Social Science Statistics Module I

Gwilym Pryce

Lecture 2
Calculating z-Scores

Slides available from *Statistics & SPSS* page of www.gpryce.com

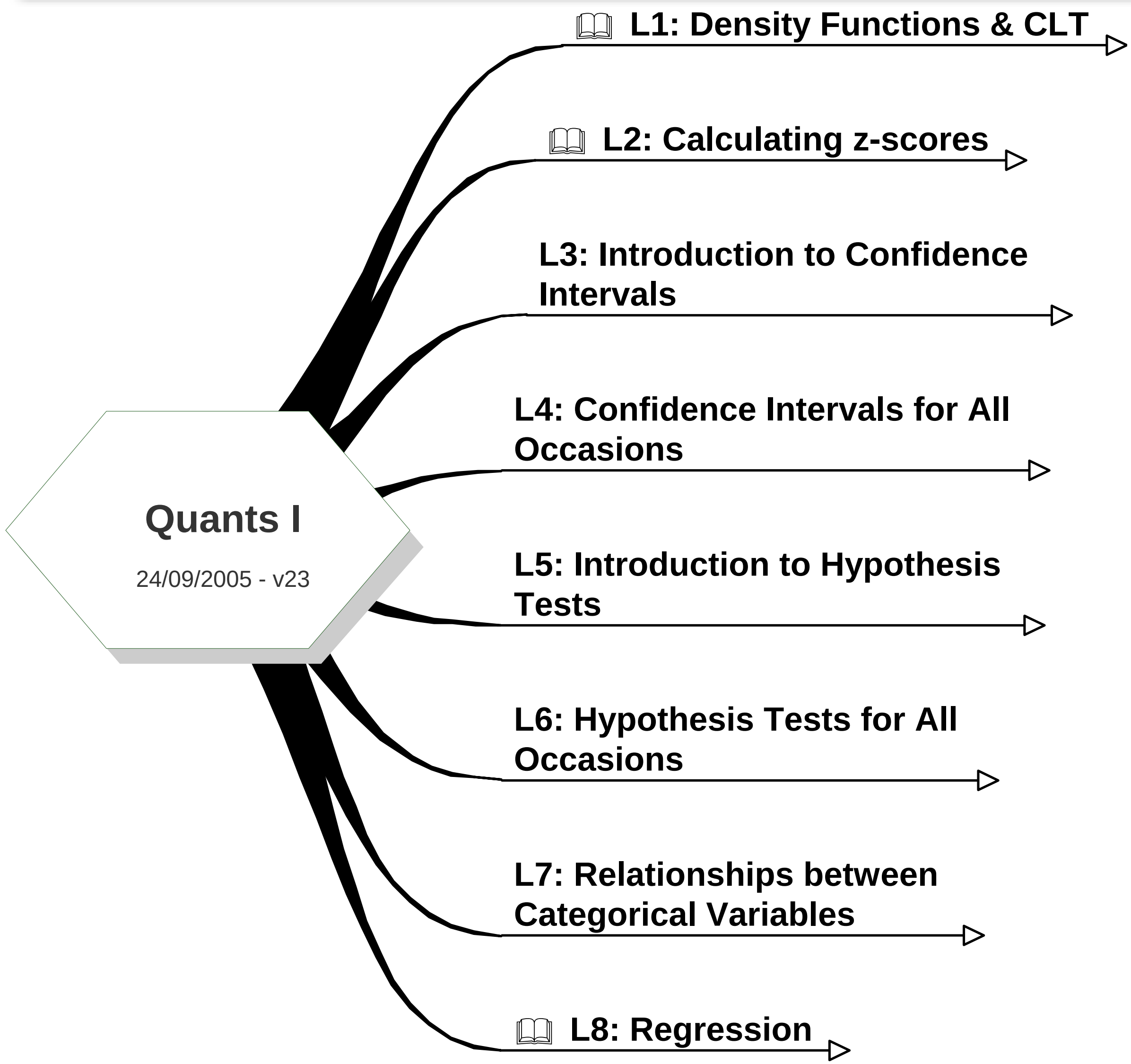
University of Glasgow

Expectations & Support:

1. **Independent learning:**
   - this is a PG course and a degree of independent learning is assumed.
   - do the reading, attend labs, review the lectures, make use of the computer labs/online help in your own time.
2. **Lab Overview & Feedback:**
   - Please feedback to the tutors & Class Reps how you think that is going, how it could be improved.
   - Tutors and Class Reps will then report back to me how things are going each week.
3. **Talk to tutors if you are struggling:**
   - Let the tutors know if you are struggling (assuming you have done the reading, attended labs etc.)
   - Tutors cannot guarantee extra support, but it might be possible to arrange extra tutorials etc.

4. **Departmental Support:**
   - Struggling students should enquire whether their own dept has support to offer.
   - All the grad school courses are only intended to constitute a generic training component;
   - Individual depts & supervisors should supplement with additional training & support as necessary.
5. **Support from Maths Advisor Shazia Ahmed, University's Maths Adviser:**
   - If you have gone through steps 1 to 4, Shazia has agreed to run one-on-one sessions with students that are struggling with particular mathematical or statistical concepts (though she has made it clear that she cannot advise on SPSS problems, nor will she do the assignment for you).
   - Students who have particular problems in this regard can contact her directly: **Shazia Ahmed**, Maths Adviser, Student Learning Service, McMillan Reading Room, **Tel: 330 5631** Fax: 330 8063
6. **Tutor of Last Resort:**
   - Students who have gone through steps 1 to 5 above, and who still feel they are not receiving enough support, can email me directly
   - I will try to arrange individual or small group meetings for people who have tried all other avenues.
     - You will need to demonstrate that you have gone through steps 1 to 5.

