HW3 - Factors and friends

Ryo Iwata

11 March, 2024

Problem 1

You are researching how GAD65 single nucleotide polymorphisms modify the enzyme's function with a goal of understanding a putative link to a broad spectrum of psychiatric conditions. You run an experiment in which you introduced mutations that result in single amino acid changes. To evaluate these changes, you measure enzyme activity in droplet microarray wells with a consistent concentration of the enzyme with a fluid substate containing both glutamate (GAD65's substrate) and pyridoxal phosphate drawn from a separate experiment. The substrate contains variable concentrations of glutamate (enzymatic substrate) and pyridoxal phosphate (enzyme cofactor).

1.A Load data and prepare data for analysis

```
# Reading in the data
raw_gad65_data <- readxl::read_excel("./data/HW3data_.xlsx", .name_repair = "minimal")
```

```
fixed_gad65_data <- raw_gad65_data |>
  # Removing all empty nan only columns
  select(-where(is.logical)) |>
  # Turning categorical variables into factors
  mutate(across(c(WellID, `Plate ID`, `Enzyme Variant`), as_factor)) |>
  # Removing rows with empty data
  na.omit()
```

1.B Considering *only* the enzyme variant, conduct an ANOVA analysis: (10 pt)

1.B.1 Write out reasonable scientific hypothesis in clear terms that this approach can test (2 pt)

Answer: When considering only enzyme variants, there is a significant difference in the enzyme activity in at least one of the tested GAD65 enzyme variants compared to the other tested GAD65 enzyme variants.

1.B.2 Write out your statistical hypothesis (null hypothesis and alternate) and the formula you are fitting. (2 pt)

```
kable(distinct(fixed_gad65_data, `Enzyme Variant`))
```

Enzyme Variant
A121C
A121T
A121R

Enzyme Variant
A121L
A121F

The hypotheses we care about are:

$$H_0 : \mu_{A121C} = \mu_{A121T} = \mu_{A121R} = \mu_{A121L} = \mu_{A121F}$$

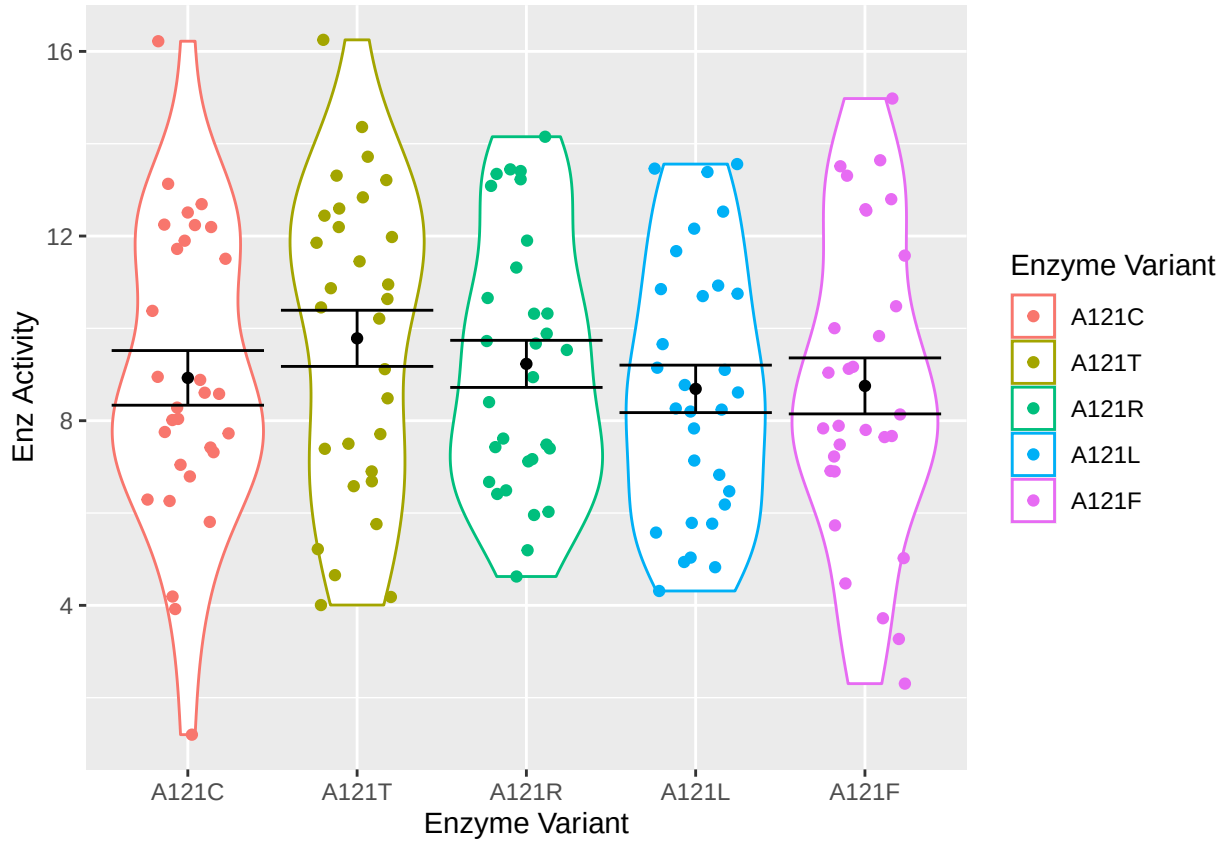
$$H_1 : \mu_{A121C} \neq \mu_{A121T} \neq \mu_{A121R} \neq \mu_{A121L} \neq \mu_{A121F}$$

