Stat E-150 Section #5

Neha Dhawan

dhawan@g.harvard.edu

Office hours by request for one-one help

Two- way ANOVA

- 2 categorical predictors
  - Each predictor has own # of levels

Main effect

Main effect

Bit of review

- **Factorial Crossing:** In ANOVA, two factors are crossed if all combinations of levels of two factors appear in the design
  - Factorial cross because we want to see the interaction of the two factors
- **Balanced design** is if the factors are fully crossed and there are equal numbers of observations per cell(equal #s of experimental units)

Factorial crossing 2-way ANOVA

Factorial Crossing

Factor A: Depression intervention **Treatments**

Factor B: Gender

	Control	CBT	Meds
Male	Male/ control **15**	Male/CBT **15**	Male/ meds **15**
Female	Female / control **15**	Female/CBT **15**	Female/ meds **15**

Balanced Design

Previous example

Researchers at MGH investigated whether the distribution of birthweights differed among mothers of different races (white, black, hispanic, and other). They also want to know whether smoking or not during pregnancy also affects baby birthweight.

0. Ask your research questions: What is it you want to know?

Does the mother’s race affect the birthweight of the baby?

Does smoking or not during pregnancy affect baby’s birthweight?

Does smoking during pregnancy affect baby’s birthweight for mom’s of a particular race?

1. Choose a form for the model: Identify the variables and their types; Examine graphs to help identify the appropriate model

Birthweight- response- quantitative

Race- predictor- categorical

- Race: 4 levels= hispanic, black, white and other

Smoking- predictor- categorical

- Binary- two levels= 0 (no), 1(yes)

Two-way ANOVA!

3. Assess how well the model fits the data: **Verify assumptions,** examine the residuals, compute significance, refine model as needed

- **Random**- random sample
- Distribution of errors
  - **Zero Mean-** always true due to analysis
  - **Constant Variance**
    - Residual plot
    - Rule of thumb: largest std devi/ smallest (gets crazy to do)
    - Levene's test of equal variance
  - **Normality** - NPP
  - **Independence**- errors independent of each other; presume with random assumption

Levene's test

$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2 = \sigma_7^2 = \sigma_8^2$

H_a: Not all variances are equal

Levene's Test of Equality of Error Variances[a]

Dependent Variable: BirthWeightGm

F	df1	df2	Sig.
.942	7	1437	.473

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

Since we got a large p-value (.473), we fail to reject the null hypothesis and can conclude that all the variances are equal

Rule of thumb

- Get a value <2 OK
- Value >2 concern

Largest StDevi = 748
Smallest StDevi 312
= 2.397

OK!

Descriptive Statistics

Dependent Variable: BirthWeightGm

Smoke	MomRace	Mean	Std. Deviation	N
0	black	3157.9349	649.04032	289
	hispanic	3360.0000	509.23476	162
	other	3350.0250	516.21677	42
	white	3389.3703	639.52686	743
	Total	3330.0699	629.12379	1236
1	black	2984.1750	748.39756	42
	hispanic	2494.8000	.	1
	other	3118.5000	312.62223	6
	white	3124.7016	589.98674	160
	Total	3093.2699	619.46805	209
Total	black	3135.8869	663.76382	331
	hispanic	3354.6920	512.16380	163
	other	3321.0844	498.84447	48
	white	3342.4744	638.77085	903
	Total	3295.8199	633.03095	1445

2. Fit the model to the data: Use the sample data to create statistics, which estimate the values of the model parameters

Q. Does the mother's race affect the birthweight of the baby?

H_0: $\mu_{black} = \mu_{hispanic} = \mu_{other} = \mu_{white}$
H_a: the means are not all equal

Since $p<.05$, we can reject the null hypothesis, and conclude that the **mean** birthweight differs among the different races.

But we don't know how they differ....**post hoc**!

Post hocs note

When don't you run a post hoc test?

- When you have an interaction
- Only have 2 levels in a factor

When DO you run a post hoc test?