Introduction & Overview:



University of Glasgow

Introduction:

- We have looked at the characteristics of density functions, & one that particularly interests us, the normal distribution
- Though we have already looked briefly at the standard normal curve, today we shall look in depth at the practicalities of calculating z scores and using them to work out probabilities.

University of Glasgow

Aims & Objectives

- Aim
    - To consider the practicalities of the standard normal curve
- Objectives
    - by the end of this lecture students should be able to:
        - Work out probabilities associated with z scores
        - Work out z_i from given probabilities
        - Derive z_i and associated probability from given values of a normally distributed variable x
        - Apply z_i scores to sampling distributions

University of Glasgow

Plan

- Review of last week's lecture
- **1. Find probabilities from z_i**
  - Tables
  - SPSS macro command
- 2. Find z_i from a given probability
  - *z* that bounds upper or lower tail area
  - ± *z* that bounds central area
- 3. Find z_i & probabilitiy from $x_i \sim N(\mu,\sigma)$
- 4. Applying *z* scores to sampling distributions

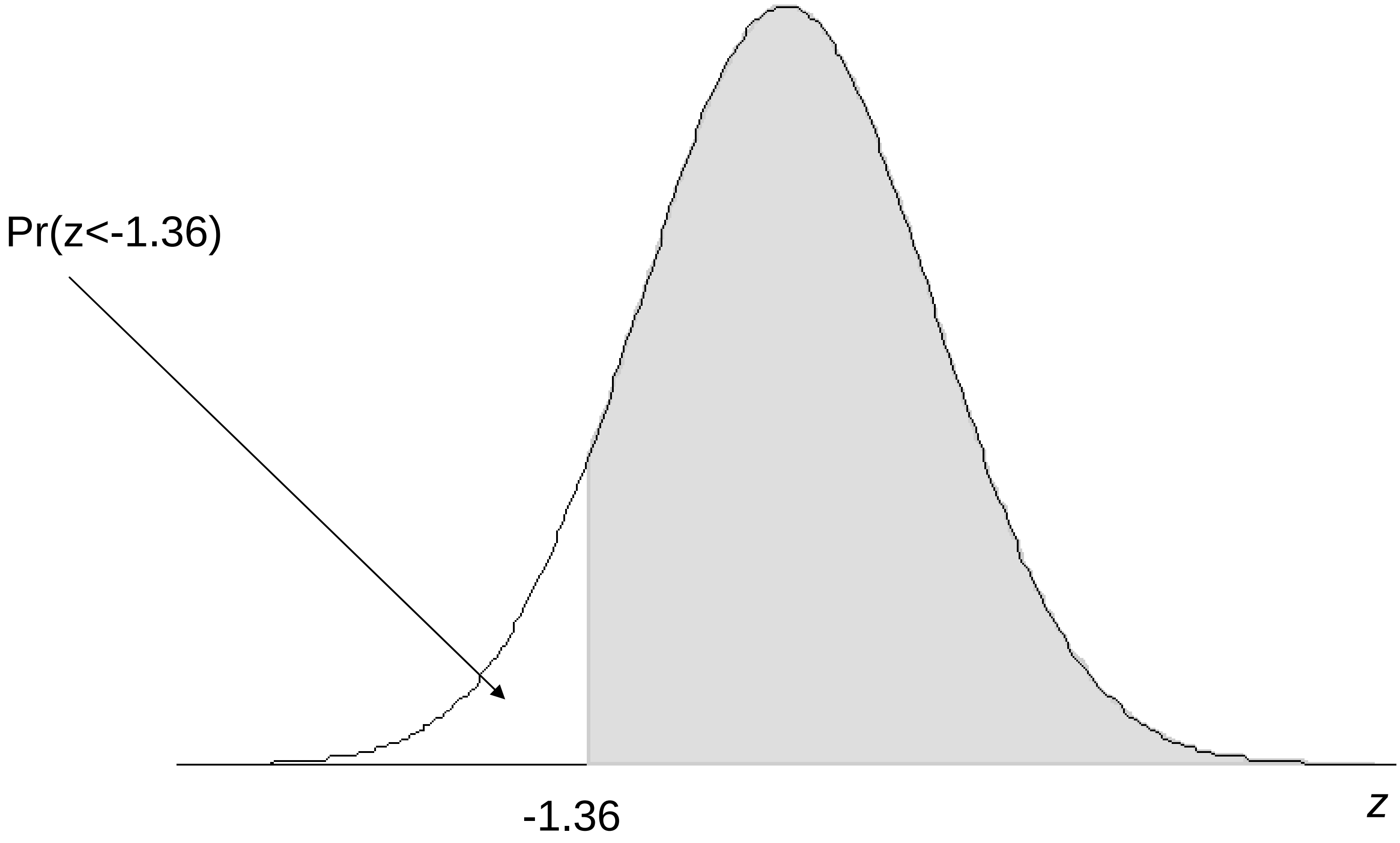
1. Find probabilities from z_i

1.1 Using Published tables

- Most stats books have z-score tables which allow you to find $Prob(z < z_i)$
  - Integer and first decimal listed in the row heading
  - Second decimal listed in the column heading
  - Probability value is located at the intersection of this row and column
- Or sometimes published tables list:
  - $Prob(0 < z < z_i)$
  - $Prob(z < z_i < 0)$
- Symmetry of the normal curve means that its easy to find any probability from any of these.

e.g. 1: Prob(z < -1.36)

