Part III. The General Linear Model.
Chapter 10.3   One way ANOVA, Fixed Effects

ReCap. Part I (Chapters 1,2,3,4) ReCap Part II (Ch 5, 6, 7) ReCap Part III (Ch 9) 10.1 Single Sample t-test 10.2 Two Sample t-test 10.3 One way ANOVA, Fixed Effects 10.4 One way ANOVA, Random Effects	"Let the computer find out" is a poor strategy for researchers who do not bother to think clearly about the problem of interest and its scientific setting. The sterile analysis of 'just the numbers' will continue to be a poor strategy for progress in the sciences.

on chalk board

p117 in Burnham, K.P., D.R. Anderson. 1998. Model Selection and Inference. NewYork: Springer.

ReCap Part I (Chapters 1,2,3,4)
Quantitative reasoning: Example of scallops,
which combined models (what is the relation of scallop density to substrate?)
with statistics (how certain can we be?)
ReCap Part II (Chapters 5,6,7)
Hypothesis testing uses the logic of the null hypothesis to make a decision about an unknown population parameter.
Estimation is concerned with the specific value of an unknown population parameter.
ReCap (Ch 9) The General Linear Model is more useful and flexible than a collection of special cases.
Regression is a special case of the GLM. We have seen an examples with the explanatory variable X fixed, with the explanatory measured with error, and for a non-linear (exponential and power law) relations of response to explanatory variable.
ReCap (Ch 10) ANOVA is a special case of the general linear model.
The explanatory variable is on nominal scale.
The t-test is a special case of ANOVA. The explanatory variable has two categories.

Today: Single Factor ANOVA - Fixed Effects

Wrap-up. GLM - ANOVA. Explanatory variable is categorical.
        For fixed factors we calculate the evidential support for the overall model. Planned comparison among 3 or more means will be more informative then unplanned (*a posteriori*) comparisons.

Application of GLM. One way ANOVA, fixed effects.

Sokal and Rohlf (1995, Box 9.4) present pea section lengths in ocular units (0.114 units/mm). The purpose of the experiment was to test the effects of four sugars and one control on growth as measured by pea section length. Instead of asking "What is the best test?" or What test do I use?" we ask: "What model do I use to analyze the effects?"

1.   Construct model

Verbal model. Pea section lengths in treated groups differ from the control (untreated) group and among themselves.

From the verbal model we see that we have one response variable (pea section length) and an explanatory variable (four sugars and one control).
We also see that the explanatory variable consists of groups—it is a categorical variable.

Graphical model A simple graphical model consists of the mean for each group, along with a display of the data.
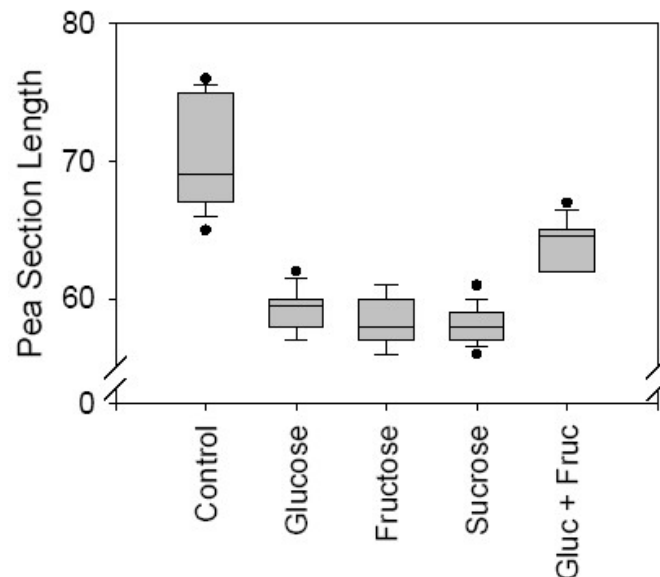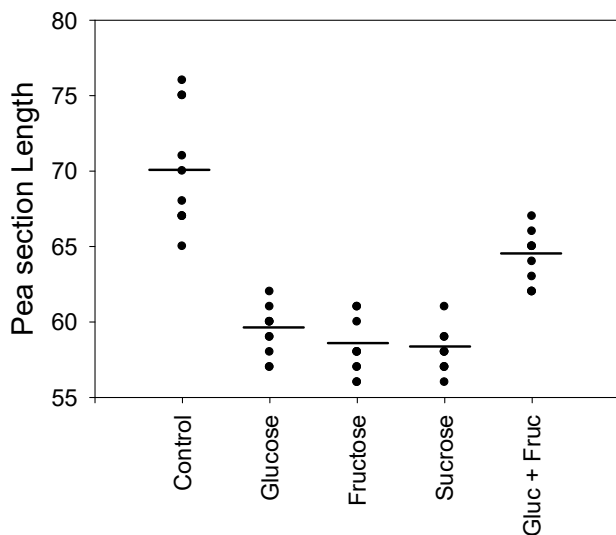
Which error model do we use? To gain some sense of the distribution of the data within each group we can use a boxplot, which shows the first, second, and third quartile. The second quartile, inside the box, is the median. The whiskers beyond the box are the 10th and 90th percentiles. The dots are outliers. The boxplot suggests some degree of heterogeneity (larger box for control) and some influence by outliers. A normal error model is a reasonable place to start.

