The Trial of the Pyx: The substandard deviation

Unit 2 Lecture 1

Jonathan Auerbach
STAT 489 Pre-Cap Prof Development
jauerba@gmu.edu

September 28, 2021

Learning Objectives

After this lecture, you will be able to:

1. Describe the The Trial of the Pyx.
2. Calculate the standard deviation and standard error.
3. State the root-n rule and explain why the standard deviation does not measure the sampling variation of the average.
4. Graph a histogram using `ggplot`.

These slides use the following R packages

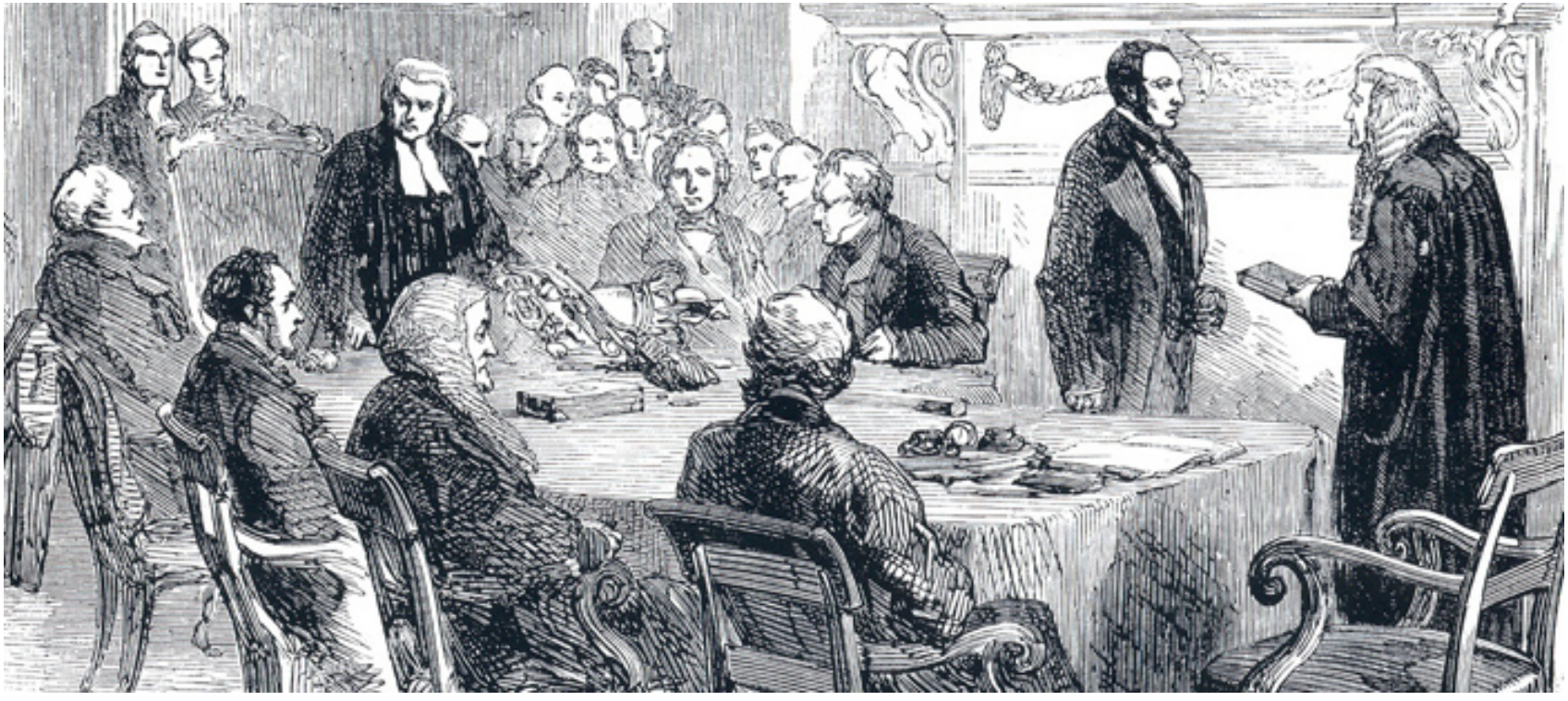
Setup:

```
library("knitr")
library("kableExtra")
library("HistData")
library("tidyverse")
theme_set(theme_bw(base_size = 18))
```

The Trial of the Pyx

- The British monarch routinely inspects the quality of coins produced by the Royal Mint.
  - An ancient tradition dating back to at least 1282.
  - Inspection is important since if coins have too much gold, one could extract for a profit. If too little gold, the currency might be debased.
- Historically, monarch took one coin from every 15 pounds produced.
  - The coins were put in a box called the pyx.
  - At trial, an independent jury weighed them against a standard.
  - The Mint reimbursed the monarch for the number of coins that fell below a tolerance weight (remedy), extrapolated to all coins minted.
- Initially, the jury conducted the trial as if the error of n coins weighed together was n times the error of one coin.
  - In fact, the error increases by $\sqrt{n}$ times the error of one coin. (Likewise, the error of the average is $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$ times the error of one coin.) This is called the root-n rule
  - Without the $\sqrt{n}$ term, the tolerance is too lenient for quality control.