Formula that is being fitted (beta also ok to use, using B for simplicity)

$$\hat{y} = \beta_0 + \beta_{A121T} x_{A121T?} + \beta_{A121R} x_{A121R?} + \beta_{A121L} x_{A121L?} + \beta_{A121F} x_{A121F?}$$

1.B.3 Make a figure which shows both the 'raw' data and summary relevant to this model/analysis (3 pt)

```
fixed_gad65_data |> ggplot(aes(x=`Enzyme Variant`,
                               y=`Enz Activity`,
                               color = `Enzyme Variant`)) +
  geom_violin() +
  geom_jitter(width = 0.25) +
  stat_summary(fun = mean, color="black", geom = "point") +
  stat_summary(geom = "errorbar", fun.data = mean_se, color="black")
```



1.B.4 Fit your model; identify and interpret the results (3 pt (1pt each for: model, results and interpretation)). Use Anova from car for your ANOVA table.

```
variant_to_activity_model = lm(`Enz Activity` ~ `Enzyme Variant`, fixed_gad65_data)

variant_to_activity_car = car::Anova(variant_to_activity_model)

kable(variant_to_activity_car)
```

	Sum Sq	Df	F value	Pr(>F)
Enzyme Variant	24.027	4	0.6206	0.64853
Residuals	1403.482	145	NA	NA

```
variant_to_activity_pvalue <-
  pull(filter(variant_to_activity_car,
              row.names(variant_to_activity_car)=="`Enzyme Variant`"),
       "Pr(>F)")

variant_to_activity_fstat <-
  pull(filter(variant_to_activity_car,
              row.names(variant_to_activity_car)=="`Enzyme Variant`"),
       "F value")

variant_to_activity_Df <-
  pull(filter(variant_to_activity_car,
              row.names(variant_to_activity_car)=="`Enzyme Variant`"),
       "Df")

print("F-statistic and associated p-value for all slope terms")
```

```
## [1] "F-statistic and associated p-value for all slope terms"
```

```
print(paste("F(",
            variant_to_activity_Df,
            ",",
            df.residual(variant_to_activity_model),
            ") = ",
            variant_to_activity_fstat,
            ", p = ",
            variant_to_activity_pvalue))
```

```
## [1] "F( 4 , 145 ) =  0.620596876015268 , p =  0.648530942510027"
```

Assuming significance with a p-value less than 0.05, we can not reject the null hypothesis that the mean enzyme activity of different enzyme variants are not significantly different from each other when considering only enzyme variant.

1.C Considering only the enzyme variant and glutamate concentration, fit a GLM model that includes both: (15 pt)

1.C.1 Write out a reasonable scientific hypothesis in clear terms (2 pt)

Answer: There is a significant difference in the enzyme activity of at least one of the tested GAD65 enzyme variants compared to the other tested GAD65 enzyme variants after controlling for the effect of the concentrations of glutamate.

1.C.2 Write out your statistical hypothesis and the formula you are fitting. (3 pt)

The hypotheses we care about are:

$$H_0 : \mu_{A121C} = \mu_{A121T} = \mu_{A121R} = \mu_{A121L} = \mu_{A121F}$$
$$H_1 : \mu_{A121C} \neq \mu_{A121T} \neq \mu_{A121R} \neq \mu_{A121L} \neq \mu_{A121F}$$