The explanatory variable in this example is a <u>fixed effect</u>. We infer only to these particular sugars. With a random effect variable, we infer beyond the categories (units) at hand to a population of similar units. In the next section of this chapter we will see examples of random effect variables.

<u>Formal model</u> $Len = \beta_o + \beta_{Trt} \cdot Trt + \varepsilon_{normal}$

2. Execute analysis.

Place data in model format.
Data for oneway ANOVA often come in a table format-- a column of values for each category of the explanatory variable. The name of the variable is not always stated. In this case we had to create a name -- *Trt*. Model format consists of one column of numbers for each variable in the model. Our model has two variables. We stack all 50 values of the response variable taken from 5 columns into a single column *Len*. We place the names of the categories into a second column, *Trt,* so that a value and its category name are in the same row.

2. Execute analysis.
Code model statement in statistical package according to the GLM

$$Len = \alpha + \beta_{Trt} \cdot Trt + \varepsilon$$

```
MTB> ANOVA 'Len' = 'Trt'
MTB> GLM 'Len' = 'Trt'
SUBC> fits c4;
SUBC> res c5.
```

Fits and residuals from:

      a. model statement output of fitted values and residuals (as above)
      b. direct calculation of parameters (five means)
      c. parameters reported by GLM routine
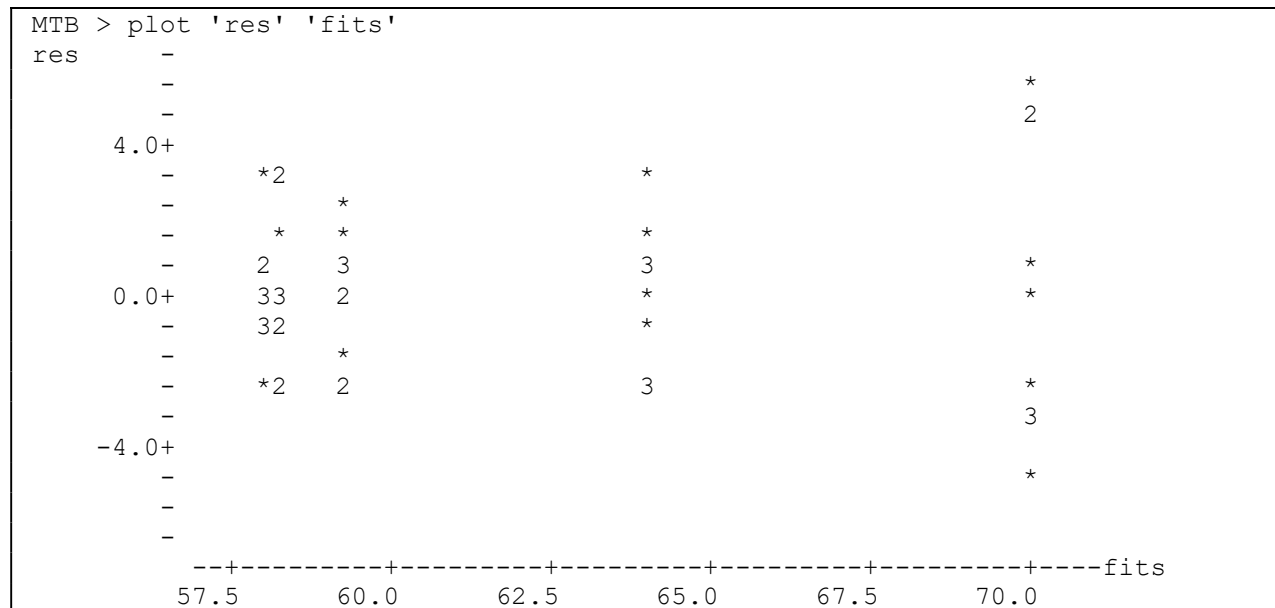
$\beta_o = 61.94$ ocular units $= 7.06$ mm