Trial of the Pyx (1854)



Two representative bags from 1848 Royal Mint study (long)

```
Pyx %>%
  filter(Bags == "1 and 2") %>%
  mutate(`Deviation (approximate)` = factor(Deviation,
                    labels = c(-.3, -.23, -.15, -.05,
                               .05, .15, .23, .3))) %>%
  relocate(count, .after = last_col()) %>%
  kable()  %>%  kable_styling(position = "center")
```

Bags	Group	Deviation	Deviation (approximate)	count
1 and 2	near std	Below -R	-0.3	34
1 and 2	near std	(-R to -.2)	-0.23	57
1 and 2	near std	(-.2 to -.l)	-0.15	172
1 and 2	near std	(-.1 to 0)	-0.05	630
1 and 2	near std	(0 to .l)	0.05	597
1 and 2	near std	(.1 to .2)	0.15	366
1 and 2	near std	(.2 to R)	0.23	116
1 and 2	near std	Above R	0.3	28

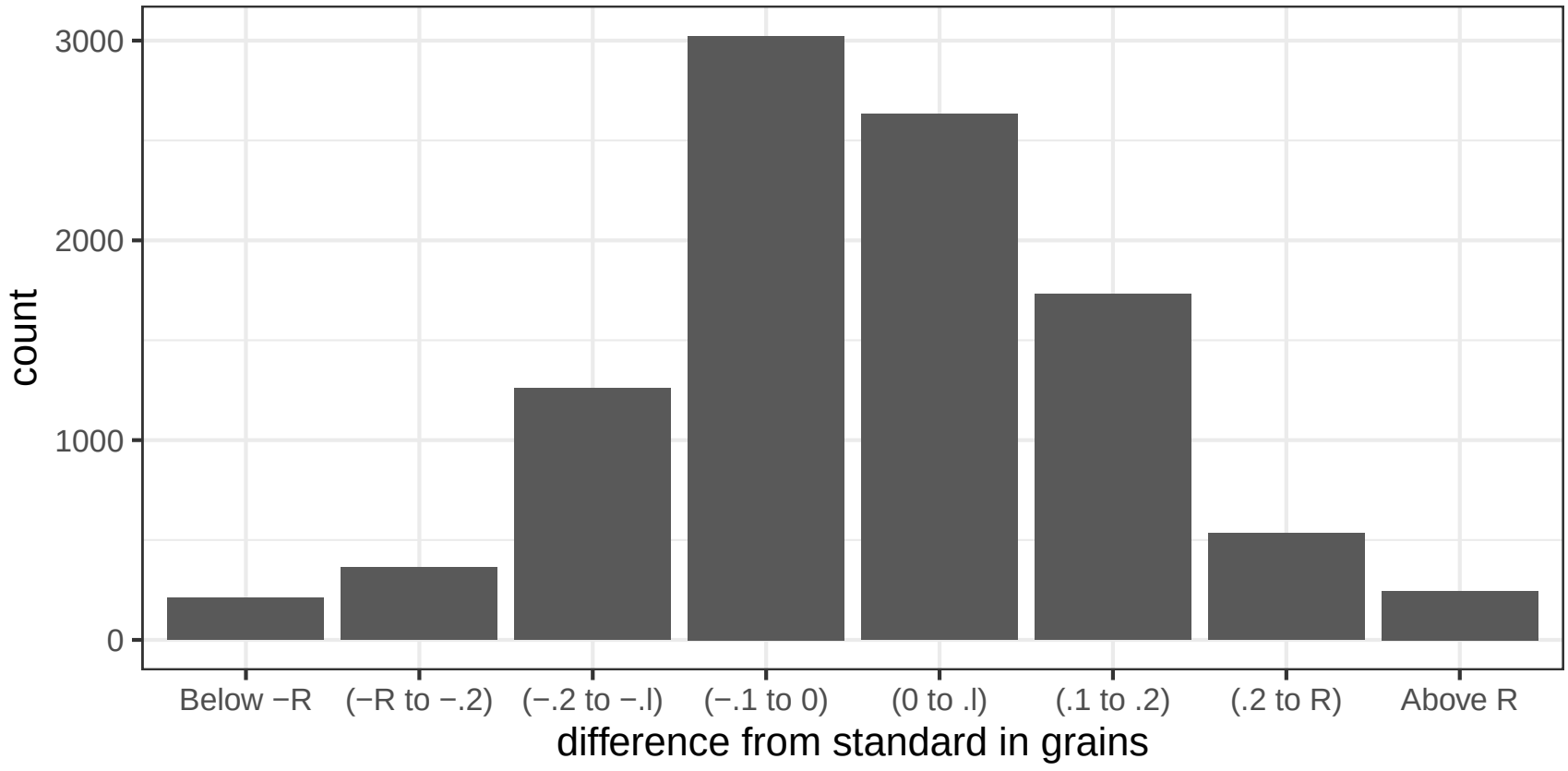
Ten representative bags from 1848 Royal Mint study (wide)

```
Pyx %>%
  pivot_wider(names_from = Deviation,
              values_from = count) %>%
  select(-Group) %>%
  kable() %>%
  kable_styling(font_size = 6)
```