- On a main effect when you don't have an interaction
- Have more than 2 levels in a factor and need to see how groups differ

Q. Does smoking or not during pregnancy affect baby's birthweight?

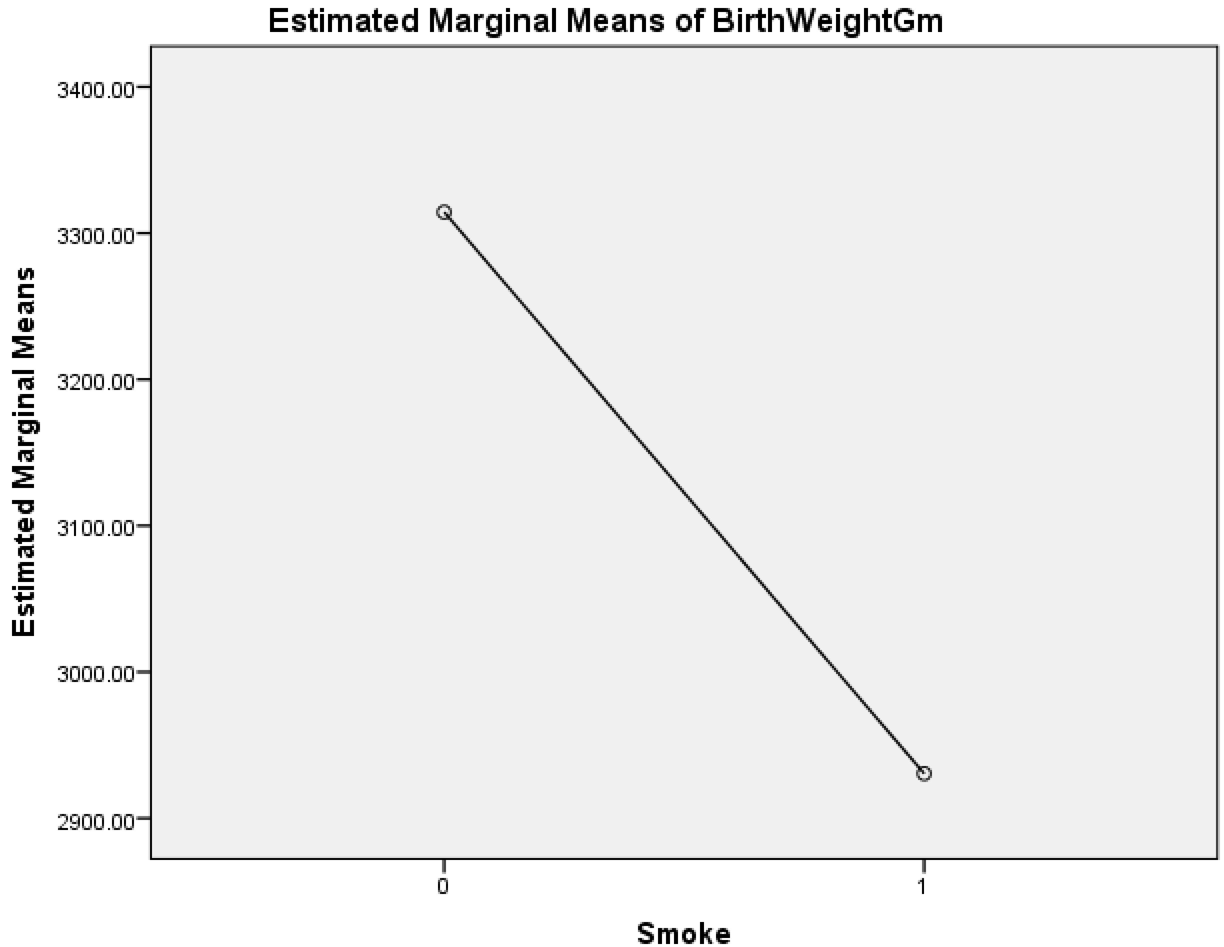
H_0: $\mu_0 = \mu_1$

H_a: the means are not all equal

Since p<.05, we can reject the null hypothesis, and conclude that the **mean** birthweight differs among mothers who smoke or not.

Only two levels so don't need post hoc!

We can tell how they differ!



Based on the graph, the mean birthweight for babies to moms that smoke is lower than for moms that do not smoke during pregnancy.

