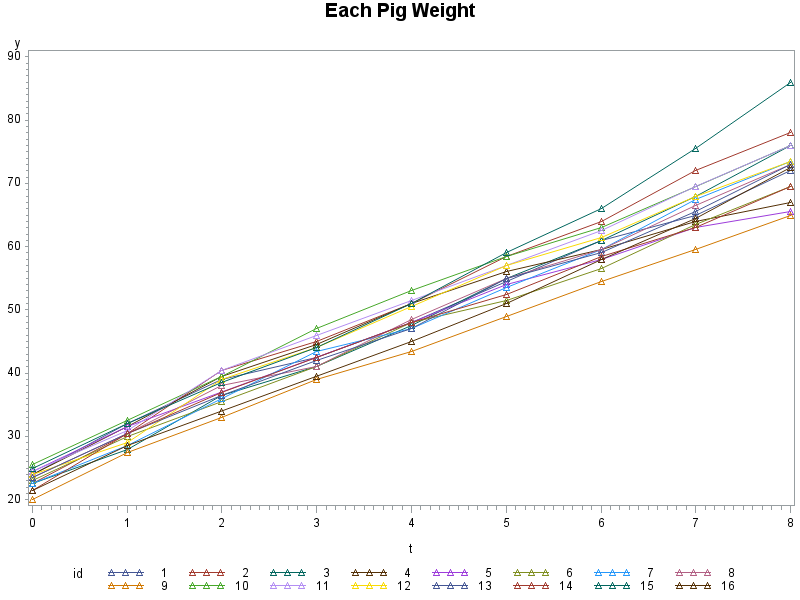
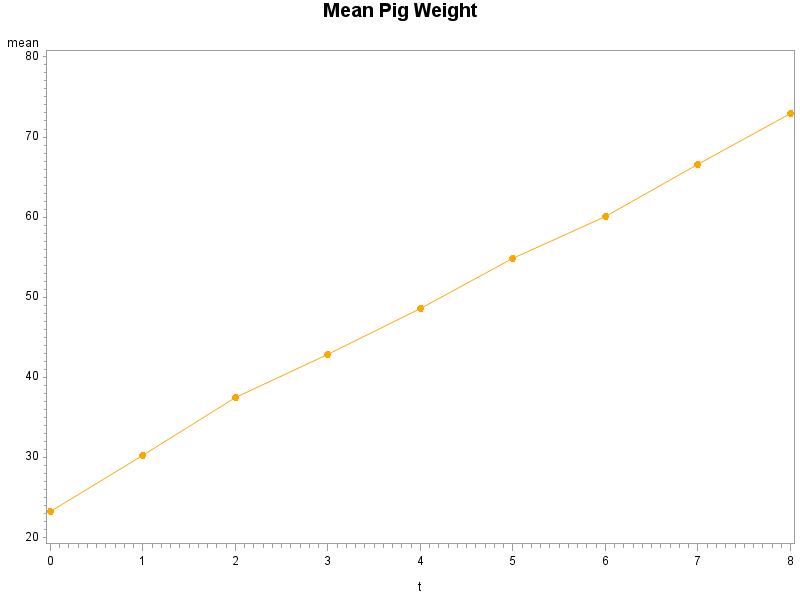
1. **Descriptive Analysis**

**Create two plots, one showing the trajectories of each pig’s weight over time and the other showing the trajectory of the mean weight over time. Comment on the notable features in these plots.**

The trajectories for each pig’s weights are very similar to one another, they all started around 20-25, and then increase linearly over the 8 successive weeks. The variances increase over week, the variance at the end of the 9th week is noticeably larger than the baseline variance.

|  |  |  |
| --- | --- | --- |
| Analysis Variable : y | | |
| Week | **N** | **Mean** |
| 0 | 16 | 23.2500000 |
| 1 | 16 | 30.2500000 |
| 2 | 16 | 37.5000000 |
| 3 | 16 | 42.8125000 |
| 4 | 16 | 48.6562500 |
| 5 | 16 | 54.8125000 |
| 6 | 16 | 60.1250000 |
| 7 | 16 | 66.5312500 |
| 8 | 16 | 72.8750000 |

****

The trajectory of the mean weights has the same pattern as individual trajectories. The mean weight started at 23.25 and then increases linearly over weeks.

1. **Obtaining and Interpreting a Linear Mixed Effects Model for Bodyweight**

**Using PROC MIXED, fit a model for weight over (continuous) time which includes subject-specific intercepts and slopes as random effects.**

* The Fixed Part:
* The Random Part:

* Where is the week from baseline, e.g., is the second week in the study period.
* is the average weight at baseline week
* is the average constant rate of change over time/week
* is the subject-specific random intercept, is the subject-specific random slope.

Assuming ) , ) , and . is between-subject intercepts variation, is between-subject slopes variation, is between-subject co-variation of intercepts and slopes.

Or we can re-write as, ( where

* And assuming within-subject error . is the within-subject error variation.

