1. **Problem**
In a deck of strange cards, there are 751 cards. Each card has an image and a color. The amounts are shown in the table below.

	orange	teal	white
cat	42	81	44
flower	94	58	43
pig	71	70	32
tree	49	98	69

(a) What is the probability a random card is both a tree and white?
(b) What is the probability a random card is teal?
(c) What is the probability a random card is orange given it is a tree?
(d) What is the probability a random card is a tree given it is orange?
(e) What is the probability a random card is either a tree or orange (or both)?
(f) What is the probability a random card is a tree?

2. **Problem**
A farm produces 4 types of fruit: A, B, C, and D. The fruits' masses follow normal distributions, with parameters dependent on the type of fruit.

Type of fruit	Mean mass (g)	Standard deviation of mass (g)
A	73	4
B	107	15
C	83	12
D	148	7

One specimen of each type is weighed. The results are shown below.

Type of fruit	Mass of specimen (g)
A	75.68
B	101.6
C	83.36
D	141.1

Which specimen is the most unusually small (relative to others of its type)?

3. **Problem**

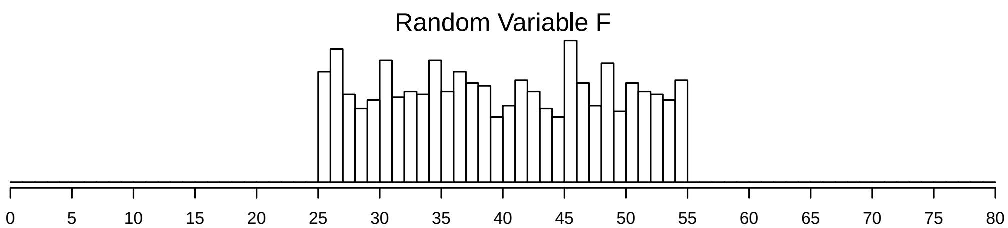
We can estimate the mean of **symmetric** distributions.

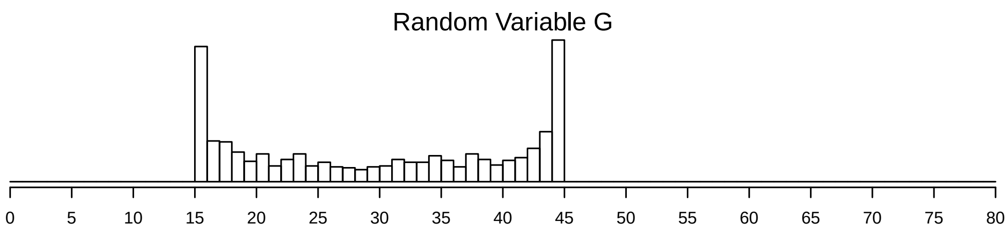
$$\bar{x} \approx \frac{\max(x) + \min(x)}{2}$$

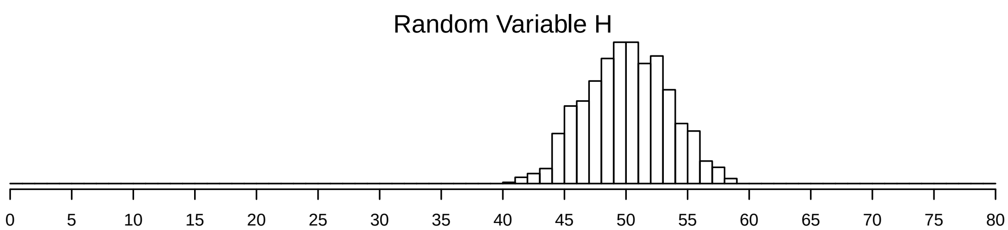
We can **roughly** estimate the standard deviation of certain distributions.

Shape	SD estimate
bell	range/6
uniform	range/4
bimodal	range/2

Three random variables (F, G, and H) were measured 1000 times each. The resulting histograms show the three distributions.







(a) Estimate the mean of F.

(b) Estimate the mean of G.

(c) Estimate the mean of H.

(d) Estimate the standard deviation of F.

(e) Estimate the standard deviation of G.

(f) Estimate the standard deviation of H.