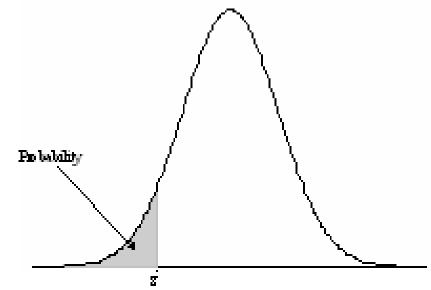
- First Draw diagram
- Then look-up the first decimal place in the row heading
- Then find the second decimal place in the column heading


Pr(z<-1.36)
-1.36
z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.40	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.30	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.20	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.10	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.00	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.90	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.80	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.70	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.60	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.50	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.40	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.30	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.20	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.10	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.00	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.90	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.80	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.70	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.60	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.50	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.40	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.30	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.20	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.10	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.00	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-.90	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-.80	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-.70	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-.60	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-.50	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-.40	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-.30	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-.20	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-.10	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table A: Standard Normal Probabilities

Each entry in the body of the table is the area under the standard normal curve to the left of *z*.



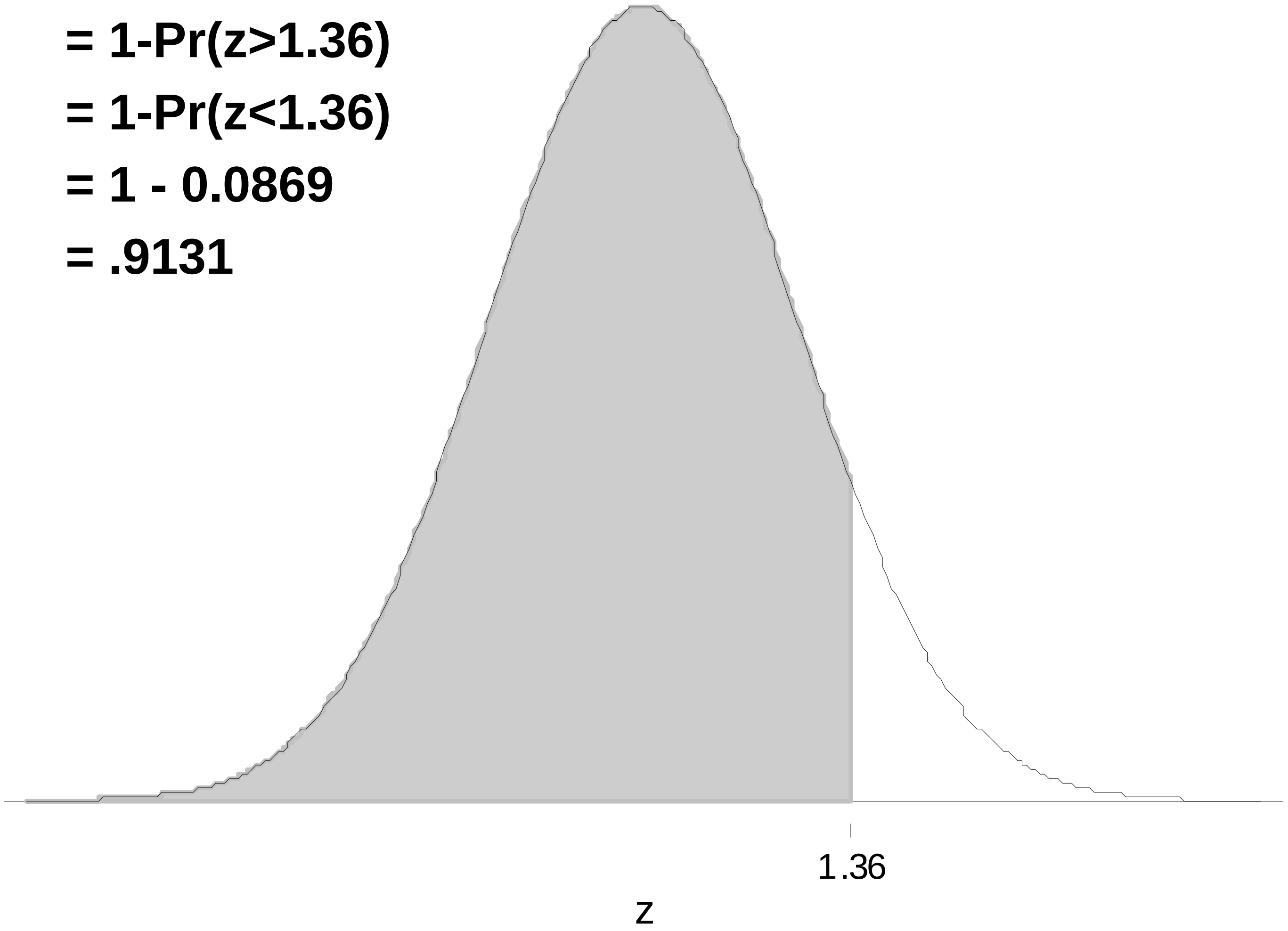
1. Pr(z < -1.36) = 0.0869

e.g. 2: Prob(z < 1.36)

- First Draw diagram
- Then work out how you can compute the area you need from the diagram
  - E.g. if z is positive, use symmetry of normal curve.
- Then look-up the first decimal place in the row heading
- Then find the second decimal place in the column heading

e.g.2: Prob(z < 1.36)

Pr(z < 1.36) = 1-Pr(z>1.36)
= 1-Pr(z<1.36)
= 1 - 0.0869
= .9131



University of Glasgow

Exercises:

1. $\Pr(z > -2.897)$
2. $\Pr(-2 < z < 2)$
3. $\Pr(-2 < z < 2)$

4. Find *zi* s.t. Prob($z < zi$) = 0.06
5. Find the value of *zi* such that Prob($-zi < z < zi$) = 0.99
6. Find the value of *zi* such that Prob($-zi < z < zi$) = 0.95

Using the macro commands:

After ***drawing the diagram***:

- 1st open a new data sheet & enter 0 in first cell
- 2nd Macros
- 3rd Open syntax window
- 4th type in command

pz_lt_zi (1.36) . calculates the probability that z is less than 1.36 & will result in the following output:

```
Prob(z < zi) for a given zi
          ZI        PROB
     1.36000      .91309
```

Prob(z < 1.36)

pz_gt_zi **calculates the probability that z is greater than zi:**

- e.g. **pz_gt_zi (-2.897) .**
- will result in the following output:

```
Prob(z > zi) for a given zi
         ZI        PROB
   -2.89700      .99812
```

- which says that 99.812% of z lie above –2.897.
- Q/ how would you find this probability using the table?