Bags	Below -R	(-R to -.2)	(-.2 to -.l)	(-.1 to 0)	(0 to .l)	(.1 to .2)	(.2 to R)	Above R
1 and 2	34	57	172	630	597	366	116	28
3	11	17	100	412	172	218	57	13
4	20	22	135	350	184	222	50	17
5	30	102	107	289	209	184	50	29
6	32	27	267	210	236	144	56	28
7	47	65	141	380	157	135	50	25
8	11	21	110	215	361	156	71	55
9	10	38	103	228	425	140	36	20
10	14	13	126	309	290	168	50	30

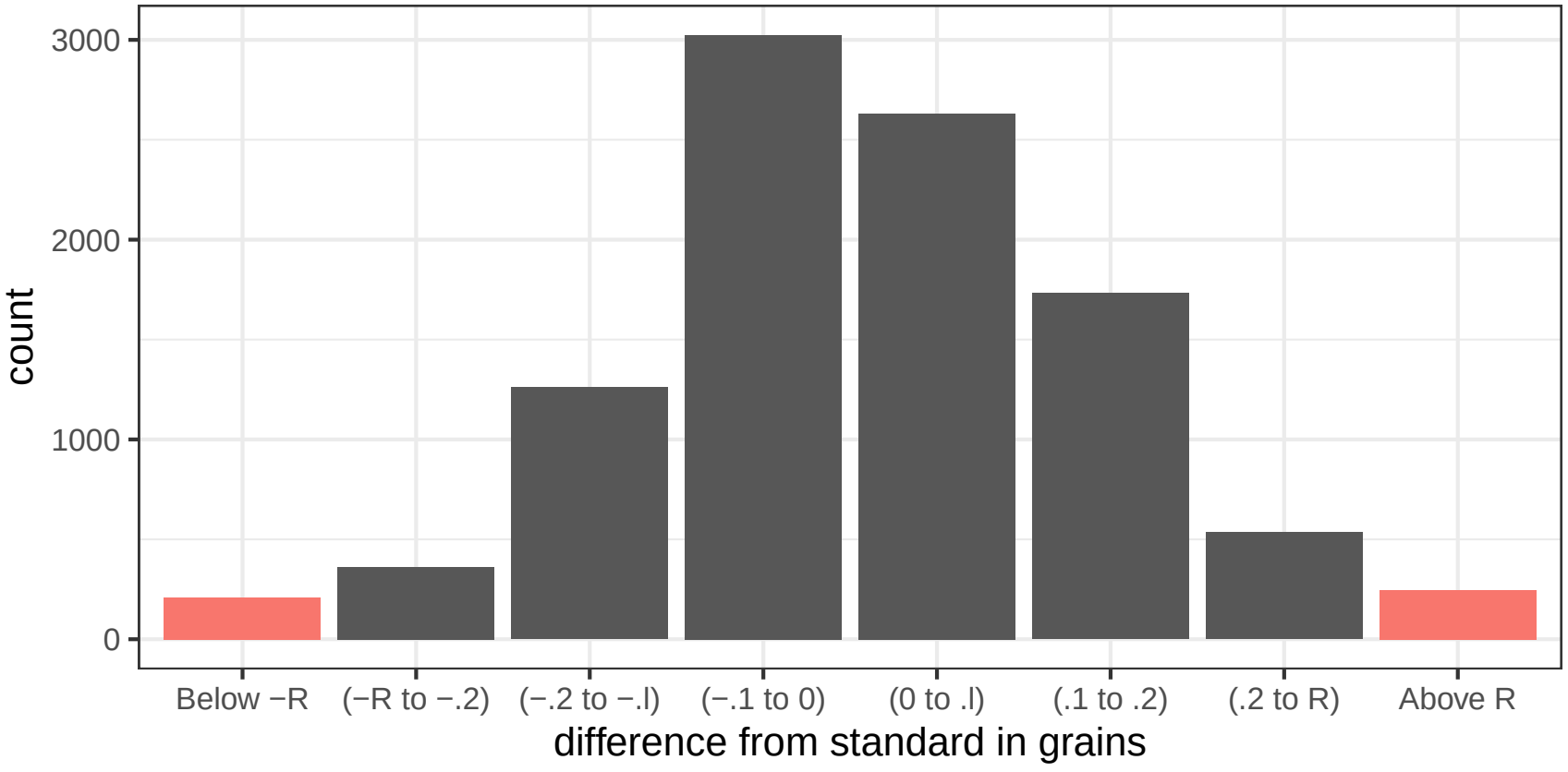
Ten representative bags from 1848 Royal Mint study

```
(pyx_hist <-
   ggplot() +
    geom_bar(aes(x = Deviation, weight = count),
             data = Pyx) +
    labs(x = "difference from standard in grains"))
```