4. **Problem**
   Let X be normally distributed with mean 89.8 and standard deviation 22.3. Please calculate the probabilities shown below and also shade a corresponding region under the density curve.

   (a) What's the probability that X is less than 65.3? **Draw a sketch**.

   (b) What's the probability that X is more than 99.6? **Draw a sketch**.

   (c) What's the probability that X is between 65.3 and 99.6? **Draw a sketch**.

5. **Problem**
   Let random variable W have mean $\mu_w = 46$ and standard deviation $\sigma_w = 8$. Let random variable X represent the **average** of $n = 196$ instances of W.

   (a) Determine the expected value of X. $\mu_x = ?$
   (b) Determine the standard deviation of X. $\sigma_x = ?$
   (c) Using normal approximation, determine $P(X < 45.7)$.
   (d) Using normal approximation, determine $P(X > 46.66)$.
   (e) Using normal approximation, determine $P(|X - \mu_x| < 0.3028571)$.
   (f) Using normal approximation, determine $P(|X - \mu_x| > 0.12)$.

6. **Problem**
   As an ornithologist, you wish to determine the average body mass of *Vermivora peregrina*. You randomly sample 13 adults of *Vermivora peregrina*, resulting in a sample mean of 12.91 grams and a sample standard deviation of 1 grams. Determine a 90% confidence interval of the true population mean.

7. **Problem**
   A student is taking a multiple choice test with 300 questions. Each question has 3 choices. You want to detect whether the student does better than random guessing, so you decide to run a hypothesis test with a significance level of 0.1.

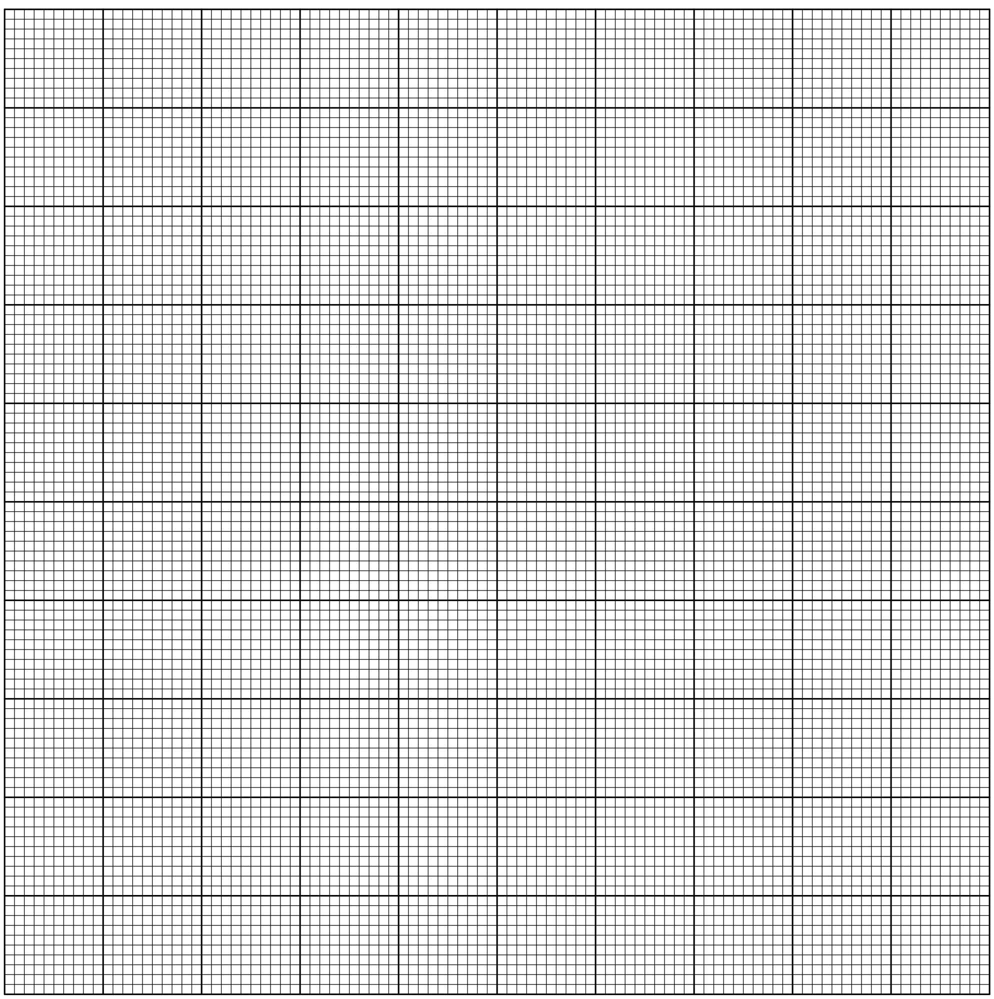
   Then, the student takes the test and gets 111 questions correct.