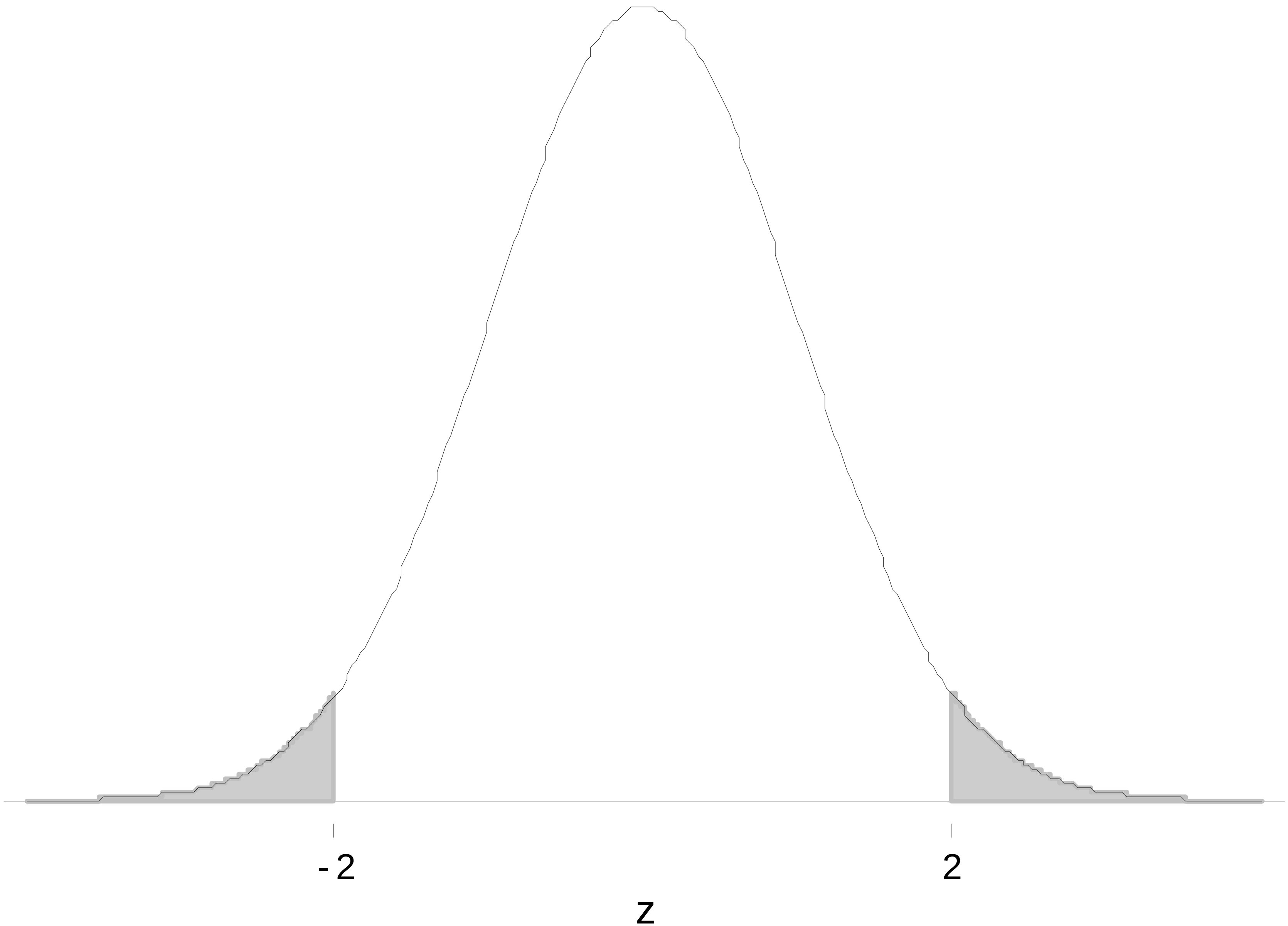

-2.897
z

pz_lg_zi **calculates the probability that z is less than ziL or greater than ziU**

- e.g. **pz_lg_zi zil=(-2) ziu=(2).** will result in the following output:

```
Prob((z &lt; ziL) OR (z &gt; ziU)) for a given zi
        ZIL          ZIU        PROB
   -2.00000     2.00000      .04550
```

- which can be interpreted as telling us that just 4.55% of z lie outside of the range –2 to 2.
- Q/ How would you find the probability that z lies above 2 and/or below -2 using the tables?