Roughly five percent exceed remedy (R ≈ .26)

```
pyx_hist +
  geom_bar(aes(x = Deviation, weight = count,
    fill = Deviation %in% c("Below -R", "Above R")),
    data = Pyx) + theme(legend.position = "none") +
  scale_fill_manual(values = c("#575757", "#F8766D"))
```



Roughly five percent exceed remedy (R ≈ .26)

```
Pyx <- Pyx %>%
  mutate(Exceed = Deviation %in% c("Below -R", "Above R"))

Pyx %>%
  summarize(`Proportion Exceeding Remedy` =
              weighted.mean(Exceed, count)) %>%
  kable(digits = 2) %>%  kable_styling(position = "center")
```

Proportion Exceeding Remedy
0.05

```
Pyx %>% group_by(Group) %>%
  summarize(`Proportion Exceeding Remedy` =
              weighted.mean(Exceed, count)) %>%
  kable(digits = 2) %>%  kable_styling(position = "center")
```

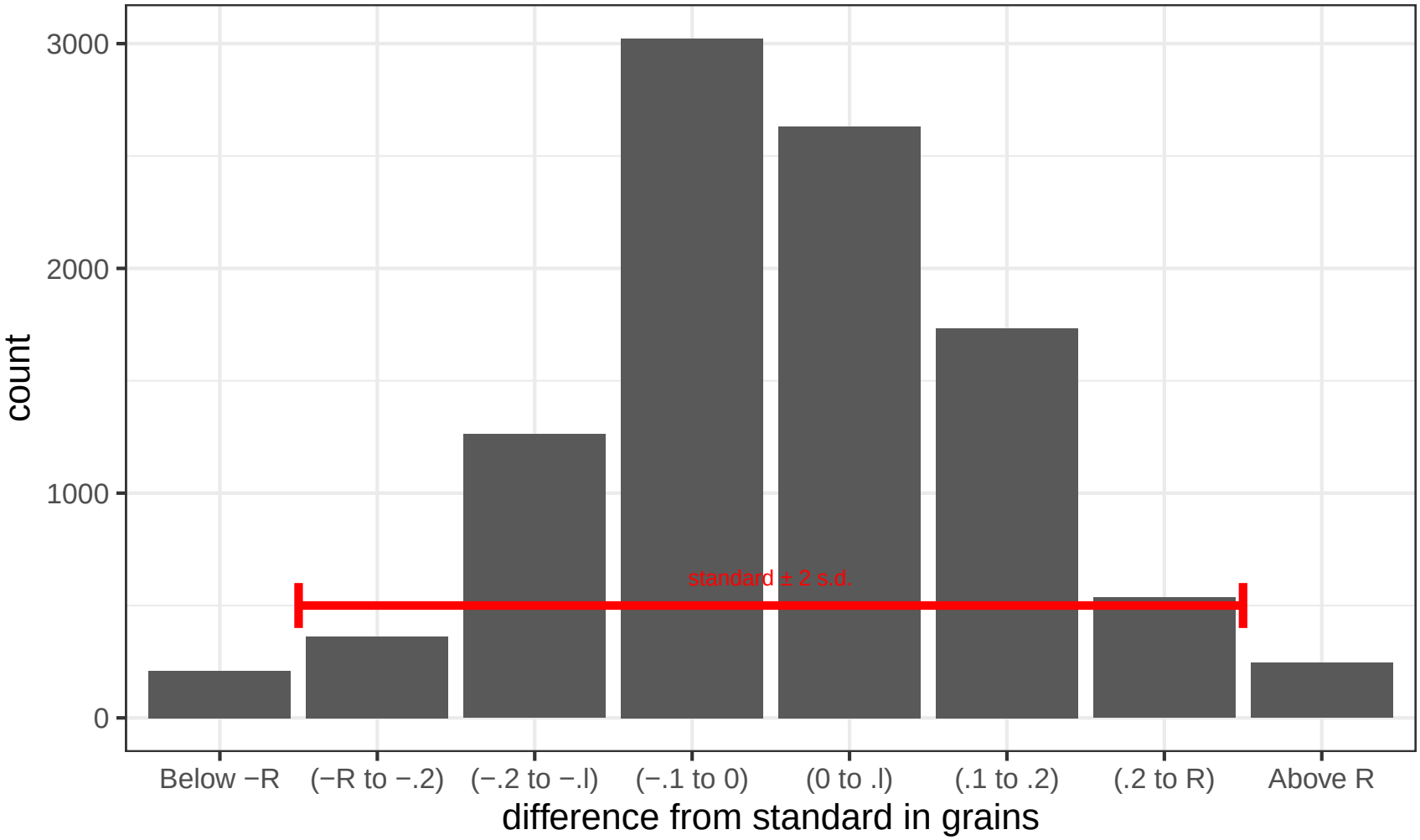
Group	Proportion Exceeding Remedy
below std	0.06
near std	0.03
above std	0.05