   (a) What kind of hypothesis test is appropriate?
   (b) State the hypotheses
   (c) Determine the p-value.
   (d) Decide whether we reject or retain the null hypothesis.
   (e) Do we think the student did significantly better than random guessing?

8. **Problem**
You have collected the following data, and started calculating some of the statistics:

x	y	xy
2.2	1.1	2.42
8.4	2.3	19.32
6	1.8	10.8
9	2.4	21.6
4.4	1.6	7.04
6.4	1.9	12.16
$\sum x = 36.4$	$\sum y = 11.1$	$\sum x_i y_i = 73.34$
$\bar{x} = 6.07$	$\bar{y} = 1.85$	
$s_x = 2.53$	$s_y = 0.476$	

Please plot the data and a corresponding regression line.



9. **Problem**
   Let each trial have a chance of success $p = 0.4$. If 46 trials occur, what is the probability of getting more than 11 but less than 21 successes?

   In other words, let $X \sim \text{Bin}(n = 46,\ p = 0.4)$ and find $P(11 < X < 21)$.

   Use a normal approximation along with the continuity correction.

10. **Problem**
A null hypothesis claims a population has a mean $\mu = 22.0$. You decide to run two-tail test on a sample of size $n = 12$ using a significance level $\alpha = 0.02$.

You then collect the sample:

21.4	30.4	19.7	23.2	22.7
27.3	26.1	24.7	22.7	21
26.2	29.7			

(a) Determine the p-value.

(b) Do you reject the null hypothesis?

1. (a) $P(\text{tree and white}) = \frac{69}{751} = 0.0919$
   (b) $P(\text{teal}) = \frac{81+58+70+98}{751} = 0.409$
   (c) $P(\text{orange given tree}) = \frac{49}{49+98+69} = 0.227$
   (d) $P(\text{tree given orange}) = \frac{49}{42+94+71+49} = 0.191$
   (e) $P(\text{tree or orange}) = \frac{49+98+69+42+94+71+49-49}{751} = 0.563$
   (f) $P(\text{tree}) = \frac{49+98+69}{751} = 0.288$

---

2. We compare the z-scores. The smallest z-score corresponds to the specimen that is most unusually small.

Type of fruit	formula	z-score
A	$z = \frac{75.68-73}{4}$	0.67
B	$z = \frac{101.6-107}{15}$	-0.36
C	$z = \frac{83.36-83}{12}$	0.03
D	$z = \frac{141.1-148}{7}$	-0.99

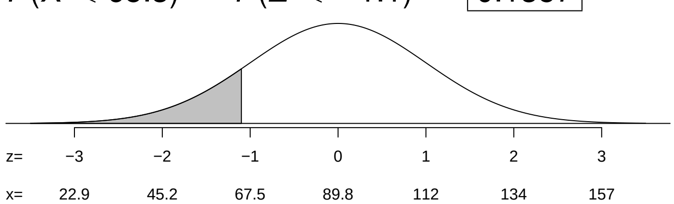
Thus, the specimen of type D is the most unusually small.

