One of these will make you laugh later.

Rough Review Thus Far

- **Trying to be scientific**
  - Esp. falsifiability (testable ideas – hypotheses!)
  - Careful measurements & procedures
- **Want to describe population**
  - Typically can't study it directly
  - Generalize from a sample
  - Make *inferences* to population
- **Statistical inferences are systematic**
  - Level of Measurement matters (focusing on interval, for now)
  - Key element: distance/difference between values

Z'S & P'S

SOC424 – Statistics w/ Dr. Ellis Godard

Where We're Going…

- **DONE: Descriptive Statistics (Summarize Data)**
  - Central Tendancy: mode, median, mean
  - Dispersion: ratios, ranges, variance, & std.dev.
- **COMING: Inferential Statistics (Generalize Beyond)**
  - Significance Testing: to examine differences between groups or among variables, in terms of the likelihood that *observed* differences are due to chance (bivariate)
  - Estimation: to generalize from sample statistics to population parameters, on the basis of the sample size and the amount of observed variance (univariate)
    - Requires a sample **randomly selected** from a population in which the variable is **distributed normally**

Ideas for Today (not quite an "outline")

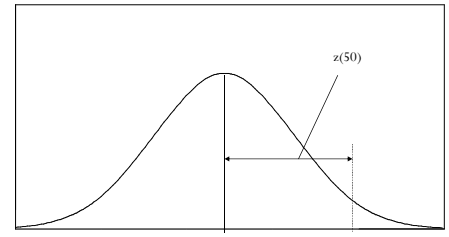
- **Overview Revisited** *(same ole)*
- **Z-scoring** *(last lab)*
  - *How many standard deviations difference?*
- **Associated Probabilities** *(% in "tail")*
  - *What's the probability of exceeding that difference (i.e. that z)?*
  - Finding them in table A
  - Lots of examples, about different parts of a curve
  - Homework Hints
- **Exercise**

Best. Lab. Ever.

Clarification of Book Formulas

- **Inside front cover, formula for z**
  - Denominator is std. dev. *only* to calculate z for a particular case or value, as you did (or will do) in lab 7
  - For a test statistic, or to make inferences, you need the standard deviation of the *sampling* distribution (= standard *error* - next lecture; you'll love it ☺)

EXAMPLE, using Math SATs



Normal Distribution → μ = 400 400 + $z(50)$

Standard Normal Distribution → 0 z

Normality, Technically

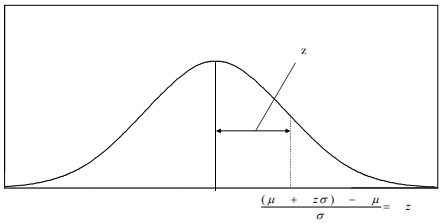
- **The normal distribution is:**
  - really a family of distributions that depend on ? and ?
  - has certain known probabilities associated with it

$$\mu = \frac{1}{\sigma\sqrt{2\pi}}(e)^{-(x-\mu)^2/2\sigma^2}$$

Where: μ = the height of the curve above any given value of the variable, x.
π = the ratio of the circumference of a circle to its diameter (3.1415…)
e = the base of the system of natural logarithms and
μ and σ = mean and standard deviation of the variable x

Standard Normal Distribution

If a variable has a normal distribution, and its values are converted to z-scores, then the corresponding distribution of z-scores is the standard normal distribution (see alt. wording in Agresti and Finlay, p. 94)



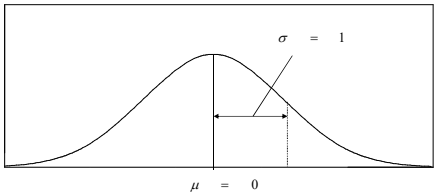
Normality, Practically

- **Characteristics**
  - mean = median (i.e. symmetric)
  - mean = mode (i.e. bell-shaped)
- **Variations**
  - Skewness measures symmetry (0 if symmetric, negative if skewed left, positive if skewed right)
  - Kurtosis measures how flat the curve is (0 if perfectly "normal", positive if pointy, and negative if flattened)
- **Implication: Areas under the curve are "constant"**
  - 68.26 % falls w/i 1 standard deviation of the mean
  - 95.44 % falls w/i 2 standard deviations of the mean
  - 99.72% falls w/i 3 standard deviations of the mean