-2
2
z

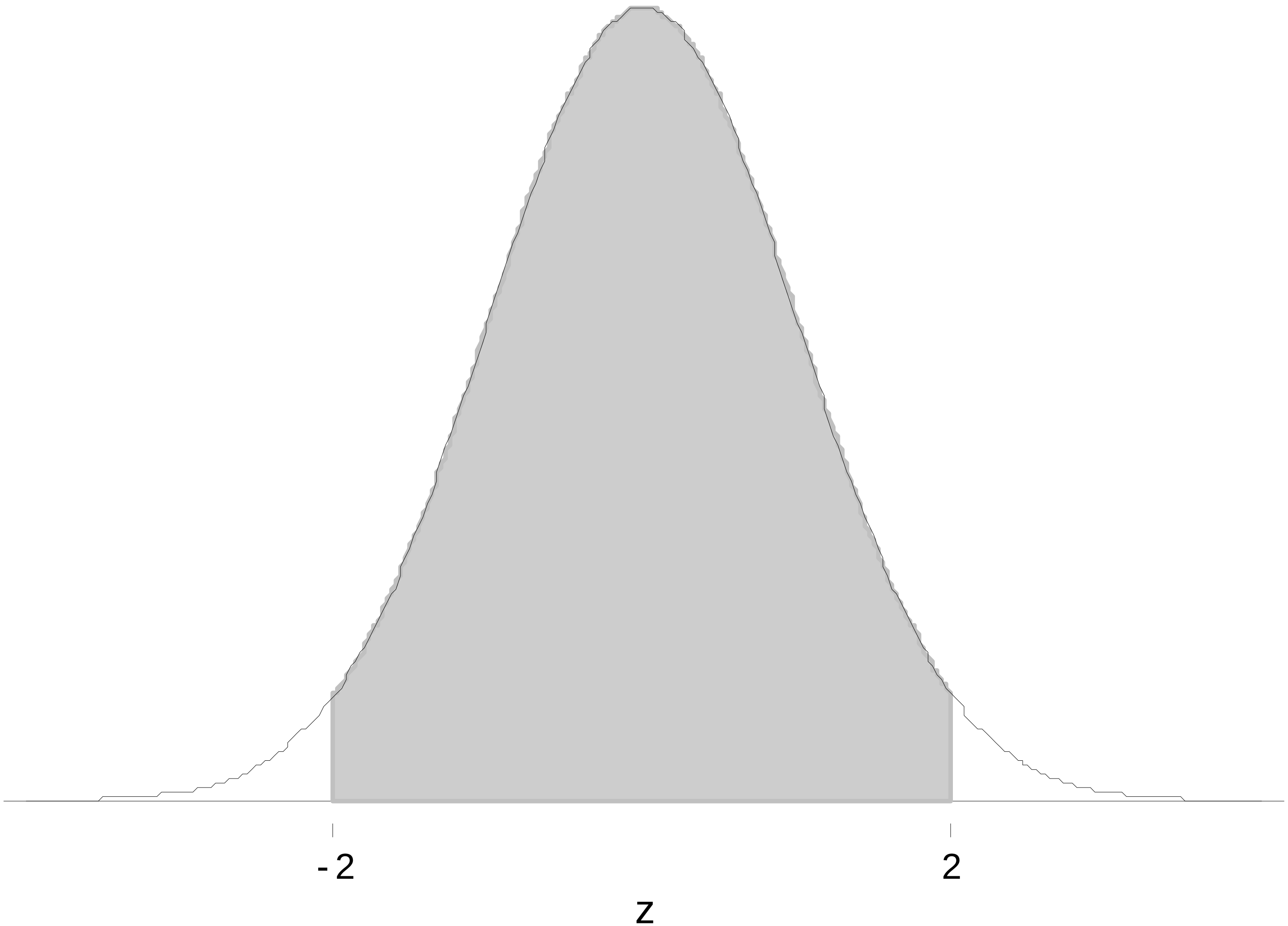
pz_gl_zi **calculates the probability that z is greater than ziL AND less than ziU**

- e.g. **pz_gl_zi zil=(-2) ziu=(2).**

results in the following output:

```
Prob(ziL < z < ziU)) for a given zi
        ZIL          ZIU        PROB
   -2.00000      2.00000      .95450
```

- which tells us that 95.45% of z lie in the range –2 to 2.
- Q/ How would you find the proportion of z that lies between -2 and 2?


-2
2
z

University of Glasgow

Plan

- Review of last week's lecture
- 1. Find probabilities from z_i
    - Tables
    - SPSS macro command
- **2. Find z_i from a given probability**
    - z that bounds upper or lower tail area
    - $\pm z$ that bounds central area
- 3. Find z_i & probabilitiy from $x_i \sim N(\mu, \sigma)$
- 4. Applying z scores to sampling distributions

