

1. (15 points) You might find it useful to sketch the distributions.
 - a. For a Standard Normal Distribution, what is the area closer to the mean than 1.45? Area to the left of -1.45 is .0735 so the area in both tails farther than 1.45 is .1471; subtract this from one to find the area in the middle, 0.8529.
 - b. For a Standard Normal Distribution, what is the area to the right of 2? 0.0228
 - c. For a Normal Distribution with mean 5 and standard deviation 7.6, what is area to the right of 14.1? This is equivalent to the area to the right of the standardized value $\frac{14.1-5}{7.6} = 1.197$, which is .1156.
 - d. For a Normal Distribution with mean 1 and standard deviation 7.8, what is area in both tails farther from the mean than 11? Equivalent to area for standard normal farther from zero than $\frac{11-1}{7.8} = 1.2821$ so $2*0.0999 = .1998$.
 - e. For a Normal Distribution with mean -5 and standard deviation 1.6, what is area in both tails farther from the mean than -2.6? Equivalent to $\frac{-2.6-(-5)}{1.6} = 1.5$, so 0.1336.
 - f. For a Normal Distribution with mean -1 and standard deviation 9.8, what values leave probability 0.157 in both tails? A standardized value of ± 1.4152 leaves .0785 probability in each tail (so .157 in both); this standardized value is equivalent to (-14.87, 12.87).
2. (15 points) In a medical study, people were randomly assigned to use either antibacterial products or regular soap. In total 592 people used antibacterial soap; 586 used regular soap. It was found that 33.1% of people using antibacterial products got a cold; 32.3% of people using regular soap got colds.
 - a. Test the null hypothesis that there is no difference in the rates of sickness for people using regular or antibacterial soap. (What is the p-value?) The standard error for the difference uses the variance of the first, which is $0.331(1 - .331) = .2214$ and the variance of the second, $.323(1 - .323) = .2187$, so the standard error of the difference is $\sqrt{\frac{.2214}{592} - \frac{.2187}{586}} = 0.0273$. The difference is 0.2927 of a standard deviation, so that there is a p-value of 0.7698 – there is almost a 77% chance that, if there were actually no difference we might still measure such a tiny difference.
 - b. Create a 95% confidence interval for the difference in sickness rates. What is the 90% confidence interval? The 99% interval? A 95% confidence interval for the difference is $0.008 \pm 1.96*0.027 = .008 \pm .054 = (-.046, .062)$. A 90% confidence interval uses 1.64 instead of 1.96 so $(-.037, .053)$; a 99% confidence interval uses 2.58 so $(-.063, .079)$.
 - c. Every other study has found similar results. Why do you think people would pay more for antibacterial soaps? Answers will vary. **Many students missed the basic finding, that people washing with antibacterial soap got sick MORE often!**
3. (15 points) A study of workers and managers asked both how much management listened to workers' suggestions (on a scale of 1-7 where "1" indicates that they paid great attention). Managers averaged a 2.50 (standard deviation of 0.55); workers answered an average 2.08 (standard deviation of 0.76) – managers ignore their workers even more often than the employees realize. There were 137 workers and 14 managers answering.

- a. Test the null hypothesis that there was no difference between workers and managers: how likely is it that there is actually no difference in average response? (What is the p-value?) The difference is $2.50 - 2.08 = 0.42$. The standard error of the difference is $\sqrt{\frac{.76^2}{137} + \frac{.55^2}{14}} = 0.16$. So the difference is $.42/.16 = 2.61$ standard errors from zero, so there is a small chance that this is just random; the chance is under 1%.
 - b. Create a 95% confidence interval for the difference between workers and managers. What is the 90% confidence interval? The 99% interval? The 95% confidence interval for the difference is $0.42 \pm 1.96 * .16 = .42 \pm .31 = (.11, .73)$. The 90% CI uses 1.64 so $(.16, .68)$; the 99% CI uses 2.56 so $(.01, .83)$.
4. (15 points) A recent survey by Intel showed that 53% of parents (561 were surveyed) were uncomfortable talking with their children about math & science. Previous surveys found that 57% of parents talked with their kids about sex & drugs.
- a. Test the null hypothesis that parents are as comfortable talking about math & science as sex & drugs; that the true value of parents uncomfortable with math and science is not different from 57%. What is the p-value? The standard error of the survey is $\sqrt{\frac{.53(1-.53)}{561}} = .021$, so the difference of 0.04 is 1.9 standard errors from zero. The p-value is .058.
 - b. Create a 95% confidence interval for the true fraction of parents who are uncomfortable with math & science. What is the 90% confidence interval? The 99% interval? The 95% CI is $.04 \pm 1.96 * .021 = (-.001, .081)$; 90% is $(.005, .075)$; 99% is $(-.014, .094)$.
5. (15 points) The New York Times reported on educational companies that over-sell their products and gave the example of "Cognitive Tutor" (CT) that helps math students. The CT students improved by 17.41 (standard error of 5.82); the regular students improved by 15.28 (standard error of 5.33). There were 153 students in the new program and 102 regular students.
- a. Test the null hypothesis that there is no difference between regular students and those in the CT group. What is the p-value for this difference? The difference is $(17.41 - 15.28) = 2.13$. The standard error of the difference is $\sqrt{\frac{5.82^2}{153} + \frac{5.33^2}{102}} = .707$, so the difference is just over 3 standard errors from zero. The p-value is .003.
 - b. Create a 95% confidence interval for the difference between regular and CT students. What is the 90% confidence interval? The 99% interval? The CIs are 90: $(.97, 3.29)$, 95: $(.74, 3.52)$, 99: $(.32, 3.94)$.
6. (20 points) Use the ATUS data (available from Blackboard) on the time that people spend in different activities.
- a. Among households with kids, what is the average time spent on activities related to kids? There are 47, 298 people in households with kids, spending an average of 66.74 minutes, with standard deviation of 99.98 minutes.
 - b. Among households with kids, how much time to men and women spend on activities related to kids? Form a hypothesis test for whether there is a statistically significant difference between the time that men and women spend with kids. What is the p-value for the hypothesis of no difference? What is a 95% confidence interval for the difference in time? There are 19,945 men and 27,353 women. The males average 41.05 minutes with kids (std dev 77.45); the women average 85.46 minutes (std dev 109.90). The difference is $85.46 - 41.05 =$