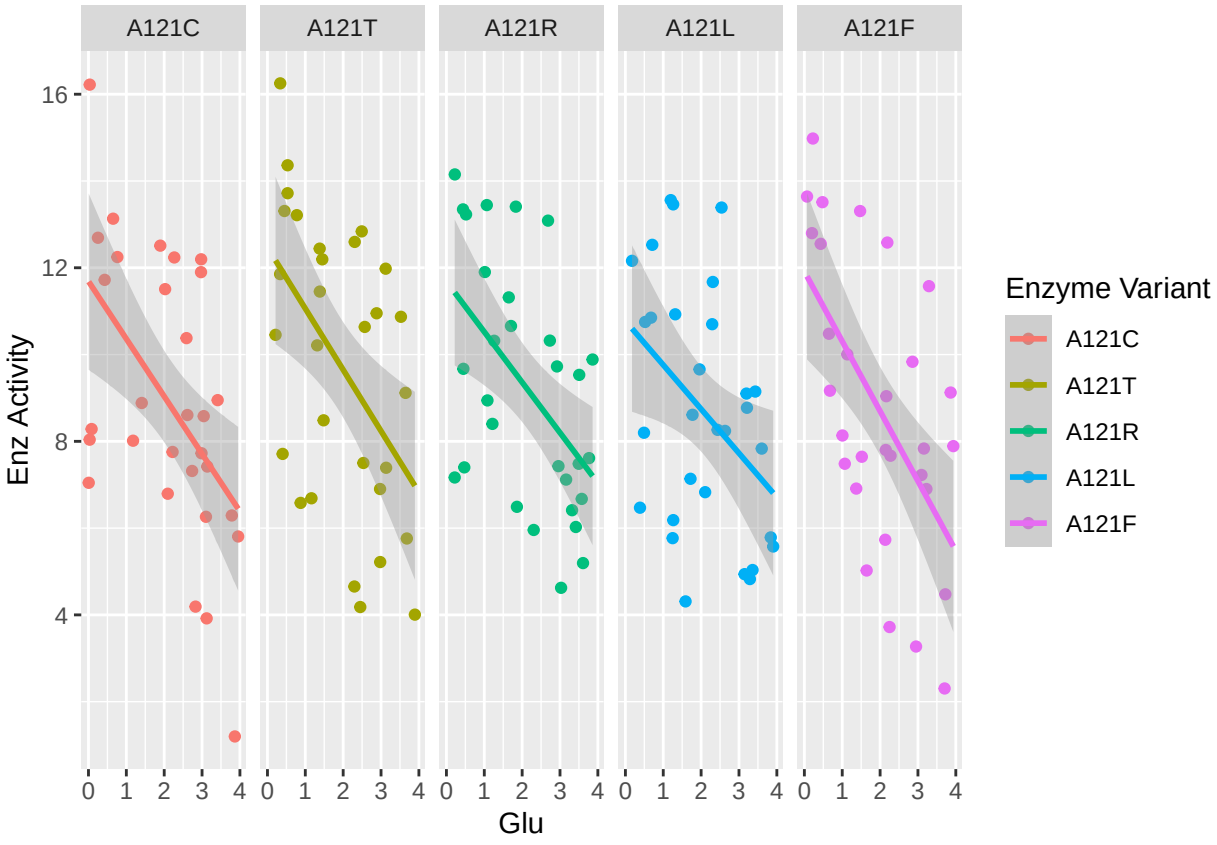
Formula that is being fitted

$$\hat{y} = \beta_0 + \beta_{A121T} x_{A121T?} + \beta_{A121R} x_{A121R?} + \beta_{A121L} x_{A121L?} + \beta_{A121F} x_{A121F?} + \beta_{Glu} x_{Glu}$$

1.C.3 Make a figure which includes all relevant variables relevant to this model and *only* those variables (5 pt)

```
fixed_gad65_data |> ggplot(aes(x=Glu, y=`Enz Activity`, color=`Enzyme Variant`)) +
  geom_point() +
  facet_wrap(vars(`Enzyme Variant`), nrow=1) +
  stat_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



1.C.4 Fit your model; present the key components in one or more concise, well-formatted table(s), and interpret the results (5 pt)

```
variant_and_glu_to_activity_model =
  lm(`Enz Activity` ~ `Enzyme Variant` + Glu, fixed_gad65_data)

variant_and_glu_to_activity_car = car::Anova(variant_and_glu_to_activity_model)
variant_and_glu_to_activity_tidy = tidy(variant_and_glu_to_activity_model)

kable(variant_and_glu_to_activity_car)
```

	Sum Sq	Df	F value	Pr(>F)
Enzyme Variant	19.915	4	0.69099	0.59933
Glu	365.930	1	50.78685	0.00000
Residuals	1037.552	144	NA	NA

```
variant_and_glu_pvalue_for_variant <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="`Enzyme Variant`"),
       "Pr(>F)")

variant_and_glu_fstat_for_variant <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="`Enzyme Variant`"),
       "F value")

variant_and_glu_df_for_variant <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="`Enzyme Variant`"),
       "Df")

print("F-statistic and associated p-value for enzyme variant slope terms")
```

```
## [1] "F-statistic and associated p-value for enzyme variant slope terms"
```

```
print(paste("F(",
            variant_and_glu_df_for_variant,
            ",",
            df.residual(variant_and_glu_to_activity_model),
            ") = ",
            variant_and_glu_fstat_for_variant,
            ", p = ",
            variant_and_glu_pvalue_for_variant))
```

```
## [1] "F( 4 , 144 ) =  0.690987159431602 , p =  0.599334439526976"
```

```
variant_and_glu_pvalue_for_glu <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="Glu"),
       "Pr(>F)")

variant_and_glu_fstat_for_variant <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="Glu"),
       "F value")
```

```
variant_and_glu_df_for_variant <-
  pull(filter(variant_and_glu_to_activity_car,
              row.names(variant_and_glu_to_activity_car)=="Glu"),
       "Df")

print("F-statistic and associated p-value for glutamate concentation slope terms")
```

```
## [1] "F-statistic and associated p-value for glutamate concentation slope terms"
```

```
print(paste("F(",
            variant_and_glu_df_for_variant,
            ",",
            df.residual(variant_and_glu_to_activity_model),
            ") = ",
            variant_and_glu_fstat_for_variant,
            ", p = ",
            variant_and_glu_pvalue_for_glu))
```