$\beta_o + \beta_{Trt} = [\ 70.1 \ 59.3 \ 58.2 \ 58.0 \ 64.1 \ ]$ ocular units

3.   Evaluate model.
Evaluate straight line assumption.
      No straight lines (regression) used, so we skip this.
Evaluate error model.   Plot residuals vs fitted values.

```
MTB > plot 'res' 'fits'
res    -
       -                                              *
       -                                              2
    4.0+
       -       *2                        *
       -            *
       -        *    *                    *
       -        2    3                    3            *
    0.0+        33   2                    *            *
       -        32                        *
       -            *
       -       *2    2                    3            *
       -                                              3
   -4.0+
       -                                              *
       -
       -
          --+---------+---------+---------+---------+---------+----fits
           57.5      60.0      62.5      65.0      67.5      70.0
```

3. Evaluate error model.

<u>Homogeneity?</u>
  The plot confirms the impression gained from the boxplot. Residuals extend over a wider range in the control than the treated groups.   The variability does not increase with increase in fitted value, so alternative error models offer no remedy. We will proceed on a judgment that the observed heterogeneity will have little effect on the overall ANOVA (comparison of all five groups).

<u>Normal errors</u> ?
Histogram looks normal.
This is onfirmed by the normal score plot (not shown), where the residuals fall close to the normal error line.

```
MTB > hist c9;
 SUBC> incr 2.
 Histogram of res   N = 50

 Midpoint    Count
    -6.00        1   *
    -4.00        3   ***
    -2.00       13   *************
     0.00       20   ********************
     2.00        9   *********
     4.00        3   ***
     6.00        1   *
```

It is interesting to note that if we check this assumption on 'the data' before computing the residuals, we would conclude that the assumptions are clearly violated. In particular the distribution skews strongly to large values. The outliers identified in the boxplot appear as a stack of values in the upper tail of the distribution. Evaluating the response variable produces the wrong conclusion.

```
MTB > hist c6
 Histogram of len   N = 50

 Midpoint    Count
       56        3   ***
       58       14   **************
       60        8   ********
       62        8   ********
       64        2   **
       66        5   *****
       68        5   *****
       70        1   *
       72        1   *
       74        0
       76        3   ***
```

<u>Independent errors?</u> We have no information on the order in which observations were made. The observations in the order shown in Sokal and Rohlf (1995) are independent at lag 1 (no correlation with adjacent observation).

In this example we were looking for large violations because sample size n is well over 30. The violations were judged to be minor. Checking the assumptions before computing the residuals produced the wrong conclusion, that a better error model was needed.   A medical analogy would be chemotherapy for a lump, before determining whether the lump is benign or cancerous.

4. Partition df and SS according to the model.

49 = dftotal	n – 1	50 – 1 = 49	
4 = df_{Trt}	5 categories in factor Trt - 1 df	5 – 1 = 4	
45 = df_{res}	df_{total} – df_{Trt}	49 – 4 = 45	

4. Partition SS according to the model.

$$SS_{tot} = \text{var}(Len) \cdot df_{total} \qquad\qquad \text{Full model}$$

1322.82 $\quad SS_{tot} = 26.996 \cdot 49 = 1322.82$

$$SS_{res} = \Sigma\,(\,\text{var}(Len_i) \cdot df_i\,) \quad i = \text{groups 1 to 5}$$

245.5 $\quad SS_{res}$ $\qquad\qquad\qquad\qquad\qquad$ Reduced model

1077.3 $\quad SS_{model} = 1322.82 - 245.5 = 1077.3 \qquad$ Improvement in fit

For any GLM, we can write the model with the partitioning of df and SS

$$Len = \alpha + \beta_{Trt} \cdot Trt + \varepsilon$$

df $\qquad 49 = \qquad 4 \qquad + 45$

SS $\qquad 1323 = \quad 1077 \quad + 246$

4. Calculate likelihood ratio for omnibus (overall) model.

$LR = (\,245.5/1322.82\,)^{-50/2} = 10^{18}$

5. State population and whether sample is representative.

Population is all possible measurements, given the method of applying treatments and the protocol for taking measurements. It is taken to be representative (not biased). This is a fixed effect model, we are inferring only to these 5 categories. We are not inferring to other sugars or combinations of sugars.