R is also two standard deviations from the standard

```
Pyx %>%
  mutate(Deviation = factor(Deviation,
                            labels = c(-.3, -.23, -.15, -.05,
                                       .05, .15, .23, .3)),
         #approximate Deviation (character) by numeric
         Deviation = as.numeric(as.character(Deviation))) %>%
  summarize(
     n = sum(count),
     mean = weighted.mean(Deviation, count),
    `mean (check)` = sum(Deviation * count) / n,
     variance  = sum(count * (Deviation - mean)^2) / n,
     s.d. = sqrt(variance),
     `two s.d.` = 2 * s.d.,
     `two s.e.` = 2 * s.d. / sqrt(n)) %>%
  kable(digits = 2) %>%
  kable_styling(position = "center")
```

n	mean	mean (check)	variance	s.d.	two s.d.	two s.e.
10000	0.01	0.01	0.02	0.13	0.26	0

```
(pyx_hist <- pyx_hist + labs(y = "count") +
   annotate(geom = "text", x = 4.5, y = 625,
            label = "standard ± 2 s.d.", color = "red") +
  geom_errorbarh(aes(xmin = a, xmax = b, y = c),
    height = 200, color = "red", size = 2,
    data = tibble(a=4.5 - 3, b=4.5 + 3, c=500)))
```



This suggests the remedy was set to permit each coin a reasonable level of variation.

- The monarch expected that under acceptable conditions, the Mint would produce coins of standard weight on average.
  - Ninety-five percent of coins would deviate from the standard by less than the remedy. Five percent of coins would exceeded the remedy.
- The monarch expected that under unacceptable conditions, the variation in weights would be larger, and more coins would exceed the remedy.
  - The Mint would have then been responsible for reimbursing the monarch for additional abnormally small coins.
- But this approach would not have detected small, systematic deviations from the standard.
  - The Mint could have theoretically stolen a large amount of gold by extracting a small amount from each coin—just as long as each coin was within tolerance.

The remedy should have been set to only permit the average a reasonable level of variation.

- If the monarch wanted to test all the coins produced by the Mint, it should have set a tolerance for the combined weight of the sample using the root-n rule.
  - The average weight should be within $\frac{2\sigma}{\sqrt{n}} = \frac{R}{\sqrt{n}}$ of standard.
- Let $\{X_i\}_{i=1}^n$ denote the weights of n coins produced under acceptable conditions, each with standard deviation $\sigma = \sqrt{\text{Var}(X_i)}$. Assuming independence:
  - by Variance-Sum formula, $\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2$.
  - the standard error is then $\sqrt{\text{Var}(\frac{1}{n}\sum_{i=1}^n X_i)} = \sqrt{\frac{1}{n^2} n\sigma^2} = \frac{\sigma}{\sqrt{n}}$
- Using the remedy $\frac{R}{\sqrt{n}}$ for the average, the monarch would have identified bags with too little or too much gold content in aggregate
  - The Mint could then reimburse the monarch for the total amount of missing gold below the remedy—not just the missing gold from abnormally small coins.

Root-n rule first described by Abraham de Moivre

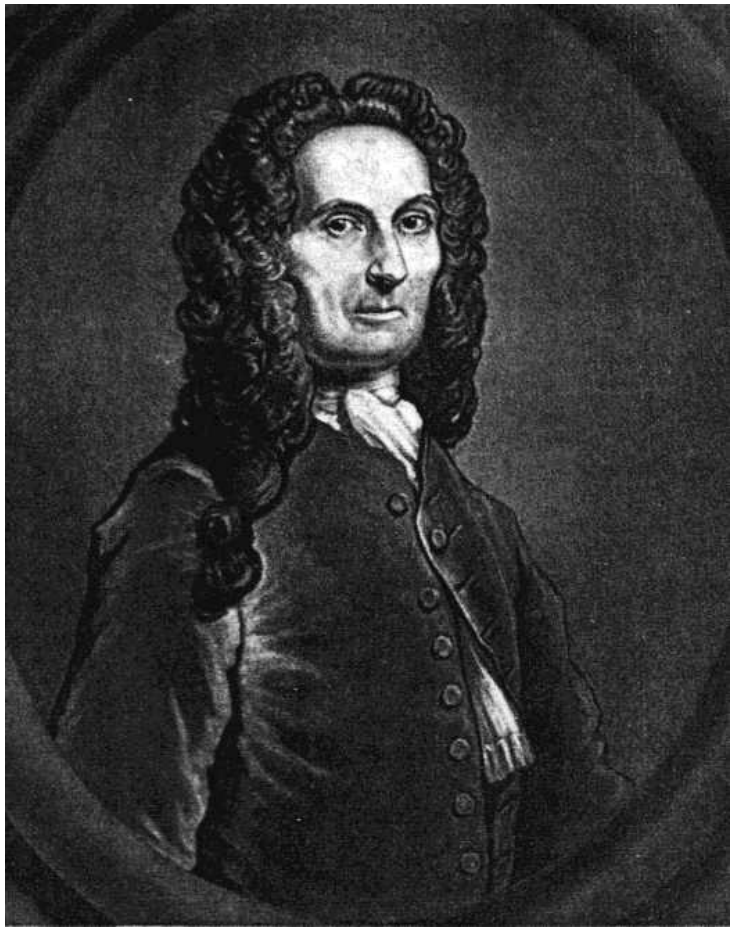
- Gerolamo Cardano (1564) noted that the accuracy of an average improves with the number of observations.
  - Jacob Bernoulli proved the first law of large numbers (1713): He showed that, as the number of coin tosses increases, the proportion landing on heads becomes arbitrarily close to the weight of the coin.
  - Today we would say the standard error of independent observations decreases as the number of trials increases.