Q. Does smoking during pregnancy affect baby's birthweight for mom's from a particular race?

Factor A: Race

		Hispanic	Black	White	Other
Factor B:	0 (no)	Hispanic/0	Black/0	White/0	Other/0
Smokes	1 (yes)	Hispanic/1	Black/1	White/1	Other/1

H_0: $\mu_{\text{Hispanic*0}} = \mu_{\text{Hispanic*1}} = \mu_{\text{Black*0}} = \mu_{\text{Black*1}} = \mu_{\text{White*0}} = \mu_{\text{White*1}}$
$= \mu_{\text{Other*0}} = \mu_{\text{Other*1}}$

H_a: the means are not all equal

Ah so many means... what does this **mean**?

It means essentially the same thing as our other hypotheses:

- Does one particular group (factorial crossed for 2 particular levels) do particularly well than the other groups?
  - In context:- Do hispanic moms who smoke give birth to particularly low birthweight babies than white mom who don't smoke?

What does an interaction really mean?

If there is an interaction:

- It means that the two factors affect one another
  - One level of factor A interacts with another level of factor B
- It means that the two factors are NOT independent of each other
- If the factors are not independent of each other, than you cannot interpret the main effects
  - CANNOT run post-hoc tests
- Graph lines intersect each other

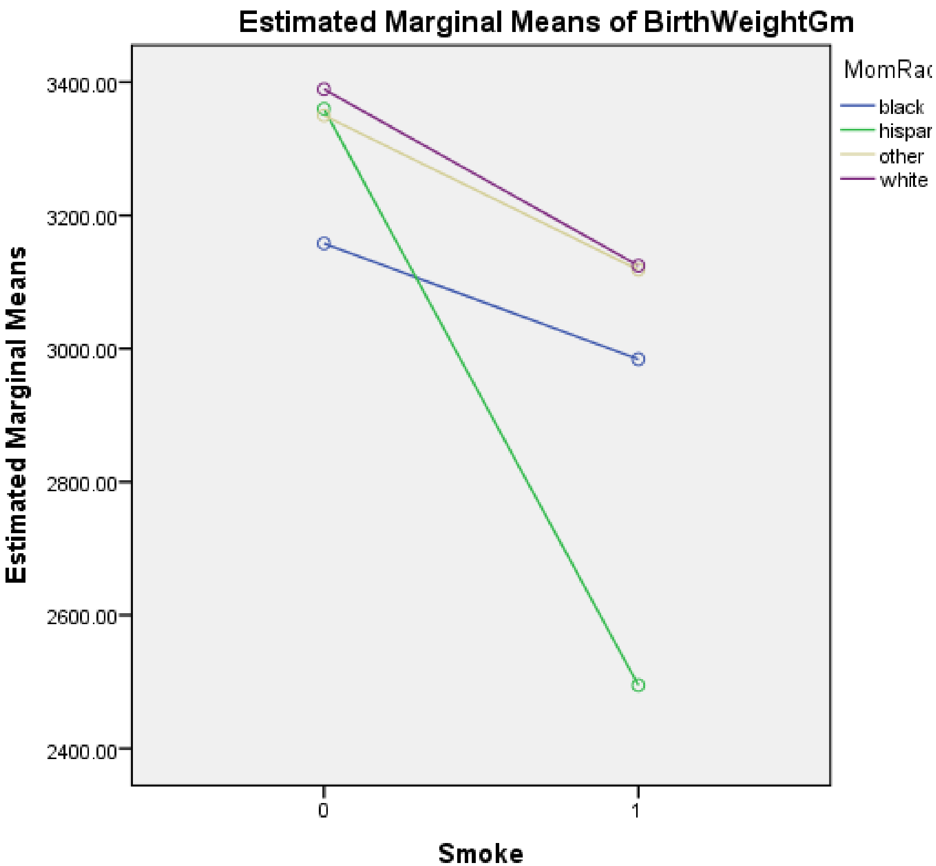
Note: we need to have at least 2 cases per cell to have an interaction

