Midterm Exam

Changing your grade

- The TA's will not give you credit for a question that has an answer different from that in our answer key.
- Why?
  - Our answers are correct.
  - If you were confused by some aspect of a question, chances are a bunch of other people were as well. We'd have to regrade everyone's exams, to be fair, and given that grading is on a curve, the result would be little or no change to anyone's grade.

Answers to the 10 (or so) most frequently missed questions

#1

- On planet gargu there are five different genders. A child can have only one of these genders. Picking a person at random, the probability of each of these genders is .2. Let's say parents want to have three children in a row. What are the odds (before they have any children) that all three children will have the same gender?

  (a) 0.8%   (b) 12.5%   (c) 1.25%   (d) 4%   (e) 50%

#1

- If you got this question right, congratulations – you were one of the few in the class to get it right!
- Nonetheless, this is a simple probability problem, I expect you to be able to solve problems like this, and you should expected to be tested more on this on the final.

The catch

- Most people said the answer was (a) 0.8%
- However, this is the (correct) probability of having 3 children who all have a *particular* gender. E.G. it's the probability of having 3 children who all have gender A
- The question doesn't ask for that. It asks for the probability of having the 3 children have the same gender – this could be any of the 5 genders

One way of solving the problem

- What is the probability of getting three kids all with a particular gender, e.g. A?
  - $(0.2)(0.2)(0.2) = 0.008$
- Many people gave this answer, 0.8%. If you gave this answer, you didn't take into account that this is just one possible gender, and you need to count all 5. (They are mutually exclusive events, so you can just add the probabilities.)
- Answer $= (5)(0.008) = 0.04 = 4\%$

Another way of solving the problem

- How many possible ways can someone have 3 children of the same gender? Call the genders A, B, C, D, and E
  - AAA, BBB, CCC, DDD, EEE = 5
- How many possible combinations of 3 children are there?
  - $5 \cdot 5 \cdot 5 = 5^3$
- What's the probability of getting 3 kids all the same gender?
  - $5/5^3 = 1/5^2 = 4/100 = 4\%$

#2

- The following lines appear in the t-table at the back of your book:

Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11

I calculate an observed value t_{obt} of -2.24, from a sample of size 10. What do I conclude?

Your choices

(a) I accept the null hypothesis $H_0: \mu = \mu_0$, if my alternative hypothesis is $H_a: \mu \neq \mu_0$, and $\alpha=0.05$.

(b) I reject the null hypothesis, $H_0 : \mu = \mu_0$, with $p<0.05$, if my alternative hypothesis is $H_a: \mu \neq \mu_0$.

(c) I reject the null hypothesis if my alternative hypothesis only applies to the upper tail.

(d) I accept the null hypothesis that $\mu < \mu_0$.

(e) I reject the null hypothesis, with $p<0.025$, if my alternative hypothesis is $H_a: \mu \neq \mu_0$.

Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11

$t_{obt} = -2.24$, N=10

- First of all, which line of the table are we dealing with?
  - Degrees of freedom = 9   (Most got this right.)

Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11

$t_{obt} = -2.24$, N=10

- What needs to be true for our t_{obt} to indicate a significant result?
  - Well, $|t_{obt}| > 1.83$, which corresponds to 5% in a *single tail* of the t-distribution
  - So, significant at the 0.05 level if $H_a : \mu < \mu_0$

Why < (lower tail)?
Because $t_{obt} < 0$

Degrees of freedom	25%	10%	5%	2.5%	1%	0.5%
9	0.70	1.38	1.83	2.26	2.82	3.25
10	0.70	1.37	1.81	2.23	2.76	3.17
11	0.70	1.36	1.80	2.20	2.72	3.11

$t_{obt} = -2.24$, N=10

- What needs to be true for our t_{obt} to indicate a significant result?
  - Well, $|t_{obt}| > 1.83$, which corresponds to 5% in a *single tail* of the t-distribution
  - So, significant at the 0.05 level if $H_a : \mu < \mu_0$
  - Or, significant at the 0.10 level if $H_a : \mu \neq \mu_0$
- These are the cases in which you *reject* the null hypothesis

If $H_a : \mu < \mu_0$, reject H_0 at the 0.05 level
If $H_a : \mu \neq \mu_0$, reject H_0 at the 0.10 level

(a) I accept the null hypothesis H_0: $\mu = \mu_0$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$, and $\alpha = 0.05$.
(b) I reject the null hypothesis, H_0 : $\mu = \mu_0$, with $p<0.05$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$.
(c) I reject the null hypothesis if my alternative hypothesis only applies to the upper tail.
(d) I accept the null hypothesis that $\mu < \mu_0$.
(e) I reject the null hypothesis, with $p<0.025$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$.