---

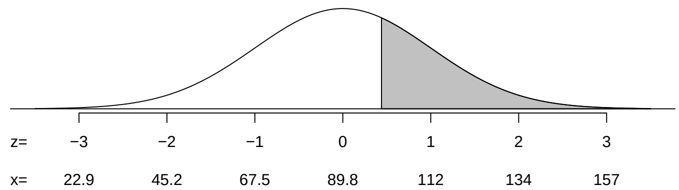
3. (a) 40
   (b) 30
   (c) 50
   (d) 7.5
   (e) 15
   (f) 3.3333333

---

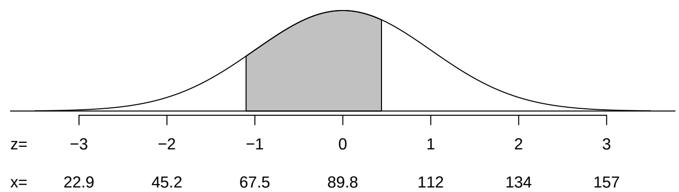
4. Notice the three probabilities will add up to 1.
   (a) $P(X < 65.3) = P(Z < -1.1) = \boxed{0.1357}$



(b) $P(X > 99.6) = P(Z > 0.44) = \boxed{0.33}$



(c) $P(65.3 < X < 99.6) = P(-1.1 < Z < 0.44) = \boxed{0.5343}$



---

5. (a) 46
   (b) 0.5714286
   (c) 0.3015
   (d) 0.123
   (e) 0.4039
   (f) 0.8337

---

6. We are given the sample size, sample mean, sample standard deviation, and confidence level.

$$n = 13$$
$$\bar{x} = 12.91$$
$$s = 1$$
$$\gamma = 0.9$$

Find the degrees of freedom.

$$\begin{aligned} df &= n - 1 \\ &= 13 - 1 \\ &= 12 \end{aligned}$$

Determine the critical t value, $t^\star$, such that $P(|T| < t^\star) = 0.9$ and $df = 12$.

$$t^\star = 1.78$$

Use the formula for bounds (mean, σ unknown).

$$\begin{aligned} LB &= \bar{x} - t^\star \frac{s}{\sqrt{n}} \\ &= 12.91 - 1.78 \times \frac{1}{\sqrt{13}} \\ &= 12.4 \end{aligned} \qquad \begin{aligned} UB &= \bar{x} + t^\star \frac{s}{\sqrt{n}} \\ &= 12.91 + 1.78 \times \frac{1}{\sqrt{13}} \\ &= 13.4 \end{aligned}$$

We are 90% confident that the population mean is between 12.4 and 13.4.

$$CI = (12.4,\ 13.4)$$

---

7. This is a right-tail (one-tail) proportion test because we only care whether the student does better than random.

Determine the null population proportion.

$$p_0 = \frac{1}{3} = 0.333$$

State the hypotheses.

$$H_0 \quad \text{claims} \quad p = 0.333$$
$$H_A \quad \text{claims} \quad p > 0.333$$

Determine the standard error.

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0.333(1-0.333)}{300}} = 0.0272$$

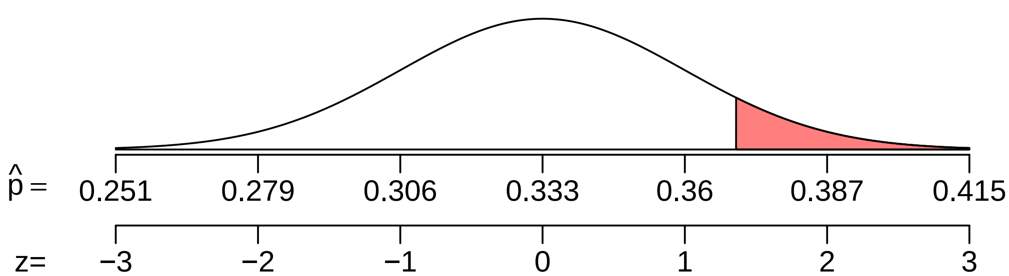
Determine the sample proportion.

$$\hat{p} = \frac{111}{300} = 0.37$$

Determine a z score. For simplicity, we ignore the continuity correction.

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.37 - 0.333}{0.0272} = 1.36$$

Make a sketch of the null's sampling distribution. The p-value is a right area.



To determine that right area, we use the z table.

$$\begin{aligned} p\text{-value} &= P(\hat{p} > 0.37) \\ &= P(Z > 1.36) \\ &= 1 - P(Z < 1.36) \\ &= 0.0869 \end{aligned}$$

Compare p-value to α (which is 0.1).

$$p\text{-value} < \alpha$$

Make the conclusion: we reject the null hypothesis.