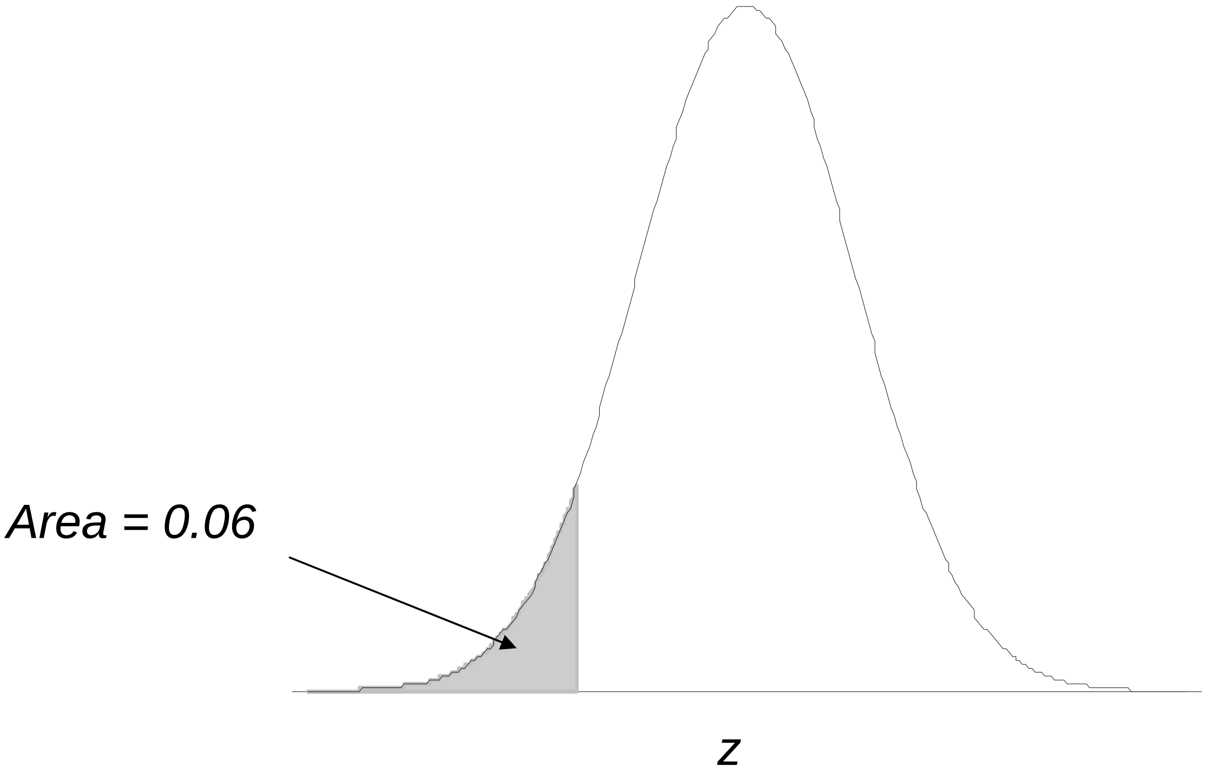
2. Find z_i from a given probability

2.1 Using tables:

- You can look up the areas in the body of the table and find the z value that bounds that area:
  - You must be careful to restate your problem in a way that fits with the probabilities reported in the table however

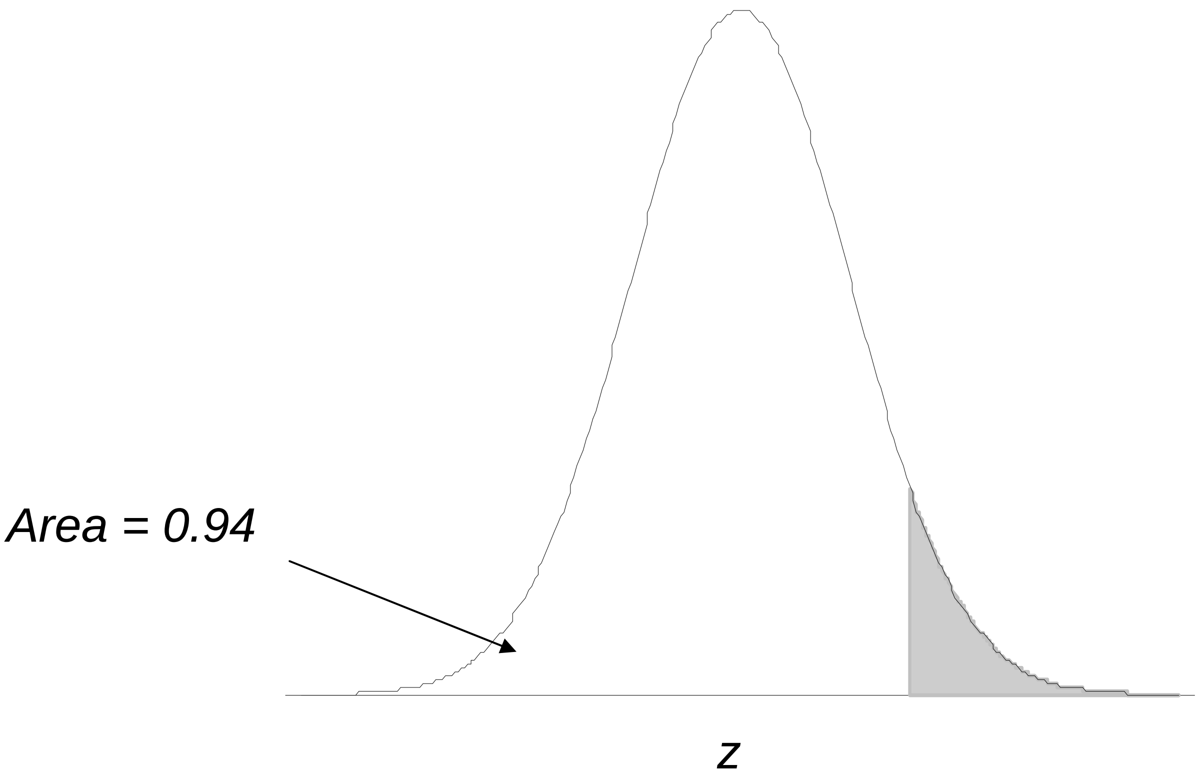
E.g. Find zi s.t. Prob(z < zi) = 0.06

- This is a small area in the left hand tail so z_i is going to be negative
  - Q/ How can we find the value of z even though table only gives positive values of z?