Standard Normal Distribution

The *standard normal distribution* is the normal distribution with parameter values:

$$\mu = 0 \quad \textit{and} \quad \sigma = 1$$



Implications Illustrated

- **Most shapes not normal, but can still *estimate***
- **If ages in a class (n=52) are $\bar{Y}$=21.7, s =1.97**
  - $\bar{Y}$ +/- s -> 19.73 to 23.67
    - An estimated 68.26% (35.49) are in that range (17 older/younger)
  - $\bar{Y}$ +/- 2s -> 17.76 to 25.64
    - An estimated 95.44% (44.62) are in that range (2 outside)
  - $\bar{Y}$ +/- 3s -> 15.79 to 27.61
    - An estimated 99.72% (51.85, i.e. all) are in that range
    - Actually, one was older (32) but sample small & not normal

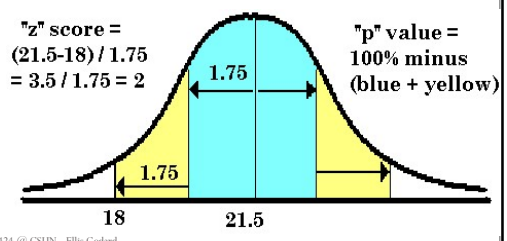
Descriptive Statistics Renewed

- **Basic Measures**
  - Central Tendency (mode, median, mean)
  - Dispersion (ratios, ranges, variance, deviations)
- **Description with Z-Scores**
  - For each # of std. devs from the mean (i.e. for each "z"), an area under the curve represents an estimated percentage of the total distribution
  - For example, if the mean is 21.5 & the std. dev. is 1.75, someone aged 18 is 2 standard deviations from the mean (their z-score is 2) and an estimated 95.44% of the values are in that interval

15 SOC424 @ CSUN - Ellis Godard

Descriptive Statistics Renewed

- Someone aged 18 is 2 standard deviations from the mean – and an estimated 95.44% of the values under the curve are in the same interval



16 SOC424 @ CSUN - Ellis Godard

EXAMPLE: What Majors Pay

Susan the English major is making $48K and Donald the engineer is making $75K.

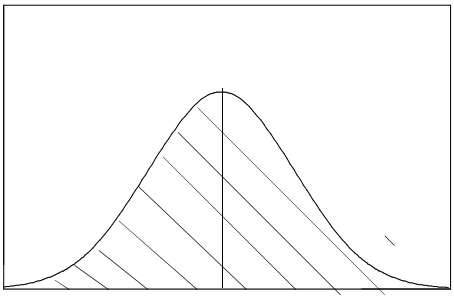
Both are making more than their majors typically do, but…

Which one has *most unusual* salary, given their major?

Major	Yi (single score)	Mu (avg. starting salary nationally)	Std. Dev.	Z Score	Tail area
English	$48,000	$42,000	$9,000	0.666	*But, professor… We only know about areas associated with a z of 1, 2, or 3. What do I do with "0.666"??*
Electrical Engineering	$75,000	$74,000	$1,000	1.0	

