Connor Fisher
Dr. Imielinski
Data 101
10/19/2025

*Preface: Ikon and Epic season passes are passes in which skiers and snowboarders can buy in which they can go to multiple resorts while only using one singular pass instead of buying a pass at each resort with an Ikon Pass you can go to resorts that are owned by the POWDR Company and with an Epic Pass you can go to resort that are owned by Vail Resorts. So a question arises, which pass is better?
Dataset: https://www.kaggle.com/datasets/rummagelabs/ikon-and-epic-resorts-season-statistics

1. **Hypothesis Testing**

Title: Ikon vs Epic, see which pass you can get more days out of.
($z = 3.94$, $p = 8.1\times10^{-5}$; perm $p < 0.0005$)

Hypotheses:
H_0: $\mu\_Ikon - \mu\_Epic = 0$
H_1: $\mu\_Ikon - \mu\_Epic \neq 0$

Code:

```
d1 <- subset(df, pass %in% c("Epic","Ikon") & !is.na(len2324), select = c(pass,len2324))
ikon <- subset(d1, pass=="Ikon")$len2324
epic <- subset(d1, pass=="Epic")$len2324
m1 <- mean(ikon); m2 <- mean(epic)
n1 <- nrow(as.data.frame(ikon)); n2 <- nrow(as.data.frame(epic))
s1 <- sample_sd(ikon); s2 <- sample_sd(epic)
se <- sqrt((s1^2)/n1 + (s2^2)/n2)
z  <- (m1-m2)/se; p_z <- 2*pnorm(-abs(z))
overall_mean <- mean(d1$len2324)
diff_days <- m1 - m2; diff_pct <- 100*diff_days/overall_mean
set.seed(42); B <- 2000
labs <- d1$pass; x <- d1$len2324
perm_diffs <- replicate(B,{sh<-sample(labs); mean(x[sh=="Ikon"])-mean(x[sh=="Epic"])})
p_perm <- mean(abs(perm_diffs) >= abs(diff_days))
```

Results:
Ikon mean = 148.76 days (n = 59)
Epic mean = 123.58 days (n = 55)
Difference = +25.18 days (≈ 18.4% of the overall mean 136.61)
$z = 3.942$, $p = 8.096\times10^{-5}$
Permutation p (B = 2000) = < 0.0005 (observed gap beat all shuffles)

Interpretation:
Resorts that accept the Ikon Pass have seasons that last around 25 days longer than Epic Pass resorts. The difference is statistically large by a two-sided z-test and also by a distribution-free permutation test, so it's very unlikely to be random label noise.

Practically, a 3–4 that Ikon passholders could take advantage of that Epic passholders could not take advantage of.
Headline Justification: The Title refers to the observed difference and p-values from both tests

2. **Confidence Intervals**
   Title: Powder Bonanza: Epic and Ikon Pass resorts average 16 feet of Snow!
   CI: Average Snowfall ≈ 193.9″ (95% CI [168.3″, 219.4″]; Width ≈ 51.1″)
   Code:
   ```
   x2 <- df$snow_in; x2 <- x2[!is.na(x2)]
   n <- length(x2); m <- mean(x2); s <- sample_sd(x2)
   zcrit <- qnorm(.975); moe <- zcrit*s/sqrt(n)
   lo <- m - moe; hi <- m + moe; width <- hi - lo
   ```
   Results:
   n = 113, mean = 193.88″, sd = 138.59″
   95% CI = [168.32″, 219.43″] (width ≈ 51.10″)
   Interpretation: If we repeated this study many more times, 95% of the intervals would contain the true mean snowfall. A 51-inch band around 194-inch signals moderate precision, credible central tendency, although with real variability across locations.
   Headline Justification: The Title alludes to the mean, bounds, and the width to support the “bonanza” claim