Needs $p<0.10$

If $H_a : \mu < \mu_0$, reject H_0 at the 0.05 level
If $H_a : \mu \neq \mu_0$, reject H_0 at the 0.10 level

(a) I accept the null hypothesis H_0: $\mu = \mu_0$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$, and $\alpha = 0.05$.
(b) I reject the null hypothesis, H_0 : $\mu = \mu_0$, with $p<0.05$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$.
(c) I reject the null hypothesis if my alternative hypothesis only applies to the upper tail.
(d) I accept the null hypothesis that $\mu < \mu_0$.
(e) I reject the null hypothesis, with $p<0.025$, if my alternative hypothesis is H_a: $\mu \neq \mu_0$.

#3

- You come upon three doors, one already open and two closed. Someone tells you that there is a prize behind one and only one of the two closed doors. What is the probability that the prize is behind one of these doors? (a) 33.33% (b) 67.67% (c) 50% (d) 0% (e) 100%
- This was intended as a bit of a trick question, and it's not worth spending much time on.
- It seems tricky because (1) in real life people don't ask questions like this; (2) the question sounds ambiguous, and might be in casual conversation, but isn't in this context.

#3

- You come upon three doors, one already open and two closed. Someone tells you that there is a prize behind one and only one of the two closed doors.
- What would you answer if I came up to you and said, "what is the probability that the prize is behind one of these doors?"
  - You've been told there is a prize behind one of these doors, so (assuming you trust the person who told you this) you should answer 100%
  - (Trusting them is a typical assumption on a probability exam. If you weren't sure if you could trust them, you could have asked.)

#3

- You come upon three doors, one already open and two closed. Someone tells you that there is a prize behind one and only one of the two closed doors. What is the probability that the prize is behind one of these doors?
  (a) 33.33%   (b) 67.67%   (c) 50%   (d) 0%   (e) 100%
- What about the other answers?
  - "I thought it was asking 'probability that *this* door has the prize', and said either 33% or 50%"
  - Alt: 'assuming the probability is the same for all doors (or all closed doors), what is that equal probability?'
  - (1) that's not what it's asking; (2) why would you assume equal probability for each door, and thus answer 33.33% or 50%?

#4

- "Which of the following is the best statement of a consequence of the central limit theorem?"
- What it's really asking: If one takes random samples of size n from a population of mean μ and standard deviation σ, then as n gets large the sampling distribution of the mean has what distribution?
  - Approximately normal
  - Mean = μ
  - Standard deviation $\sigma/\sqrt{n}$
- We went over this many times in class. Go back over your notes.

#4

- Many people answered mean = μ, standard deviation σ. I don't know where this comes from.
- Many people answered mean = μ, standard deviation $\sigma/\sqrt{n-1}$
  - This is NOT using the sample standard deviation to estimate the population standard deviation. Do not use n-1 here!
  - This is a theoretical result. We proved in class that the denominator was sqrt(n)

#5

- Which of the following is false, for a z-test of significance:
  - (a) All else being equal, a result will be more likely to be significant if n (the sample size) is larger.
  - (b) All else being equal, a result will be more likely to be significant if the standard deviation is smaller.
  - (c) If we double our number of samples, and our variance also doubles, then we have not changed the significance of our results.
  - (d) We are less likely to find a result significant if we decrease α.
  - (e) If we double our number of samples, and our standard deviation also doubles, then we have not changed the significance of our results.

Recall z-test for significance

- Compare z_{crit} to
    $z_{obt} = $ (observed – expected)/(s/sqrt(n))
- Increasing n -> z_{obt} increases -> result more likely to be significant
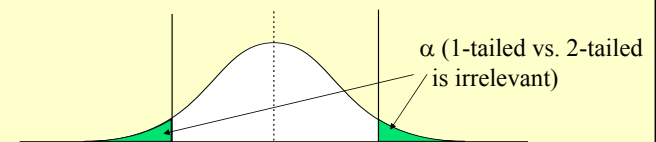- Decreasing s -> z_{obt} increases -> result more likely to be significant

#5

- Which of the following is false, for a z-test of significance:
  - (a) All else being equal, a result will be more likely to be significant if n (the sample size) is larger.
  - (b) All else being equal, a result will be more likely to be significant if the standard deviation is smaller.
  - (c) If we double our number of samples, and our variance also doubles, then we have not changed the significance of our results.
  - (d) We are less likely to find a result significant if we decrease α.
  - (e) If we double our number of samples, and our standard deviation also doubles, then we have not changed the significance of our results.

