Lecture 11

1. Proc Reg subset selection with class variables
2. Making CLASS variables and interactions
3. Computing predicted values
4. Sample size estimates for t-tests: Proc Power
5. Sample size estimates for ANOVA: Proc GLMpower

Proc Reg subset selection with class variables

NHANES data for adults aged 20–29: model systolic blood pressure (SBP) on diastolic blood pressure (DBP), height, weight, and waist-to-hip ratio—all continuous predictors—plus two categorical predictors:

- sex *(2 levels: F , M)*
- education category *(3 levels: grade school, high school, college)*

```
Proc GLM  data = corrected_nhanes20;
   class sex education_category;
   model  SBP =  ht_in  wt_lbs  educ_yrs   dbp   waist_hip
                 age  sex  education_category
                 sex*education_category / solution;
```

Parameter		Estimate		Standard Error	t Value
Intercept		81.90605641	B	4.66303904	17.56
ht_in		-0.02438108		0.05569066	-0.44
wt_lbs		0.05165548		0.00470155	10.99
educ_yrs		-0.00271070		0.06457209	-0.04
dbp		0.44551653		0.01519333	29.32
waist_hip		-0.86954111		2.39310536	-0.36
age		-0.10594330		0.04883977	-2.17
sex	F	-5.74340522	B	0.94646277	-6.07
sex	M	0.00000000	B	.	.
education_category	1_grade_school	1.03253605	B	0.80835332	1.28
education_category	2_high_school	0.05357176	B	0.80645882	0.07
education_category	3_college	0.00000000	B	.	.
sex*education_catego	F 1_grade_school	-1.52047176	B	0.94547534	-1.61
sex*education_catego	F 2_high_school	-1.35004801	B	1.05508767	-1.28
sex*education_catego	F 3_college	0.00000000	B	.	.
sex*education_catego	M 1_grade_school	0.00000000	B	.	.
sex*education_catego	M 2_high_school	0.00000000	B	.	.
sex*education_catego	M 3_college	0.00000000	B	.	.

```
NOTE: The X'X matrix has been found to be singular, and a generalized inverse
      was used to solve the normal equations.  Terms whose estimates are
      followed by the letter 'B' are not uniquely estimable.
```


Proc Reg has neither a CLASS statement nor a * for interaction terms.

Proc Reg can handle indicator (0/1) variables just fine.

However, Proc Reg cannot handle categorical variables with more than 2 levels, or categorical variables with character values, such as F and M.

This means we must create indicator variables for each level of categorical variables, and also make interaction terms.

Here is how Proc GLM does it:

Continuous-by-Class Effects Columns are constructed by multiplying the continuous values into the design columns for the class effect.

Data			Design Matrix				
				A		X*A	
X	A	μ	X	A1	A2	X*A1	X*A2
21	1	1	21	1	0	21	0
24	1	1	24	1	0	24	0
22	1	1	22	1	0	22	0
28	2	1	28	0	1	0	28
19	2	1	19	0	1	0	19
23	2	1	23	0	1	0	23

Making indicator variables for categorical predictors

We follow the Proc GLM methods to make indicator variables in a data step:

```
data add_indicators;
 set corrected_nhanes20;
 grade_school = (education_category = "1_grade_school");
 high_school = (education_category = "2_high_school");
 college = (education_category = "3_college");

 female = (sex = "F");
 female_gradeschool = female * grade_school;
 female_highschool = female * high_school;
 female_college = female*college;
```

Obs	id	sex	female	education_category	grade_school	high_school	college	female_gradeschool	female_highschool	female_college
1	37905	M	0	2_high_school	0	1	0	0	0	0
2	40846	F	1	1_grade_school	1	0	0	1	0	0
3	38425	F	1	1_grade_school	1	0	0	1	0	0
4	45394	M	0	2_high_school	0	1	0	0	0	0
5	37436	M	0	2_high_school	0	1	0	0	0	0
6	39707	M	0	1_grade_school	1	0	0	0	0	0
7	7988	M	0	1_grade_school	1	0	0	0	0	0
8	44892	M	0	1_grade_school	1	0	0	0	0	0
9	9160	F	1	1_grade_school	1	0	0	1	0	0
10	4865	M	0	2_high_school	0	1	0	0	0	0