3. **Hypothesis of Independence**
   Title: Brand Turf Wars: How do Ikon and Epic carve up the country and abroad?
   Chi-Square Test of Independence: $\chi^2 = 13.56$, $p = 0.0011$
   Code:
   ```
   d3 <- subset(df, pass %in% c("Epic","Ikon"), select = c(pass,country,st_abbr))
   east_states <-
   c("ME","NH","VT","MA","CT","RI","NY","NJ","PA","MD","DE","DC","VA","WV",
               "NC","SC","GA","FL","OH","MI","IN","IL","WI","MN","IA","MO","KY","TN")
   west_states <- c("WA","OR","CA","AK","HI","ID","MT","WY","UT","CO","NV","AZ","NM")

   Region3 <- ifelse(d3$country!="United States","International",
           ifelse(d3$st_abbr %in% east_states,"East",
           ifelse(d3$st_abbr %in% west_states,"West","East")))  # fallback keeps 3 groups

   tab <- table(d3$pass, Region3)
   chi <- chisq.test(tab, correct = FALSE)
   prop_by_brand <- prop.table(tab, margin = 1)
   ```
   Contingency Table:

	East	International	West
Epic	27	18	10
Ikon	12	20	27

   Results:

$\chi^2 = 13.562$, df = 2, p = 0.001135
Row shares (within brand):
Epic East 0.491, Int'l 0.327, West 0.182
Ikon East 0.203, Int'l 0.339, West 0.458
Interpretation: The two brands and regions in which they operate are not independent. Epic's footprint leans East, while Ikon's leans West, and both have a similar number of resorts internationally, but the difference in the U.S. produces a significant χ^2. This suggests distinct regional strategies or legacy portfolios by brand.

Headline Justification: Title reports χ^2 and p-value and reflects the direction seen in the row-wise shares.

4. **Bayesian Reasoning**
Title: Thin Air Beats Powder: Higher Altitude triples the odds of a longer season
Setup:
LR ≈ 3.00; Posterior ≈ 0.517
Snowfall still strong: LR ≈ 2.58; Posterior ≈ 0.480
H: **LongSeason** = `len2324 ≥ Q3` (top 25%)
E_1: **HighSnow** = `snow_in ≥ Q3`
E_2: **HighAltitude** = `top_ft ≥ Q3`
Code:

```
q_len  <- quantile(df$len2324, probs=.75, na.rm=TRUE)
q_snow <- quantile(df$snow_in,  probs=.75, na.rm=TRUE)
q_alt  <- quantile(df$top_ft,   probs=.75, na.rm=TRUE)

H <- as.integer(df$len2324 >= q_len)
E_snow <- as.integer(df$snow_in >= q_snow)
E_alt  <- as.integer(df$top_ft  >= q_alt)

ok_snow <- !is.na(H) & !is.na(E_snow); ok_alt <- !is.na(H) & !is.na(E_alt)
pH <- mean(H[!is.na(H)]); prior_odds <- pH/(1-pH)

pE_H_snow   <- mean(E_snow[ok_snow][H[ok_snow]==1])
pE_not_snow <- mean(E_snow[ok_snow][H[ok_snow]==0])
LR_snow <- pE_H_snow / pE_not_snow
post_odds_snow <- prior_odds * LR_snow
post_prob_snow <- post_odds_snow / (1 + post_odds_snow)

pE_H_alt   <- mean(E_alt[ok_alt][H[ok_alt]==1])
pE_not_alt <- mean(E_alt[ok_alt][H[ok_alt]==0])
LR_alt <- pE_H_alt / pE_not_alt
post_odds_alt <- prior_odds * LR_alt
post_prob_alt <- post_odds_alt / (1 + post_odds_alt)
```

Results:
Prior: P(H) = 0.263, Prior odds = 0.357
Snowfall evidence: LR = 2.582, Posterior odds = 0.922, Posterior P = 0.480
Altitude evidence: LR = 3.000, Posterior odds = 1.071, Posterior P = 0.517
Interpretation: Before seeing conditions, the chance a resort lands in the long-season quartile was 26.3%. Observing high snowfall multiplies those odds by 2.58×), while high altitude multiplies them by 3.00×. Both signals raise plausibility, with altitude being the stronger predictor for how long a resort season lasts in this dataset.
Headline Justification: Title quotes the likelihood ratios and posterior probabilities, directly reflecting the Bayes update.

5. **From z-tests to Bayes: What I learned**
   Across methods, a consistent story emerges. The z-test and permutation test show a 25-day Ikon advantage. The 95% CI mean snowfall is near 194 inches with a ~51-inch band, reflecting real geographic diversity. A pass-brand and region χ^2 test confirms that Epic and Ikon portfolios differ by geography. Finally, Bayesian reasoning guides what makes a resort season last so long and if altitude or snowfall matters more and we found out that altitude actually matters more.