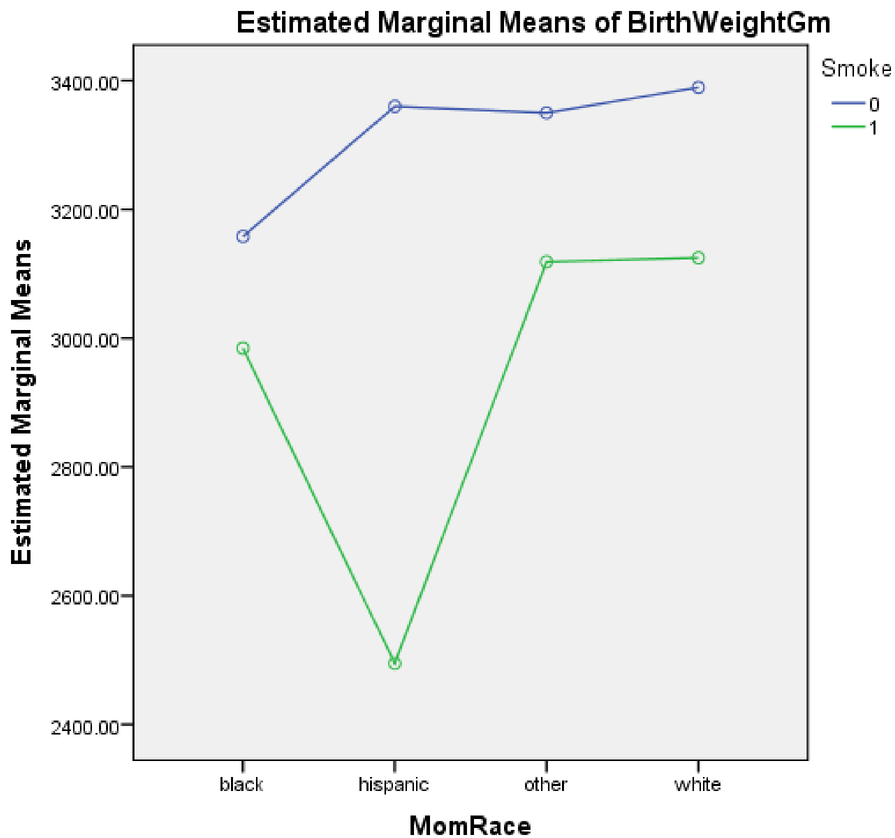
If there is NO interaction:

- The two factors ARE independent of each other
  - All treatments in one row do better than all treatment in other row
- Graph lines parallel to each other
- You can interpret the main effects normally
  - CAN run post-hocs

Inferring interaction from the graph

- Always plot both ways to know if there is an interaction; if both graphs show the lines crossing then there is an interaction





Inferring interaction from a table

	Male	Female
Boston	135	156
New York	112	133
Detroit	176	185
Houston	100	120

Dependent variable mean GRE scores

	Male	Female
Boston	156	134
New York	154	198
Detroit	165	154
Houston	132	161

Q. Does smoking during pregnancy affect baby's birthweight for mom's from a particular race?

H_0: the main effect of each factor is the same for each level of the other factor

H_a: the two factors interact

Since $p>.05$, we fail to reject the null hypothesis, and conclude that the mean baby birthweight for mothers who smoke or not do not depend on the mother's race.

Q. Does the mother's race affect the birthweight of the baby?

Posthoc to see how birthweights differ across race

H_0: $\mu_i = \mu_j$

H_a: $\mu_i \neq \mu_j$

Based on the output we can see that black moms have significantly lower birthweight babies than hispanic or white moms ($p<0.001$)

Repeated Measures ANOVA

- Blocking as a within-subjects design
- Each subject gets all treatments, each treatment is given once to all subjects, and the order of the conditions is randomized
- Benefit is that each person is their own control
- Longitudinal study with different time points

Smoking again

Let's say you want to test an intervention method to stop mothers from smoking. You randomly assign mothers to one of two intervention methods: Negative conditioning and Electrotherapy and you include a control group. You measure the number of cigarettes they smoke as they listen to jazz, rock, and hip-hop background music at different time points during the therapy.

0. Ask your research question: What is it you want to know?

Does one of the intervention method help mothers stop smoking?