5. Decide on mode of inference. Is hypothesis testing appropriate?

This is a controlled experiment where we wish to infer to similarly conducted experiments in the future and anywhere else, given the same species of pea plant. To draw the inference from the sample to the population we will use hypothesis testing.

6. State the full (null) and reduced (alternative) model pairs.

The research hypothesis is that the treatments differ in effect.

$\quad H_A:\mu_{Control} \neq \mu_{Glucose} \neq \mu_{Fructose} \neq \mu_{Sucrose} \neq \mu_{Glucose+Fuctose}$

The null hypothesis is the treatments do not differ in effect.

$\quad H_o:\mu_{Control} = \mu_{Glucose} = \mu_{Fructose} = \mu_{Sucrose} = \mu_{Glucose+Fuctose}$

This is equivalent to the following pair of hypotheses.

$\quad$ The symbol $\beta_{Trt} \cdot Trt$ stands for the collection of 5 means.

$\quad$ If the means differ, then var($\beta_{Trt} \cdot Trt$) >0 $\qquad H_A:$ var($\beta_{Trt} \cdot Trt$) > 0

$\quad$ If the means do not differ, var($\beta_{Trt} \cdot Trt$) =0 $\qquad H_A:$ var($\beta_{Trt} \cdot Trt$) = 0

H_A/H_o pairs are statements about parameters of the population. They are not statements about the estimates of the parameters, which we compute from the data.

6. State the null and alternative hypothesis pairs.

With a fixed effects model, it is natural to investigate the sources of differences among groups, if the overall test reveals a difference.

For experiments, we usually have some expectations about the direction and sometimes about the strength of the contrasts among groups. Based on these, we can undertake planned or *a priori* (before the fact) comparisons. Such comparisons use our biological understanding more effectively than unplanned or *a posteriori* (after the fact) comparisons. We can set up planned comparisons any way we like, but we need to take into account the number of comparisons.

6. H_A / H_o for planned comparisons.

In this example the natural comparison is the control versus the other four treatments. One degree of freedom used.

$$H_A: \beta_{control} \neq (1/4)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose} + \beta_{gluc+fruc})$$

A second comparison is mixed sugars versus pure sugars. One *df* used.

$$H_A: (\beta_{gluc+fruc}) \neq (1/3)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose})$$

A third comparison is within the pure sugars: polysaccharides (sucrose) versus monosaccharides (glucose and fructose). One *df* used.

$$H_A: (\beta_{sucrose}) \neq (1/2)(\beta_{glucose} + \beta_{fructose})$$

Other choices are possible.

6. Test statistic, distribution, Type I error.

State test statistic	*F*-ratio
Distribution of test statistic	*F*-distribution
Type I error	Controlled at 5%, multiple tests.

With multiple tests, we need to take into account the number of comparisons. For example, if we undertake 20 comparisons, we expect on average one false positive conclusion, i.e. failure to reject H_o (5% = one in 20). If we use a fixed criterion of 5% for multiple tests, then we expect the error rate after multiple tests (called the ***experimentwise error rate***) to exceed the 5% level. This suggests that we lower the criterion level according to the number of tests, so that the experimentwise error rate does not rise above our stated tolerance for Type I error. One solution is to divide the tolerance level by the number of tests. Thus with 3 tests we would use 5% / 3 = 1.67% to declare significance. This is called the Bonferroni method for limiting the experimentwise error rate. An alternative method is called the Dunn-Sidak method (Ury 1976 in Sokal and Rohlf 1995). For k tests, the experimentwise error rate is $\alpha_{expwise} = 1 - (1 - \alpha)^k$ hence we use $\alpha = 1 - (1 - \alpha_{expwise})^{1/k}$ for our error rate.

For three *a priori* tests we would use $\alpha = 1 - (1 - 0.05)^{1/3} = 0.017$ to achieve a 5% error rate.

7. Move df, and SS to ANOVA table. Compute MS and F

Source	DF	SS	MS	F	P
trt	4	1077.3	269.330	49.37	0.000
Error	45	245.5	5.456		
Total	49	1322.8	[26.996]		

Calculate Type I error from F-distribution.