```
## [1] "F( 1 , 144 ) =  50.786845009417 , p =  4.56721100008259e-11"
```

Assuming significance with a p-value less than 0.05, we can not reject the null hypothesis that the mean enzyme activity of different enzyme variants are not significantly different from each other when controlling for glutamate concentration. F(4, 144) = 0.69, p = 0.60 We can reject the null hypothesis that the Glumate concentration has a statistically significant effect on the enzyme activity. F(1, 144) = 50.79, p = <0.001

1.D. Considering only the enzyme variant and pyridoxal phosphate concentration, fit a new GLM and interpret it: (15 pt)

1.D.1 Write out a reasonable scientific hypothesis in clear terms (2 pt)

Answer: There is a significant difference in the enzyme activity in at least one of the tested GAD65 enzyme variants compared to the other tested GAD65 enzyme variants after controlling for the effect of the concentrations of pyridoxal phosphate.

1.D.2 Write out your statistical hypothesis and the formula you are fitting. (3 pt)

The hypothesis we care about is still:

$$H_0 : \mu_{A121C} = \mu_{A121T} = \mu_{A121R} = \mu_{A121L} = \mu_{A121F}$$

$$H_1 : \mu_{A121C} \neq \mu_{A121T} \neq \mu_{A121R} \neq \mu_{A121L} \neq \mu_{A121F}$$

Formula that is being fitted

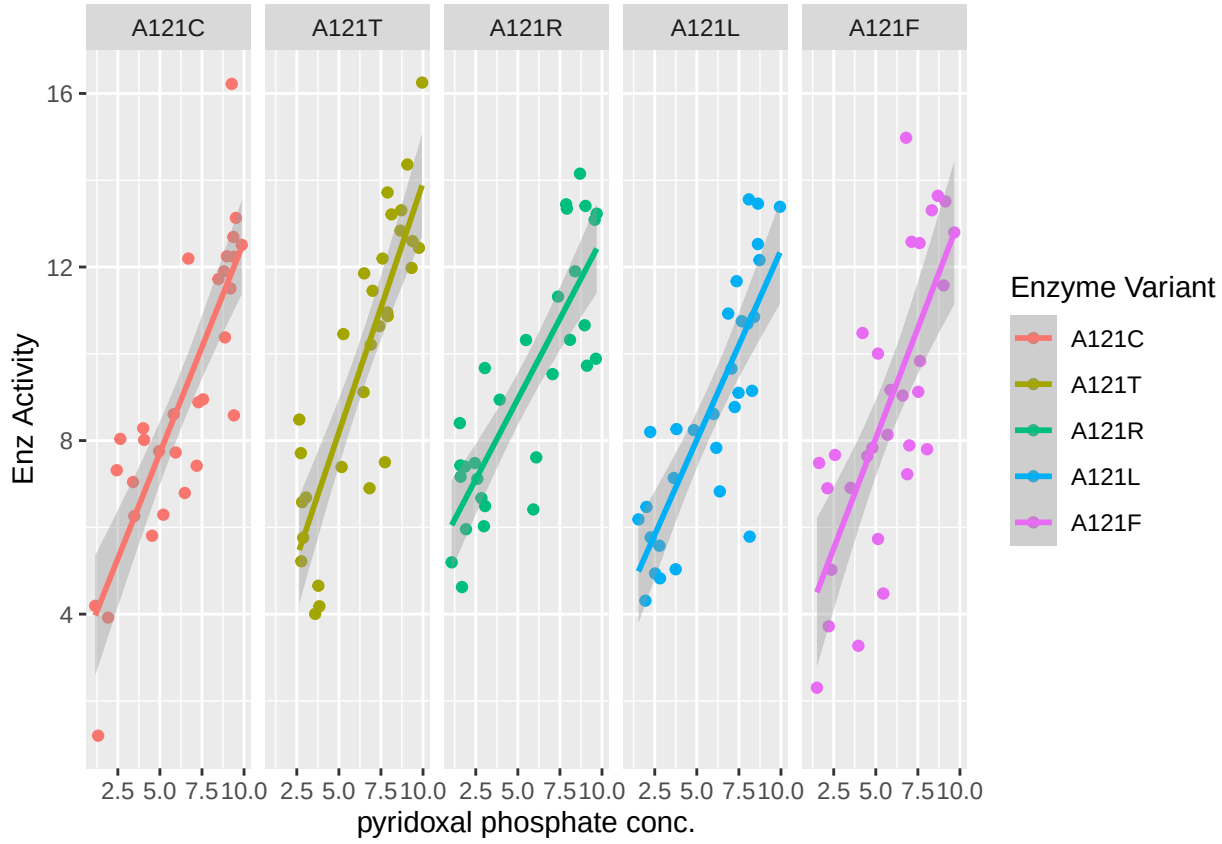
$$\hat{y} = \beta_0 + \beta_{A121T} x_{A121T?} + \beta_{A121R} x_{A121R?} + \beta_{A121L} x_{A121L?} + \beta_{A121F} x_{A121F?} + \beta_{PLP} PLP$$

1.D.3 Make a figure which shows all relevant variables relevant to this model and *only* those variables (5 pt)

```
fixed_gad65_data |> ggplot(aes(x=`pyridoxal phosphate conc.`,
                               y=`Enz Activity`,
                               color=`Enzyme Variant`)) +
  geom_point() +
```

```
  facet_wrap(vars(`Enzyme Variant`), nrow=1) +
  stat_smooth(method = "lm")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



1.D.4 Fit your model; identify, present and interpret the results (5 pt)

```
variant_and_PLP_to_activity_model =
  lm(`Enz Activity` ~ `Enzyme Variant` + `pyridoxal phosphate conc.`,
     fixed_gad65_data)

variant_and_PLP_to_activity_car = car::Anova(variant_and_PLP_to_activity_model)
variant_and_PLP_to_activity_tidy = tidy(variant_and_PLP_to_activity_model)

kable(variant_and_PLP_to_activity_car)
```