1. **Choose a form for the model: Identify the variables and their types**; Examine graphs to help identify the appropriate model

Factor 1– intervention

- Three levels; ECT, Negative conditioning, and Control
- Between subjects

Factor 2– Music

- Three levels; jazz, rock, hip-hop
- Within subjects

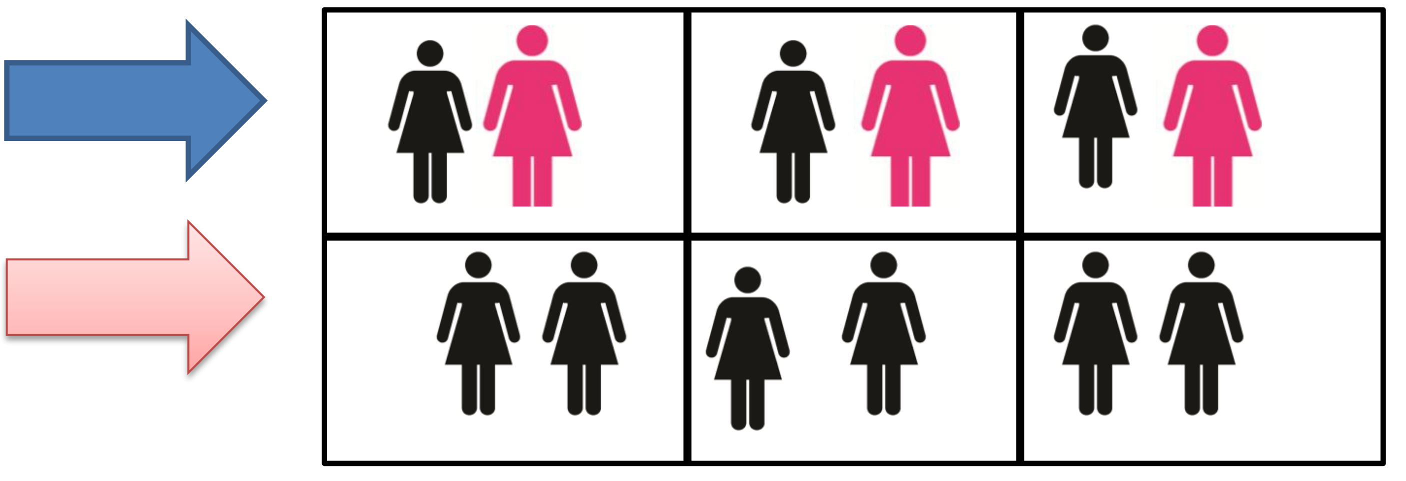
Within subj design; Repeated Measures ANOVA!

Factor A: Music

Factor B: Intervention	Jazz	Rock	Hip-hop
ECT	Jazz/ECT	Rock/ECT	Hip-hop/ECT
NCond	Jazz/NCond	Rock/NCond	Hip-hop/NCond
Control	Jazz/Control	Rock/Control	Hip-hop/Control

ECT

NCond



Sphericity Assumption

- Variances between the differences of all combinations of related levels are equal
- Mauchly’s test of spehricity
- Still need to test normality and randomness
- Want to fail to reject like Levenes’s test
- If reject hypothesis then use Greenhouse-geisser correction

Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
music	.868	3.695	2	.158	.883	1.000	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept + intervention
Within Subjects Design: music

