

where μ: overall mean

$\sum_{i=1}^{I} \alpha_i = 0$: diff. effects

$\sum_{j=1}^{J} \beta_j = 0$: diff effects

$\sum_{i=1}^{I} \delta_{ij} = \sum_{j=1}^{J} \delta_{ij} = 0$: crossed diff. effect

we assume $\varepsilon_{ijk} \sim N(0, \sigma^2) \Rightarrow Y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \delta_{ij}, \sigma^2)$, thus

$$L(\mu, \alpha_i, \beta_j, \delta_{ij}) = \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} (Y_{ijk} - \mu - \alpha_i - \beta_i - \delta_{ij})^2 \right\}$$

$$l = \log L = -\frac{IJK}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - \mu - \alpha_i - \beta_i - \delta_{ij})^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum_i \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})$$

$$= \frac{1}{\sigma^2} \left[Y_{...} - IJK\mu - JK \sum_i \alpha_i - IK \sum_j \beta_j - K \sum_i \sum_j \delta_{ij} \right] =$$

$$= \frac{1}{\sigma^2} \left[Y_{...} - IJK\mu \right]$$

$$\frac{\partial l}{\partial \mu} = 0 \iff Y_{...} - IJK\mu = 0$$

$$\hat{\mu} = \frac{Y_{...}}{IJK} = \bar{Y}_{...}$$

$$\frac{\partial l}{\partial \alpha_i} = \frac{1}{\sigma^2} \left[\sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij}) \right]$$

$$= \frac{1}{\sigma^2} \left(Y_{i..} - JK\mu - JK\alpha_i - K \sum_j \beta_j - K \sum_j \delta_{ij} \right)$$

$$= \frac{1}{\sigma^2} \left(Y_{i..} - JK\mu - JK\alpha_i \right)$$

$$\frac{\partial l}{\partial \alpha_i} = 0 \Rightarrow Y_{i..} - JK\mu - JK\alpha_i = 0$$

$$JK\alpha_i = Y_{i..} - JK\mu$$

$$\hat{\alpha}_i = \bar{Y}_{i..} - \hat{\mu}$$

$$= \bar{Y}_{i..} - \bar{Y}_{...}$$

$$\frac{\partial l}{\partial \beta_j} = \frac{1}{\sigma^2} \sum_i \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})$$

$$= \frac{1}{\sigma^2} \left(Y_{\cdot j \cdot} - IK\mu - K\sum_i \alpha - IK\beta_j - K\sum_i \delta_{ij} \right)$$

$$= \frac{1}{\sigma^2} \left(Y_{\cdot j \cdot} - IK\mu - IK\beta_j \right)$$

$$\frac{\partial l}{\partial \beta_j} = 0 \iff Y_{\cdot j \cdot} - IK\mu - IK\beta_j = 0$$

$$IK\beta_j = Y_{\cdot j \cdot} - IK\mu$$

$$\hat{\beta}_j = \bar{Y}_{\cdot j \cdot} - \hat{\mu}$$

$$= \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots}$$

$$\frac{\partial l}{\partial \delta_{ij}} = \frac{1}{\sigma^2} \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \delta_{ij})$$

$$= \frac{1}{\sigma^2} \left(Y_{ij\cdot} - K\mu - K\alpha_i - K\beta_j - K\delta_{ij} \right)$$

$$\frac{\partial l}{\partial \delta_{ij}} = 0 \Rightarrow Y_{ij\cdot} - K\mu - K\alpha_i - K\beta_j - K\delta_{ij} = 0$$

$$\delta_{ij} = \frac{Y_{ij\cdot}}{K} - \mu - \alpha_i - \beta_j$$

$$\hat{\delta}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{\cdots} - \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdots} - \bar{Y}_{\cdot j \cdot} + \bar{Y}_{\cdots}$$

$$= \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j \cdot} + \bar{Y}_{\cdots}$$

Chapter 12

(21) $SS_{TOT} = SS_W + SS_B$

$$\sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{i.})^2 + J\sum_{i=1}^{I}(\bar{Y}_{i.}-\bar{Y}_{..})^2$$

where $\bar{Y}_{i.} = \frac{1}{J}\sum_{j=1}^{J} Y_{ij}$ and $\bar{Y}_{..} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} Y_{ij}$

To test $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_I = 0$ we use the test statistic

$F = \frac{SS_B/(I-1)}{SS_W/I(J-1)}$ If the null hypothesis is true $\Rightarrow F \approx 1$ and should be larger if H_0 is false