- The root-n rule was first described by Abraham de Moivre in *The Doctrine of Chances* (2nd edition, 1738), which stated

  > *the Square-root of the number which denotes how many experiments have been, or are designed to be taken ... will be as it were the Modulus by which we are to regulate our Estimation.*

- The term standard deviation was coined by Karl Pearson (1894), who was also the first to represent it with the symbol σ. Udny Yule introduced the term standard error three years later.

Abraham de Moivre (1736) and *The Doctrine of Chances* (1st edition, 1718)



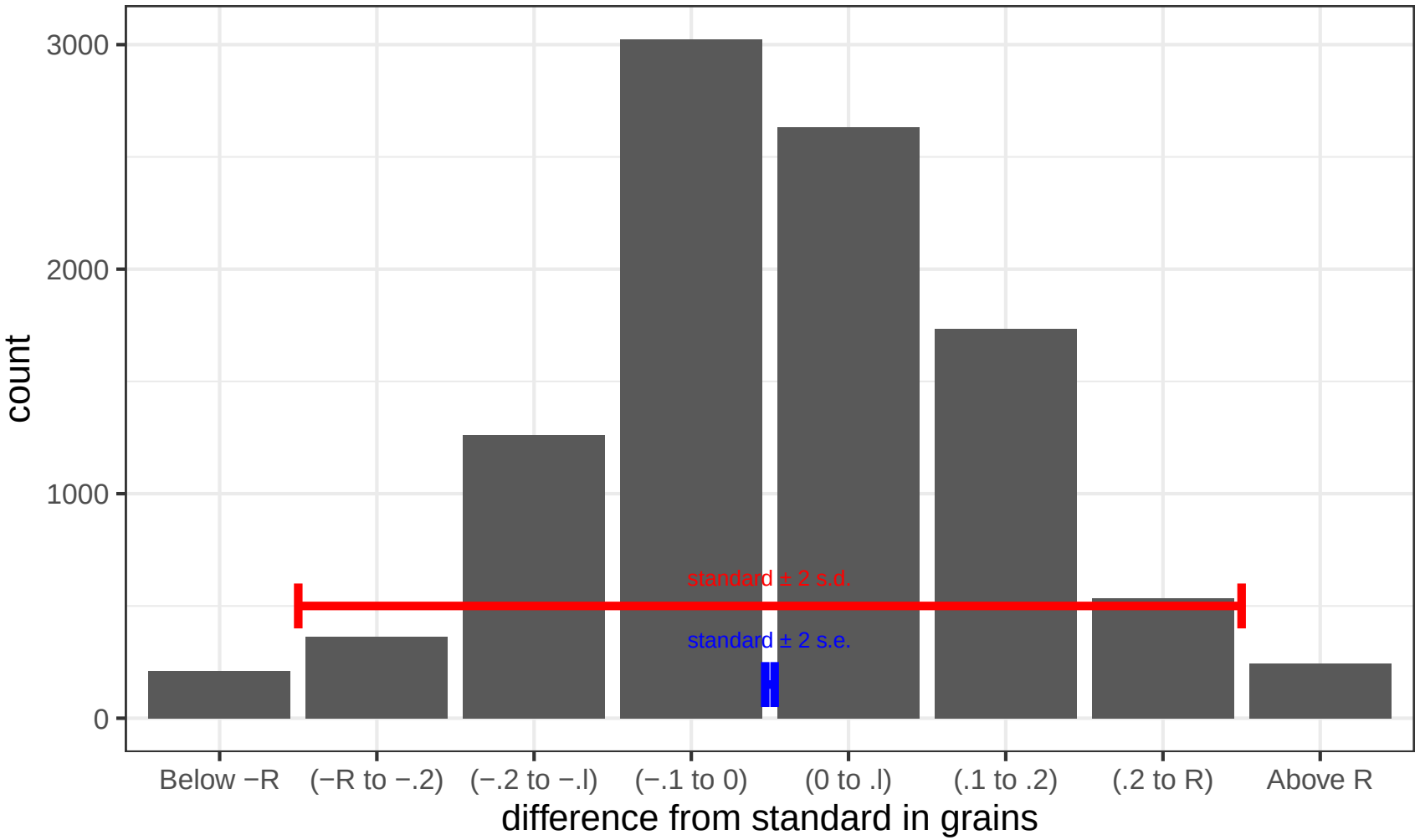
THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability
of Events in Play.

By *A. De Moivre.* F. R. S.