Now fit the same model we used in Proc GLM in Proc REG:

```
Proc REG  data = add_indicators;
    model sbp =  ht_in wt_lbs educ_yrs dbp waist_hip age
           female grade_school high_school
        college female_gradeschool female_highschool female_college;
```

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	B	81.90606	4.66304	17.56	<.0001
ht_in	1	-0.02438	0.05569	-0.44	0.6616
wt_lbs	1	0.05166	0.00470	10.99	<.0001
educ_yrs	1	-0.00271	0.06457	-0.04	0.9665
dbp	1	0.44552	0.01519	29.32	<.0001
waist_hip	1	-0.86954	2.39311	-0.36	0.7164
age	1	-0.10594	0.04884	-2.17	0.0301
female	B	-5.74341	0.94646	-6.07	<.0001
grade_school	B	1.03254	0.80835	1.28	0.2016
high_school	B	0.05357	0.80646	0.07	0.9470
college	0	0	.	.	.
female_gradeschool	B	-1.52047	0.94548	-1.61	0.1079
female_highschool	B	-1.35005	1.05509	-1.28	0.2008
female_college	0	0	.	.	.

These regression coefficients are exactly the same as those from Proc GLM.

What are the Bs?


```
NOTE: Model is not full rank. Least-squares solutions for the parameters are
      not unique. Some statistics will be misleading. A reported DF of 0 or B
      means that the estimate is biased.
NOTE: The following parameters have been set to 0, since the variables are a
      linear combination of other variables as shown.

college =  Intercept - grade_school - high_school

female_college =  female - female_gradeschool - female_highschool
```

The problem is caused by the extra indicators that are linear combinations of other indicators, as identified by SAS.

Let X be the $p \times n$ matrix of p predictors (independent variables), with n observations.

To find the least squares estimates of the regression coefficients, we need to find the inverse of the matrix $(X^{\mathsf{T}} X)$.

$(X^{\mathsf{T}} X)$ will not have an inverse if some of the columns of X are linear combinations of other columns.

See Gilbert Strang (2005) *Linear Algebra and Its Applications (4th ed.)*

Now we can use automatic subset selection in Proc Reg:

```
proc reg data=add_indicators;

   model sbp =  ht_in wt_lbs educ_yrs dbp waist_hip age
          female grade_school high_school college
         female_gradeschool female_highschool female_college
       / selection=cp ;
```

```
                              C(p) Selection Method

             Number of Observations Read                         3507
             Number of Observations Used                         3369
             Number of Observations with Missing Values           138

Number in
  Model        C(p)  R-Square  Variables in Model
      6      2.3803    0.4617  wt_lbs dbp age female grade_school female_college
      5      3.1058    0.4613  wt_lbs dbp age female high_school
      7      4.2378    0.4617  ht_in wt_lbs dbp age female grade_school
                               female_college
      7      4.2713    0.4617  wt_lbs dbp age female grade_school
                               female_gradeschool female_college
      7      4.2713    0.4617  wt_lbs dbp age female grade_school
                               female_highschool female_college
      7      4.2713    0.4617  wt_lbs dbp age female grade_school
                               female_gradeschool female_highschool
      7      4.2713    0.4617  wt_lbs dbp age grade_school female_gradeschool
                               female_highschool female_college
      7      4.3233    0.4617  wt_lbs dbp waist_hip age female grade_school
                               female_college
```

Best models have $C_p \le p$. Why do several models have the same C_p?

The model with "best" C_p is

```
Number in
  Model        C(p)  R-Square  Variables in Model
      6      2.3803    0.4617  wt_lbs dbp age female grade_school female_college
```

What does it mean to have only 1 of the 3 education categories?

Hierarchical model: every term involved in an interaction also appears as a main effect. Helps with interpretation.

Is this a hierarchical model?