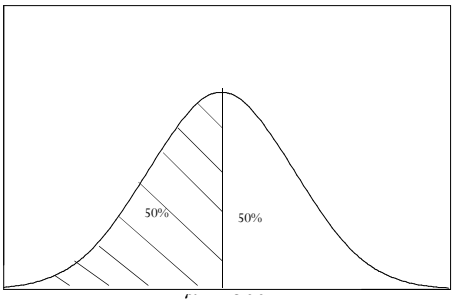
17 SOC424 @ CSUN - Ellis Godard

Entire Curve = 100%



18 SOC424 @ CSUN - Ellis Godard

Median/Mean/Mode splits it 50/50



19 SOC424 @ CSUN - Ellis Godard

FINITE EXAMPLES

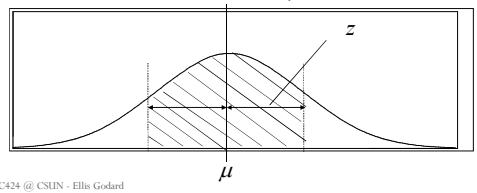
Within 1 standard deviation (z=1), 66.26%

33.13% on each side

Within 2 standard deviations (z=2), 95.44%

1.96 would be a bit less (95%)

Within 3 standard deviations, 99.72%



20 SOC424 @ CSUN - Ellis Godard

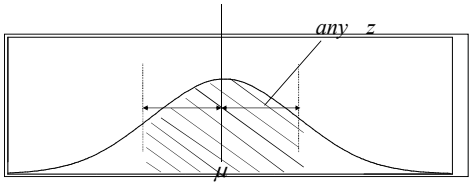
Admin Normal Stats Curves Z&P Tables Examples Hints Cases

INFINITE POSSIBILITIES

For *any* value of z (1, 2, 3, .666, 5.78…) there is an associated percentage of cases within that range

Table A will give you the % ("p") for any z

But that "p" is the tail, not the middle

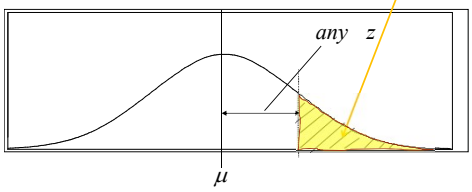


21 SOC424 @ CSUN - Ellis Godard

Admin Normal Stats Curves Z&P Tables Examples Hints Cases

TAIL PROBABILITIES

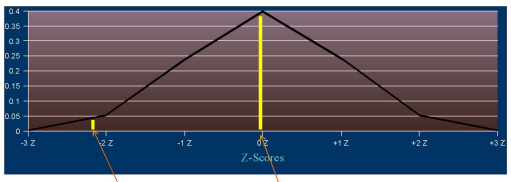
For any value of z, there is a percentage *beyond* that z – an associated probability (a "p value") which describes the area under the curve which has values *further* from the mean than z…



22 SOC424 @ CSUN - Ellis Godard

Admin Normal Stats Curves Z&P Tables Examples Hints Cases

Probabilities Associated with the Standard Normal Distribution



The probability of getting a value more than this far from the mean is low – it's already pretty far, and few are farther

The chance of getting a value farther away, than not very far away, is pretty big (close to 50% on each side)

23 SOC424 @ CSUN - Ellis Godard

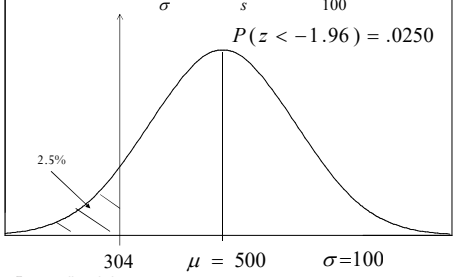
Admin Normal Stats Curves Z&P Tables Examples Hints Cases

