Stats 110B

HW5 Suggested Solutions

http://www.stat.ucla.edu/~dinov/courses_students.dir/03/Spr/Stat110B.dir/STAT110B.html
http://www.stat.ucla.edu/~dinov/courses_students.dir/03/Spr/Stat110B.dir/assignments.html

1. Problem 1

(a) $x_{1\bullet} = 34.3$, $x_{2\bullet} = 39.6$, $x_{3\bullet} = 33$, $x_{4\bullet} = 41.9$, $x_{\bullet\bullet} = 148.8$, $\Sigma\Sigma x_{ij}^2 = 946.68$

$SST = \Sigma\Sigma x_{ij}^2 - 1/(IJ)x_{\bullet\bullet}^2 = 946.68 - 1/24\,(148.8)^2 = 24.12$

$SSTr = 1/J\ \Sigma x_{i\bullet}^2 - 1/(IJ)x_{\bullet\bullet}^2 = 8.98$

$SSE = SST - SSTr = 24.12 - 8.98 = 15.14$

ANOVA table :

SOURCE	DF	SS	MS	F
Treatment	3	8.98	2.99	3.95
Error	20	15.14	0.757	
Total	23	24.12		

Compared F with $F_{0.05,3,20} = 3.10$

$F > 3.10$, therefore we reject $H_o : \mu_1 = \ldots = \mu_4$ and we conclude that at least two of the grains differ with respect to the true average thiamin content.

(b) We have to assume normality, equal variance of all cases and independence of each trial.

2. Problem 2:

SOURCE	DF	SS	MS	$F_o \sim \boldsymbol{F}(5, 23)$	P-value (Use SOCR resource)
Regression	5	50	10	11.5	0.000011997693485873053
Error	23	20	0.86963		
Total	28	70			

Since the p-value is close to zero, we would reject $H_o : \beta_1 = \ldots = \beta_5 = 0$, which means at least one of the βs should be included in the model, and the response Y is in a linear relation with at least one of the 5 predictors.

3. Problem 3

(a) $x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

where α_i is the fixed effect(angle)

β_j is the random effect(connector)

(b) $H_o : \alpha_1 = \ldots = \alpha_4$

H_a : at least one pair is different

We can then construct an ANOVA table

Source	Df	SS	MS	F
Angle	3	58.16	19.3867	2.5565
Connector	4	246.97	61.7425	8.1419
Error	12	91	7.583	
Total	19	396.13		

$f_A = 2.5565 < F_{0.05,3,12} = 3.49$
Therefore we do not reject the null and conclude that the force required to cause separation is not influenced by angle of pull.

(c) treating connector as a random effect, we have
$H_o : \sigma^2_\beta = 0$
$H_a : \sigma^2_\beta \neq 0$
From the ANOVA table we have $f_B = 8.1419 < F_{0.05,4,12} = 3.26$
Therefore we reject the null and conclude that there is difference between the connectors.

(d) $w = Q_{0.05,\ 5,12}\ (7.583/4)$^0.5 = 4.51

$x_{\bullet 1}$-bar	$x_{\bullet 2}$-bar	$x_{\bullet 3}$-bar	$x_{\bullet 4}$-bar	$x_{\bullet 5}$-bar
43.9	43.7	42.125	38.725	49.575

If the difference between the two means are less than 4.51, we would underline them, indicating no significant difference.
Here, connector 5 seems to be significantly different from all the others, while connector 4 seems to be different from connectors 1 and 2 apart from 5. Other than that, there is no significant difference between the pairs.