$F \sim F_{(I-1,\, I(J-1))}$ because is the ratio of two χ^2 divided by their degrees of freedom Here $I=4$
$J=5$

ANOVA table

Source	df	SS	MS	F	p-value ($p(F>2.27)$)
Groups	3	27234.2	9078.07	2.27	0.1196
Error	16	63953.6	3997.1		Critical value $F_{(3,16)}(1-0.05)$
Total	22	91187.8			3.2389
					$\Rightarrow$ we don't reject H_0

For $SS_B = J\sum_{i=1}^{I}(\bar{Y}_{i.}-\bar{Y}_{..})^2$ $\quad \bar{Y}_{1.} = \frac{1}{5}\sum_{j=1}^{5} Y_{1j} = \frac{1}{5}(279+338+334+198+303) = 290.4$

$\bar{Y}_{2.} = 323.2$

$\bar{Y}_{..} = \frac{1}{5\cdot 4}\sum_{i=1}^{4}\sum_{j=1}^{5} Y_{ij} = 314.9$ $\quad \bar{Y}_{3.} = 274.8$

$\bar{Y}_{4.} = 371.2$

$SS_B = 5[(290.4-314.9)^2 + (323.2-314.9)^2 + (274.8-314.9)^2 + (371.2-314.9)^2]$
$= 27234.2$

$SS_W = \sum_{i=1}^{I}\sum_{j=1}^{J}(Y_{ij}-\bar{Y}_{i.})^2 = (279-290.4)^2 + \dots + (303-290.4)^2 + (378-323.2)^2 + \dots + (286-323.2)^2$
$+ (172-274.8)^2 + \dots + (250-274.8)^2 + (381-371.2)^2 + \dots + (318-371.2)^2$

$= 63953.6$

Kruskal-Wallis test

R_{ij} = rank of the Y_{ij}

I	II	III	IV
6	17	1	18
14	5	12.5	16
11	19	12.5	15
2	4	7	20
9	8	3	10

$\bar{R}_{i\cdot} = \frac{1}{J_i} \sum_{j=1}^{J_i} R_{ij}$

$\bar{R}_{1\cdot} = \frac{1}{5}(6+14+11+2+9) = 8.4$

$\bar{R}_{2\cdot} = \frac{1}{5}(17+5+19+4+8) = 10.6$

$\bar{R}_{3\cdot} = \frac{1}{5}(1+12.5+12.5+7+3) = 7.2$

$\bar{R}_{4\cdot} = \frac{1}{5}(18+16+15+20+10) = 15.8$

$$\bar{R}_{\cdot\cdot} = \frac{1}{N} \sum_{i=1}^{I}\sum_{j=1}^{J} R_{ij} = \frac{N+1}{2} = 10.5$$

$$SS_B = \sum_{i=1}^{I} J(\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2 = 5[(8.4-10.5)^2 + (10.6-10.5)^2 + (7.2-10.5)^2 + (15.8-10.5)^2] = 217$$

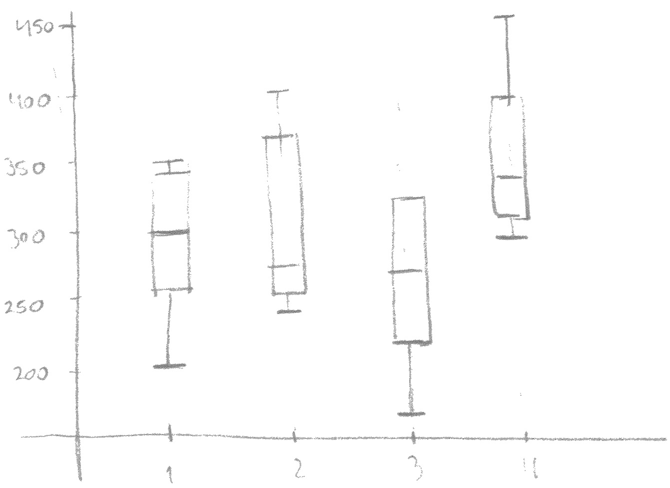
$$K = \frac{12}{N(N+1)} SS_B = \frac{12}{(20)(21)} 217 = 6.2$$

