Sample exam B

Stephanie Ranft S2459825

April 14, 2020

Statistics 2
PSBE2-07

A small study is performed using scores attained by various people categorised into groups. The coding is done as follows:

Group 1		
Score	C1	C2
6	-0,75	-0,5
7	-0,75	-0,5
8	-0,75	-0,5
8	-0,75	-0,5

Group 2		
Score	C1	C2
8	1	0
9	1	0
9	1	0

Group 3		
Score	C1	C2
5	0	1
6	0	1

Figure 1

Questions 1–3 will deal with this study.

1. For the model $\widehat{\text{score}} = B_0 + B_1 C_1 + B_2 C_2$, compute B_1.
   (a) -1.833
   (b) -1.639
   (c) 1.333
   (d) 5.500

2. What is the correct interpretation of B_0?
   (a) Sample mean of the reference group.
   (b) Sample mean of all groups.
   (c) Population mean of all groups.
   (d) Population mean of the reference group.

3. SPSS provides $pr_2^2 = 0.77$. What is the best interpretation of pr_2^2?
   (a) When comparing group 2 with the other two groups combined, 77% of the variance in score can be explained.
   (b) When comparing group 3 with the other two groups combined, 77% of the variance in score can be explained.
   (c) When comparing group 2 with group 1, 77% of the variance in score can be explained.
   (d) When comparing group 3 with group 1, 77% of the variance in score can be explained.

4. A random and representative sample is taken from a population. For each individual in the sample, the following are recorded: sex (0 = male or 1 = female), education (0 = below average, 1 = above average), and the dependent variable income. The total sample of size $n = 200$ is distributed as follows:

	Below	Above	
Male	40	60	100
Female	60	40	100
	100	100	200

Table 1

Which hypothesis is not appropriate?

(a) H_0 : the group distributions are equal.
(b) $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$.
(c) $H_0 : \chi^2 = 0$.
(d) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$.

5. For a sample, the following information is collected: $n = 47$, mean = 12, median = 6, and variance = 9. After standardization, the 15th observation is transformed to a value of 1. What was the original value in the sample of x_{15}?

(a) $x_{15} = -3$.
(b) $x_{15} = -1$.
(c) $x_{15} = 15$.
(d) $x_{15} = 21$.

A moderated regression analysis, with dependent variable *dv* and centered independent variables *idv* (with $sd = 1.29$) and *mod* ($sd = 0.41$), provided the following SPSS output, to be used in questions 6–7:

Coefficients[a]

	B	Std. Error	Beta	t	Sig.
(Constant)	-,300	,333		-,900	,369
idv	6,293	,354	,619	17,788	,000
mod	3,227	,342	,331	9,447	,000
idv*mod	6,096	,400	,536	15,232	,000

a. Dependent Variable: dv

Figure 2

6. Consider the following two claims, and answer whether they are true:

A: "The simple slope for $mod = 0.5$ is equal to 9.34."

B: "When $mod = -1.889$, *dv* will be [approximately] the same for low values of *idv* as for high values of *idv*."

(a) Claim A is true. Claim B is true.
(b) Claim A is true. Claim B is false.
(c) Claim A is false. Claim B is true.
(d) Claim A is false. Claim B is false.

7. External software provided $\text{Cov}(B_{13}) = 0.120$. Provide the standard error of the simple slope for the regression of *dv* on *idv* at $mod = 0.2$.

(a) 0.43.
(b) 0.69.
(c) 0.42.
(d) 0.99.

8. Compute the simple regression equation of dv on for idv for mod at the default ‘high’ level.

   (a) $\widehat{dv} = 0.81 + 8.38idv$.
   (b) $\widehat{dv} = 1.02 + 8.79idv$.
   (c) $\widehat{dv} = 2.93 + 12.39idv$.
   (d) $\widehat{dv} = 7.82 + 11.09mod$.

The data in the table below shows group means in a 2 × 2 design (factors A and B). The numbers in parentheses indicate the sample sizes.

	Factor B	
Factor A	B_1	B_2
A_1	23 ($n = 12$)	33 ($n = 18$)
A_2	15 ($n = 22$)	10 ($n = 14$)

Table 2

Use this setting to answer questions 9–11.

9. What is the weighted mean of the two groups of B_1 defined by factor A?

   (a) 17.8.
   (b) 20.3.
   (c) 23.0.
   (d) 19.0.

10. What is the weighted mean of the groups?

    (a) 17.8.
    (b) 20.3.
    (c) 23.0.
    (d) 19.0.