(c) If we double our number of samples, and our variance also doubles, then we have not changed the significance of our results.

- The key thing to look at here is the SE=s/sqrt(n). Since we're talking about variance, let's square it:
  - $SE^2 = s^2/n$
- If we double the number of samples, and the variance doubles, $SE^2 = 2s^2/(2n) = s^2/n$
- Nothing changes in the equation for t_{obt}, so the significance of our results doesn't change.
- Many people said that (c) was false, but it is true.

(d) We are less likely to find a result significant if we decrease α.

α (1-tailed vs. 2-tailed is irrelevant)

- *Decreasing* α moves the critical value *farther out*, and thus we are less likely to find a result significant. True.

(e) If we double our number of samples, and our standard deviation also doubles, then we have not changed the significance of our results.

- Again, look at SE=s/sqrt(n).
- If we double the number of samples, and the standard deviation doubles,
  $$SE = 2s/sqrt(2n) = sqrt(2)\ s/sqrt(n)$$
  $$\neq s/sqrt(n)$$
- False

#5

- Which of the following is false, for a z-test of significance:
  - (a) All else being equal, a result will be more likely to be significant if n (the sample size) is larger.
  - (b) All else being equal, a result will be more likely to be significant if the standard deviation is smaller.
  - (c) If we double our number of samples, and our variance also doubles, then we have not changed the significance of our results.
  - (d) We are less likely to find a result significant if we decrease α.
  - (e) If we double our number of samples, and our standard deviation also doubles, then we have not changed the significance of our results.

#6

- What is the sampling distribution?
- Check your notes from class. It's the distribution of values taken by a *statistic* in all possible samples of the same size from the same population.
- It is not the distribution of values taken by a *parameter* in all possible samples of the same size from the same population.
  - Tricky because we are talking about a theoretical distribution, which might make you think of parameters instead of statistics
  - But the sampling distribution is explicitly the distribution we expect to see if we compute the sample *statistic* a whole bunch of times.
  - And anyhow, the *parameter* isn't going to take on a *distribution of values* in all possible *samples*. A parameter just has a true value, and it doesn't change from sample to sample.

#7

- If we do a significance test, and get a p of 0.049, which of the following is false:
  - (a) If the null hypothesis is actually true, the probability of getting an effect as large as, or larger than, the effect we observed is approximately 5%
  - (b) If we ran the experiment 200 times, we would expect to get an effect smaller than the effect we observed approximately 190 times, assuming the null hypothesis is really true.
  - (c) If $\alpha = 0.05$, we reject the null hypothesis.
  - (d) If $\alpha = 0.05$, then the null hypothesis is not true.
  - (e) If $\alpha = 0.01$, then we maintain the null hypothesis as viable.

This one's easy, it's (d). We can never tell from our significance test that the null hypothesis is not true, we can only say it's unlikely we would get the results we got if the null hypothesis were true.

#8

- Suppose H_0 is $\mu=5$, and H_a is $\mu\neq5$. We obtain a sample mean of 7, with a standard deviation of 4, for a sample size of 16. Suppose the relevant value of t_{crit} is 2.13, for a two-tailed hypothesis test with a=0.05. What is t_{obt}, and is the result significant at the $\alpha=0.05$ level?
  - (a) $t_{obt} = t_{crit} + 2$; significant
  - (b) $t_{obt} = (7 - 5) / 4$; not significant
  - (c) $t_{obt} = 2$; not significant
  - (d) $t_{obt} = 7/(4 / sqrt(16))$; significant
  - (e) I can't compute t_{obt} – I don't know the population standard deviation, σ.

t_{obt} = (observed – expected)/(s/sqrt(n)) = (7-5)/(4/sqrt(16)) = 2
This is insignificant, and (c) is the correct answer.
You *can* compute t_{obt} – the whole point of using the t-distr
is that you don't need to know the population σ.