**a. Obtain and interpret parameter estimates and associated 95% confidence intervals for the fixed effects in the model for the mean trajectory.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Solution for Fixed Effects | | | | | |
| Effect | **Estimate** | **Standard Error** | **Pr > |t|** | **Lower** | **Upper** |
| Intercept | 24.2285 | 0.4510 | <.0001 | 23.2672 | 25.1898 |
| t | 6.0766 | 0.1408 | <.0001 | 5.7765 | 6.3766 |

The estimated intercept, , has interpretation as the average weight at baseline week is 24.2285 kg, which is statistically significant. And there is a 95% chance that the true average weight at baseline would be between 23.2672 kg and 25.1898 kg.

The estimate of the population average slope, , indicates that the population-average constant rate of change is 6.0766 kg per week, which is statistically significant. And there is a 95% chance that the true average rate of change would be between 5.7765 and 6.3766 kg / week.

**b. Obtain and interpret parameter estimates for the variances and correlation in the model.**

There’re 4 parameters for the variances and correlation in the linear mixed effects model.

|  |  |  |
| --- | --- | --- |
| Covariance Parameter Estimates | | |
| Cov Parm | **Subject** | **Estimate** |
| UN(1,1) | **id** | 2.6389 |
| UN(2,1) | **id** | -0.2200 |
| UN(2,2) | **id** | 0.2899 |
| Residual |  | 1.6299 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated G Correlation Matrix | | | | |
| Row | **Effect** | **id** | **Col1** | **Col2** |
| 1 | Intercept | 1 | 1.0000 | -0.2516 |
| 2 | t | 1 | -0.2516 | 1.0000 |

* is the variance of subject-specific/random intercepts, indicating substantial variability from pig to pig in body weight at baseline week. For example, approximately 95% of pigs have body weight between 21.04454 and 27.41246.
* =0.2899 is the variance of subject-specific/random slopes, indicating small variability from pig to pig in rate of change over time. For example, approximately 95% of pigs have rate of change between 5.02129 and 7.13191 kg/week.
* is the correlation of subject-specific intercepts and slopes, indicating pigs that started at a higher body weight than average would be expected to have smaller weight increase rate than average.
* is the variance of within-subject error, indicating substantial within-pig variability in body weight over the study period.

**c. Obtain and interpret the 90% normal range for trends over time in body weight among pigs in the population sampled (i.e. for the pig-specific random effects for trend over time).**

=0.2899 is the variance of subject-specific random slopes, so the 90% normal range for trends over time in body weight among pigs in the population sampled is . It means approximately 90% of pigs have rate of change over time between 5.190893 and 6.962307 kg /week.

**3. The investigator who provided the data is interested in designing a randomized clinical trial to evaluate an additive to the standard pig feed which might increase the rate of growth over time in bodyweight in pigs. He feels that an increase in bodyweight of 0.2 kg per week above that observed in the study for which the data are provided would be important, and would like to design a study to have 90% power to detect this increase using a two-sided 0.05 level of significance.**

**a. What sample size would be needed in the randomized trial if the growth rate in the control group (without the additive) was the same as observed in the study for which the data are provided? Assume that the duration of the study from the first to last measurement is 8 weeks and that measurements are obtained every 4 weeks using the same technique as in the study for which the data are provided (so a measurement at times 0, 4 and 8 weeks). Also assume that equal numbers are randomized to each of the intervention (with the additive) and control groups and that the variance components observed in the study for which the data provided are reasonable choices for what would be found in the proposed trial. Show how you derived your answer.**

* Interested in comparing an additive treatment and the control group
* Plan to randomize an equal number of subjects, N, to each group
* Plan to take n=3 repeated measurements, at week 0, 4, and 8 (
* Reasonably assume linear trends based on the data provided
* Effect of interest is the difference in rate of change , suppose investigators want to detect minimum
* Based on observed data, investigators posit that between-subject variability in the rate of change =0.2899, and the within-subject error variability
* Investigators desire to have 90% power when conducting a 2-sided test at the 5% significance level (i.e., and ).

Given these specifications,

The projected N in each group is

Thus, to ensure power of at least 90% investigators will need to have a total of 360 pigs, randomizing an equal number of (180) to each group.

**b. What do you notice about the data provided that could affect the sample size needed? Briefly justify your answer.**

Let denote estimate of the slope for the ith subject. Variability of , as shown above:

Variability of is too large relative to the square of the minimum effect size , in other words, ratio of noise to signal is too large. When noise is large relative to signal, we need large sample to have the power to detect an effect.

There are two sources of variability in the variability of :

* A within-subject variance component,
* A between-subject variance component,

Ways to reduce the magnitude of **:**

* Increase length of study
* Increase number of repeated measures: now only have 3 which is too small
* The natural heterogeneity of the study population ,the major source of variability, but investigators have little control over this

So the most promising way to decrease the required sample size is to increase number of repeated measures and increase length of study.