$p = 10^{-17}$ =FDIST(49.37,4,45) from spreadsheet function

Type I error is calculated from the likelihood ratio so we expect the Type I error to be as small as the likelihood ratio was large.

8. Recompute p-value if necessary.

Recalculation is hardly necessary, given the magnitude of the likelihood ratio and the consequently minuscule Type I error. The *a priori* tests are of concern as they are all single degree of freedom tests, which are susceptible to deviations from homogeneity. To protect ourselves from poor estimates of Type I error due to homogeneous errors, we could use a randomized p-value on the single degree of freedom tests. However, we saw little evidence of heterogeneity in the residual *vs* fit plot.

9. Report statistical conclusion about model terms.

We reject the null hypothesis of no difference in pea section length among the 5 groups (control and 4 treatments). $F_{4,45} = 49.37$, $p = 10^{-17}$

Reject H_o: $\mu_{Control} = \mu_{Glucose} = \mu_{Fructose} = \mu_{Sucrose} = \mu_{Glucose+Fuctose}$

We cannot however "accept" the H_A.

~~Accept~~ H_A: $\mu_{Control} \neq \mu_{Glucose} \neq \mu_{Fructose} \neq \mu_{Sucrose} \neq \mu_{Glucose+Fuctose}$

The alternative hypothesis survives the test.

Decision against a fixed Type I error is not evidence for H_A

We use the LR to report the evidence:

The alternative is 10^{18} times more likely than the null.

10.   Report science conclusions and interpret parameters of biological interest

Where are the differences, among the 5 groups ?

Two approaches   *A priori* Planned comparisons. Use prior knowledge.

A posteriori. Hunting. See Burnham and Anderson quote.

Statistical packages tend to foster the use of *a posteriori* comparisons.

In most situations, one can develop *a priori* comparisons that use information more effectively. Often these are one-tailed (directional) comparisons and hence have more capacity (power) to detect differences. Planned comparisons are more informative than uplanned comparisons.

We look at the three *a priori* comparisons. To achieve an experimentwise error rate of 5% will use $\alpha = 0.017$ from the Dunn-Sidak formula.

10.   Report and interpret parameters of biological interest

First planned comparison: Growth in treated media differs from that in untreated.

H_A: $\beta_{control} \neq (1/4)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose} + \beta_{gluc+fruc})$

When we compare control versus the 4 treatments, the resulting ANOVA table uses the mean square error based on two groups, rather than on 5 groups (as in previous analysis).

Source	DF	SS	MS	F
aprior1	1	832.3	832.32	81.45
Error	48	490.5	10.22	
Total	49	1322.8	27.00	

```
MTB > let k1 = 245.5/45
MTB > let k2 = 832.32/k
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k2 k4
K2       152.564    # F-ratio
K4        0         # p-value
```

For *a priori* comparisons we calculate the F-ratio based on the MS_{error} estimate from 5 groups with 45 degrees of freedom.
$MS_{error} = 245.5/45$
$F = 152.64$ instead of $F = 82.45$

Heterogeneity of the residuals was a concern for this test. However, the concern matters little, given size of the F-ratio.

The strength of the difference is of interest so we calculate the effect size. The estimate of suppression of growth by sugar relative to control is $(70.1 - 59.9)/70.1 = 15\%$

```
MTB > invcdf .975 k5;
SUBC> t 45.
MTB > print k5
K5       2.01409
MTB > let k6 = 59.9 - k5*.472
MTB > let k7 = 59.9 + k5*.472
MTB > let k8 = 70.1 - k5*1.26
MTB > let k9 = 70.1 + k5*1.26
MTB > print k6-k9
K6       58.9494     # lower 95% CI
K7       60.8507     # upper 95% CI
K8       67.5622     # lower 95% CI
K9       72.6377     # upper 95% CI
```

The confidence limits for the treated group   LL = 58.95 UL = 60.85
The confidence limits for the control group   LL = 67.56 UL = 72.64
The confidence limits do not overlap.

8

10. Report and interpret parameters of biological interest

<u>Second planned comparison</u>: Growth in mixed versus differs from that in pure sugar.