We think the student did better than random guessing typically allows.

(a) Right tail (one-tail) proportion test
(b) Hypotheses: H_0 claims $p = 0.333$ and H_A claims $p > 0.333$.
(c) The p-value is 0.0869
(d) We reject the null hypothesis.
(e) We think the student did better than random guessing typically allows.

---

8. Remember the formula for the correlation coefficient.

$$r = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y}$$

$$r = 0.997$$

The regression line has the form

$$y = a + bx$$

So, a is the y-intercept and b is the slope. We have formulas to determine them:

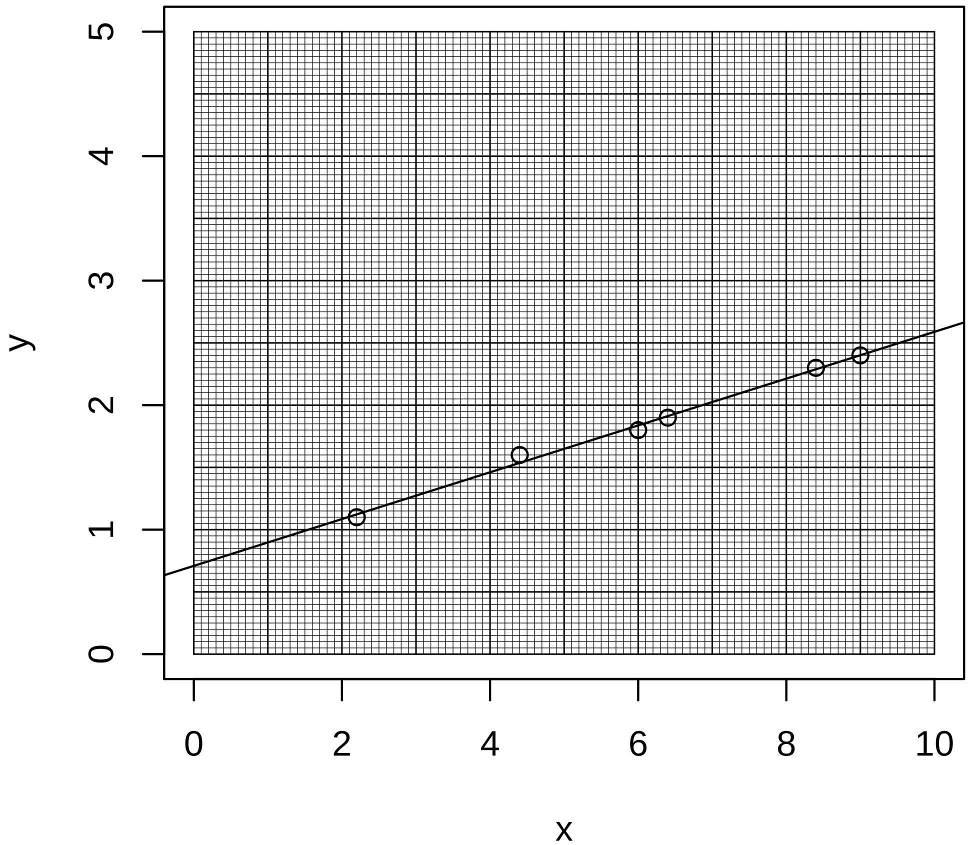
$$b = r\frac{s_y}{s_x} = 0.997 \cdot \frac{0.476}{2.53} = 0.188$$

$$a = \bar{y} - b\bar{x} = 1.85 - (0.188) \cdot 6.07 = 0.709$$

Our regression line:

$$y = 0.709 + (0.188)x$$

Make a plot.