11. Which interpretation is most correct?

    (a) There is a main effect for both factors.
    (b) There is an interaction effect.
    (c) There is a main effect for factor A, and an interaction effect.
    (d) There are two main effects, and an interaction effect.

12. In a one-way ANOVA, the following regression model is estimated:
    $\hat{Y} = B_0 + B_1C_1 + B_2C_2 + B_3C_3 + B_4C_4 + B_5C_5$.
    How many levels of the independent variable exist?

    (a) 4
    (b) 5
    (c) 6
    (d) It is not possible to answer.

13. Which of the following assumptions is not required in an analysis of variance?

    (a) The scores of the dependent variable are normally distributed.
    (b) The scores of the independent variables are normally distributed.
    (c) The variances of the scores of the dependent variable between the groups are equal.
    (d) There is a relationship between the independent and the dependent variables.

14. A researcher wants to know if caffeine in sports drink improves performance. She chooses a between-groups design with two factors: Caffeine (five conditions: C1 through C5) and Gender (two conditions). One hundred and twenty athletes, divided over the experiment conditions, do an endurance test. The dose of caffeine is systematically increased in equal steps from conditions C2 through C5; no caffeine is added to the sports drink in condition C1.

    The following research questions are examined:

    I Does caffeine have an effect on performance?

    II Do men and women perform differently on the endurance test?

    III Is the change in performance according to caffeine different between men and women?

    How many variables should be included in a regression model in order to answer all three research questions above?

    (a) 5
    (b) 9
    (c) 7
    (d) 17

15. In a multiple regression analysis, the percentage of explained variance can always be calculated by:

    (a) Adding the squared correlations between each independent variable and the dependent variable.
    (b) Adding the squares of the semipartial correlations in the final model.
    (c) Adding the squares of the standardized regression coefficients.
    (d) Squaring the correlation between the observed and the estimated Y values.

16. A researcher has scores on two predictors (A and B, both categorical with two levels each) and one dependent variable (continuous).

    The researcher wants to test whether the main effects are significant. Which analytical strategy is the best to choose among the four options given?

    (a) A regression analysis in which the product between the independent and dependent variables is included in the model.
    (b) A regression analysis with all effects (main and interaction) added which relies on unweighted effects coding for both factors.
    (c) A regression analysis with only the interaction effect added which relies on contrast coding for both factors.
    (d) An analysis of variance where the difference between the two A-conditions is tested across factor B, followed by an analysis of variance where the difference between the two B-conditions is tested across factor A.

A researcher wants to regress Y (continuous variable) on A (categorical variable with four levels: N, P, S, and C).

Group	Mean	Sample size	Standard deviation
N	31	44	3.4
P	62	56	2.8
S	38	37	4.3
C	42	51	3.6

Table 3

Suppose the researcher wants to use contrast coding so that each of the following contrasts is directly tested via the corresponding regression coefficient:
Contrast 1: $H_0 : \frac{\mu_N+\mu_P+\mu_S}{3} = \mu_C$
Contrast 2: $H_0 : \frac{\mu_N+\mu_P}{2} = \mu_S$
Contrast 3: $H_0 : \mu_N = \mu_P$.

Use this setting to answer questions 17–19.

17. What is this method of contrasting called?
    (a) Deviation.
    (b) Simple.
    (c) Helmert.
    (d) Difference.

18. Calculate s_p.
    (a) 12.18
    (b) 3.49
    (c) 3.53
    (d) 3.45

19. Which is the correct statistic for testing contrast 2?
    (a) $t_{184} = 12.630$.
    (b) $t_{185} = 12.630$.
    (c) $t_{184} = 11.194$.
    (d) $t_{185} = 11.194$.

END OF EXAM

Solutions:
The regression equations for each group are as follows:

$$\mu_1 = \beta_0 - 0.75\beta_1 - 0.5\beta_2; \qquad \mu_2 = \beta_0 + \beta_1; \qquad \mu_3 = \beta_0 + \beta_2.$$

Using sample estimates, we have

$$\bar{y}_1 = B_0 - 0.75B_1 - 0.5B_2; \qquad \bar{y}_2 = B_0 + B_1; \qquad \bar{y}_3 = B_0 + B_2.$$