4. Problem 4:

(a) n =14
$\Sigma x_i = 3300$, $\Sigma y_i = 5010$, $\Sigma x_i^2 = 913{,}750$, $\Sigma y_i^2 = 2{,}207{,}100$, $\Sigma x_i y_i = 1{,}413{,}500$
estimate of $\beta_1 = s_{xy}/s_{xx} = (1{,}413{,}500 - 3300(5010)/14)/(913750 - (3300)^2/14)$
$= 1.7114323$
estimate of β_0 = y-bar – estimate of β_1 (x-bar) = 5010/14 – 1.7114323(3300/14)
= -45.5519
equation : y-head = -45.5519 + 1.7114x

(b) estimate = -45.5519 + 1.7114(225) = 339.51

(c) x down by 50, estimated expected change = (β_1-head)(-50) = -50(1.7114323)
= -85.57(y down by 85.57)

(d) No. The reason is that the value 500 is outside the range of the x values in the data. If we use x = 500, there is the risk of extrapolation.

5. Problem 5:

$H_o : p_1 = p_2 = p_3 = p_4 = 0.25$
H_a : at least one of the probabilities is not 0.25
df = 3, n = 1361, expected count for each season = 1361/4 = 340.25
$\chi^2 = \Sigma(\text{observed count} - \text{expected count})^2/\text{expected count}$
$= ((328 - 340.25)^2 + \ldots + (327 - 340.25)^2)/340.25$
$= 4.0345$
comparing with the chi-squared table, p-value > 0.1
Therefore we do not reject the null. So the data fails to indicate a seasonal relationship with the incidence of violent crime.

6. Problem 6 : H_o : type of car and commuting distance are independent.

H_a : they are dependent

df = (3-1)(4-1) = 6

expected count =(row total)(column total)/(total count)

$\chi^2 = ((10.19 - (52)(49)/250)^2)/ ((52)(49)/250) + \ldots\ldots$

$\ldots\ldots + ((11.40 - (38)(75)/250)^2)/ ((38)(78)/250)$

$= 14.15 > \chi^2_{0.05, 6} = 12.592$

Therefore we reject the null and the two variables are not independent.

7. Problem 7: We assume that the two distributions have the same shape and spread.

$H_o : \mu_1 - \mu_2 = 0$

$H_a : \mu_1 - \mu_2 < 0$

Where μ_1 denote the population mean if the unpolluted source

$w = \Sigma r_i$, where r_i is the rank of x_i

Polluted :	21.3	18.7	23	17.1	16.8	20.9	19.7
Rank :	11	7	12	3	2	10	8

Unpolluted :	14.2	18.3	17.2	18.4	20
Rank :	1	5	4	6	9

w = 1 + 5 + 4 + 6 + 9 =25

compared with 5(5 + 7 + 1) – 47(from table with $\alpha = 0.01$) = 18

reject the null if $w \leq 18$ (because we have $H_a : \mu_1 - \mu_2 < 0$)

in this case we do not reject the null and conclude that the true average fluoride concentration for the two sources are the same.

8. Problem 8: $H_o : \mu_1 - \mu_2 = 0$

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

Water sample

Analyst 1 :	31.4	37.0	44.0	28.8	59.9	37.6
Analyst 2 :	28.1	37.1	40.6	27.3	58.4	38.9
Difference :	3.3	-0.1	3.4	1.5	1.5	-1.3
Rank :	5	1	6	3.5	3.5	2
Sign	+	-	+	+	+	-

$s_+ = 5 + 6 + 3.5 + 3.5 = 18$

reject if $s_+ > c$ or $s_+ < (6)(6+1)/2 - c$ $c \approx 19$ at $\alpha = 0.05$

$s_+ > 19$ or $s_+ < 2$

in this case we do not reject H_o and conclude that there are no differences between the Nitrogen concentrations measured by the two analysts.

9. Problem 9: H_o : all distributions are the same

1		2		3	
80	7	70	2.5	63	1
92	11	81	8	76	5

87	10	78	6	70	2.5
83	9	74	4		

Column rank total : 37 20.5 8.5

$H = 12/(11(12))\,(37^2/4 + 20.5^2/4 + 8.5^2/3) - 3(11+1) = 6.8542$

$\chi^2_{0.05,2} = 5.992$

H > 5.992, we reject the null and conclude that there is therapy effect on reading Comprehension.