#9

- Which is not true of t distributions?
  - (a) They have more probability (or area) in the tails than the standard normal or z distribution.
  - (b) The mean = 0 and the area under the curve = 1 for all t distributions.
  - (c) They are calculated using the sample standard deviation instead of the population standard deviation.
  - (d) There is a different t distribution for each degrees of freedom.
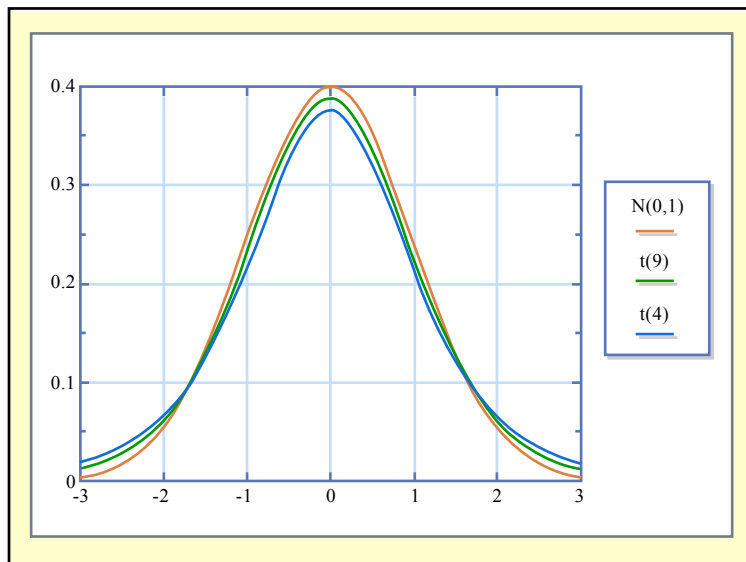 - (e) They are normal.

Figure by MIT OCW.

#9

- Which is not true of t distributions?
  - (a) They have more probability (or area) in the tails than the standard normal or z distribution. True
  - (b) The mean = 0 and the area under the curve = 1 for all t distributions. True
  - (c) They are calculated using the sample standard deviation instead of the population standard deviation. True
  - (d) There is a different t distribution for each degrees of freedom. True
  - (e) They are normal. False

#10

- The mean and variance of z-scored data are always
  - (a) mean = 1, variance = 0
  - (b) mean = μ, variance = σ^2
  - (c) mean = 0, variance = -1
  - (d) mean = 0, variance = 1
  - (e) none of the above; it depends upon the distribution of the data before z-scoring.

A lot of people said (b), and a lot said (e). The answer is (d) – the mean of z-scored data is always 0, and the variance is always 1. It does not depend upon the distribution of the data before scoring. Answer (b) is meaningless.

A bonus problem – many got this right, but still a fair number didn't

- Let $Y = x_1 + x_2 + \ldots + x_N$, where each xi is a random variable with distribution X. You want to do a MATLAB simulation to get an estimate of the distribution of Y, for different values of N. For each value of N, you do the "experiment" M times, i.e. you generate M samples of Y.
- Fill in the blanks: "As you increase ___, your estimate of the distribution gets closer to the true distribution of Y; as you increase ___, the distribution of Y becomes more normal.

times you run an experiment vs. # numbers added to get a random variable

$Y = x_1 + x_2 + \ldots + x_N$, where each x_i is a random variable with distribution X.
For each value of N, you do the "experiment" M times, i.e. you generate M samples of Y.

Answer: (1) Ignore the issue of a MATLAB simulation for a moment. Y is a random variable (not a single number!) which is the sum of N random variables. By the central limit theorem, Y approximates a normal distribution for sufficiently large N.

(2) OK, now consider the MATLAB experiment. We are using MATLAB to estimate the distribution of Y, by generating M samples of Y and looking at their distribution. As you increase M (the number of samples from the distribution) you get a better estimate of the distribution.

Yet one more problem

- Again, many of you got this right, but for those of you who didn't…

AGE of women	18-29	30-64	65+	TOTAL
Married	7,842	43,808	8,270	59,920
~Married	13,930	7,184	751	21,865
Widowed	36	2,523	8,385	10,944
Divorced	704	9,174	1,263	11,141
TOTAL	22,512	62,689	18,669	103,870

- Using the table above, what is the probability that a woman is married, given that she is 30-64?
  - (a) (43,808 / 59,920) (59,920 / 103,870) / p(married)
  - (b) 43,808 / 103,870
  - (c) (59,920 / 103,870) (43,808 / 59,920) / p(age 30-64)
  - (d) 43,808 / 59,920
  - (e) 62,689 + 59920 – 43,808

AGE of women	18-29	30-64	65+	TOTAL
Married	7,842	43,808	8,270	59,920
~Married	13,930	7,184	751	21,865
Widowed	36	2,523	8,385	10,944
Divorced	704	9,174	1,263	11,141
TOTAL	22,512	62,689	18,669	103,870

- To be tricky, I did this using Bayes' rule, instead of the straightforward way.
- Straightforward answer = 43808/62689
- Bayes' rule:
  p(married | 30-64)
  = p(30-64|married) p(married) / p(30-64)

  = (59920/103870) (43808/59920) / p(30-64)

Any other exam questions you'd like to go over?