	Sum Sq	Df	F value	Pr(>F)
Enzyme Variant	23.609	4	1.7281	0.14695
pyridoxal phosphate conc.	911.653	1	266.9175	0.00000
Residuals	491.830	144	NA	NA

```
variant_and_PLP_pvalue_for_variant <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`Enzyme Variant`"),
```

```
      "Pr(>F)")

variant_and_PLP_fstat_for_variant <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`Enzyme Variant`"),
       "F value")

variant_and_PLP_df_for_variant <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`Enzyme Variant`"),
       "Df")


print("F-statistic and associated p-value for enzyme variant terms")
```

```
## [1] "F-statistic and associated p-value for enzyme variant terms"
```

```
print(paste("F(",
            variant_and_PLP_df_for_variant,
            ",",
            df.residual(variant_and_PLP_to_activity_model),
            ") = ",
            variant_and_PLP_fstat_for_variant,
            ", p = ",
            variant_and_PLP_pvalue_for_variant))
```

```
## [1] "F( 4 , 144 ) =  1.72807052414683 , p =  0.146953276991127"
```

```
variant_and_PLP_pvalue_for_glu <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`pyridoxal phosphate conc.`"),
       "Pr(>F)")

variant_and_PLP_fstat_for_variant <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`pyridoxal phosphate conc.`"),
       "F value")

variant_and_PLP_df_for_variant <-
  pull(filter(variant_and_PLP_to_activity_car,
              row.names(variant_and_PLP_to_activity_car)=="`pyridoxal phosphate conc.`"),
       "Df")


print("F-statistic and associated p-value for Pyridoxal phosphate slope terms")
```

```
## [1] "F-statistic and associated p-value for Pyridoxal phosphate slope terms"
```

```
print(paste("F(",
            variant_and_PLP_df_for_variant,
            ",",
            df.residual(variant_and_PLP_to_activity_model),
            ") = ",
            variant_and_PLP_fstat_for_variant,
            ", p = ",
            variant_and_PLP_pvalue_for_glu))
```

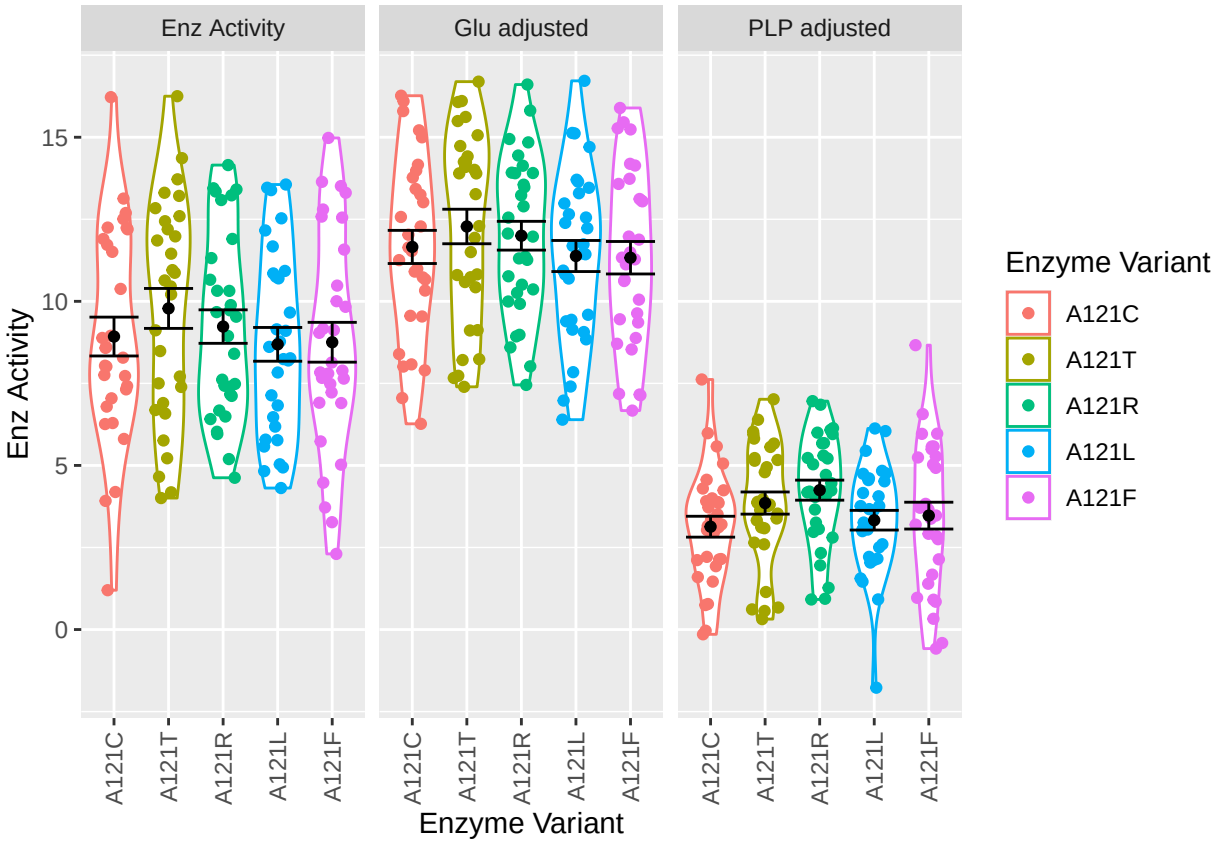
```
## [1] "F( 1 , 144 ) =  266.917537036642 , p =  1.33618287642415e-34"
```