Calculating predicted values

Continuing with the model with “best” C_p

```
Number in
  Model        C(p)  R-Square  Variables in Model
      6     2.3803    0.4617  wt_lbs dbp age female grade_school female_college
```

To calculate predicted values we need the fitted regression equation:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	79.78739	1.48050	53.89	<.0001
wt_lbs	1	0.05265	0.00369	14.26	<.0001
dbp	1	0.44286	0.01488	29.75	<.0001
age	1	-0.11817	0.04740	-2.49	0.0127
female	1	-7.09551	0.28900	-24.55	<.0001
grade_school	1	1.03568	0.30668	3.38	0.0007
female_college	1	1.54684	0.65828	2.35	0.0188

Suppose we want to calculate a predicted value for this new case:

wt_lbs	dbp	age	female	grade_school	female_college	
150	70	25	0	0	0	*male HS*

It is easy to calculate a predicted value by hand:

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	79.78739	1.48050	53.89	<.0001
wt_lbs	1	0.05265	0.00369	14.26	<.0001
dbp	1	0.44286	0.01488	29.75	<.0001
age	1	-0.11817	0.04740	-2.49	0.0127
female	1	-7.09551	0.28900	-24.55	<.0001
grade_school	1	1.03568	0.30668	3.38	0.0007
female_college	1	1.54684	0.65828	2.35	0.0188

$$79.78739 + 0.05265 * 150 + 0.44286 * 70 - 0.11817 * 25 = 115.731$$

Harder to get a standard error or confidence interval for the prediction, because this depends on the location of the new case relative to the center of the data.

Predicted values

To get predictions from any of the regression procedures, create a new data set with the explanatory variables for the new cases and then SET it on top of the data.

```
Data B;     response is omitted, only explanatory variable values
   input  ht_in  wt_lbs dbp age female grade_school
         high_school female_college;
   cards;
 65 150 70 25 0 0 1 0    male HS
 65 150 70 25 1 0 1 0    female HS
 65 150 70 25 1 0 0 1    female college
 ;

Data pred;
   set B add_indicators;
```


The three cases added to data set `pred` do not affect the model fit because they are missing the response.

```
Proc REG  data = pred;
   model sbp = wt_lbs dbp age female grade_school female_college;
   output out = Q1  P  = prediction  STDP  = SE_fit;
```

There are two kinds of predictions from a regression: a predicted *mean* and a predicted *observation*. The predictions are identical, but the SE is larger for a predicted observation.

Usually we are interested in a predicted mean.

`STDP` gives SE(predicted mean).

`STDI` gives SE(predicted observation)

In simple linear regression with one predictor x, the standard error of the predicted mean at x_p is

$$\text{Root MS(Error)} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{(n-1)\text{SD}(x)^2}}$$

so predictions farther from the center of the data have bigger standard errors.

When we predict a new observation, the variability is essentially the sum of the variability of the mean plus the variability around the mean.

Standard error of prediction for a new observation at x_p

$$\text{Root MS(Error)} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{(n-1)\text{SD}(x)^2}}$$


```
proc print data=Q1 (obs=5);
  var sbp prediction SE_fit ;
```

Obs	sbp	prediction	SE_fit
1	.	115.731	0.30067
2	.	108.636	0.30216
3	.	110.182	0.58642
4	.	.	.
5	109	107.263	0.22644

Why is the 4th row missing a prediction?

To compare the first 3 models from the subset screening, we can compute similar predictions:

```
proc Reg data = pred;
  model sbp  =  wt_lbs dbp age female high_school;
  output out = Q2   P = prediction2   stdp = SE_fit2 ;

proc Reg data = pred;
  model sbp  =  ht_in wt_lbs dbp age female
      grade_school female_college;
  output out = Q3   P = prediction3   stdp = SE_fit3 ;

data c;
  merge Q1 Q2 Q3;    unusual merge with no BY variable

proc print data = c (obs = 3);
  var prediction prediction2 prediction3 SE_fit SE_fit2  SE_fit3;
```


Obs	prediction	prediction2	prediction3	SE_fit	SE_fit2	SE_fit3
1	115.731	115.676	115.884	0.30067	0.32935	0.35274
2	108.636	108.709	108.610	0.30216	0.31170	0.30381
3	110.182	109.711	110.183	0.58642	0.20147	0.58645

The predictions are almost identical, although the SE vary.