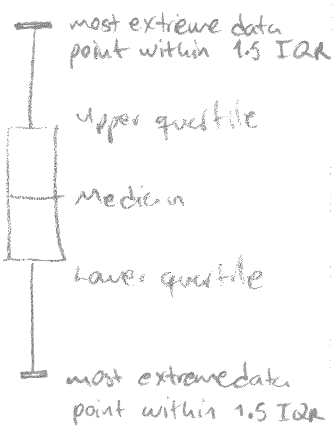
$K \sim \chi^2_{(I-1)} = \chi^2_{(3)}$ $(I=3,\ J_i \geq 5$ or $I>3$ and $J_i \geq 4)$

Critical value at $\alpha = 0.05$ is $\chi^2_{(3)}(1-0.05) = 7.8147$

or P-value $P(K > 6.2) = 0.1023$

So we don't have enough evidence to reject H_0





(26)

Drugs \ Dogs	1	2	3	4	5	6	7	8	9	10	
1	0.28	0.51	...							0.33	0.434
2	0.30	0.39								0.32	0.469
3	1.07									0.30	0.853
	0.55	0.75	0.77	0.45	0.64	0.7	0.56	0.55	0.56	0.32	0.585

Can be considered a two-way layout with $I=3$ levels in the factor "drug" $J=10$ levels in the factor "dog" and no interaction between dogs and drug (randomized block designs)

We want to test effect α of drugs: $H_A: \alpha_1 = \alpha_2 = \alpha_3 = 0$

Let SS_A: sum of squares due to drug type

SS_B: sum of squares due to dog

SS_{AB}: sum of squares due to interaction

$$SS_A = JK \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 = 1.08$$

$$SS_B = IK \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2 = 0.52$$

$$SS_{AB} = K \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 = 3.2$$

$$MS_A = \frac{SS_A}{I-3} = \frac{1.08}{2} = 0.54$$

$$MS_B = \frac{SS_B}{J-1} = \frac{0.52}{9} = 0.057$$

$$MS_{AB} = \frac{SS_E}{(I-1)(J-1)} = \frac{3.2}{18} = 0.18$$

Source	df	SS	MS	F
Drugs	2	1.08	0.54	3
Dogs	9	0.52	0.057	0.3167
"Interaction"	18	3.2	0.18	
Total	29			

$$F = \frac{MS_A}{MS_{AB}} = 3 \qquad \text{critical value } F_{(I-1,(I-1)(J-1))}(1-0.05) = 3.55$$

$\Rightarrow$ don't reject H_0 at 0.05 significance level

Friedman's test

- Rank the data within each subject

Drug \ Dog	1	2	3	4	5	6	7	8	9	10	
1	1	2	3	2	1	2	1	3	1	3	1.9
2	2	1	1	3	2	1	3	1	2	2	1.8
3	3	3	2	1	3	3	2	2	3	1	2.4

$\bar{R}_{..} = 2 \qquad \sum(\bar{R}_{i.} - \bar{R}_{..}) = (1.9-2)^2 + (1.8-2)^2 + (2.4-2)^2 = 0.21$

$Q = \frac{12J}{I(I+1)} \sum(\bar{R}_{i.} - \bar{R}_{..})^2 = \frac{12 \cdot 10}{3 \cdot 4} \, 0.21 = 2.1$

$Q \sim \chi^2_{(I-1)} = \chi^2_{(2)}$ critical value is 5.99 so we don't reject H_0

On randomized block tests

- Under the two-way layout model

	SS_A	SS_B	SS_{AB}	SS_E
df	$I-1$	$J-1$	$(I-1)(J-1)$	$IJ(K-1)\sigma^2$

$E(MS_A) = \sigma^2 + \frac{JK}{I-1}\sum_{i=1}^{I} \alpha_i^2$

$\Rightarrow$ if MS_A / MS_E large suggests some of the α_i not being zero

$E(MS_B) = \sigma^2 + \frac{IK}{J-1}\sum_{j=1}^{J} \beta_j^2$

$E(MS_{AB}) = \sigma^2 + \frac{K}{(J-1)(J-1)}\sum_{i=1}^{I}\sum_{j=1}^{J} \delta_{ij}^2$

$E(MS_E) = \sigma^2$

- Under randomized block design (same degrees of freedoms as in two-way layout)

$E(MS_A) = \sigma^2 + \frac{J}{I-1}\sum_{i=1}^{I} \alpha_i^2$

$\Rightarrow$ MS_A / MS_{AB} large suggests some of the α_i not being zero

$E(MS_B) = \sigma^2 + \frac{I}{J-1}\sum_{i=1}^{I} \beta_i^2$

$E(MS_{AB}) = \sigma^2$