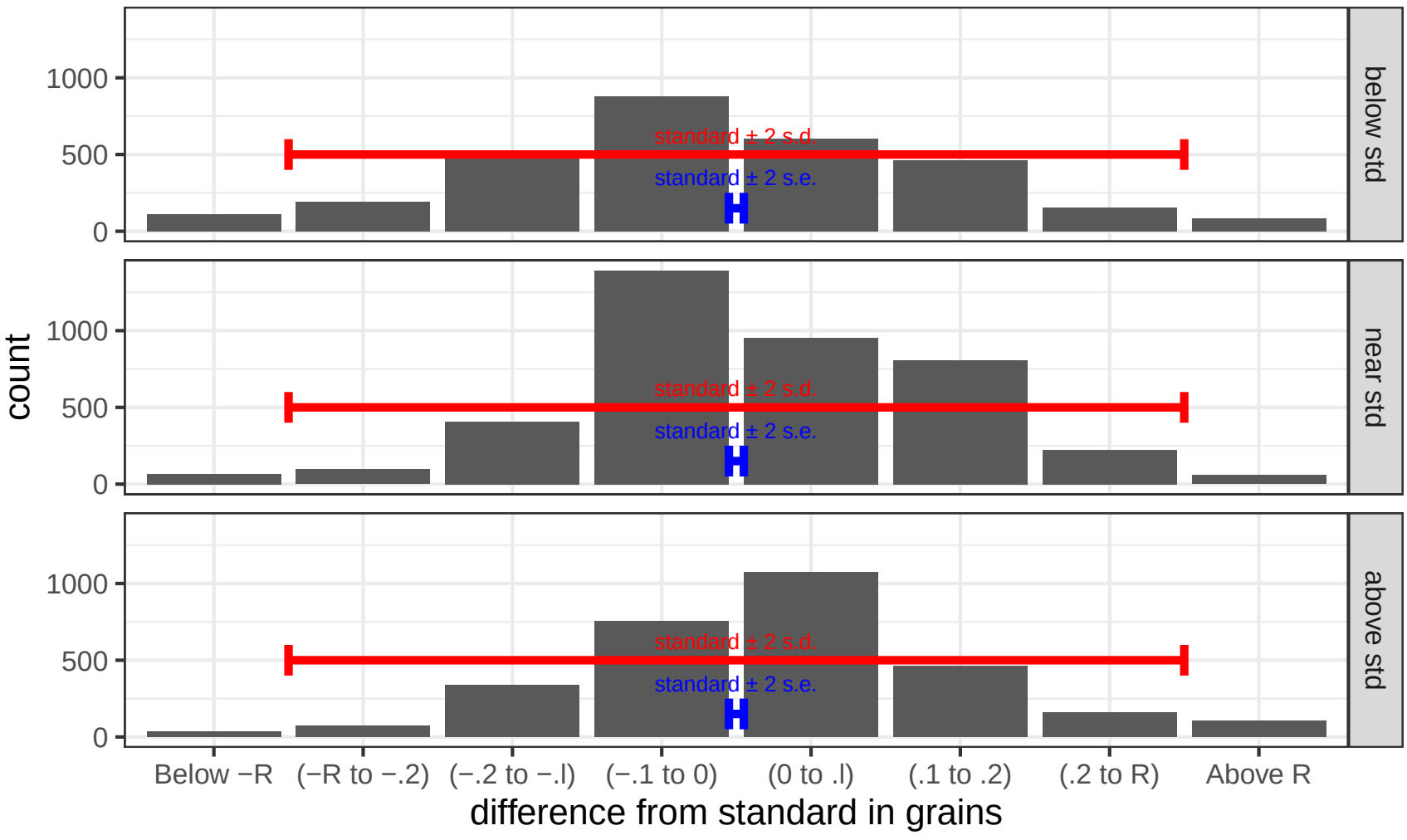
LONDON:
Printed by *W. Pearson,* for the Author. MDCCXVIII.

Source: https://en.wikipedia.org/wiki/Abraham_de_Moivre#/media/File:Abraham_de_moivre.jpg

