Gurtej Singh STA 106 HW1

1.) Let treatments -> all possible combinations of the two cat. variable factor levels.

*Simultaneous Format

	A	B	C	D	F
U	UA	UB	UC	UD	UF
E	EA,	EB	EC	ED	EF

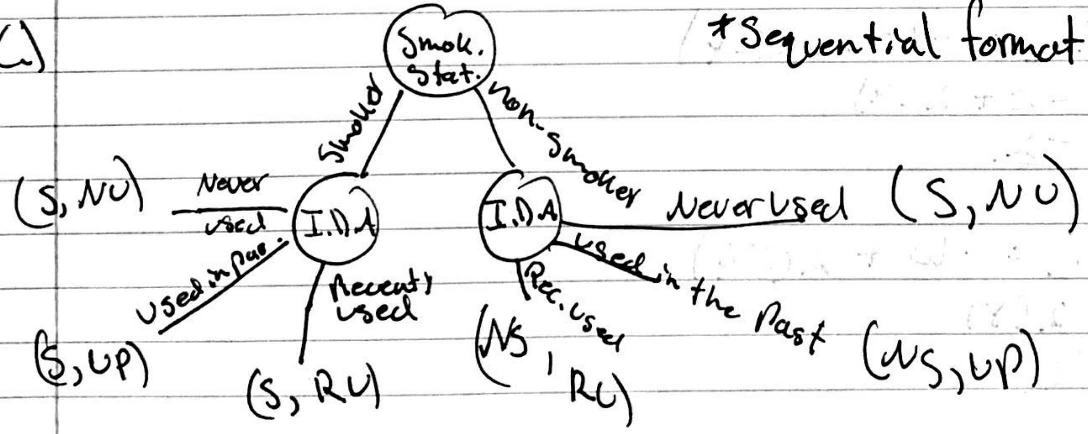
} 10 treatments.

b.)

	has Cancer = h	Does not = n
(Y)oung	Yh	Yn
(M)iddle	Mh	Mn
(O)ld	Oh	On

} 6 possible treatments.

c.) *Sequential format :)



6 treatments possible.

1.D.)

Intelligence

		H	M	L
Type	B	BH	BM	BL
	G	GH	GM	GL
	D	DH	DM	DL

9 types of treatments

2.) $\mu_Y = 4$, $\sigma_Y = 8$; find μ_U & σ_U

a.) $U_1 = 3 + 4Y$

$\mu_{U_1} = 3 + 4(4) = 19$; $\mu_{U_1} = 19$

$\sigma_{U_1} = 3 + 4\sigma_Y$

$= 4(8)$; $\sigma_{U_1} = 32$

B.) $U_2 = -10 + 2Y$

$\mu_{U_2} = -10 + 2(\mu_Y)$

$= -10 + 2(4)$

$= \underline{-2}$

$\sigma_{U_2} = -10 + 2(\sigma_Y)$

$= 2(8)$

$= \underline{16}$

C.) $U_3 = \frac{1}{4} - Y$

$\mu_{U_3} = \frac{1}{4} - (\mu_Y)$

$=$ [redacted] $-4 + 1/4$

$=$ [redacted] $\underline{-3.75}$

$\sigma_{U_3} = \frac{1}{4} - \sigma_Y$

$= -(8)$ [redacted]

$= \underline{-8}$

Gurteg Singh

3.) $\overline{Y}$ $\quad \mu = 20 \quad \sigma = 5 \quad n = 10$

if we can say all Y_i are <u>independent</u>

$$\frac{1}{n}\sum_{i}^{n} E(Y_i) = \cancel{\frac{1}{n}\cdot n}\cdot \mu$$

$$\therefore \mu_{\overline{Y}} = 20$$

Given $\overline{Y} \sim N(\mu_{Y_i}, \sigma_{Y_i}/\sqrt{n})$

$$\sigma_{\overline{Y}}^2 = (\sigma_{Y_i}/\sqrt{n})^2$$
$$= \frac{25}{10} \rightarrow \underline{2.5}$$

b.) Given $\sum Y_i \sim N(n\mu_{Y_i}, \sqrt{n}\ \sigma_{Y_i})$

$$\sum_{i=1}^{10} Y_i$$

$$\mu_{\sum Y_i} = 10 \cdot [redacted]\ 20$$
$$= [redacted]\ \underline{200}$$

$$\sigma_{\sum Y_i}^2 = (\sqrt{n}\cdot\sigma_{Y_i})^2$$

$$n\cdot\sigma_{Y_i}^2 = [redacted]\ \underline{250}$$

c.) $Y^* = [redacted]\alpha + b\overline{Y}$

$$\mu_{Y^*} = \alpha + b(\mu_{\overline{Y}})$$
$$\mu_{Y^*} = [redacted]\alpha + 20b$$

$$\sigma^2_{Y^*} = b^2\sigma^2_{\overline{Y}}$$

$$b^2 \cdot 2.5$$
$$\sigma^2_{Y^*} = 2.5b^2$$

D.) $y^{*} = 5 - 2 \sum_{i=1}^{10} y_i$

$\mu_{y^{*}} = 5 - 2(\mu_{\Sigma y_i})$

$\mu_{y^{*}} = 5 - 2(200)$

$= \underline{-395}$

$\sigma^2_{y^{*}} = -2\,\sigma^2_{\Sigma y_i}$

$= (-2)^2 (\sigma^2_{\Sigma y_i})$

$= 250 \cdot 4$

$= \underline{1000}$

Gurtej Singh STA 106 HW. 1

4.) $\bar{Y}_1 - \bar{Y}_2$ (normal distribution)

a.) $\mu_{1-2} = \mu_1 - \mu_2$

$\sqrt{\frac{\sigma_1^2}{100} + \frac{\sigma_2^2}{100}} = \sigma_{1+2}$

Specifically $\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{100}}\right)$

b.) $\bar{Y}_1 + \bar{Y}_2$ (Normal distribution)