Predicted values are much more stable than regression coefficients when we make small changes in regression model.

95% confidence interval for a prediction is

$$\hat{y} \;\pm\; t(0.05, \text{error df}) * \text{SE}$$

Estimating Sample Size

Suppose we are planning a trial to compare two treatments A and B.

Our statistical analysis plan is a two-sample t-test to compare μ_A and μ_B at the $\alpha = 0.05$ level, so $p < .05$ will be significant. Null hypothesis is $H_0 : \mu_A = \mu_B$.
Or, if we let $\delta = \mu_A - \mu_B$, then $H_0 : \delta = 0$.

		Truth	
		No difference ($\delta = 0$)	Real difference ($\delta \neq 0$)
Significance Test	t-test finds $\delta = 0$	OK	Type 2 error: *test finds no difference but there is one*
Significance Test	t-test finds $\delta \neq 0$	Type 1 error: *test asserts real difference but there isn't one*	OK

Assuming “No difference" is true, how can we set our chance of being correct (upper left square)?

Assuming “Real difference" is true, the chance of *not* making a Type 2 error is called the **power** of the test and it is the chance of finding a significant difference when there really is one. People like to have power of at least 80%.

Once we select α, the only way to increase power is to increase the sample size.

Is it worthwhile to do a study that has power of 50%?

Is it ethical?

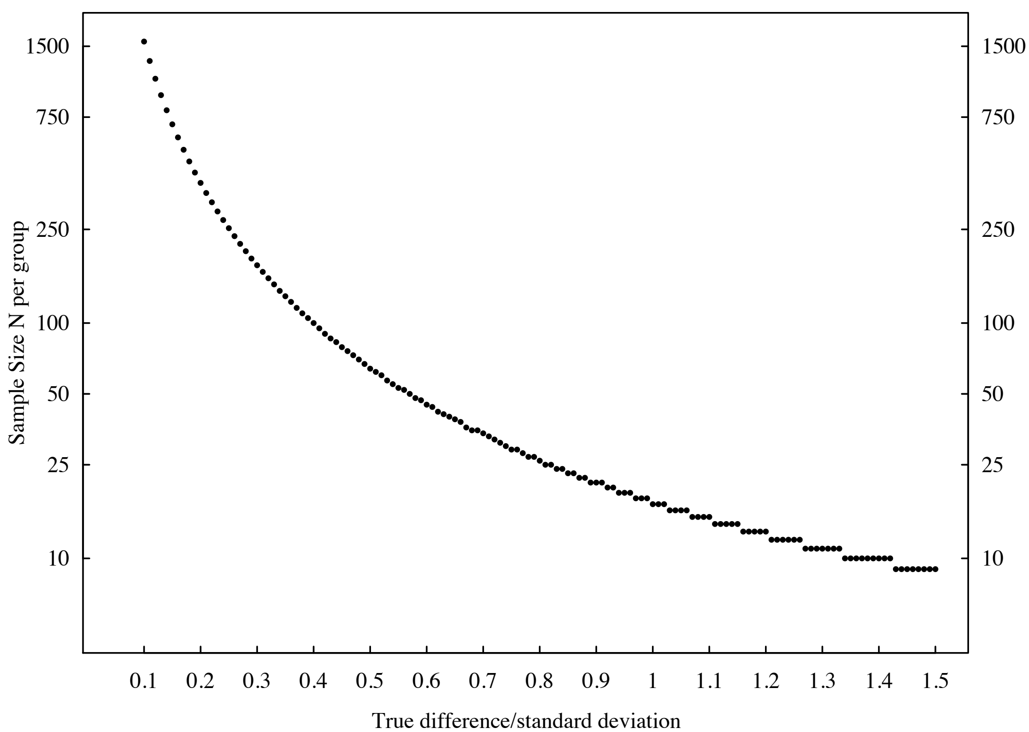
What about pilot studies?