**c. As the experimental additive is currently difficult to produce, the investigator is interested in knowing whether more frequent measurements or a longer study would markedly reduce the sample size. Develop a table which shows the sample size requirement if measurements were obtained every 4 weeks over an 8 week period (as done for part a), every 4 weeks over a 16 week period, every week over an 8 week period, or every week over a 16 week period. Briefly comment on any additional considerations that might be important in choosing among these four possible designs.**

|  |  |  |
| --- | --- | --- |
| Design | N | Total Sample Size 2N |
| every 4 weeks over an 8 week period | 179.118->180 | 360 |
| every 4 weeks over a 16 week period | 157.704->158 | 316 |
| every week over an 8 week period | 166.627->167 | 334 |
| every week over a 16 week period | 154.450->155 | 310 |

Additional considerations that might be important in choosing among these four possible designs:

* The cost of feeding more pigs
* The cost of taking measurements
* The cost of extending the study period

Apparently, either extending the study period or more repeated measurements can decrease the required sample size, which could lead to less cost of feeding more pigs. But extending the study period and repeated measurements both would increase the cost. So there’s a trade-off between the reduce the cost of feeding more pigs and the increase cost of taking repeated measurements and extending study period.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Appendix: SAS Code

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*import dataset;

**data** pig;

infile 'C:\data\Projects\APCD High Cost\Longitudinal\pigweight.txt';

input id wt1 wt2 wt3 wt4 wt5 wt6 wt7 wt8 wt9 ;

**proc** **print**;**run**;

\* transpose from wide format to long format;

**data** pig\_L;

set pig;

y=wt1;t=**0**;tc=**0**;output;

y=wt2;t=**1**;tc=**1**;output;

y=wt3;t=**2**;tc=**2**;output;

y=wt4;t=**3**;tc=**3**;output;

y=wt5;t=**4**;tc=**4**;output;

y=wt6;t=**5**;tc=**5**;output;

y=wt7;t=**6**;tc=**6**;output;

y=wt8;t=**7**;tc=**7**;output;

y=wt9;t=**8**;tc=**8**;output;

drop wt1 wt2 wt3 wt4 wt5 wt6 wt7 wt8 wt9;

**proc** **print**;**run**;

\*Q1: Plot each subject's response trajectories;

**proc** **gplot** data=pig\_L;

symbol1 interpol=join value=triangle;

symbol2 interpol=join value=triangle;

symbol3 interpol=join value=triangle;

symbol4 interpol=join value=triangle;

symbol5 interpol=join value=triangle;

plot y\*t=id;

title 'Each Pig Weight';

**run**;

\* Plot mean response profiles;

**proc** **means** data=pig\_L n mean nway;

var y;

class t;

output out=meandata mean=mean;

**proc** **print** data=meandata;

**run**;

**proc** **gplot** data=meandata;

symbol1 color=orange interpol=join value=dot;

plot mean\*t;

title 'Mean Pig Weight';

**run**;

\*Q2:Obtain and Interpreting a Linear Mixed Effects Model;

**proc** **mixed** data=pig\_L ;

class id ;

model y=t /s chisq cl;

random intercept t/type=un subject=id G Gcorr vcorr=**1**;

**run**;

\*Q3: Calculate Sample Size and Power;

%let n=3;

%let tao=8;

%let label=every 4 weeks over an 8 week period;

**data** sample;

title="&label.";

a=(**12**\*(&n.-**1**)\***1.6299**)/(&tao.\*&tao.\*&n.\*(&n.+**1**))+**0.2899**;

N=(**1.96**+**1.282**)\*(**1.96**+**1.282**)\***2**\*a/**0.04**;size=**2**\*N;drop a;

**proc** **print**;

**run**;

%let n=5;

%let tao=16;

%let label=every 4 weeks over a 16 week period;

**data** sample;

title="&label.";

a=(**12**\*(&n.-**1**)\***1.6299**)/(&tao.\*&tao.\*&n.\*(&n.+**1**))+**0.2899**;

N=(**1.96**+**1.282**)\*(**1.96**+**1.282**)\***2**\*a/**0.04**;size=**2**\*N;drop a;

**proc** **print**;

**run**;

%let n=9;

%let tao=8;

%let label=every week over an 8 week period;

**data** sample;

title="&label.";

a=(**12**\*(&n.-**1**)\***1.6299**)/(&tao.\*&tao.\*&n.\*(&n.+**1**))+**0.2899**;

N=(**1.96**+**1.282**)\*(**1.96**+**1.282**)\***2**\*a/**0.04**;size=**2**\*N;drop a;

**proc** **print**;

**run**;

%let n=17;

%let tao=16;

%let label=every week over a 16 week period;

**data** sample;

title="&label.";

a=(**12**\*(&n.-**1**)\***1.6299**)/(&tao.\*&tao.\*&n.\*(&n.+**1**))+**0.2899**;

N=(**1.96**+**1.282**)\*(**1.96**+**1.282**)\***2**\*a/**0.04**;size=**2**\*N;drop a;

**proc** **print**;

**run**;