Calculate the means and rearrange to solve for B_0, B_1, and B_2:

$$7.25 = B_0 - 0.75B_1 - 0.5B_2; \qquad 8.67 = B_0 + B_1; \qquad 5.5 = B_0 + B_2.$$

$$\implies B_1 - B_2 = 8.67 - 5.5 = 3.167 \implies B_1 = 3.167 + B_2$$

$$\implies B_0 = 8.67 - B_1 = 8.67 - [3.167 + B_2] = 5.5 - B_2.$$

$$\implies 7.25 = B_0 - 0.75B_1 - 0.5B_2 = [5.5 - B_2] - 0.75[3.167 + B_2] - 0.5B_2 = 3.125 - 2.25B_2.$$

$$\implies B_2 = -\frac{7.25 - 3.125}{2.25} = -1.833; \quad B_1 = 3.167 - 1.833 = 1.333; \quad B_0 = 5.5 + 1.833 = 7.333.$$

Note that,

$$7.333 = \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3}{3} \implies \bar{y} = 7.333 + \overbrace{(1.333 \times -0.75 - 1.833 \times -0.5)}^{-0.0833} z_1 + 1.333z_2 - 1.833z_3,$$

where $z_1 = 1$ if in group 1, $z_2 = 1$ if in group 2, and $z_3 = 1$ if in group 3. That is, $\bar{y}_1 = 7.333 - 0.833 = 7.25$, $\bar{y}_2 = 7.333 + 1.333 = 8.67$, and $\bar{y}_3 = 7.333 - 1.833 = 5.5$.

1. Using the above, $B_1 = 1.333$ (c).

2. B_0 is the sample estimate of β_0, which is the mean of the population across all three groups (b).

3. The partial regression coefficient pr_2 measures the linear association of group 1 with group 2, when the effects of group 3 has be partialled out of **both**. The squared value pr_2^2 is a measure of explained variance (c).

4. The regression equation looks like: $Y = \alpha + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3 + \varepsilon$, where

		z_1	z_2	z_3
Male	Below	0	0	0
	Above	1	0	0
Female	Below	0	1	0
	Above	0	0	1

Table 4

Therefore (d).

5. $(x_{15} - \bar{x})/s_x = z_{15}$, therefore $x_{15} = 1 \times \sqrt{9} + 12 = 15$; it is (c).

6. We have,

$$\overline{dv}\,(mod = 0.5) = -0.3 + 6.293idv + 3.227 \times 0.5 + 6.096idv \times 0.5 = 1.3135 + 9.341idv.$$

The effects of idv on dv are the same for any value of idv when

$$6.293idv + 6.096idv \times mod = 0 \implies mod = -\frac{6.293}{6.096} = -1.032.$$

The effects of mod on dv are the same for any value of mod when

$$3.227mod + 6.096idv \times mod = 0 \implies idv = -\frac{6.293}{3.227} = -1.889.$$

Therefore (b).

7. We have,

$$
\begin{aligned}
dv &= \alpha + \beta_1 idv + \beta_2 mod + \beta_2 idv \times mod + \varepsilon = (\alpha + \beta_2 mod) + (\beta_1 + \beta_3 mod)\, idv + \varepsilon. \\
\implies dv\,(mod = 0.2) &= (\alpha + 0.2\beta_2) + (\beta_1 + 0.2\beta_3)\, idv + \varepsilon. \\
\implies \operatorname{Var}(\beta_1 + 0.2\beta_3) &= \operatorname{Cov}(\beta_1 + 0.2\beta_3, \beta_1 + 0.2\beta_3) = \operatorname{Var}(\beta_1) + 2 \times 0.2 \operatorname{Cov}(\beta_1, \beta_3) + (0.2)^2 \operatorname{Var}(\beta_3) \\
&= 0.354^2 + 2 \times 0.2 \times 0.12 + (0.2)^2 \times 0.4^2 = 0.1797.
\end{aligned}
$$

Therefore the standard error for the simple slope for the regression of dv on idv when $mod = 0.2$ is $\sqrt{0.1797} = 0.42$; it is (c).

8. The default 'high' level is $mod = 1$, therefore (c):

$$
\begin{aligned}
\widehat{dv} &= -0.3 + 6.293 idv + 3.227 mod + 6.096 idv \times mod \\
&= (-0.3 + 3.227 mod) + (6.293 + 6.096 mod)\, idv. \\
\implies \widehat{dv}\,(mod = 1) &= (-0.3 + 3.227) + (6.293 + 6.096)\, idv = 2.927 + 12.389 idv.
\end{aligned}
$$