Estimating sample size for *t*-tests

The ingredients we need:

- study design: two samples, one sample, paired t-test
- estimate of difference δ, or smallest difference of practical importance
- estimate of variability σ between experimental units
- two-sided or one-sided test, usually two-sided
- α, cutoff value for significance, usually 0.05
- required power, usually 80%

The ratio (δ/σ) is called the **effect size**, which has no units and so can be compared across endpoints and experiments.



Required number in each group for a 2-sample, 2-sided t-test at α = .05 and power = 80%.

Sample size for a 2-sample, 2-sided t-test at $\alpha = .05$ and power = 80%.

δ/σ	N per group
1.0	17
.8	25
.6	50
.4	100

Proc Power to estimate sample size

Example 1. A researcher wants to find a difference greater than $\delta = 50$ mg between two treatments, A and B. Volunteers will be randomly assigned in equal numbers to either A or B.

She believes that the standard deviation of the response is $\sigma = 60$ mg (an effect size of $\delta/\sigma = 0.83$).

She plans to do a 2-sided t-test at $\alpha = .05$ and wants power = 80% .

```
Proc Power;
      twosamplemeans   meandiff  = 40 45 50 55
      stddev = 60  alpha = 0.05   power = 0.8
      npergroup = .  ;
```

```
The POWER Procedure
Two-sample t Test for Mean Difference

Fixed Scenario Elements

Distribution              Normal
Method                     Exact
Alpha                       0.05
Standard Deviation            60
Nominal Power                0.8
Number of Sides                2
Null Difference                0
```

Computed N Per Group

Index	Mean Diff	Actual Power	N Per Group
1	40	0.808	37
2	45	0.801	29
3	50	0.807	24
4	55	0.806	20

Power calculation: minimum detectable difference

Fix N, power, α and σ: we can find the corresponding δ, the minimum detectable difference in that outcome. For some reason, this is called a *power calculation*.

In our example 1, we chose two groups of $N = 24$. For secondary endpoints cholesterol, with SD of 30 mg/dL, and diastolic blood pressure, with SD of 10 mm Hg, we find:

```
Proc Power;
  twosamplemeans  meandiff = .    stddev= 30 10
  alpha = 0.05 power=0.8 npergroup=24;
```

The POWER Procedure
Two-sample t Test for Mean Difference

Index	Std Dev	Mean Diff	
1	30	24.78	*cholesterol*
2	10	8.26	*DBP*

Example 2. We want to compare two treatments, A and B. Volunteers will be randomly assigned in equal numbers to either A or B. The response will be measured at baseline and after 6 months at the end of the study, and we want to compare *changes from baseline.* We would like find a difference of about $d = 20$ between two treatments. We plan to do a 2-sided test at $\alpha = .05$ and want power = 80% .

In the literature, we found an value of 36 for the standard deviation of the response. But this SD applies to the single measurement at baseline or to the final measurement, *but not to the difference between them.*

Baseline and final measurements from the same person will be correlated.
We can use an estimate based on the variance of the difference of correlated random variables:

$$\text{SD of differences} = (\text{SD of one measurement})\sqrt{2(1-\rho)},$$

where ρ is the population correlation between measurements from the same person.

$$\begin{aligned}
\mathrm{Var}(X-Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)-2\mathrm{Cov}(X,Y)\\
&= \sigma_X^2+\sigma_Y^2-2\sigma_X\sigma_Y\rho\\
&= 2\sigma^2+\sigma^2-2\sigma^2\rho\\
&= 2\sigma^2(1-\rho)\\
\mathrm{SD}(X-Y) &= \sigma\sqrt{2(1-\rho)}
\end{aligned}$$

If you had paired data to estimate ρ, you could estimate the SD of the differences directly and would not need to estimate ρ. Thus, ρ must usually be guessed. A conventional choice is $\rho = 0.6$, which assumes moderate correlation.

What happens as ρ gets close to 1?

For our example:

$$\text{SD of differences} = (\text{SD of one measurement})\sqrt{2(1-\rho)} = 36\sqrt{2(1-0.6)} = 32.2,$$

so the SD of changes, 32.2, is smaller than the SD of a single measurement, 36.