44.41; the standard error of this difference is 0.86 so the difference is huge and p-value about zero! A 95% Confidence interval for the difference is $44.41 \pm 0.86 * 1.96 = 44.41 \pm 1.69 = (42.72, 46.10)$.

- c. Why do you think that we would find these results? Explain (perhaps with some further empirical results from the same data set).

Answers will vary.

7. (25 points) Use the PUMS data (available from Blackboard) on the residents of NYC. Consider the time (in minutes) spent by people to travel to work; this variable has name JWMNP.

- a. How many men and women answered this question? What variables do you think would be relevant, in trying to explain the variation in commuting times?

There are 66,256 men and 66,087 women answering; their answers are very close – 39.34 for men and 39.44 for women. Answers will vary for which are the relevant variables.

- b. Form a linear regression with the dependent variable, "JWMNP Travel Time to Work," and relevant independent variables.

Answers will vary. Example is below.

- c. Which independent variables have coefficients that are statistically significantly different from zero?

Answers will vary. Example is below.

Here is the SPSS output for a simple linear regression, explaining time commuting. All of the variables are statistically significantly different from zero except for Native American, Hispanic, kids under 6, all of the educational measures, and all of the poverty measures.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.437 ^a	.191	.190	23.223

a. Predictors: (Constant), commute_subway, if family income less than 200 percent of poverty line, nativeamerican, educ_collassoc, asianamerican, female, boro_bk, educ_somcoll, kids_under6, boro_si, raceother, Age, educ_hs, commute_bus, boro_bx, foreign_born, educ_adv, africanamerican, if family income less than 100 percent of poverty line, kids_under17, Hispanic, boro_qns, commute_car, educ_coll, if family income less than 150 percent of poverty line

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	9135976.012	25	365439.040	677.593	.000 ^a
Residual	3.878E7	71905	539.319		
Total	4.792E7	71930			

a. Predictors: (Constant), commute_subway, if family income less than 200 percent of poverty line, nativeamerican, educ_collassoc, asianamerican, female, boro_bk, educ_somcoll, kids_under6, boro_si, raceother, Age, educ_hs, commute_bus, boro_bx, foreign_born, educ_adv, africanamerican, if family income less than 100 percent of poverty line, kids_under17, Hispanic, boro_qns, commute_car, educ_coll, if family income less than 150 percent of poverty line

ANOVA^b

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	9135976.012	25	365439.040	677.593	.000 ^a
Residual	3.878E7	71905	539.319		
Total	4.792E7	71930			

a. Predictors: (Constant), commute_subway, if family income less than 200 percent of poverty line, nativeamerican, educ_collassoc, asianamerican, female, boro_bk, educ_somcoll, kids_under6, boro_si, raceother, Age, educ_hs, commute_bus, boro_bx, foreign_born, educ_adv, africanamerican, if family income less than 100 percent of poverty line, kids_under17, Hispanic, boro_qns, commute_car, educ_coll, if family income less than 150 percent of poverty line

b. Dependent Variable: Travel time to work

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.611	.557		20.844	.000
	Age	.116	.007	.057	15.637	.000
	female	-.795	.180	-.015	-4.422	.000
	africanamerican	5.539	.244	.089	22.667	.000
	nativeamerican	2.137	1.527	.005	1.399	.162
	asianamerican	2.800	.308	.035	9.098	.000
	raceother	2.226	.344	.028	6.464	.000
	Hispanic	.037	.290	.001	.127	.899
	kids_under6	.419	.298	.006	1.404	.160
	kids_under17	.729	.232	.014	3.146	.002
	foreign_born	1.810	.200	.035	9.062	.000
	educ_hs	-.564	.331	-.009	-1.704	.088
	educ_somcoll	.416	.355	.006	1.174	.240
	educ_collassoc	.681	.430	.007	1.581	.114
	educ_coll	.585	.346	.010	1.692	.091
	educ_adv	-.519	.359	-.008	-1.446	.148
	boro_bx	9.924	.328	.131	30.244	.000
	boro_si	20.007	.431	.181	46.468	.000
	boro_bk	9.493	.258	.167	36.844	.000
	boro_qns	12.302	.270	.212	45.582	.000

if family income less than 100 percent of poverty line	-.014	.475	.000	-.030	.976
if family income less than 150 percent of poverty line	-.613	.481	-.008	-1.273	.203
if family income less than 200 percent of poverty line	-.406	.360	-.006	-1.126	.260
commute_car	1.023	.286	.018	3.581	.000
commute_bus	19.409	.337	.248	57.555	.000
commute_subway	20.466	.261	.390	78.546	.000

a. Dependent Variable: Travel time to work