9. This is calculated as

$$
\bar{y}^w_{B_1} = \frac{\bar{y}_{B_1,A_1} \times n_{B_1,A_1} + \bar{y}_{B_1,A_2} \times n_{B_1,A_2}}{n_{B_1,A_1} + n_{B_1,A_2}} = \frac{23 \times 12 + 15 \times 22}{12 + 22} = 17.824,
$$

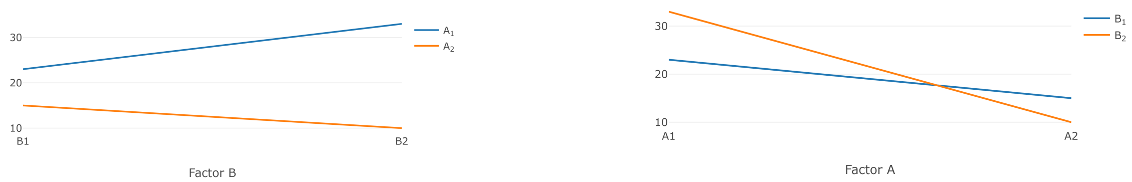
therefore it is (a).

10. This is calculated as

$$
\begin{aligned}
\bar{y}^w &= \frac{\bar{y}_{B_1,A_1} \times n_{B_1,A_1} + \bar{y}_{B_1,A_2} \times n_{B_1,A_2} + \bar{y}_{B_2,A_1} \times n_{B_2,A_1} + \bar{y}_{B_2,A_2} \times n_{B_2,A_2}}{n_{B_1,A_1} + n_{B_1,A_2} + n_{B_2,A_1} + n_{B_2,A_2}} \\
&= \frac{23 \times 12 + 15 \times 22 + 33 \times 18 + 10 \times 14}{12 + 22 + 33 + 10} = 20.303.
\end{aligned}
$$

therefore it is (b).

11. The effects diagram can be displayed in either of the following ways, and both display two main effects and an interaction effect; it is (d).



(a) (b)

Figure 3

12. There are five beta coefficients plus the intercept means that there are six groups, therefore it is (c).

13. (b).

14. This a 6×2 two-way ANOVA: $I = 5$ and $J = 2$, so the df for factor *Caffeine* is $4 = I - 1$, for factor *Gender* is $1 = J - 1$, and the interaction effect is $5 = (I - 1)(J - 1)$. Summing these together:

$$
(I - 1) + (J - 1) + (I - 1)(J - 1) = 4 + 1 + 5 = 9 = 5 \times 2 - 1 = IJ - 1.
$$

Recall that the df error is $n - IJ = (n - 1) - (IJ - 1)$! Therefore it is (b).

15. Recall the ballentine diagram for two IVs and one DV:

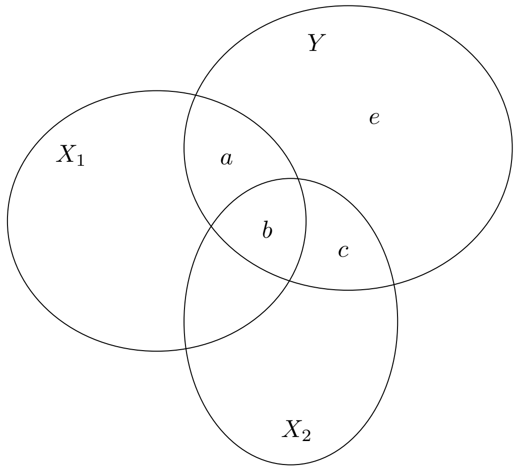


Figure 4

Var$(Y) = a + b + c + e = R^2 + e$, therefore it is (b).

16. (d).

17. (c).

18. The mean squared error is calculated as:

$$MSE = \frac{s_N^2 (n_N - 1) + s_P^2 (n_P - 1) + s_S^2 (n_S - 1) + s_C^2 (n_C - 1)}{n_N + n_P + n_S + n_C - 4}$$
$$= \frac{3.4^2 \times 43 + 2.8^2 \times 55 + 4.3^2 \times 36 + 3.6^2 \times 50}{43 + 55 + 36 + 50} = 12.184.$$

Therefore $s_p = \sqrt{MSE} = \sqrt{12.184} = 3.490$ and it is (b).

19. Firstly, the contrast estimate c is calculated as:

$$c = \frac{31 + 62}{2} - 37 = 9.5 \qquad \Longrightarrow \qquad t = \frac{c}{s_p \sqrt{\frac{0.5^2}{44} + \frac{0.5^2}{56} + \frac{(-1)^2}{37}}} = 12.630 \sim t\,(df = n - I = 184)\,.$$

Therefore it is (a).