Area = 0.06
z

E.g. Find z_i s.t. Prob($z < z_i$) = 0.06

- Table only gives positive values of z.
  - But because the normal distribution is symmetrical, we can look at the upper tail of the same area & know that the z value will be of the same absolute value.
  - I.e. Find z_i s.t. Prob($z > z_i$) = 0.94
  - The tables only give us the Prob($z < z_i$)
    - But we can simply switch the sign on z_i:
    - i.e. $-z_i$ s.t. Prob($z > -z_i$) = $+z_i$ s.t. Prob($z < +z_i$)



Area = 0.94
z

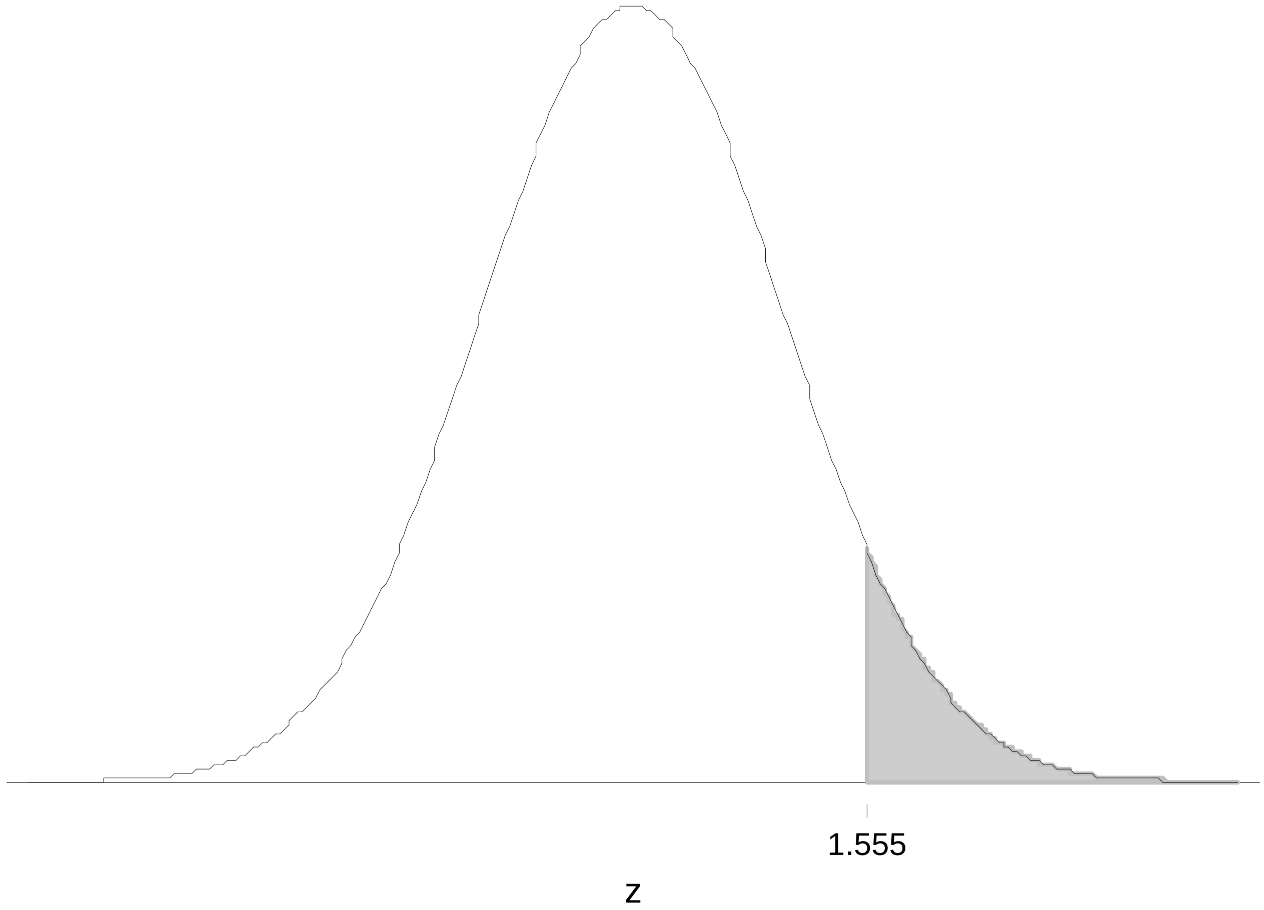
2.2 Use zi_lt_zp and zi_gt_zp Macros:

zi_lt_zp p = (0.06).

```
Value of zi such that Prob(z < zi)
  = PROB when PROB is given
            ZI          PROB
    -1.55477         .06000
```

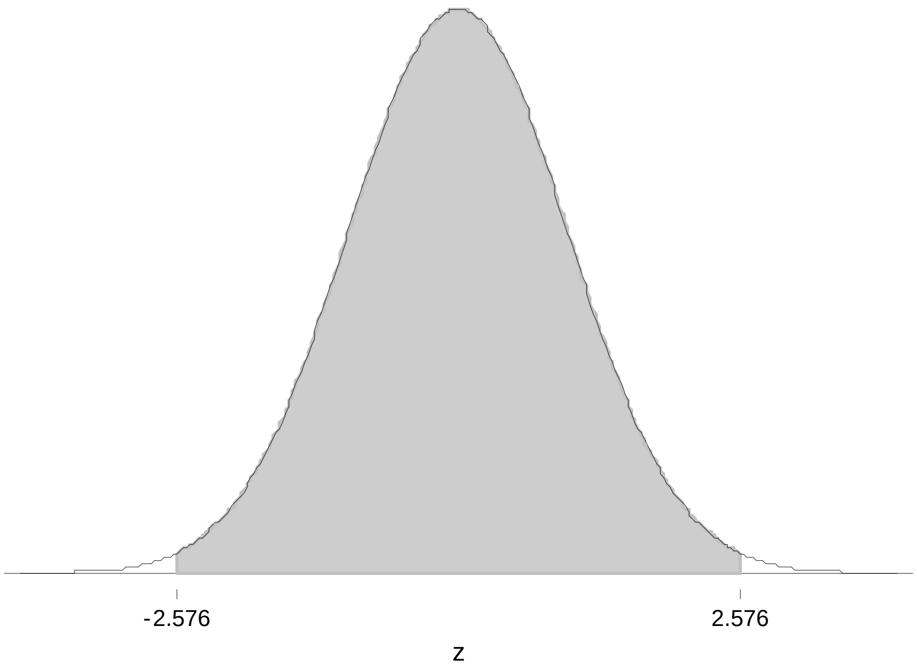
zi_gt_zp p = (0.06).

```
ValuValue of zi such that Prob(z > zi) = PROB when PROB is given
          ZI          PROB
     1.55477        .06000
```


