Midterm 1 review

Duzhe Wang

the Department of Statistics, UW-Madison

Announcements

- This week's office hours: 6-8pm Wednesday at R1475 MSC.
- Midterm 1 will cover all materials through MSE.
- You are allowed to take a two-side note and calculator. Statistical tables will be provided.
- No multiple choices problems. 5 big problems in total.
- Do we grade on a curve? I don't know either...

Outline

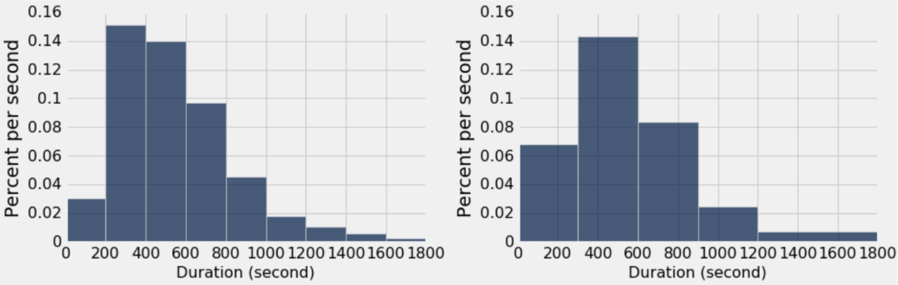
1. Descriptive statistics
2. Probability
3. Discrete RV
4. Continuous RV
5. Estimation

What you should know

- Sample mean, sample standard deviation
- Median
- Q_1, Q_3
- range
- IQR
- **histogram: you need understand how to extract information from the histogram.**

Practice problem

Two histograms of bike trip durations below were generated using different bins.



Practice problem

Write the proportion of trips that fall into each range of durations below. If it is not possible to tell from the histograms, instead write **Not enough information**.

- Between 200(inclusive) and 400 (exclusive) seconds.
- Between 300 (inclusive) and 900 (exclusive) seconds.
- Between 400 (inclusive) and 900 (exclusive) seconds.
- Between 200 (inclusive) and 300 (exclusive) seconds.

Practice problem

R was used to get summary statistics on data on the average commute time. For each of 51 states:

```
commute = c(15.2, 15.4, 16.5, 16.9, 17.5, 17.5, 18.1, 18.9, 19.1, 19.4, 19.5, 19.7,
  19.9, 20.3, 20.4, 21, 21.2, 21.6, 21.7, 21.8, 21.8, 22.1, 22.1, 22.5, 22.6,
  22.7, 22.7, 22.9, 23, 23.2, 23.3, 23.3, 23.4, 23.4, 23.6, 23.7, 23.8, 24.5,
  24.6, 24.7, 24.8, 24.8, 25.8, 26, 26.1, 26.5, 27, 28.4, 28.5, 30.2, 30.4)
> summary(commute)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  15.20   20.10   22.70   22.43   24.55   30.40
```

Practice problem

- What is the IQR of the commute times?
- Find the Q_3 of the following data set: 8.7, 9.2, 8.7, 8.0, 8.5, 10.1, 7.5, 7.8, 8.8, 8.0

Outline

1. Descriptive statistics
2. Probability
3. Discrete RV
4. Continuous RV
5. Estimation

What you should know

- Probability uses population information to describe samples in long run.
- Statistics uses sample information to make uncertain claims about population.
- Random process, outcome, sample space, event, probability
- P(E)=sum of probabilities of outcomes in E
- $0 \leq P(E) \leq 1$
- P(not E)=1-P(E)
- A and B are independent if occurrence of one doesn't change the probability of the other, then P(A and B)=P(A)P(B).

Practice problem

We have a fair four-sided die with sides numbered 1-4. If we roll it three times and got at least two 4, what is the probability of rolling a 3 for the first time and a 4 for the second time?

Outline

1. Descriptive statistics
2. Probability
3. Discrete RV
4. Continuous RV
5. Estimation

What you should know

- values can be put in sequence.
- pmf: p(x)=p(X=x)
- mean: $\mu = E(X) = \sum_x xp(x)$.
- mean properties: E(c)=c, E(cX)=cE(X), E(X+c)=E(X)+c, E(X+Y)=E(X)+E(Y)
- variance: $\sigma^2 = E[(X-\mu)^2] = \sum_x (x-\mu)^2 p(x)$.
- Variance properties: VAR(c)=0, $VAR(cX) = c^2 VAR(X)$, VAR(X+c)=VAR(X). For independent X and Y, VAR(X+Y)=VAR(X)+VAR(Y).

What you should know

- Bernoulli RV: $P(Y=1)=\pi, P(Y=0)=1-\pi$.
  $\mu=\pi, \sigma^2=\pi(1-\pi)$.
- Binomial RV: $X \sim Bin(n,\pi)$ is of success in n independent Bernoulli trials, each with P(success)$=\pi$.
- $P(X=x)=\frac{n!}{x!(n-x)!}\pi^x(1-\pi)^{n-x}$ for $x=0,...,n$.
- $\mu=n\pi$, $\sigma^2=n\pi(1-\pi)$.

Practice problem

A class of students took a quiz whose score distribution is in the table below.

score	1	2	3	4
proportion of class with score	0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{2}$

(a) What is the mean score?

(b) What is the variance of the scores?

Practice problem

Suppose each ticket in a lottery has a $\frac{1}{8}$ chance of being a winner. What is the probability of having exactly 4 winners in a randomly selected group of 10 tickets?

Practice problem

A grandmother will put a pile of money into two uncertain investments:

- a stock whose return has a mean of 6% and a standard deviation of 1%
- a bond whose return has a mean of 4% and a standard deviation of 0.5%

Suppose she puts half her money in the stock and half in the bond. Let $R_s =$ the return on the stock and $R_b =$the return on the bond(each as a percentage). Then $R = \frac{1}{2}R_s + \frac{1}{2}R_b$.

- What is the expected value of her return?
- What is the standard deviation of her return?

Outline

1. Descriptive statistics
2. Probability
3. Discrete RV
4. Continuous RV
5. Estimation

Continuous RV

- values fill interval.
- $P(a \leq X \leq b)$= area under $f(x)$ between a and b.
- cumulative distribution function $F(x) = P(X \leq x)$.
- Normal distribution: $N(\mu, \sigma^2)$, μ is the mean and σ^2 is the variance.
- If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$.
- If $Z \sim N(0, 1^2)$, then $X = Z\sigma + \mu \sim N(\mu, \sigma^2)$.
- $P(X < x) = P(Z = \frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma})$.
- **You should know how to use Z table.**

Practice problem

Suppose NBA player weights are $N(221, 15^2)$.

- Find the weight such that 20% of players weight less than that weight.
- A random group of 5 NBA players cross a play-ground bridge together, even though its breaking strength is only 1000 pounds. What is the probability that it breaks.

Outline

1. Descriptive statistics
2. Probability
3. Discrete RV
4. Continuous RV
5. Estimation

What you should know

- Simple random sample
- $X_1, ..., X_n$ are iid with mean μ and variance σ^2, then $E(\bar{X}) = \mu$, $VAR(\bar{X}) = \frac{\sigma^2}{n}$.
- Estimator, Bias, MSE
- Practice problem: review the example of how to calculate MSE for different estimators we talked about in lecture.

Good luck!