Assuming significance with a p-value less than 0.05, we can not reject the null hypothesis that the mean enzyme activity of different enzyme variants are not significantly different from each other when controlling for pyridoxal phosphate concentration. F(4 , 144) = 1.73 , p = 0.15 We can reject the null hypothesis that the pyridoxal phosphate concentration has a statistically significant effect on the enzyme activity. F(1 , 144) = 266.91, p = <0.001

1.E. Using your models from C and D, calculate activity after removing the estimated contribution of (1) glutamate and (2) pyridoxal phosphate concentrations (note: create 2 new variables). Recreate figure 1.B.3 for each of these with a separate facet for each (will require some data carpentry). (10 pt)

```
# Getting the glutamate coefficient
glu_coef_for_variant_and_glu_to_activity <-
  pull(filter(variant_and_glu_to_activity_tidy,
              term == "Glu"),
       estimate)

# Getting the glutamate coefficient
PLP_coef_for_variant_and_glu_to_activity <-
  pull(filter(variant_and_PLP_to_activity_tidy,
              term == "`pyridoxal phosphate conc.`"),
       estimate)

# Adjusting for Glu or PLP
glu_and_PLP_adjusted_gad65_data <-
  fixed_gad65_data |>
  mutate(`Glu adjusted` =
           `Enz Activity` -
           Glu * glu_coef_for_variant_and_glu_to_activity,
         `PLP adjusted` =
           `Enz Activity` -
           `pyridoxal phosphate conc.` * PLP_coef_for_variant_and_glu_to_activity) |>
  pivot_longer(cols=c(`Enz Activity`, `Glu adjusted`, `PLP adjusted`),
               values_to = "Enz Activity")
```

```
glu_and_PLP_adjusted_gad65_data |> ggplot(aes(x=`Enzyme Variant`,
                                y=`Enz Activity`,
                                color = `Enzyme Variant`)) +
  geom_violin() +
  geom_jitter(width = 0.25) +
  stat_summary(fun = mean, color="black", geom = "point") +
  stat_summary(geom = "errorbar", fun.data = mean_se, color="black") +
  facet_wrap(vars(name)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



1.E.1 Describe this figure and provide an interpretation of each facet. (5 pt)

```
variant_and_glu_to_activity_tidy |>
  select(term, estimate) |>
  filter(term == "Glu") |>
  kable()
```

term	estimate
Glu	-1.3111

```
variant_and_PLP_to_activity_tidy |>
  select(term, estimate) |>
  filter(term == "`pyridoxal phosphate conc.`") |>
  kable()
```

term	estimate
pyridoxal phosphate conc.	0.92838

Answer: The `Enz Activity` facet is a violin plot of the distribution with associated points of all the enzyme activity levels grouped by enzyme variant. The `Glu adjusted` facet is similar to the `Enz Activity` facet, but the estimated contribution for the glutamate concentration was removed from each enzyme activity. This is seen in the violin plot with all the distributions having an overall higher enzyme activity for each enzyme variant when comparing to the `Enz`

`Activity` facet. This is because glutamate has a negative effect on enzyme activity(as seen with the negative slope). So adjusting for it would increase the enzyme activity because all glutamate concentrations are positive. The `PLP adjusted` facet is similar to the `Glu adjusted`, but adjusting for the pyridoxal phosphate concentration. The enzyme activity in this facet is smaller than that of `Enz Activity` because pyridoxal phosphate concentration has a positive effect on enzyme activity (as seen with the positive slope). So adjusting for it would decrease the enzyme activity because all pyridoxal phosphate concentrations are positive

1.E.2 How does that interpretation relate to your interpretation of the statistical models you ran above?. (5 pt)

Answer: The interpretation of the facets `Enz Activity, Glu adjusted, and PLP adjusted` relates to the findings from the statistical models analyzed previously, specifically highlighting how glutamate and pyridoxal phosphate concentrations impact enzyme activity.

Glutamate Adjustment: The statistical model showed that glutamate concentration had a significant negative effect on enzyme activity. This effect was quantified in the Glu variable within the model, indicating that as glutamate concentration increases, enzyme activity decreases. This relationship was visually represented in the Glu adjusted facet, where enzyme activity levels appear higher after the removal of glutamate's estimated contribution. The adjustment tries to account for glutamate's negative effect, thereby isolating the impact of the enzyme variant alone on activity.

Pyridoxal phosphate Adjustment: The statistical model showed that Pyridoxal phosphate concentration had a significant positive effect on enzyme activity. This effect was quantified in the pyridoxal phosphate variable within the model, indicating that as pyridoxal phosphate concentration increases, enzyme activity increases This relationship was visually represented in the Glu adjusted facet, where enzyme activity levels appear lower after the removal of pyridoxal phosphate's estimated contribution. The adjustment tries to account for pyridoxal phosphate's positive effect, thereby isolating the impact of the enzyme variant alone on activity.

1.F. Considering the enzyme variant, glutamate and pyridoxal phosphate concentrations, conduct an ANCOVA/General Linear Model analysis: (40 pt)

1.F.1 Write out a reasonable scientific hypothesis in clear terms (3 pt)

Answer: There is a significant difference in the enzyme activity in at least one of the tested GAD65 enzyme variants compared to the other tested GAD65 enzyme variants after controlling for the effect of the concentrations of glutamate and pyridoxal phosphate.

1.F.2 Write out your statistical hypotheses and the formula you are fitting. (4 pt)

The hypotheses we care about are:

$$H_0 : \mu_{A121C} = \mu_{A121T} = \mu_{A121R} = \mu_{A121L} = \mu_{A121F}$$
$$H_1 : \mu_{A121C} \neq \mu_{A121T} \neq \mu_{A121R} \neq \mu_{A121L} \neq \mu_{A121F}$$

Formula that is being fitted

$$\hat{y} = \beta_0 + \beta_{A121T} x_{A121T?} + \beta_{A121R} x_{A121R?} + \beta_{A121L} x_{A121L?} + \beta_{A121F} x_{A121F?} + \beta_{Glu} x_{Glu} + \beta_{PLP} x_{PLP}$$

1.F.3 Fit your model; present and interpret the results (5)