There's a p for the part beyond any z

(distance, in std deviations)

$$z = \frac{Y_i - \mu}{\sigma} \approx \frac{Y_i - \bar{Y}}{s} = \frac{304 - 500}{100} = 1.96$$

$$P(z < -1.96) = .0250$$

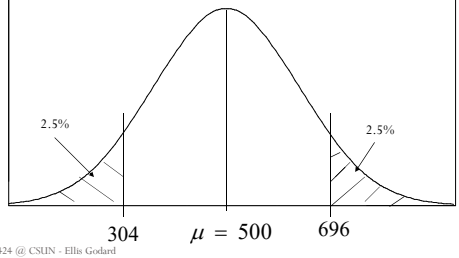


24 SOC424 @ CSUN - Ellis Godard

Admin Normal Stats Curves Z&P Tables Examples Hints Cases

To talk about both tails, double the p

$$P(|z| > 1.96) = P(z < -1.96) + P(z > 1.96)$$
$$= 2.5\% + 2.5\% = 5\%$$

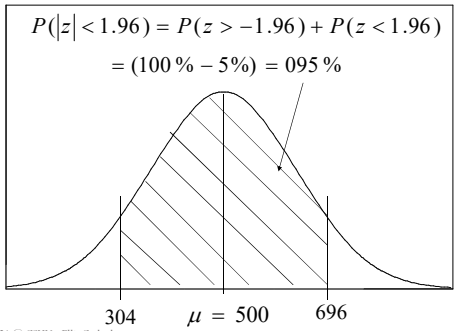


25 SOC424 @ CSUN - Ellis Godard

Admin Normal Stats Curves Z&P Tables Examples Hints Cases

For the middle, subtract tails from 100% (1)

$$P(|z| < 1.96) = P(z > -1.96) + P(z < 1.96)$$
$$= (100\% - 5\%) = 095\%$$



26 SOC424 @ CSUN - Ellis Godard

For a slice, subtract from 50% (.5000)

$P(-1.96 < z < 0) = .5000 - .0250 = .4750$

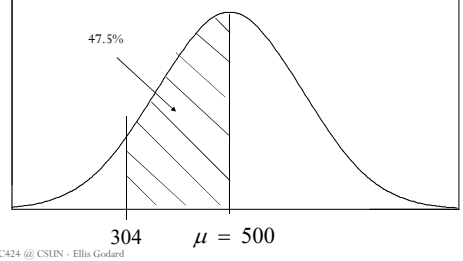


Table A

- Don't go find your own
- Use the one in the book
- It's also on Canvas
- And right here….
- There are other ways to lay out zs & ps.
- If you use one of the others, your answers will probably be wrong…
- …and you won't be ready for Table B & more
- PLEASE don't try to teach yourself in your own way ☹ - follow the steps, please! ☺
- USE THIS TABLE !!

Tables

Second Decimal Place of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000233									
4.0	.0000317									
4.5	.00000340									
5.0	.000000287									

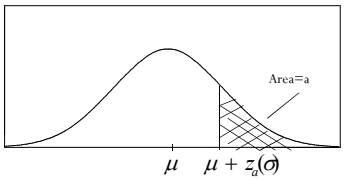
Source: R. E. Walpole, *Introduction to Statistics* (New York: Macmillan, 1968).

Next step: Implications

- **Consider *tails beyond* certain distances, & *areas between* different values**
  - Just as a z of 1 is associated w/ about 68% and a z of 2 is associated w/ about 95%, any z score has some associated probability of being within that distance from the mean
  - Often, we are concerned *just* with the "tails"
    - z=1, p=68.26% > ~32% in 2 tails, ~16% (0.16) in each
    - z=2, p=95.44% > ~5% in 2 tails, so ~2.5% (0.025) in each

A BIT MORE NOTATION…

The symbol z_a represents the z-score from Table A corresponding to the right-hand tail area of a.



Sidebar on Probability Tables

- **All tests rest on probability distributions**
  - E.g., 2 z's is associated w/ 95.44%, because that's the area under the curve of the z distribution within two z's (standard deviations) from the mean.
  - Other tests use other distributions (e.g. "t", F, etc.)
- **SPSS provides "exact" probability values**
  - Typically to four or five decimal places
  - P almost always represents an estimate of the *risk of being wrong in rejecting the null hypothesis*
- **For hand calculations, we use tables that summarize *some* values**
  - E.g., 2 z's being associated with 95.44% is part of the table that summarizes the distribution of z's
  - There's a different table for each test.