Experimental Design, II

9.07
3/30/2004

Recall from last time we talked about the goal of experiments (and thus experimental design):

- Determine whether a relationship is likely to exist between
  - One or more independent variables (factors) and a dependent variable, or
  - Two or more dependent variables
- Minimize the possibility that the results you get might be due to a hidden confounding factor
- Maximize the power of your test for this relationship, while keeping the probability of a Type I error to a minimum
- Quantify your uncertainty in the results
- Wide range of applicability of the results

In particular, we talked about ways to minimize the possibility of confounding factors:

- Use a controlled experiment vs. an observational study, where feasible
- Use a contemporaneous comparison group, instead of an historical one
- Randomize assignment to groups
- Where appropriate, use blinding (e.g. double-blind experimental design)
- Where appropriate, use placebos

These are all techniques aimed at minimizing the difference between the experimental groups other than the experimental conditions you want to test.

Many of the things we did to reduce confounding are useful for other reasons as well

- Minimizing irrelevant differences between experimental groups reduces variance, and thus can make our statistical test more powerful
- Proper randomization allows us to accurately quantify our uncertainty in the results (because our statistical methods make strong assumptions about the kind of randomization involved in collecting the data)

Another useful technique: replication

- Allows us to quantify uncertainty
  - With just one data point, we have no idea of the variability, & thus can't say anything about our uncertainty
- Increasing n reduces the SE
  - More precision in estimating the parameters of the population
  - Increases the power of the statistical test
- Ideally, pick n according to your desired probability of Type I and Type II errors
  - Avoid tests with not enough power, and tests which will find tiny, unimportant differences significant

Increase the power by increasing the "signal-to-noise" ratio

- In addition to minimizing the variability of responses (see techniques for reducing confounding), we can also increase the power by increasing the raw effect size (e.g. $m_1 - m_2$)
  - Give a bigger drug dose, so you expect a bigger effect
  - Make the task more difficult, so there's more difference between good performance and poor (avoid threshold effects)

Do a better statistical test

- Figure out what statistical test you will do before you collect the data!!
  - Are you likely to satisfy the assumptions of the test?
- With a small change to your experimental protocol, could you run a more powerful test?

More general design issues

- Between-subjects designs
- Within-groups designs
- Factorial designs

Between subjects designs

Between subjects designs

- Two or more treatment conditions
- Subjects not exposed to all treatment conditions
- We get the comparison with other treatment conditions by comparing *between* subjects

Randomized post-test design

	Independent variable	Dependent variable
S1 S2 S3	Treatment 1	Measure A
S4 S5 S6	Treatment 2	Measure A

Randomized pre-test post-test design

	Pre-test	Independent variable	Post-test
S1 S2 S3	Measure A	Treatment 1	Measure A
S4 S5 S6	Measure A	Treatment 2	Measure A

Benefits and issues of pre-test, post-test design

- Benefits:
  - Evaluate the assumption that the two groups are the same before treatment
  - Look at how much difference the treatment makes
- Issues:
  - Takes time
  - Performance on pre-test may affect performance on post-test (practise, etc.) in a way which interferes with your study

Matched pairs

	Independent variable	Dependent variable
S1A S2B S3C	Treatment 1	Measure A
S4A S5B S6C	Treatment 2	Measure A

Within subjects designs

Within subjects designs

- Two or more treatment conditions
- Each subject is exposed to each treatment condition
- Comparison between conditions may be made *within* each subject

Within subjects design

	Independent variable	Dependent variable
S1 S2 S3	Treatment 1	Measure A
S1 S2 S3	Treatment 2	Measure A

Examples

- Each subject does a reaction time task both when they have had 6+ hours of sleep, and when they have had < 6 hours of sleep.
- Each subject tries 3 new chocolate bars, and rates (from 1 to 10) how much they like each of them.
- Repeated measures experiments we talked about before

Benefits and issues

- Benefits:
  - Fewer participants
  - Reduce irrelevant subject variability
    - E.G. due to some subjects just being better at reaction time tasks
- Issues:
  - Order effects!
    - Rating of chocolate may depend upon which tried first
    - Reaction time may be faster on second condition because of learning, or slower because of fatigue

Dealing with order effects: Counterbalancing

- Varying the order of presentation of the independent variable
- This way, if one subject does condition A, then condition B, some other subject does condition B before condition A, and any order effects will (hopefully) get averaged out.