```
variant_ancova_model =
  lm(`Enz Activity` ~ `Enzyme Variant` + `Glu` + `pyridoxal phosphate conc.`,
     fixed_gad65_data)

variant_ancova_car = car::Anova(variant_ancova_model)
variant_ancova_tidy = tidy(variant_ancova_model)

kable(variant_ancova_car)
```

	Sum Sq	Df	F value	Pr(>F)
Enzyme Variant	22.987	4	3.0439	0.01918
Glu	221.848	1	117.5051	0.00000
pyridoxal phosphate conc.	767.570	1	406.5552	0.00000
Residuals	269.982	143	NA	NA

Answer: Assuming significance with a p-value less than 0.05, we can reject the null hypothesis that the mean enzyme activity of different enzyme variants are not significantly different from each other when controlling for the effects of Glu and pyridoxal phosphate concentrations. F(4 , 144) = 3.04, p = 0.019 We can reject the null hypothesis that the glutamate concentrations does not have a statistically significant effect on the enzyme activity. F(1 , 144) = 117.51, p = <0.001 We can reject the null hypothesis that the pyridoxal phosphate concentration does not have a statistically significant effect on the enzyme activity. F(1 , 144) = 406.56, p = <0.001

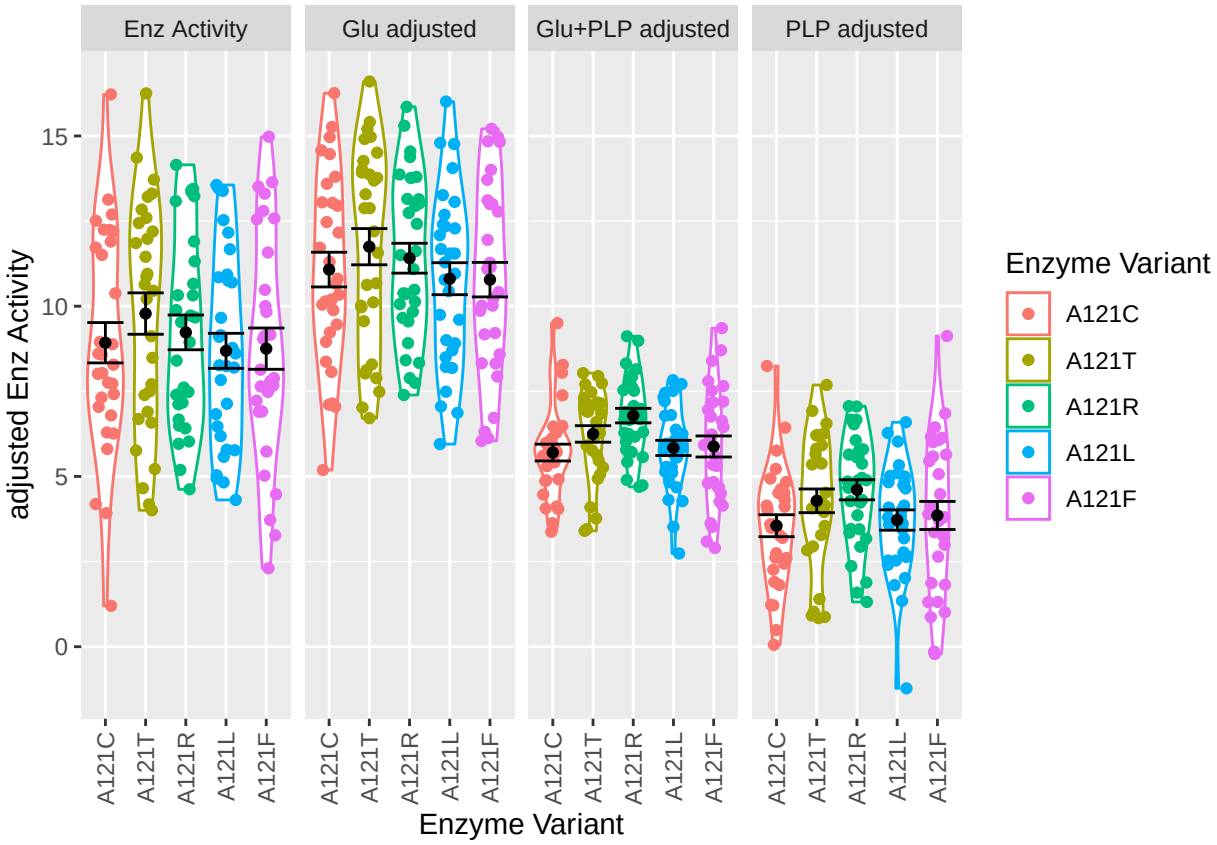
1.F.4 Using your model from F, calculate activity after removing the estimated (model fitted) contribution of only glutamate and pyridoxal phosphate concentrations and both substrate and pyridoxal phosphate concentrations (note: 3 separate new adjusted-activity variables). Recreate figure 1.B.3 for each of these with a separate facet for each. Discuss (compare and contrast) to your figure from E. (16 pt)