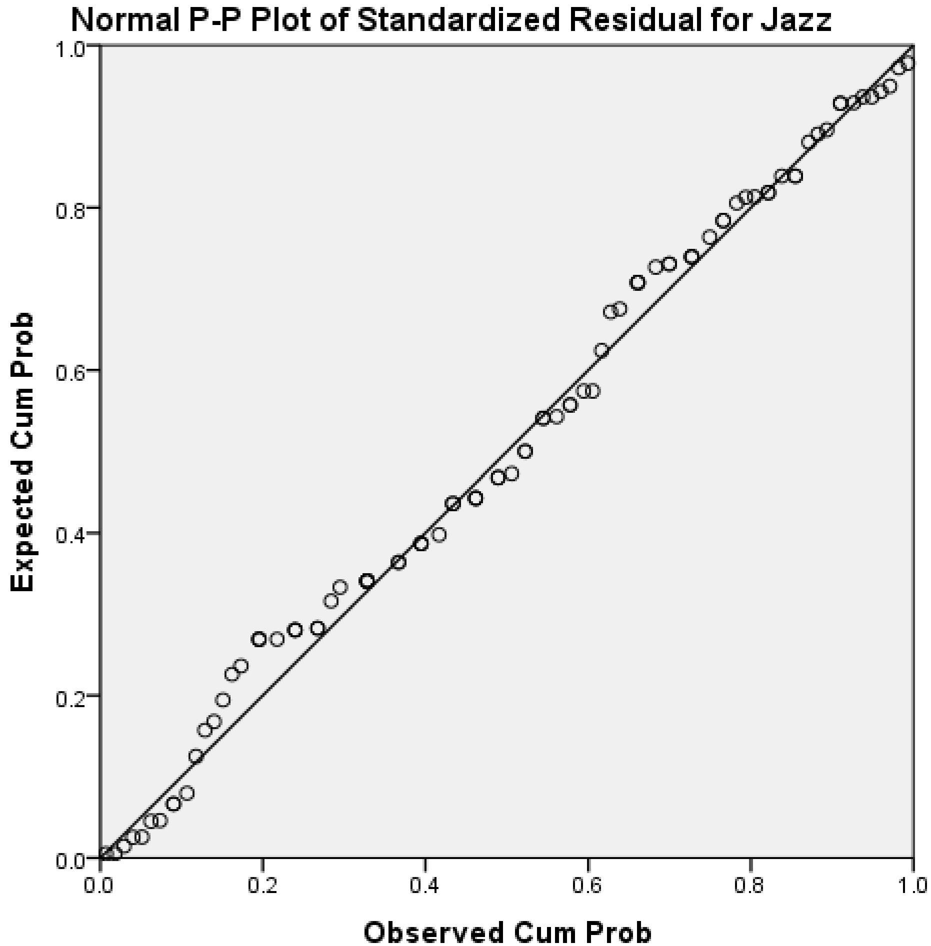
b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

H_0: the variances of the differences are equal

H_a: the variances of the differences are not equal

Since we got a large p-value (.158), we fail to reject the null hypothesis and can conclude that all the variances of the differences are equal. We meet our assumption of Sphericity.

Normality



Normal P-P Plot of Standardized Residual for Jazz
1.0
0.8
0.6
0.4
0.2
0.0
Expected Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob

Test the factors

Q. Is there a difference between the intervention groups? (Between subjects)

H_0: $\mu_{ECT} = \mu_{NCond} = \mu_{Control}$

H_a: the means are not all equal

Since $p<0.05$ (Sphericity assumed), we can reject the null hypothesis, and conclude that the **mean** number of cigarettes smoked significantly differ amongst intervention types.

But this doesn't tell us **how** they differ! Posthoc!

PostHoc

difference between the intervention groups (Between subjects)

H_0: $\mu_i = \mu_j$

H_a: $\mu_i \neq \mu_j$

Based on the output we can see that mean number of cigarettes smoked are significantly lower for ECT(Group 1) in comparison to NCond and Control (Group 2 and 3) and NCond is significantly higher than ECT but lower than Controls (all $p<0.001$)

Test the factors

Q. Is there a difference between the type of music? (Within subjects)

H_0: $\mu_{Jazz} = \mu_{rock} = \mu_{hip\text{-}hop}$

H_a: the means are not all equal

Since $p<0.001$ (Sphericity assumed), we can reject the null hypothesis, and conclude that the **mean** number of cigarettes smoked significantly differ amongst music type.

But this doesn't tell us **how** they differ! Posthoc!

PostHoc

difference between the intervention groups (within subjects)

H_0: $\mu_i = \mu_j$

H_a: $\mu_i \neq \mu_j$

Based on the output we can see that mean number of cigarettes smoked are significantly lower for Rock(Group 2) in comparison to jazz and hip-hop (Group 1 and 3) and Jazz is significantly higher than rock but lower than hip-hop(all $p<0.001$)

But wait, what about the interaction?!