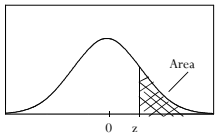
What Z-Score Table (A) *looks like*

Second Decimal Place of z

z	.00	.01	...	.05	...	.09
0.0	.5000	.4960	...	.4801	...	.4641
0.1	.4602	.4562	...	.4404	...	.4227
...	...	...	...	...	...	...
1.0	.1587	.1562	...	.1469	...	.1379
...						
2.0	.0228	.0222	...	.0202	...	.0183
....						
3.0	.00135					
...						

p value for z=1.09

Before & after decimal (e.g. 1.9 from 1.96)



Any Z Score <-> Any P Value

- **All tables are organized basically same**
  - Pick the appropriate column and row
  - Their intersection provides the needed value, usually a "critical value" for the test (e.g. t or F)
  - BUT the z-table provides probability values
  - P value for a z of 0.12 = 0.4522 (45.22%)

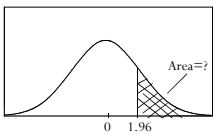
	Second decimal place		
z	.00	.01	.02
0.0	.5000	.4960	.4920
0.1	.4602	.4562	.4522
0.2	.4207	.4168	.4129

Zs that aren't on the table

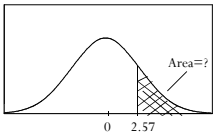
- ***Don't* average two (e.g. 1.64 and 1.65)**
- ***Don't* just say it's "between 1.64 & 1.65"**
- **Use the *smaller z* and *bigger p***
  - This is usually the most conservative
  - Don't want to *exaggerate* an observed difference

- **Also…**
  Don't be scared of a very *big* z;
  it just means there's a much smaller p

EXAMPLES: from z to p

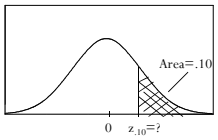


What is the probability that a value falls more than 1.96 standard deviations above the mean? That is, if z=1.96, what is "area"? Or, alternately, what is P(z>1.96)?)

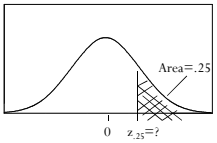


What is the probability that a value is 2.57 standard deviations above the mean? i.e. if z=2.57, what is "area" -- that is, P(z > 2.57)?

EXAMPLES: from p to z



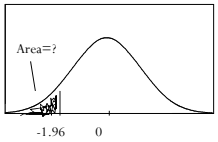
If the "area" above z is 0.10, what is z?



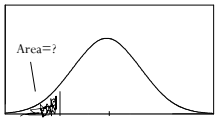
If the "area" above z is 0.25, what is z?

EXAMPLES: left-hand tails

For Negative Values of z, areas are found by symmetry…



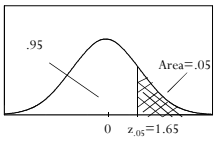
If z= -1.96, what is P(z < -1.96)?



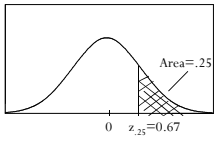
If z= -2.57, what is P(z < -2.57)?

EXAMPLES: non-tail areas

Since the total area under the curve is equal to 1.00…



P(z > 1.65) + P(z<1.65) = 1.00
P(z < 1.65) =1.00 - P(z > 1.65)
= 1.00 - 0.05
= 0.95



What is P(z < 0.67)?

EXAMPLE: What Majors Pay

Susan the English major is making $48K and Donald the engineer is making $75K.

Both are making more than their majors typically do, but…

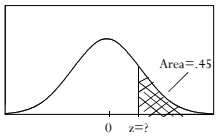
Which one has *most unusual* salary, given their major?