---

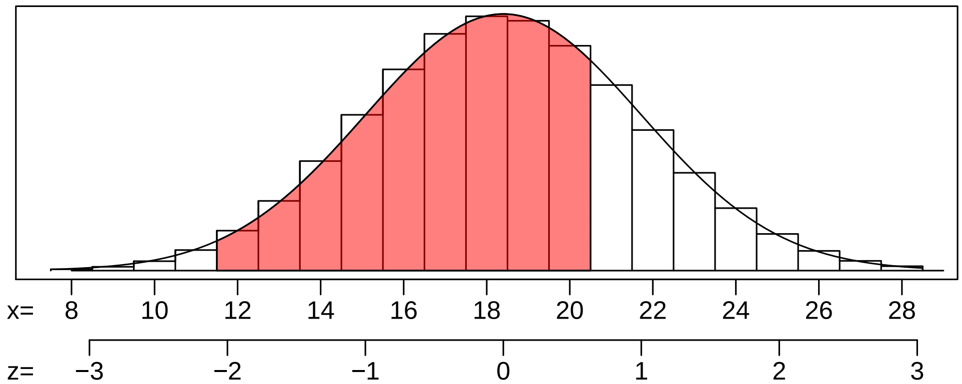
9. Find the mean.

$$\mu = np = (46)(0.4) = 18.4$$

Find the standard deviation.

$$\sigma = \sqrt{np(1-p)} = \sqrt{(46)(0.4)(1-0.4)} = 3.3226$$

Make a sketch, specifically try to picture whether you need to add or subtract 0.5 for the continuity correction.



Find the z scores.

$$z_1 = \frac{11.5 - 18.4}{3.3226} = -2.08$$

$$z_2 = \frac{20.5 - 18.4}{3.3226} = 0.63$$

Find the percentiles (from z-table).

$$\ell_1 = 0.0188$$

$$\ell_2 = 0.7357$$

Calculate the probability.

$$P(11 < X < 21) = 0.7357 - 0.0188 = 0.717$$

---

10. State the hypotheses.

$$H_0 \quad \text{claims} \quad \mu = 22$$
$$H_A \quad \text{claims} \quad \mu \neq 22$$

Find the mean and standard deviation of the sample.

$$\bar{x} = 24.592$$

$$s = 3.413$$

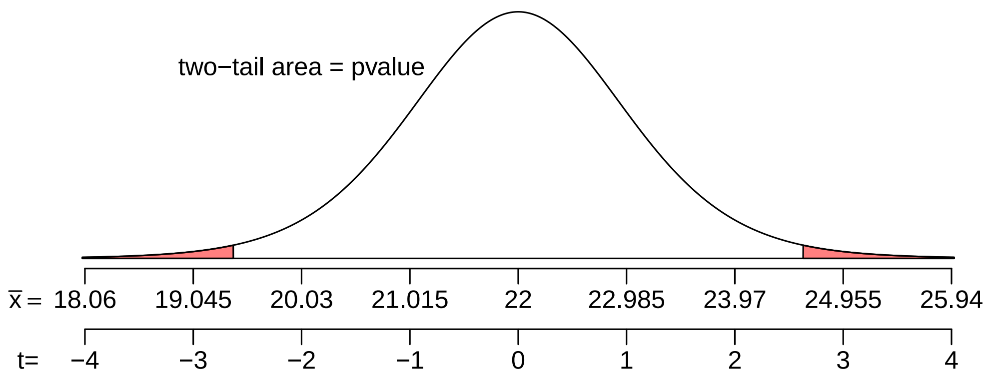
Determine the degrees of freedom.

$$df = 12 - 1 = 11$$

Find the standard error.

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3.413}{\sqrt{12}} = 0.985$$

Make a sketch of the null's sampling distribution.



Find the t score.

$$t = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{24.592 - 22}{0.985} = 2.63$$

Find the p-value.

$$p\text{-value} = P(|T| > 2.63)$$

We can't get an exact value with our table, but we can determine an interval that contains the p-value. (Look at row with df = 11.)

$$P(|T| > 2.72) = 0.02$$

$$P(|T| > 2.33) = 0.04$$

Basically, because t is between 2.72 and 2.33, we know the p-value is between 0.02 and 0.04.

$$0.02 < p\text{-value} < 0.04$$

Compare the p-value and the significance level ($\alpha = 0.02$).

$$p\text{-value} > \alpha$$

No, we do not reject the null hypothesis.

(a) $0.02 < p$-value < 0.04

(b) No, we do not reject the null hypothesis.