Use $\hat{\sigma} = 32.2$ with a minimum difference of $\delta = 20$ in Proc Power to compute the sample size to compare changes from baseline.

```
Proc Power;
  twosamplemeans  meandiff = 20   stddev = 36 32.2
  alpha = 0.05 power=0.8 npergroup=.;
```

Index	Std Dev	Actual Power	N Per Group	
1	36.0	0.801	52	
2	32.2	0.803	42	*uses SD of changes*

Example 3. In the last example, our endpoint was change from baseline. Suppose that we want to detect a *within-group* change of at least $\delta = 25$. What test?

Earlier we used SD = 36 for both baseline and final measurements, but in fact the paper says the baseline SD was 42 while the final SD was 30. We'll use the same estimate of correlation, $\rho = 0.6$.

```
Proc Power;
  pairedmeans  meandiff=25 corr=.6 stddev=36     average SD
               alpha = 0.05 power=0.8 npairs = . ;

  pairedmeans  meandiff=25  corr=.6
              pairedstddevs  = 42 | 30   first SD, second SD
               alpha = 0.05 power=0.8 npairs = . ;
```



```
                    The POWER Procedure
             Paired t Test for Mean Difference
Estimate using average  of 2 SDs:

             Mean Difference                  25
             Standard Deviation               36
             Correlation                     0.6

                    Actual        N
                     Power    Pairs
                     0.827       16

Estimate using 2 SDs:

            Mean Difference                   25
            Standard Deviation 1              42
            Standard Deviation 2              30
            Correlation                      0.6

                    Actual        N
                     Power    Pairs
                     0.813       17
```

Sample size for testing correlation

Example 4. In the literature, we found a reported correlation between SBP and DBP of $r = .54$ in healthy young adults. For a study of blood pressure measurements in a clinical population, how many subjects would we need to have power of 80% to detect a similar correlation at the $\alpha = 0.05$ level?

Proc Power assumes that we will compute a Pearson correlation, and that the variables are bivariate Normal. Try a range of true correlations: $\rho = .3, .4, .5, .6$.

```
Proc Power;
  onecorr   corr =  .4 .45  .5  .55 .6
        power=.8  ntotal=. ;
```

Will sample size increase or decrease with ρ?


```
                    The POWER Procedure
          Fisher's z Test for Pearson Correlation

                 Fixed Scenario Elements

Distribution                        Fisher's z transformation of r
Method                                        Normal approximation
Nominal Power                                                  0.8
Number of Sides                                                  2
Null Correlation                                                 0
Nominal Alpha                                                 0.05
Number of Variables Partialled Out                               0


                     Computed N Total

                            Actual    Actual        N
           Index     Corr    Alpha     Power    Total

               1     0.40   0.0500     0.802       46
               2     0.45   0.0499     0.806       36
               3     0.50   0.0499     0.814       29
               4     0.55   0.0498     0.807       23
               5     0.60   0.0497     0.813       19
```

Test options for Proc Power *(from SAS Help document)*

ONESAMPLEMEANS	one-sample t test, confidence interval precision, or equivalence test
TWOSAMPLEMEANS	two-sample t test, confidence interval precision, or equivalence test
PAIREDMEANS	paired t test, confidence interval precision, or equivalence test
ONEWAYANOVA	one-way ANOVA including single-degree-of-freedom contrasts
ONESAMPLEFREQ	tests of a single binomial proportion
PAIREDFREQ	McNemar's test for paired proportions
TWOSAMPLEFREQ	chi-square, likelihood ratio, and Fisher's exact tests for two independent proportions
MULTREG	tests of one or more coefficients in multiple linear regression
ONECORR	Fisher's z test and t test of (partial) correlation
TWOSAMPLESURVIVAL	log-rank, Gehan, and Tarone-Ware tests for comparing two survival curves

Sample size for ANOVA: Proc GLMPower

GLMPower estimates sample size or power for general linear models. It needs a data set with estimated means for each combination of class variables. It is possible to add continuous predictors with an estimated correlation with the response.