H_A: $(\beta_{gluc+fruc}) \neq (1/3)(\beta_{glucose} + \beta_{fructose} + \beta_{sucrose})$

Source	DF	SS	MS	F
df2	1	48.13	48.133	6.11
Error	38	299.47	7.881	
Total	39	347.60	8.913	

We compare the two groups (one with 10 and one with 30 observations) to obtain the MS due to the two groups.

```
MTB > let k2 = 48.133/k1
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k2 k4
K2      8.82275        # Fratio
K4      0.00475895     # pvalue
```

We use the MS_{error} from all 5 groups to recompute the F-ratio. The F-ratio is 8.82 instead of 6.11, as in the ANOVA table.

The confidence limits are on the order of +/− 2 units.

```
MTB > print k6 - k9
K6  59.3870      # lower 95% CI  pure sugar
K7  61.6790      # upper 95% CI
K8  57.0997      # lower 95% CI  mixed sugar
K9  58.9003      # upper 95% CI
```

Mixed glucose + fructose reduces growth by $(60.533 – 58)/60.533) = 4\%$ relative to pure sugars

<u>Third planned comparison</u>: Growth in polysaccharide differs from that in monosaccharides.

H_A: $(\beta_{sucrose}) \neq (1/2)(\beta_{glucose} + \beta_{fructose})$

Source	DF	SS	MS	F	P
df3	1	190.82	190.817	58.94	0.000
Error	28	90.65	3.237		
Total	29	281.47	9.706		

```
MTB > let k1 = 245.5/45
MTB > let k2 = 190.817/k1
MTB > cdf k2 k3;
SUBC> f 1 45.
MTB > let k4 = 1-k3
MTB > print k3 k4
k3      34.9766     # Fratio
K4      0.000000417 # pvalue
```

The monosaccharides suppress growth more than the polysaccharide (sucrose).

```
MTB > describe c11;
 SUBC> by c12.
     df3       N     MEAN    MEDIAN   TRMEAN   STDEV    SEMEAN
  Gluc      1    10   59.300   59.500   59.250   1.636    0.517
  Fruc      2    10   58.200   58.000   58.125   1.874    0.593
  Sucr      3    10   64.100   64.500   64.000   1.792    0.567


 MTB > aoveoneway c22 c23 c25
                           INDIVIDUAL 95 PCT CI'S FOR MEAN
                           BASED ON POOLED STDEV
  LEVEL       N     MEAN    STDEV  --+---------+---------+---------+----
  glucose    10    59.300   1.636        (---*----)
  fructose   10    58.200   1.874  (----*---)
  sucrose    10    64.100   1.792                           (---*----)
```

10.   Report and interpret parameters of biological interest

Conclusions from the 3 planned comparisons of parameters (means).

We reject the hypothesis that a 2% sugar solution has no effect on growth.

$F_{1,45} = 152.564$ $\alpha = 0.017$, $p < 0.0001$

We reject the hypothesis that Mixed glucose + fructose has the same effect as pure sugars

$F_{1,45} = 8.82$ $\alpha = 0.017 > p = 0.00476$

We reject the hypothesis that monosaccharides (fructose, glucose) have the same effect as the polysaccharide (sucrose).

$F = 34.98$ $\alpha = 0.017 > p = 0.00000417$

Conclusions based on confidence limits for the 3 planned comparisons.

A 2% sugar solution (4 groups) reduces growth by $(70.1 - 59.9)/70.1 = 15\%$

Control: mean(Len) = 70.1 units, 95% CI = 67.6 to 72.6units, n = 10

With sugar: mean(Len) = 59.9 units,   95% CI = 58.9 to 60.9 units, n = 40

Mixed glucose + fructose reduces growth by $(60.533 - 58)/60.533) = 4\%$ relative to pure sugars.

Mixed: mean(Len) = 58 units, 95% CI = 57.1 to 58.9, units, n = 10

Pure sugar: mean(Len) = 60.5 units,   95% CI = 59.4 to 61.8 units, n = 30

Sucrose (a polysaccharide) reduces growth less than glucose or fructose (monosaccharides).

Sucrose: mean(Len) = 64.1 units, 95% CI = 62.6 to 65.6 units, n = 10

Fructose: mean(Len) = 58.2 units, 95% CI = 56.7 to 59.7 units, n = 10

Glucose: mean(Len) = 59.3 units, 95% CI = 57.8 to 60.8 units, n = 10

For these three planned comparisons, confidence limits do not overlap and so conclusions were readily drawn from examination of the confidence limits.