H_0: the main effect of each factor is the same for each level of the other factor

H_a: the two factors interact

Since $p<.05$, we reject the null hypothesis, and conclude that the mean number of cigarettes mothers smoked during each music trial depends on the intervention group

Test time!!

Suppose we are interested in analyzing the effect of gender and age on income. We will treat the ages as categories: ages 18 - 29, 30 - 39, 40 - 49, and 50 or higher

a) What analysis would you use for this data and what are the questions you want to address?

		Gender	
		Female	Male
Age category	18 - 29	Income for n subjects Female, 18-29	Income for n subjects Male, 18-29
	30 - 39	Income for n subjects Female, 30-39	Income for n subjects Male, 30-39
	40 - 49	Income for n subjects Female, 40-49	Income for n subjects Male, 40-49
	≥ 50	Income for n subjects Female, ≥ 50	Income for n subjects Male, ≥ 50

Tests of Between-Subjects Effects

Dependent Variable:rincom2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	2236.798[a]	7	319.543	15.360	.000
Intercept	121426.954	1	121426.954	5836.953	.000
agecat4	1350.516	3	450.172	21.640	.000
sex	842.073	1	842.073	40.478	.000
agecat4 * sex	60.316	3	20.105	.966	.408
Error	14583.002	701	20.803		
Total	149966.000	709			
Corrected Total	16819.800	708			

a. R Squared = .133 (Adjusted R Squared = .124)

b) What can you conclude about the main effects from the output?

c) Is there a significant interaction?

Multiple Comparisons

rincom2

(I) 4 categories of age	(J) 4 categories of age	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
18-29	30-39	-2.1172*	.49334	.000	-3.4997	-.7348
	40-49	-3.9165*	.51013	.000	-5.3461	-2.4870
	50+	-3.2930*	.54550	.000	-4.8216	-1.7643
30-39	18-29	2.1172*	.49334	.000	.7348	3.4997
	40-49	-1.7993*	.44207	.001	-3.0381	-.5605
	50+	-1.1757	.48245	.116	-2.5277	.1762
40-49	18-29	3.9165*	.51013	.000	2.4870	5.3461
	30-39	1.7993*	.44207	.001	.5605	3.0381
	50+	.6235	.49961	.669	-.7765	2.0236
50+	18-29	3.2930*	.54550	.000	1.7643	4.8216
	30-39	1.1757	.48245	.116	-.1762	2.5277
	40-49	-.6235	.49961	.669	-2.0236	.7765

Based on observed means.
The error term is Mean Square(Error) = 20.803.
*. The mean difference is significant at the .05 level.

d) What can you conclude from the post hoc test?

Solutions

a) We would use **Two-way ANOVA** to analyze our data because we have a two **categorical predictors/factors.** The research questions we want to address are:

- **Are there significant mean differences for income between male and female employees?**
- **Are there significant mean differences for income by age category among employees?**
- **Is there a significant interaction on income between gender and age category?**

(8)

Solutions

b) **H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$**

H_a: the means are not all equal

Since **p-value <0.05**, we **reject the null hypothesis**. The data provides **evidence of a significant difference in the mean income amongst the different age categories**. (5)

H_0: $\mu_1 = \mu_2$

H_a: the means are not all equal

Since **p-value <0.05**, we **reject the null hypothesis**. The data provides **evidence of a significant difference in the mean income amongst men and women employees** (5)

Solutions

c) **H_0: the main effect of each factor is the same for each level of the other factor**

H_a: the two factors interact

Since **p-value >0.05**, we **fail to reject the null hypothesis**. The data provides **evidence that there is no significant interaction between age category and gender; mean income between age categories does not depend on gender** (5)

d.) **H_0: $\mu_i = \mu_j$**

H_a: $\mu_i \neq \mu_j$

Mean income between the 18-29 group differed significantly in mean income compared to the other age categories. In addition, the mean income for employees 30-39 years of age differed significantly from those 40-49 years of age (all $p<0.05$). (6)

The End!

Questions?

Email dhawan@g.harvard.edu for feedback and any future changes