Suppose we plan to compare e treatment groups, *A*,*B*, *C*, and expect the standard deviation $\sigma = 10$. Make data set with the expected mean values for the groups:

```
data group_means;
  input trt $ y;
  cards;
  A 40
  B 50
  C 60
  ;
```

```
proc glmpower data=group_means;
  class trt;
  model y = trt;
  power stddev = 10
        ntotal = .     find sample size
        power  = .80 ;
  contrast "A vs B" trt -1 1 0;   compare A to B
```


Dependent Variable	y
Error Standard Deviation	10
Nominal Power	0.8
Alpha	0.05

Computed N Total

Index	Type	Source	Test DF	Error DF	Actual Power	N Total
1	Effect	trt	2	15	0.805	18
2	Contrast	A vs B	1	48	0.815	51

Why is the sample size so much larger to compare A to B, than to find a difference between treatment groups?

Sample size estimate for two-factor ANOVA

Physical Capacity Evaluation (PCE), a new method for rating physical ability and impairment in the elderly (*Am J Public Health,* 1995; 85:558-560).
The paper reported results by gender for three age groups: age 65 to 74 ($n = 89$), age 75 to 84 ($n = 121$), and age 85 to 97 ($n = 79$).

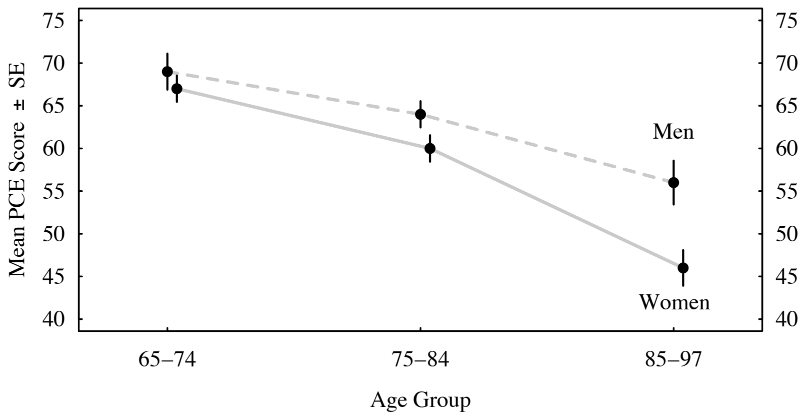


TABLE 1—Subjects' Global Health Assessment Questionnaire (HAQ) and Physical Capacity Evaluation (PCE) Scores, by Age Group and Sex

	Age, y						
	65–74		75–84		85–97		
	Men (n = 38)	Women (n = 51)	Men (n = 61)	Women (n = 60)	Men (n = 34)	Women (n = 45)	Total Sample (n = 289)
HAQ score[a]							
Median	0.00	0.38	0.12	0.50	0.38	1.38	0.38
Mean	0.38	0.52	0.35	0.68	0.59	1.33	0.63
SD	0.61	0.66	0.54	0.69	0.59	0.82	0.72
Score of 0, %	58	47	43	27	24	11	35
PCE score[b]							
Median	73	70	66	62	57	48	64
Mean	69	67	64	60	56	46	61
SD	13	11	12	12	15	14	15

[a]0 = no disability, 3 = unable to do any task.
[b]0 = worst function, 100 = best function.

Suppose we want to carry out a similar experiment with these groups.

Create a data set with the expected group means (cell means) for the 6 age-gender groups using the results from the paper:

```
data cellmeans;
  input gender $  age_group $  PCE;
  cards;
        Men         65-74        69.0
        Men         75-84        64.0
        Men         85-97        56.0
        Women       65-74        67.0
        Women       75-84        60.0
        Women       85-97        46.0
```


Because there are three F-test hypotheses, there are 3 sample sizes to estimate:

- n_1 per group to detect the gender main effect
- n_2 per group to detect the age main effect
- n_3 per group to detect the age-gender interaction

```
Proc GLMPower  data=cellmeans;
   class gender age_group;
   model PCE = gender   age_group   gender*age_group;  all 3 hypotheses

   power  stddev = 13   ntotal = .   power = .80 ;  solve for total sample n
```