```
glu_coef_ancova <- pull(filter(variant_ancova_tidy,
                               term == "Glu"),
                        estimate)

PLP_coef_ancova <- pull(filter(variant_ancova_tidy,
                               term == "`pyridoxal phosphate conc.`"),
                        estimate)
```

```
glu_and_PLP_adjusted_gad65_data <- augment(variant_ancova_model) |>
  mutate(`Glu adjusted` =
           `Enz Activity` - Glu * glu_coef_ancova,
         `PLP adjusted` =
           `Enz Activity` - `pyridoxal phosphate conc.` * PLP_coef_ancova,
         `Glu+PLP adjusted` =
           `Enz Activity` -
           `pyridoxal phosphate conc.` * PLP_coef_ancova -
           Glu * glu_coef_ancova)

glu_and_PLP_adjusted_ancova_gad65_pivoted <-
  glu_and_PLP_adjusted_gad65_data |>
  pivot_longer(cols=c(`Enz Activity`,
                      `Glu adjusted`,
                      `PLP adjusted`,
```

```
                      `Glu+PLP adjusted`),
                    values_to = "adjusted Enz Activity")
```

```
glu_and_PLP_adjusted_ancova_gad65_pivoted |> ggplot(aes(x=`Enzyme Variant`,
                                  y=`adjusted Enz Activity`,
                                  color = `Enzyme Variant`)) +
  geom_violin() +
  geom_jitter(width = 0.25) +
  stat_summary(fun = mean, color="black", geom = "point") +
  stat_summary(geom = "errorbar", fun.data = mean_se, color="black") +
  facet_wrap(vars(name), nrow=1) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```




```
variant_ancova_tidy |>
  select(term, estimate) |>
  filter(term == "Glu" | term == "`pyridoxal phosphate conc.`") |>
  kable()
```

term	estimate
Glu	-1.03185
pyridoxal phosphate conc.	0.86103

Answer: The `Enz Activity` facet is the same plot as in 1E. The `Glu adjusted` and `PLP adjusted` facets are similar to the corresponding ones in 1E. With the direction of change compared to `Enz Activity` being similar to the corresponding ones in 1E with `Glu adjusted` being greater and

`PLP adjusted` being smaller for the similar reasons to that of `1E`. As seen above, the coefficients are still negative for glutamate and positive for pyridoxal phosphate concentration for this model. The addition of the `Glu+PLP adjusted` facet builds upon these two by removing both the estimated contribution of glutamate and pyridoxal phosphate concentrations. Because adjusting for glutamate increased and adjusting for pyridoxal phosphate decreased, we see that when adjusting for both the overall distributions are in between the two adjustments for each corresponding variant. We also see the distribution is less varied when removing the contribution of both as seen by the smaller standard error lines. This would be the goal for this correction, so that enzyme activity is better reflecting the effects of enzyme variants independent glutamate pyridoxal phosphate concentrations.

1.F.5 Calculate the mean values for each factor level after correcting for only glutamate, only pyridoxal phosphate, and glutamate and pyridoxal phosphate using the model from F. Compare these values to the estimates from your model for each factor level. (12 pt)

```
glu_and_PLP_adjusted_gad65_data |>
  summarise(`Model Estimates` = mean(.fitted),
            `Glu adjusted` =mean(`Glu adjusted`),
            `PLP adjusted`=mean(`PLP adjusted`),
            `Glu+PLP adjusted` =mean(`Glu+PLP adjusted`),
            .by = "Enzyme Variant") |>
  kable()
```

Enzyme Variant	Model Estimates	Glu adjusted	PLP adjusted	Glu+PLP adjusted
A121C	8.9273	11.076	3.5534	5.7018
A121T	9.7842	11.749	4.2844	6.2488
A121R	9.2309	11.410	4.6084	6.7873
A121L	8.6895	10.808	3.7197	5.8378
A121F	8.7537	10.780	3.8543	5.8806

Answer: The Glu adjustment increases enzyme activity levels, which shows glutamate's negative relationship with enzyme activity. The PLP adjustment decreases enzyme activity levels, which shows pyridoxal phosphate's positive relationship with enzyme activity. When adjusting for both Glu and PLP, the enzyme activity distributions lie between those observed when adjusting for each factor separately. All these trends are seen across all factor levels or enzyme variants.

1. Bonus! : Go back (revisit, do not duplicate here) to your plots above and use [theme] (https://ggplot2.tidyverse.org/reference/theme.html) (and any other ggplot functions you care to use) to polish them by adjusting theme and mapping characteristics (e.g., the fonts, font sizes, colors, color scales, etc.) of your plots. List each change, indicate what you changed, and clearly yet concisely explain the benefit of each change. (up to +5 pt; half a point per well-justified change)

1. Rotated the tick values for the variants so they are vertical and not overlapping with `theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))`
2. Made the facet wraps into one row for easier comparison with `facet_wrap(vars(name), nrow=1)`