Major	Yi (single score)	Mu (avg. starting salary nationally)	Std. Dev.	Z Score	Tail area
English	$48,000	$42,000	$9,000	0.666	0.2514
Electrical Engineering	$75,000	$74,000	$1,000	1.0	0.1587

HOMEWORK HINTS

Find the z value for which the probability that a normal variable exceeds equals …

$$\mu + z\sigma$$

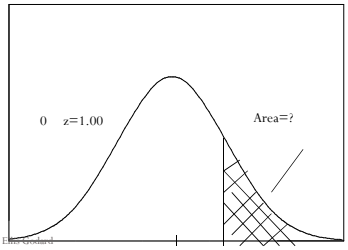


0.450.
This is the same as asking what is the area associated with $z_{0.45}$.

Use Table A to find that $z_{0.45}$=0.12
(area actually falls between 0.12 and 0.13, but for now, use the smaller of the two – we'll talk about why in 3 more lectures…)

HOMEWORK HINTS

Always draw a picture first!



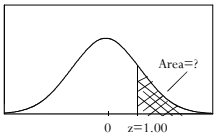
Examples for labs (as needed?)

- **If the mean jail sentence is 7 years, with a standard deviation of 2.5 years, and assuming that jail sentences are normally distributed*…**
  - What % of sentences are *shorter* than 7 years?
  - What % of sentences are *longer* than 7 years?
  - What % of sentences are between 4.5 and 9.5 years?
  - What % of sentences are between 2 and 12 years?
  - What % of sentences are shorter than 2 years?
  - What % of sentences are longer than 14.5 years?

*They aren't.

HOMEWORK HINTS

For a probability distribution, find the probability that a measurement is …



…more than one standard deviation above the mean. (That is, what is the area associated with z=1?)

$$P(z > 1.00) = 0.1587$$

…more than one standard deviation below the mean?

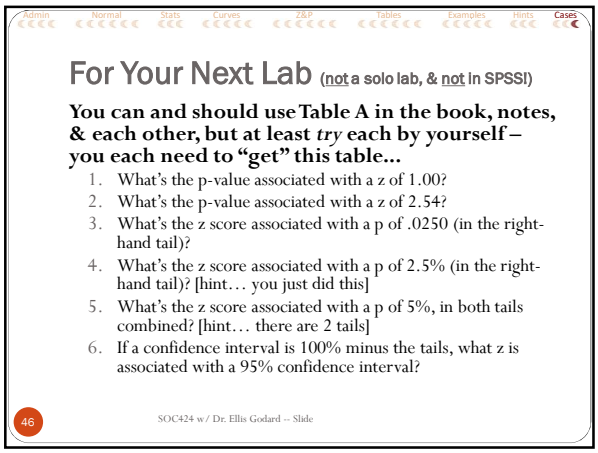
Hint: Always draw a picture first!

Examples for labs as needed? II

- **If the average grade is 84, with a standard deviation of 6, assuming that the grades are normally distributed*…**
  - What % of grades are C- or lower (below 72)?
  - What % of grades are an A- (use 90 to 93**)
  - What's the standard deviation if 15.87% earn an A (90 or higher)?

*They aren't.

** Due to rounding, it would really be 89.5 to 92.4.


Admin Normal Stats Curves Z&P Tables Examples Hints Cases
For Your Next Lab (not a solo lab, & not in SPSS!)
You can and should use Table A in the book, notes, & each other, but at least try each by yourself – you each need to "get" this table…
1. What's the p-value associated with a z of 1.00?
2. What's the p-value associated with a z of 2.54?
3. What's the z score associated with a p of .0250 (in the right-hand tail)?
4. What's the z score associated with a p of 2.5% (in the right-hand tail)? [hint… you just did this]
5. What's the z score associated with a p of 5%, in both tails combined? [hint… there are 2 tails]
6. If a confidence interval is 100% minus the tails, what z is associated with a 95% confidence interval?
46
SOC424 w/ Dr. Ellis Godard -- Slide