1.555
z

2.3 ± *z* that bounds central area

- Find the value of z_i such that Prob($-z_i < z < z_i$) = 0.99
- Q/ How would you do this using tables?



Using Tables:

- First find half of the central area:
  - Area of half of central area = 0.99 / 2 = 0.495
  - Then add 0.5 = 0.495 + 0.5 = 0.995
- Then find z value associated with that area:
  - Look up 0.995 in the body of the table
  - $z = 2.575$

zi_gl_zp p=(0.99) .

```
Value of zi such that Prob(-zi < z < zi)
   = PROB, when PROB is given
        ZIL          ZIU         PROB
   -2.57583    2.57583      .99000
```

- which tells you that the central 99% of z values are bounded by + and – 2.576

University of Glasgow

Plan

- Review of last week's lecture
- 1. Find probabilities from z_i
    - Tables
    - SPSS macro command
- 2. Find z_i from a given probability
    - z that bounds upper or lower tail area
    - $\pm z$ that bounds central area
- **3. Find z_i & probabilitiy from $x_i \sim N(\mu, \sigma)$**
- 4. Applying z scores to sampling distributions

3. Find z_i & probabilitiy from x_i ~N(μ,σ)

- For a particular value x_i of a normally distributed variable x, we can calculate the standardised normal value, z_i, associated with it by subtracting the population mean, and dividing by the population standard deviation,

$$z_i = \frac{x_i - \mu}{\sigma}$$

- So, we can standardise any value of *x* provided we know the population mean and population standard error of the mean.
- And once you have standardised a value (i.e. converted it to a *z*-score), then you can use it to calculate probabilities under the standard normal curve knowing that these probabilities correspond to probabilities under the original distribution of *x*.

You know that the height of all 18 year old males is normally distributed with a mean of 1.8m and a standard deviation of 1.2m. What proportion of 18year olds are < 2m tall?

x_i = height of 18 year olds
= 2

μ = population mean of x
= mean height of all 18yr olds
= 1.8

σ = population standard deviation of x.
= 1.2

University of Glasgow

Because height is normally distributed, we know that: Prob(height < 2m) = Prob($z < z_i$)

- where z_i is the standardised value for $x = 2$. First we need to calculate z_i:

$$z_i = \frac{x_i - \mu}{\sigma} \quad = \quad \frac{2 - 1.8}{1.2} = \frac{0.2}{1.2} = 0.1667$$

- Now that we have calculated $z_i = 0.1667$, we can calculate Prob(height < 2m)= Prob(z < 0.1667)
- Using the pz_lt_zi macro, we get:

pz_lt_zi (0.1667).

```
 Prob(z < zi) for a given zi
         ZI         PROB
     .16670      .56620
```

- That is, 56.62% of 18year old males are less than 2m tall.

University of Glasgow

Plan

- Review of last week's lecture
- 1. Find probabilities from z_i
    - Tables
    - SPSS macro command
- 2. Find z_i from a given probability
    - *z* that bounds upper or lower tail area
    - ± *z* that bounds central area
- 3. Find z_i & probabilitiy from $x_i \sim N(\mu, \sigma)$
- **4. Applying *z* scores to sampling distributions**

4. Applying z scores to sampling distributions

- "*nature's questionable tendency to normalcy*" $\Rightarrow$ limited use of z scores
- … if it were not for the CLT:
    - Sampling distributions of means are always normally distributed provided *n* is large.
- $\Rightarrow$ following formula for z:

$$z_i = \frac{\bar{x}_i - \mu}{\sigma_{\bar{x}}}$$

where :

μ = population mean

$\bar{x}_i$ = sample mean

$\sigma_{\bar{x}}$ = standard deviation of all the sample means, called the standard error of the mean

z_i = z score

If the sample mean inside leg of gerbils is 2.7cm, the population mean is 3cm and the standard error of the mean is 4, what is the z-score for the sample mean? What proportion of all possible large samples of gerbils have inside leg measurements of less than 2.7cm?

$$z_i = \frac{\bar{x}_i - \mu}{\sigma_{\bar{x}}} = \frac{2.7 - 3}{4} = -0.075$$

University of Glasgow

Prob(sample mean < 2.7) = Prob(z < -0.075)

zi_lt_zp p = (-0.075).

```
 Prob(z < zi) for a given zi
            ZI          PROB
        -.07500       .47011
```

- That is, 47.01% of all possible sample mean inside leg lengths are less than 2.7cm.

- So the CLT + *z* score allow us to say something about sample means.

Summary:

- 1. How to find probabilities from z_i
  - Tables
  - SPSS
- 2. How to find z_i from a given probability
  - z that bounds upper or lower tail area
  - $\pm z$ that bounds central area
- 3. How to find z_i & probabilitiy from $x_i \sim N(\mu,\sigma)$
- 4. How to z scores can be applied to sampling distributions

University of Glasgow

Reading:

- Pryce, chapter 3
- M&M section 1.3 and chapter 5.