The value for the standard deviation is the average of the group SDs from the paper.

```
Dependent Variable                  PCE
Error Standard Deviation             13
Nominal Power                       0.8
Alpha                              0.05
```

Computed N Total

Index	Source	Test DF	Error DF	Actual Power	N Total
1	gender	1	186	0.807	192
2	age_group	2	36	0.863	42
3	gender*age_group	2	564	0.802	570

To detect the gender main effect, we need $192/2 = 96$ per gender group or $192/6 = 32$ per age-gender group.

For the age main effect, we need $42/3 = 14$ per age group.

To detect the age-gender interaction, we need $570/6 = 95$ in each of the 6 groups.

Which effect is the most important?

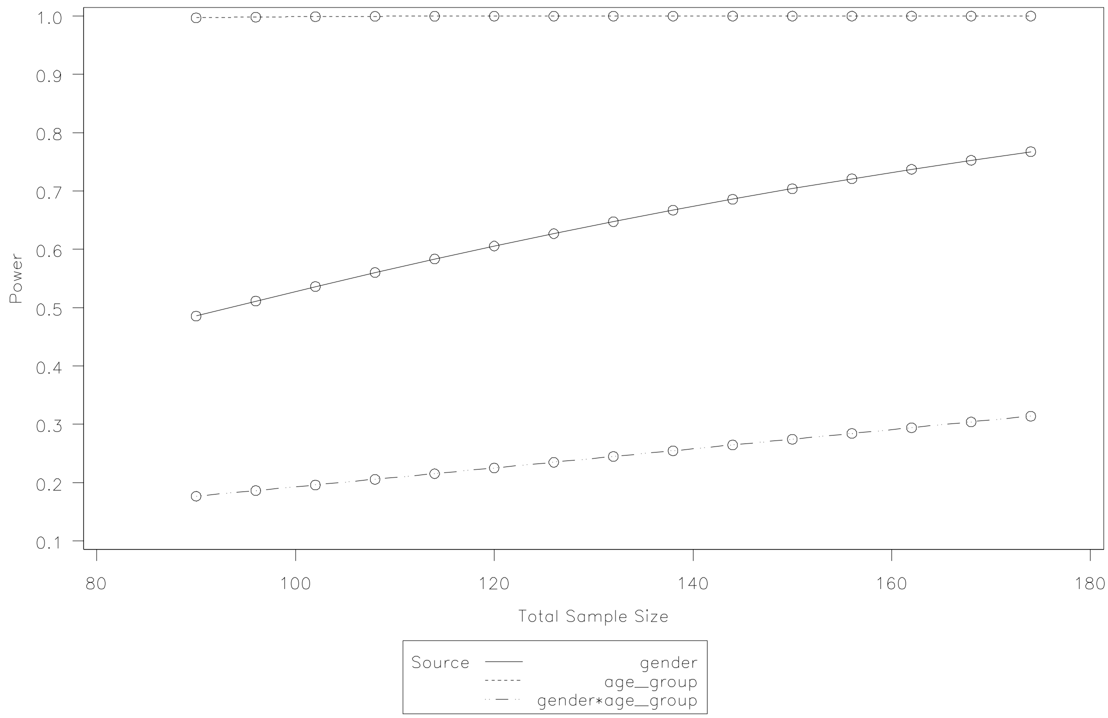
Power calculations for ANOVA

We can also fix a sample size, or range of sample sizes, and calculate power to detect each of the fixed effects, given a set of expected values for the cell means.

Using the same observed means as before, we try a range of *total* sample sizes from 90 to 180, and get a graph of power against sample size:

```
Goptions reset=all vsize=5in ftext=simplex ctext=black lfactor=1.5;

Proc GLMPower data=cellmeans;
  class gender age_group;
  model PCE = gender  age_group  gender*age_group;
  power stddev = 13
        ntotal = 120
        power  = .  ;     solve for power
  plot x=n min=90 max=180;
```



GLMPower also produces tables with estimated power at each n.