$\mu_1 + \mu_2 = \mu_{1+2}$ specifically

$\sqrt{\frac{\sigma_1^2}{100} + \frac{\sigma_2^2}{100}} = \sigma_{1+2}$ $\bar{Y}_1 + \bar{Y}_2 \sim N\left(\mu_1 + \mu_2, \sqrt{\left(\frac{\sigma_1^2 + \sigma_2^2}{100}\right)}\right)$

c.) $\frac{\bar{Y}_1 + \bar{Y}_2}{2} - \bar{Y}_3$ (Normal distribution)

$= \frac{\mu_1 + \mu_2}{2} - \mu_3$

$\mu_{1+2-3} = \frac{\mu_1 + \mu_2}{2} - \mu_3$

$\frac{\sigma_1^2 + \sigma_2^2}{2} + \sigma_3^2$

$= \frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2 + \sigma_3^2$

$= \sqrt{\frac{1}{4}(\sigma_1^2) + \frac{1}{4}(\sigma_2^2) + \sigma_3^2} = \sigma_{1+2+3}$

Specifically $\frac{\bar{Y}_1 + \bar{Y}_2}{2} - \bar{Y}_3 \sim N\left(\frac{\mu_1 + \mu_2}{2} + \mu_3, \sqrt{\left(\frac{\sigma_1^2 + \sigma_2^2}{4} + \sigma_3^2\right)}\right)$

D.) $\overline{Y_1} + \overline{Y_3} - 2\overline{Y_2}$

$\mu_1 + \mu_3 - 2\mu_2 = \mu_{1+3-2}$

$\sigma_1^2 + \sigma_3^2 + 2\sigma_3^2 = \sigma^2_{1+3+2}$

$\sqrt{\sigma_1^2 + \sigma_3^2 + 4(\sigma_3^2)} = \sigma_{1+3+2}$

Specifically, not normally distributed)

5a.) $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

Let i be the categorical variable for smokers & non-smokers respectively $\{i = 1, 2\}$

B.) $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(s_1)^2}{n_1} + \frac{(s_2)^2}{n_2}}} \rightarrow \frac{150 - 139.18}{\sqrt{\frac{(27.49)^2}{266} + \frac{(27.49)^2}{234}}}$$

$$= \frac{10.82}{\sqrt{2.841 + 3.229}} = \frac{10.82}{2.4637}$$

$t_{234+266} = 4.3917$ ~~(233)~~

P-val < 0.0001

c.) $Pr(\mu_1 > |t_s|) =$ [redacted] % ; By t-table.

This means that the probability of observing our data or something more extreme is 0.0016%, very low!

d.) Given an α value of 5%, we can reject the null hypothesis, that the mean systolic blood pressures of smokers are equal to that of non-smokers.

Gurtej Singh

6.) A Type I error in this problem would have been incorrectly rejecting the claim that the mean systolic blood pressures of smokers are equal to those of non-smokers, even though it was true, and both are equal.

b.) A type II error would have occurred if we rejected the alternative hypothesis - that the mean systolic smoker blood pressures are not equal to those of non-smokers when in fact it was true.

c) $(1-\alpha)100\% = 99\%$ CI

$\alpha = 0.01$

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{(1-\alpha/2,\, n_1+n_2-2)} \cdot \sqrt{Sp^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$150_{.03} - 139.18 \pm \underbrace{t_{(0.995,\, 498)}}_{=2.5851} \cdot \underbrace{2.4637}_{=S}$$

$$10.85 \pm 6.36685$$

$$= \underline{(4.483,\, 17.216)}\ \text{C.I } 99\%$$

d.) With 99% confidence, we can say that the true mean systolic blood pressure of smokers is between 4.483 to 17.216 greater than that of non-smokers. Or we could say that the mean systolic blood pressure of non-smokers is 4.483 to 17.216 lower than those of smokers (w 99% confidence).

c.) The largest difference between the two groups would arise when comparing smokers to non-smokers, and ~~[illegible]~~ would be 17.216 greater for smokers (mean systolic blood pressures) than non-smokers.

with 99% confidence

Gurtej Singh

2.) This would be a mixed study because the subjects chosen per high school are not random, and this could influence the outcome. However, we randomly assign these individuals a different type of treatment, and as such, is a combination of observational & experimental.

b.) The primary variable of interest would be the students' performance on the standardized math test.

c.) The explanatory variables are the school sections & type of curriculum.

Factor levels:

[redacted] schools → [redacted] sections (4 per)

Curriculum → Standard, computer-based

d.) Treatments: 24 possible.

12 sections that teach with standard.
12 sections that teach with computer based.

ie.	section	1	2	3	4	[redacted]
Curric-	Computer	1C	2C	3C	4C	[redacted]
ulum	Standard	1S	2S	3S	4S	[redacted]

a.) This study is a mixed aswell since the patients chosen are not random, and could-for example be chosen from locations that may influence theyre fitness status which therefor influences the results. However, each patient is randomly assigned to a doctor, and is therefor, partly experimental.

b.) The primary variable of interest is the # of days post surgery each patient takes to recover (succesful rehabilitation).

c.) The explanatory variables include
- prior fitness status (Physical)
- the doctor they were paired with

Factor-levels:

Prior fitness status
- below average
- average
- above average

Doctor
- 1
- 2
- 3

D.) Possible Treatments = 9

Fitness \ Doctor	1	2	3
below	below, 1	below, 2	below, 3
average	averge, 1	average, 2	average, 3
above	above, 1	above, 2	above, 3

Scanned with CamScanner