Counterbalancing

- Full randomization:
  - If there are M conditions in the experiment, there are M! possible orderings of those conditions.
  - Need M! subjects for full randomization
- Latin square design
  - Ensure that each condition appears once in each position in the order, and that each subject sees each condition once
  - Only need M orderings, M subjects
  - There are a bunch of possible orderings (latin squares) that meet this criterion. Use randomization to pick one.

Latin square design

	Trial 1	Trial 2	Trial 3
S1	Chocolate 1	Chocolate 2	Chocolate 3
S2	Chocolate 2	Chocolate 3	Chocolate 1
S3	Chocolate 3	Chocolate 1	Chocolate2

So far we have talked about single-factor experiments

- Subjects were exposed to different *levels* of a single independent variable
  - Different chocolate bars
  - Doing a task with good or poor sleep
  - Undergoing a treatment, or not
- Subjects might also be exposed to more than one *factor*, i.e. more than one independent variable
  - This is useful for increasing the applicability of the results – multiple factors affect the dependent variable in real life, so it's useful to know how these factors interact.

Factorial designs

Factorial designs

- A design in which participants are exposed to more than one independent variable
- E.G. Look at the effects of aspirin and beta carotene on preventing heart attacks
  - Factors (i.e. independent variables):
    1. aspirin, 2. beta carotene
  - Levels of these factors that are tested:
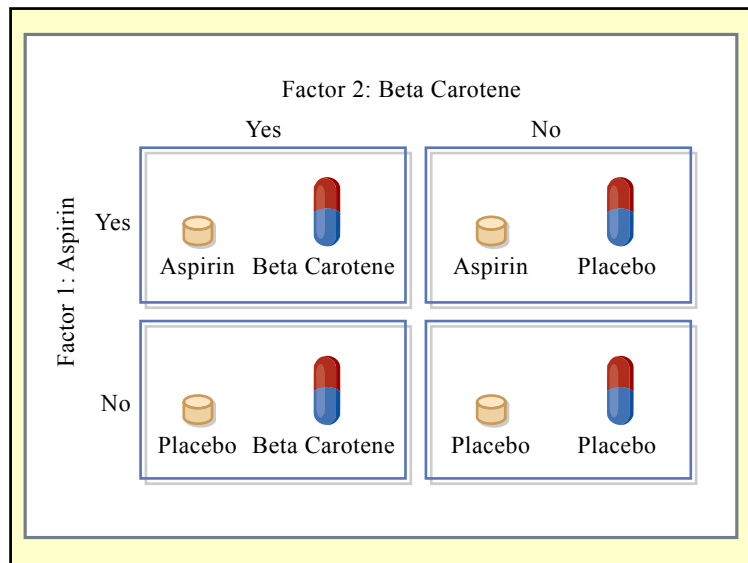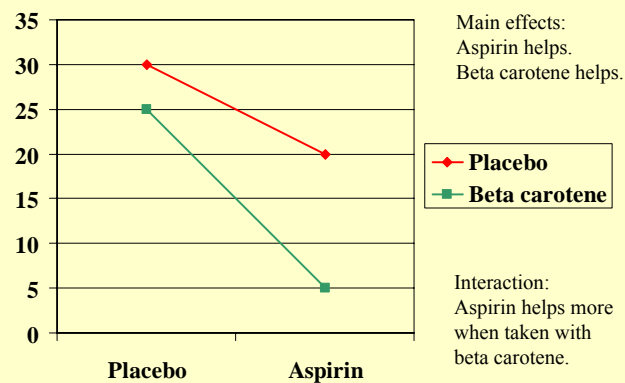   1. (aspirin, placebo), 2. (beta carotene, placebo)
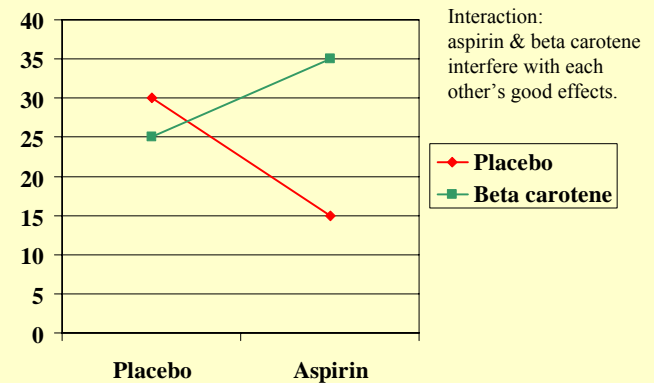
Figure by MIT OCW.