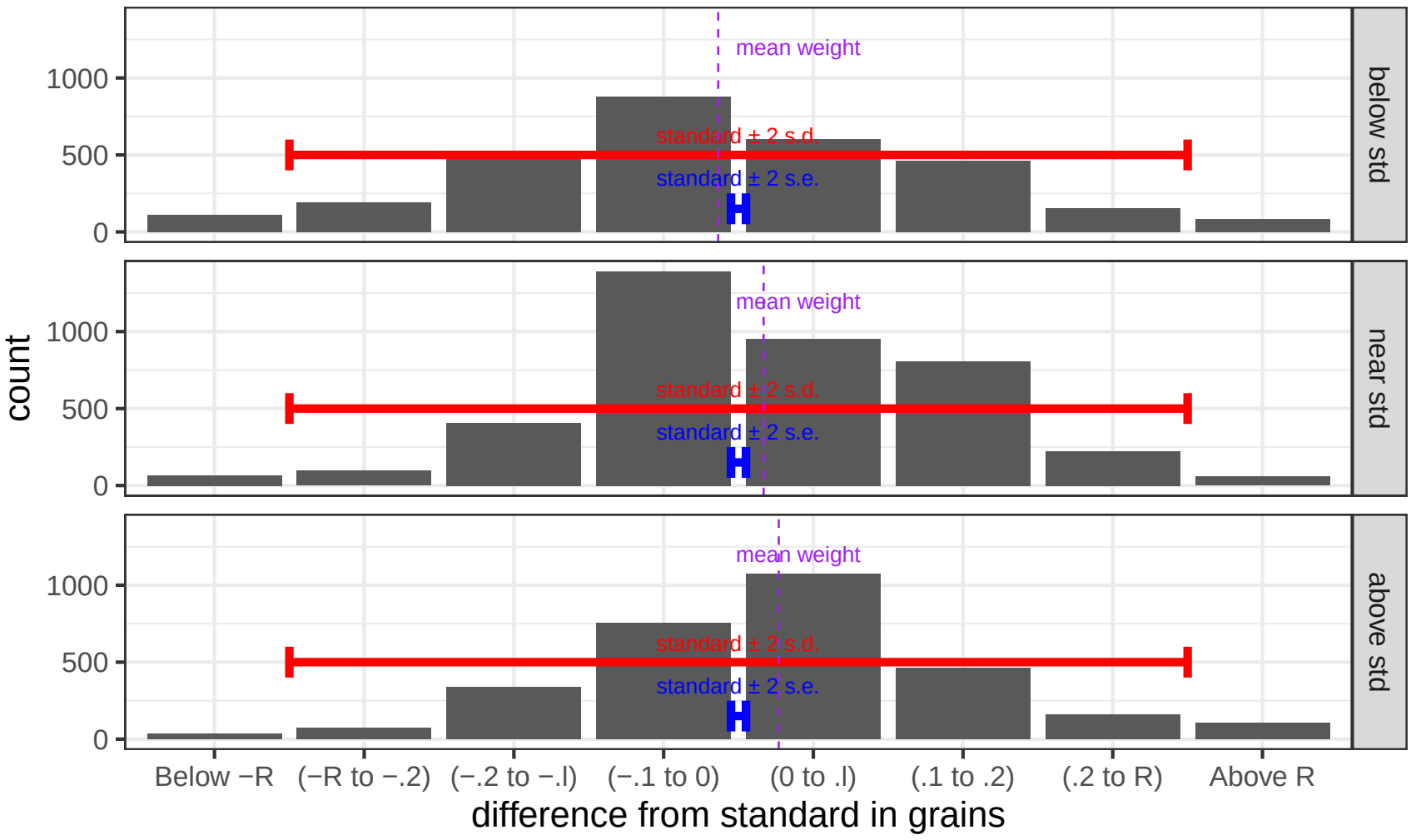
```
pyx_hist +
  annotate(geom = "text", x = 4.5, y = 350,
           label = "standard ± 2 s.e.", color = "blue") +
  geom_errorbarh(aes(xmin = a, xmax = b, y = c),
    height = 200, color = "blue", size = 2,
    data = tibble(a=4.5 - .03, b= 4.5 + .03, c= 150))
```



```
(pyx_hist <- pyx_hist + facet_grid(Group ~ .) +
   annotate(geom = "text", x = 4.5, y = 350,
            label = "standard ± 2 s.e.", color = "blue") +
  geom_errorbarh(aes(xmin = a, xmax = b, y = c),
    height = 200, color = "blue", size = 2,
    data = tibble(a=4.5 - .05, b= 4.5 + .05, c= 150)))
```



```
pyx_hist + geom_vline(aes(xintercept = mean), linetype = 2,
    color = "purple", data = Pyx %>% group_by(Group) %>%
    mutate(mean =
           weighted.mean(as.numeric(Deviation), count))) +
  annotate("text", x = 4.9, y = 1200,
           label = "mean weight", color = "purple")
```



Was the Trial of the Pyx effective?

- The Mint reimbursed the monarch for the number of coins that fell below remedy (R), extrapolated to all coins minted.
  - But a two standard deviation tolerance is too large to control for average quality.
  - An unscrupulous Mint could extract gold from each coin without any one coin falling below R. In comparison, the average weight would deviate by more than two standard errors and is thus the more appropriate measure.
- However, it is unlikely the Mint took advantage of the crown.
  - There is no evidence the Mint knew the root-n rule.
  - The trial primarily served to inspire public confidence in the currency.
- The Trial could have been just as effective with fewer coins
  - The coins were assayed by fire, a destructive process.
  - The root-n rule means that the first few observations are much more valuable than the last few.

References

1. Friendly, Michael. "HistData: Data sets from the history of statistics and data visualization." R package version 0.7-5 (2014).

2. Great Britain (1848). "Report of the Commissioners Appointed to Inquire into the Constitution, Management and Expense of the Royal Mint." In Vol 28 of House Documents for 1849.

3. Stigler, Stephen. "Eight centuries of sampling inspection: the trial of the Pyx." Journal of the American Statistical Association 72.359 (1977): 493-500.

4. Walker, Helen M. "Studies in the history of statistical method: With special reference to certain educational problems." (1929).