Outcomes of a factorial design

- Main effects
  - What effect does aspirin have on heart attacks, independent of the level of beta carotene?
  - What effect does beta carotene have on heart attacks, independent of the level of aspirin?
- Interaction(s)
  - The influence that two or more independent variables have on the dependent variable, beyond their main effects
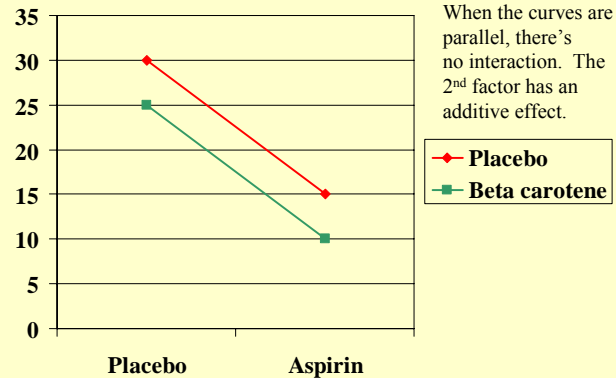 - How does beta carotene *interact* with aspirin, as far as preventing heart attacks?

Main effects:
Aspirin helps.
Beta carotene helps.

Interaction:
Aspirin helps more when taken with beta carotene.

Interaction:
aspirin & beta carotene interfere with each other's good effects.

No interaction

35		
30		
25		
20		
15		
10		
5		
0		
	Placebo	Aspirin

When the curves are parallel, there's no interaction. The 2nd factor has an additive effect.

◆ Placebo
■ Beta carotene

Use of factorial experiments

- Are there order effects?
- Are there subject effects?
- Does one factor modulate the effect of another?

Become familiar with the terminology

- A *factor* is a class of independent variable
  - E.G. gender, time-of-exam, dose of drug A
- Each factor is said to have some number of *levels* in a given experiment
  - Gender will have two levels, male & female
  - Time of exam might have 4 levels: 6pm, 7pm, 8pm, and 9pm
  - Dose of drug A might have 3 levels: no drug, normal drug dosage, and 3 x normal dose

---

Name of experimental design, for factorial designs

- Depends upon:
  - Number of factors (independent variables)
  - Number of levels of each factor studied
  - Whether the experimental design is between- or within- groups

Experimental design names

- If there is only one independent variable, it's a one-way design
  - Does coffee drinking affect exam performance?
- Two independent variables -> two-way design
  - Does coffee and/or time of exam affect performance?
- Three independent variables -> three-way
  - Does coffee and/or gender and/or time of exam affect performance?

Experimental design names

- If there are 3 levels of factor 1, and 4 levels of factor 2, then it's a *3 x 4* design
  - E.G. cups of coffee x time of exam
  - Cups of coffee = (1, 2, or 3), Time of exam = (6pm, 7pm, 8pm, 9pm)
- If there are 3 levels of factor 1, 2 levels of factor 2, and 4 levels of factor 3, then it's a 3 x 2 x 4 design
  - E.G. cups of coffee x gender x time of exam
  - Cups = (1, 2, 3), Time = (6, 7, 8, 9), Gender = (M, F)

Experimental design names

- Experiments can be either:
  - Between groups
  - Within groups (repeated measures)
  - Or *mixed*, if the comparison in some independent variables is *between*, and in others it is *within*.

Example

- Consider our earlier example of the effect on exam performance of coffee, gender, and time of exam.
- Each subject only has one gender – this is a between-groups factor
- Probably only gave each subject the exam once, at one time, after one particular level of # cups of coffee. So these are both between-groups factors as well.
- So, this is a 3-way (3x2x4) between-groups design.